6. Martingales in Discrete Time

6.1. Definitions and examples

In this chapter we shall study stochastic processes indexed by the non-negative integers $n = 0, 1, 2, \ldots$. Throughout, it will be helpful to view processes dynamically, that is, to take n to be a measure of time. In fact, the index n does mark time in most applied models, and the time interpretation is natural for questions about how a process evolves as n increases.

At time $n = 0$, the most one knows about the future evolution of a stochastic process is its law, which is just a prescription for computing probabilities of all possible events. As time progresses one learns which events actually occur, hence obtaining information that may be relevant to making predictions of the future. It is standard in stochastic process theory to model this accumulation of knowledge by a *filtration*. Given a probability space $(\Omega, \mathcal{F}, I\!P)$, a *filtration* is just an increasing sequence $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \cdots$ of sub-σ-algebras of $\mathcal{F}$. The idea is that, at each time n, $\mathcal{F}_n$ is a model of all possible events that can be observed up to time n; more colloquially, $\mathcal{F}_n$ represents the past up to and including time n. The basic example is the filtration generated by a stochastic process $X = \{X_n\}_{n \geq 0}$. This filtration is denoted by $\{\mathcal{F}_n^X\}$ and is defined by

$$\mathcal{F}_n^X := \sigma\{X_0, \ldots, X_n\}, \quad n \geq 0;$$

in other words, $\{\mathcal{F}_n^X\}$ is the set of all possible events concerning $X_1, \ldots, X_n$. Clearly, these σ-algebras increase with n and so form a filtration. The filtration $\{\mathcal{F}_n^X\}$ is natural for the study of X if the information one learns is restricted solely to observation of X. However, more information is often available, for example, if X is just one component of a multi-variate process that is being observed. Given a filtration $\{\mathcal{F}_n\}$, a stochastic process X is said to be *adapted to* $\{\mathcal{F}_n\}$ if $\mathcal{F}_n^X \subset \mathcal{F}_n$ for every time n; adaptedness holds if and only if X_n is $\mathcal{F}_n$-measurable for every n and means, informally, that the information available in the filtration at time n includes at least an observation of the entire past history of X.

The information represented by the σ-algebras in a filtration will be exploited by conditioning. Suppose time n is the present and consider a future time $m > n$, at which it is desired to assess the probability that X_m falls in a set U. To take advantage of the observations up to time n, one should use not $I\!P(X_m \in U)$, but the conditional probability $I\!P(X_m \in A/\mathcal{F}_n)$. Likewise, at time n, knowing $\mathcal{F}_n$, the expected value of X_m is $E[X_m/\mathcal{F}_n]$; according to Proposition B.5 in Chapter 5, $E[X_m/\mathcal{F}_n]$ is the $\mathcal{F}_n$-measurable function which, out of the class of all $\mathcal{F}_n$-measurable functions Z minimizes $E[(X_m - Z)^2]$. Thus $E[X_m/\mathcal{F}_n]$ may be thought of as the best (in the mean-square sense) prediction of the value of X_n based on $\mathcal{F}_n$.

Filtrations and conditioning are the primary elements going into the definition of martingales. Martingale processes are best motivated as general models of what constitute fair, favorable, or unfavorable games of chance. We shall continue to use the language of gambling games throughout the chapter to talk about martingales, because it supplies good intuition, but, it will be clear that the theory's importance far transcends gambling applications. Thus, consider a process X_n which is supposed to represent the total accumulated fortune earned by a gambler in n plays of a game, where a negative fortune means a loss. What are the minimal conditions that should be imposed to consider the game fair, that is, one that is advantageous neither to the player or the casino? Let $\{\mathcal{F}_n\}$ denote the flow of observable information, and suppose it to include knowledge of the outcomes of all previous plays, so that X_n is $\{\mathcal{F}_n\}$-adapted. To assess favorability, take the simplest measure, namely, expected value conditioned on present information. If the game is based on betting strategy, for example, $\mathcal{F}_n$ will contain all the information that can be used at time n in deciding on the bet. Then, it is reasonable to say that the game is fair if, at each time n, the player expects neither an average gain or loss from the next play *conditional on knowledge of* $\mathcal{F}_n$; mathematically, $E[X_{n+1}/\mathcal{F}_n] = X_n$ for every time n. Likewise, a game is favorable if $E[X_{n+1}/\mathcal{F}_n] \geq X_n$ for every time n, and unfavorable if the inequality is reversed. Processes satisfying these types of constraints on predicted values are called martingales; the next definition establishes the standard terminology.

Definition Let $\{\mathcal{F}_n\}$ be a filtration. A stochastic process $\{X_n\}_{n\geq 0}$ is called an $\{\mathcal{F}_n\}$-martingale if
(i) For each $n \geq 0$, $E[|X_n|] < \infty$;
(ii) $\{X_n\}_{n\geq 0}$ is $\{\mathcal{F}_n\}$-adapted;
(iii) For each $n \geq 0$, $E[X_{n+1}/\mathcal{F}_n] = X_n$ a.s.

$\{X_n\}_{n\geq 0}$ is called an $\{\mathcal{F}_n\}$-submartingale if it satisfies (i), (ii), but (iii) is replaced by $E[X_{n+1}/\mathcal{F}_n] \geq X_n$ for $n \geq 0$.
$\{X_n\}_{n\geq 0}$ is called a $\{\mathcal{F}_n\}$-supermartingale if it satisfies (i), (ii), but (iii) is replaced by $E[X_{n+1}/\mathcal{F}_n] \leq X_n$.

Saying that $\{X_n\}_{n\geq 0}$ is a *martingale with respect to the filtration* $\{\mathcal{F}_n\}$ is an equivalent way of saying that $\{X_n\}_{n\geq 0}$ is an $\{\mathcal{F}_n\}$-martingale. We shall often abuse terminlogy by dropping explicit mention of the filtration and just saying "$\{X_n\}_{n\geq 0}$ is a martingale (sub-martingale, supermartingale)." In these cases, either the associated filtration should be clear from context, or explicit identification of the filtration is not important.

Note immediately that $\{X_n\}_{n\geq 0}$ is a supermartingale if and only if $\{-X_n\}_{n\geq 0}$ is a submartingale. This saves a lot of work, since we need only carry out proofs in one case or the other in order to have statements about both. A process is a martingale if and only if it is both a submartingale and a supermartingale.

Here are some basic facts that follow immediately from the definitions and basic properties of conditional expectation.

Proposition 1.1 (a) A martingale $\{X_n\}_{n\geq 0}$ is a constant expectation process: that is, $E[X_0] = E[X_1] = E[X_2] = \cdots$. Submartingales are increasing in expectation $(E[X_0] \leq E[X_1] \leq E[X_2] \leq \cdots)$, and supermartingales are decreasing in expectation.

(b) If $\{X_n\}_{n\geq 0}$ is an $\{\mathcal{F}_n\}$-martingale (respectively, sub- or supermartingale), then

$$E[X_m/\mathcal{F}_n] = X_n \quad \text{(resp., } \geq, \leq\text{) a.s. for all } m > n \geq 0.$$

(c) An $\{\mathcal{F}_n\}$-(sub)(super)martingale $\{X_n\}_{n\geq 0}$ is always a (sub)(super) martingale with respect to its own filtration $\mathcal{F}_n^X := \sigma\{X_0, \ldots, X_n\}$, $n \geq 0$.

Items (a) and (b) in this proposition are very natural. In the language of games of chance, statement(a) says that the property of a game being fair (or favorable, or unfavorable) passes from the level of conditional expectations to unconditioned expectations, while (b) says that if the game is fair (respectively, favorable or unfavorable) when looking one step ahead into the future, which is only what the martingale definitions require, it is fair (respectively, favorable or unfavorable) looking m steps into the future.

Proof: (a) Consider a martingale $\{X_n\}_{n\geq 0}$ with respect to $\{\mathcal{F}_n\}$. Then by conditioning inside the expectation, $E[X_{n+1}] = E[E[X_{n+1}/\mathcal{F}_n]] = E[X_n]$. The proof in the martingale case now follows by induction, and the case of a sub- or supermartingale is similar.

(b) Consider the martingale case only. The result is a direct consequence of the tower property of conditioning. Let $m > n$. Then $\mathcal{F}_n \subset \mathcal{F}_{m-1}$ and hence

$$E[X_m/\mathcal{F}_n] = E[E[X_m/\mathcal{F}_{m-1}]/\mathcal{F}_n] = E[X_{m-1}/\mathcal{F}_n].$$

Iteration of this procedure gives the result.

(c) This is also a consequence of the tower property. For every n, $\mathcal{F}_n^X \subset \mathcal{F}_n$. Hence

$$E\left[X_{n+1}/\mathcal{F}_n^X\right] = E\left[E[X_{n+1}/\mathcal{F}_n]/\mathcal{F}_n^X\right] = E\left[X_n/\mathcal{F}_n^X\right] = X_n,$$

the last equality holding because X_n is $\mathcal{F}_n^X$-measurable. $\diamond$

It will also be useful to have the notion of a reverse martingale, which has the martingale propery running backward in time.

Definition Let $\{\mathcal{G}_n\}$ be a *reverse filtration*, that is, a decreasing sequence of σ-algebra. The process $\{Y_n\}_{n\geq 0}$ is said to be a reverse $\{\mathcal{G}_n\}$-martingale if Y_n is $\mathcal{G}_n$-measurable and $E[Y_n/\mathcal{G}_{n+1}] = Y_{n+1}$, for each $n \geq 0$. Reverse sub- and supermartingales are defined by replacing equality by $\geq$ and $\leq$ respectively.

If one changes the time index, relabelling n by $-n$, a reverse martingale becomes a martingale starting at time $-\infty$. Note that by developing the analogue of property (b) of Propoisition 6.1 for reverse martingales, one finds that for any reverse $\{\mathcal{G}_n\}$-martingale $\{Y_n\}$,

$$Y_n = E[Y_0/\mathcal{G}_n], \qquad n \geq 0. \tag{1}$$

A reverse martingale thus represents a sequence of conditional expectations over a sequence of progressively coarser σ-algebras.

Examples and further discussion

1. Let $\{\xi_k\}_{k\geq 0}$ be i.i.d., integrable random variables, and set $\mu = E[X_k]$ (which is the same for all k). Define $\mathcal{F}_0 := \{\emptyset, \Omega\}$, and $\mathcal{F}_n := \sigma\{\xi_1, \ldots, \xi_n\}$, if $n \geq 1$. Let $S_0 := 0$, and, for $n \geq 1$, $S_n := \sum_1^n X_k$. Then $\{S_n\}$ is an $\{\mathcal{F}_n\}$ -martingale if $\mu = 0$, a submartingale if $\mu \geq 0$ and a supermartingale if $\mu \leq 0$. The easy details of checking the martingale properties are left to the reader. Notice that for this example, $\mathcal{F}_n = \mathcal{F}_n^S$ for $n \geq 1$. The process $\{S_n\}$ is called a *random walk*. In this chapter we shall derive information about random walks using martingale techniques.

2. Here is a method for constructing a large class of martingales. Let Z be an integrable random variable and let $\{\mathcal{F}_n\}$ be a filtration. Then $X_n := E[Z/\mathcal{F}_n]$, $n \geq 0$ defines an $\{\mathcal{F}_n\}$-martingale. The integrability and $I\!\!F_n$ measurability of X_n for any n follow directly from the definition and elementary properties of conditional expecation. The martingale property is a consequence of the tower property of conditioning:

$$E\left[X_{n+1}/\mathcal{F}_n\right] = E\left[E\left[Z/\mathcal{F}_{n+1}\right]/\mathcal{F}_n\right] = E\left[Z/\mathcal{F}_n\right] = X_n.$$

In the context of this example, extend the index set to $\bar{I\!\!N} := \{0, 1, 2, 3, \ldots\} \cup \{\infty\}$, linearly ordered in the natural way so that ∞ is larger than any non-negative integer. Extend the filtration $\{\mathcal{F}_n\}$ to $\bar{I\!\!N}$ by definining $\mathcal{F}_\infty := \sigma(\cup_1^\infty \mathcal{F}_n)$. In the construction above, one can assume without loss of generality that Z is $\mathcal{F}_\infty$-measurable. Indeed, if it is not, use the tower property to write,

$$X_n = E\left[Z/\mathcal{F}_n\right] = E\left[E\left[Z/\mathcal{F}_\infty\right]/\mathcal{F}_n\right]$$

and replace Z by $E[Z/\mathcal{F}_\infty]$. Now extend extend the process X_n to $\bar{I\!\!N}$ by setting $X_\infty = Z$. The extended process now forms a martingale on the extended index set in the sense that X_n is $\mathcal{F}_n$-measurable for *all* $n \in \bar{I\!\!N}$, and $E\left[X_m/\mathcal{F}_n\right] = X_n$ whenever n and m belong to $\bar{I\!\!N}$ and $m > n$.

Given a martingale $Y = \{Y_n\}$, call it *closable* if there exists an $\mathcal{F}_\infty$-measurable and integrable random variable Z such that $Y_n = E[Z/\mathcal{F}_n]$. It is natural to ask when a martingale is closable, and later in the chapter a necessary and sufficient condition for closability will be presented. Not all martingales are closable, for instance the random walk of example 1.

Although not exhaustive, the class of closable martingales provides another good, intuitive way to think about martingales, namely as a sequence of conditional expectations over successively finer σ-algebras. Compare to representation (1) of a reverse martingale.

3. We present an illustration of the general construction presented in example 2, related to the problem of differentiation of measures. Take $([0, 1), \mathcal{B}([0, 1)), \lambda)$, where

λ is Lebesgue measure, as the underlying probability space, and for each $n \geq 0$, let $\mathcal{F}_n$ be the finite σ-algebra generated by subintervals of the form $[k2^{-n}, (k+1)2^{-n})$, $0 \leq k < 2^n$. Let f be a Lebesgue integrable, Borel function on $[0, 1)$, thought of as a random variable. Then $f_n := E[f/\mathcal{F}_n]$, $n \geq 0$ is a martingale. This example makes explicit the averaging interpretation inherent in the idea of conditional expectation. By specializing equation (2) of Chapter 5 to this case, one finds

$$f_n(x) = \sum_j^{2^n-1} \left(2^n \int_{j2^{-n}}^{(j+1)2^{-n}} f(y)\, dy \right) \mathbf{1}_{[j2^{-n},(j+1)2^{-n})}(x).$$

When f is continuous, simple arguments show that these averages tend to f everywhere as $n \to \infty$. In fact, a theorem from analysis on differentiation of measures says that f_n will converge to f Lebesgue almost everywhere if f is merely Borel measurable and integrable.

Returning to the general construction $X_n = E[Z/\mathcal{F}_n]$ of example 2, it is natural to suspect that $X_n \to E[Z/\mathcal{F}_\infty]$, a.s. In fact this is true and known as Lévy's upward theorem. We will prove it later as a consequence of general theorems about martingale convergence, and thereby recover an independent proof of the convergence of f_n to f.

4. Let $\{\xi_n\}$ and S_n be as in example 1. Let $\mathcal{G}_n := \sigma(S_n, S_{n+1}, \ldots)$. We claim that

$$E\left[\xi_1/\mathcal{G}_n\right] = \frac{S_n}{n} \quad \text{for every } n \geq 1. \tag{2}$$

It follows from this that

the empirical mean process of an i.i.d. sequence is a reverse martingale.

The proof is as follows. It is a general fact that if (X, W) is a pair of random variables independent of a σ-algebra $\mathcal{H}$, then $E\left[X/W, \mathcal{H}\right] = E\left[X/W\right]$; the proof is left as an exercise. By noticing that (ξ_1, S_n) is independent of $\sigma(\xi_{n+1}, \xi_{n+2}, \ldots)$ and that $\mathcal{G}_n$ is generated by S_n and $\sigma(\xi_{n+1}, \xi_{n+2}, \ldots)$, one derives from this remark that

$$E\left[\xi_1/\mathcal{G}_n\right] = E\left[\xi_1/S_n\right].$$

Now, observe that for any $1 \leq k \leq n$, the pairs of random variables (ξ_1, S_n) and (ξ_k, S_n) have the same probability distribution. It then follows that

$$E\left[\xi_k/S_n\right] = E\left[\xi_1/S_n\right] \qquad 1 \leq k \leq n.$$

Therefore

$$S_n = E\left[S_n/S_n\right] = E\left[\sum_1^n \xi_k/S_n\right] = nE\left[\xi_1/S_n\right].,$$

and (2) follows. Later in the chapter, we shall recover the strong law of large numbers for i.i.d. sequences from a martingale convergence theorem applied to this reverse martingale.

5. Martingales for asymmetric, simple random walk. Let ξ_n be i.i.d. Bernoulli random variables with probability distribution $I\!P(\xi_n = 1) = p$ and $I\!P(\xi = -1) = q$, where $p \neq q$ and $p + q = 1$. Let S_n be the associated random walk, and $\{\mathcal{F}_n\}$ the associated filtration, in the notation of example 1. Since $p \neq q$, $\{S_n\}$ is not itself a martingale. One can build a martinagle from S_n simply by subtracting off the mean. It is left to the reader to show that $\{S_n - n(p-q)\}$ is a martingale. But there are other more interesting martingales. On the integer lattice, define the function $f(i) = \alpha + \beta(q/p)^i$, where α and β are arbitrary real constants. Then

$$f(S_n) \quad \text{is an } \{\mathcal{F}_n\}\text{-martingale.}$$

Indeed, one easily checks that $E[(q/p)^{\xi_n}] = 1$ for any n. Thus, because ξ_{n+1} is independent of $\mathcal{F}_n$ and S_n is $\mathcal{F}_n$-measurable,

$$\begin{aligned} E\left[\alpha + (q/p)^{S_{n+1}} / \mathcal{F}_n\right] &= \alpha + E\left[(q/p)^{S_n} (q/p)^{\xi_{n+1}} / \mathcal{F}_n\right] \\ &= \alpha + (q/p)^{S_n} E\left[q/p)^{\xi_{n+1}} / \mathcal{F}_n\right] = Y_n. \end{aligned}$$

6. Martingales and harmonic functions for Markov chains.

Random walks are particular examples of a class of processes called Markov chains, which are of central importance in stochastic process theory. The class of martingales constructed for the random walk in example 5 illustrate a deeper and more general procedure for constructing martingales from Markov chains. The purpose of the next discussion is to give a taste of this broader theory. It is not our purpose in this chapter to give a full account of Markov chains. Indeed, to avoid excessive digression, we formulate definitions only in a very restrictive setting, with pointers to greater generality. In particular, we restrict the discussion mostly to Markov chains on a discrete state space.

Most of the conditioning for discrete state-space chains is of the form $I\!P(A/Y)$, where Y is a discrete random variable. Recall from chapter 5, that if Y is discrete and if one defines

$$(*) \qquad h(y) = I\!P\left(A/Y = y\right) \quad \text{if } I\!P\left(Y = y\right) > 0$$

(and leaves $h(y)$ undefined otherwise), then $I\!P(A/Y) = h(Y)$. The following principle will also be useful. Suppose $\mathcal{G}$ and $\mathcal{H}$ are σ-algebras and let $\mathcal{G} \vee \mathcal{H}$ denote the σ-algebra generated by their union.

$$(**) \qquad \text{If } I\!P\left(A/\mathcal{G} \vee \mathcal{H}\right) \text{ is } \mathcal{G}\text{-measurable, then } I\!P\left(A/\mathcal{G} \vee \mathcal{H}\right) = I\!P\left(A/\mathcal{G}\right).$$

The proof is a direct consequence of the tower property and is left to the reader.

Let Λ be a countable set; we use indices i and j to denote typical elements of elements of Λ. A *stochastic matrix* indexed by $\Lambda \times \Lambda$ is a matrix $P = [p_{ij}]$, $(i,j) \in \Lambda \times \Lambda$, satisfying $0 \le p_{ij} \le 1$ for every (i,j) and $\sum_{j \in \Lambda} p_{ij} = 1$ for every i. The ideas behind a stochastic matrix is that p_{ij} represents the probability of a transition from state i to state j of a stochastic processes evolving in Λ.

Definition An Λ-valued process $\{X_n\}_{n \ge 0}$ is said to be a *time-homogenous Markov chain with probability transition matrix* P if, for every $n \ge 0$ and elements $i_0, \ldots, i_n$ and j of Λ,

$$\mathbb{P}\left(X_{n+1}=j / X_n=i_n, \ldots, X_0=i_0\right) = \mathbb{P}\left(X_{n+1}=j / X_n=i_n\right) = p_{ij}. \tag{3}$$

Random walks on the integer lattice provide a good example. Take Λ to be the set of all integers, let ξ_n, $n \ge 1$ be i.i.d. Λ-valued random variables, and let

$$r_i := \mathbb{P}(\xi_n = i) \qquad -\infty < i < \infty,$$

be the common probability mass function of the ξ_n's. Denote the associated random walk by $S_n = \sum_1^n \xi_k$. Then, since $S_{n+1} = S_n + \xi_{n+1}$ and ξ_{n+1} is independent of $S_0, S_1, \ldots, S_n$,

$$\begin{aligned} \mathbb{P}\left(S_{n+1}=j \;/\; S_n=i_n, \ldots, S_0=i_0\right) &= \mathbb{P}\left(S_{n+1}=j \;/\; S_n=i_n\right) \\ &= \mathbb{P}(\xi_{n+1}=j-i_n) = r_{j-i_n}. \end{aligned} \tag{4}$$

Hence $\{S_n\}$ is a random walk with transition matrix $[r_{j-i}]$, $-\infty < i,j < \infty$.

In the case of the Bernoulli random walk of example 5, the transition matrix works out to be $P = p_{ij}$, where

$$p_{ij} = \begin{cases} p & \text{, if } j = i+1; \\ q & \text{, if } j = i-1; \\ 0 & \text{, otherwise.} \end{cases} \tag{5}$$

Condition (3) has the following consequence, which will be important. Think of a function f on Λ as a vector $(f(i))_{\{i \in \Lambda\}}$, and let $(Pf)(i) = \sum_\Lambda p_{ij} f(j)$, $i \in \Lambda$ be the product of the transition matrix P with f. Then

$$\begin{aligned} E\left[f(X_{n+1})/X_n=i_n, \ldots, X_0=i_0\right] &= E\left[f(X_{n+1})/X_n=i_n\right] \\ &= \sum_\Lambda p_{i_n j} f(j) = (Pf)(i_n). \end{aligned} \tag{6}$$

Condition (3) in the definition of a Markov chain says that in assessing the conditional probability of a future event given the entire past history of the chain, the only relevant data is the present value X_n of the chain. In short, the future is independent of the past given the present. This is the Markov property. We pause to give it a rigorous, general definition.

Definition A process $\{Y_n\}$ taking values in a measurable space $(\Sigma, \mathcal{S})$ is said to be a Markov process if for each non-negative integer n, and each event A in the σ-algebra , $\sigma\{Y_k\,;\, k \geq n+1\}$ of events in the future at time n,

$$\mathbb{P}\left(A \,/\, Y_n, \ldots, Y_0\right) = \mathbb{P}\left(A \,/\, Y_n\right) \quad \text{a.s.} \tag{7}$$

Property (3) defining a Markov chain is a special case of the Markov property (7). If (3) is satisfied, then, by principle (*) and additivity,

$$\mathbb{P}\left(X_{n+1} \in U \,/\, X_n, \ldots, X_0\right) = \sum_{j \in U} p_{X_n, j}.$$

This is X_n-measurable and hence, by principle (**),

$$\mathbb{P}\left(X_{n+1} \in U \,/\, X_n, \ldots, X_0\right) = \mathbb{P}\left(X_{n+1} \in U \,/\, X_n\right).$$

Thus, (7) is recovered for events A of the form $\{X_{n+1} \in U\}$, looking one step ahead into the future. It turns out that (3) is sufficient to imply the Markov property (7) for all future events. In fact, a more general statement is true

Lemma 1.2 In order to check the Markov propery (7), it suffice to check that for every $n \geq 0$ and Borel U

$$\mathbb{P}\left(X_{n+1} \in U \,/\, X_n, \ldots, X_0\right) = \mathbb{P}\left(X_{n+1} \in U \,/\, X_n\right) \tag{8}$$

This Lemma implies that *a Markov chain as defined in (3) is a Markov process.*

The proof of the lemma will be deferred to an appendix, as it is not important for our immediate purposes. Suffice it to say here that the proof involves the tower propery and repeated conditioning to first prove (7) for all future events A depending on the outcomes of only a finite number of X_m's, and then extending to all events in the future by a Dynkin system argument.

Martingales have an important role in the theory of Markov processes. To illustrate in the case of Markov chains, it helps to make another definition.

Definition Let P be a stochastic matrix for a countable state space Λ. A function $f : \Lambda \to \mathbb{R}$ is called *P-harmonic* if $f = Pf$ (in the notation established above). The function f is P-subharmonic (respectively, superharmonic), if if $f \leq Pf$ (respectively, $f \geq Pf$), where $f \leq Pf$ means $f(i) \leq (Pf)(i)$ for all $i \in \Lambda$.

Here is the purpose of defining harmonic functions.

Proposition 1.2 Let $\{X_n\}_{n\geq 0}$ be a Markov chain on Λ with transition probability matrix P. If f is P-harmonic (respectively, subharmonic or superharmonic), and if $f(X_n)$ is integrable for all n, then $\{f(X_n); n \geq 0\}$ is an $\{\mathcal{F}_n^X\}$-martingale (respectively, submartingale or supermartingale.)

Coversely, if $f(X_n)$ is (sub)(super) martingale, and if for each $i \in \Lambda$ there is an n such that $P(X_n = i) > 0$, then f is a P-(sub)(super) harmonic function.

Proof: Let f be P-harmonic. Observe that, by the Markov property and (6),

$$\begin{aligned} E\left[f(X_{n+1}) \;/\; \mathcal{F}_n^X\right] &= E\left[f(X_{n+1})/X_n\right] \\ &= (Pf)(X_n), \end{aligned} \tag{9}$$

which shows that $\{f(X_n)\}$ is a martingale. Conversely, if $f(X_n)$ is a martingale, so that (9) is true, and if $I\!P(X_n = i) > 0$, then $f(i) = (Pf)(i)$. The sub- and supermartingale cases have similar proofs. ◇

For an example, return to the asymmetic, Bernoulli random walk presented in example 5, whose transition matrix is presented in equation (5). For this case, any P harmonic function must, by definition, satisfy the homogeneous, linear difference equation

$$f(i) = pf(i+1) + qf(i-1)$$

It is easy to check that the general solution is $f(i) = \alpha + \beta(q/p)^i$, where α and β are arbitrary real numbers. It then follows from Proposition 6.2 that $\alpha + \beta(q/p)^{S_n}$ is a martingale, thus recapturing, in a systematic fashion, the result of example 5.

It is interesting to do another example, generalizing the random walk. This time let the state space be the non-negative integers. We model a process that takes unit steps up or down, like a random walk, except when it is at state 0. The walk can go no lower than 0, and so from 0, it may only either stay at 0 or jump to a higher lever. Thus, for $i \geq 1$, suppose that the transition probabilities are

$$p_{ij} = \begin{cases} p_i & \text{, if } j = i+1; \\ q_i & \text{, if } j = i-1; \\ 0 & \text{, otherwise,} \end{cases}$$

where $q_i = 1 - p_i$ for each $i \geq 1$. This is similar to the transitions for the random walk of example 5, but the steps are no longer identically distributed, since the step probabilities p_i and q_i now depend on the present location i of the walk. For the moment, put no definite structure on the transition probabilities $p_{00}, p_{01}, \ldots$ out of state 0. A P-harmonic function f must satisfy at least the linear difference equation

$$f(i) = p_i f(i+1) + q_i f(i-1), \quad i \geq 1. \tag{10}$$

A constant function f is always a solution. To get any other solution, note that, from (10) and $p_i + q_i = 1$,

$$f(i+1) - f(i) = \frac{q_i}{p_i}\left(f(i) - f(i-1)\right).$$

Hence,

$$f(i) = f(0) + (f(1) - f(0)) \sum_{j=1}^{i-1} \frac{q_j \cdots q_1}{p_j \cdots p_i}, \quad i \geq 2. \tag{11}$$

There are now several cases, corresponding to what happens at the boundary. Suppose $p_{00} = 1$; this is the case of an absorbing boundary, because once the chain hits 0, it must stay there. Then the condition for P-harmonicity at $i = 0$ reads, $f(0) = f(0)$ which is trivially satisfied by any real number. As a consequence $f(0)$ and $f(1)$ may be chosen arbitrarily and we get that all P-harmonic functions are of the form,

$$f(0) = \alpha, \quad f(1) = \beta, \quad \text{and} \quad f(i) = \alpha + (\beta - \alpha) \sum_{j=1}^{i-1} \frac{q_j \cdots q_1}{p_j \cdots p_i}, \quad i \geq 2. \tag{12}$$

The second case allows for the possibility of a reflection from 0 to 1: $p_{00} = 1 - p_0$ and $p_{01} = p_0$ where $p_0 > 0$. Then, P-harmonicity requires $f(0) = (1 - p_0) f(0) + p_0 f(1)$, which means that $f(0) = f(1)$. Refering back to (11), one sees that in this case there are *no* non-constant harmonic functions. On the other hand, formula (12) will provide a subharmonic function if $\beta > \alpha$ and a superharmonic function if $\beta < \alpha$.

The final case is that of a general transition probability $p_{00}, p_{01}, \ldots$ out of 0. The reader may check as an exercise that as long as $p_{00} < 1$ there are no non-constant P-harmonic functions.

7. Convex functions of martingales Given a martingale or a submartingale, new submartingales can be built from it by convex transformations. This is a very useful fact that will be used repeatedly.

Proposition 1.3
(a) Let $\{X_n\}_{n \geq 0}$ be a martingale and let ϕ be a convex function for which $\phi(X_n)$ is integrable for every n. Then $\{\phi(X_n)\}$ is a submartingale.

(b) If $\{X_n\}_{n \geq 0}$ is a submartingale and if ϕ is an *increasing* convex function such that $\phi(X_n)$ is integrable for every n, then $\{\phi(X_n)\}$ is a submartingale.

Proof: We prove (a) and leave (b) as an exercise. The proof is just an application of the conditional Jensen inequality:

$$\phi(X_n) = \phi\left(E\left[X_{n+1} / \mathcal{F}_n\right]\right) \leq E\left[\phi(X_{n+1}) / \mathcal{F}_n\right].$$

6.2 Preservation of the martingale property for predictable gambling strategies

In this section, we present a method of constructing new martingales from a given martingale $X = \{X_n\}_{n\geq 0}$. This method is at the heart of the definition of a stochastic integral with respect to a martingale in continuous-time martingale theory, so we shall refer to the method as discrete-time stochastic integration, although this terminology is not at all standard. In the more colorful language of games of chance, discrete-time stochastic integrals may be thought of as the fortune earned by a betting strategy. We introduce the subject from this viewoint.

Let $\{X_n\}_{n\geq 0}$ denote the fortune in a game of chance in which one dollar is bet on every play. In other words, $X_n - X_{n-1}$ is the gain or loss from a dollar bet on play n. For the moment make no assumption on the law of the $\{X_n\}_{n\geq 0}$. On play n the gambler is free to choose an amount V_n to bet. Then the gain in fortune by time n is

$$(X.V)_n := \sum_{k=1}^{n} V_k \left(X_k - X_{k-1}\right).$$

The process $\{(X.V)_n\}$ is referred to as the discrete-time stochastic integral of $V = \{V_n\}$ against X.

So far, no restrictions have been place upon X and V. Assume now that the information available to the gambler immediately *after* play n is represented by a σ-algebra $\mathcal{F}_n$, and assume that $\mathcal{F}_n$ increases with n, so that $\{\mathcal{F}_n\}$ is a filtration. As a consequence, the amount V_{n+1} bet on play $n+1$ is determined knowing at most the information in $\mathcal{F}_n$, and hence should be $\mathcal{F}_n$-measurable. This measurability constraint is dignified by a name.

Definition The process $\{V_n\}$ is called $\{\mathcal{F}_n\}$-predictable if for each $n \geq 1$, V_n is $\mathcal{F}_{n-1}$-measurable for all $n \geq 1$.

The next theorem, although simple, will be absolutely basic to the development of martingale theory. It looks at how martingale properties are preserved under the operation of stochastic integration.

Theorem 2.1

(a) Let $X = \{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-martingale. If $\{V_n\}$ is an $\{\mathcal{F}_n\}$-predictable process and if, for each $n \geq 0$, there is a constant $K_n < \infty$ such that $\mathbb{P}(|V_n| \leq K_n) = 1$, then

$$\{(V.X)_n\} \quad \text{is an } \{\mathcal{F}_n\}\text{-martingale.}$$

(b) Let $X = \{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-submartingale. If $\{V_n\}$ is an $\{\mathcal{F}_n\}$-predictable process and if, for each $n \geq 0$, there is a constant $K_n < \infty$ such that $\mathbb{P}(0 \leq V_n \leq K_n) = 1$, then

$$\{(V.X)_n\} \quad \text{is an } \{\mathcal{F}_n\}\text{-submartingale.}$$

A verbal paraphrase makes the statements of this theorem intuitively obvious. Statement (a) says that whatever (bounded) predictable betting strategy is used, a fair game remains fair. Without clairvoyance one cannot change a fair game to a favorable game. Statement (b) says that if non-negative predictable bets are place on a favorable game, it remains favorable.

Proof: We do case (a) only. The assumption that $I\!\!P(|V_n| \leq K_n) = 1$ for all n serves to guarantee that $(V.X)_n$ is integrable for every n, as may easily be checked. Clearly also, the predicatibility of V implies that $V.X$ is $\{\mathcal{F}_n\}$-adapted. It remains to check the martingale property:

$$\begin{aligned} E\left[(V.X)_{n+1} / \mathcal{F}_n\right] &= E\left[(V.X)_n + V_{n+1}\left(X_{n+1} - X_n\right) / \mathcal{F}_n\right] \\ &= (V.X)_n + V_{n+1} E\left[\left(X_{n+1} - X_n\right) / \mathcal{F}_n\right] \\ &= (V.X)_n \end{aligned}$$

The first equality follows from the definition of $V.X$, the second from the predictability of V, and the last from the martingale property of X. $\diamond$

6.3 Optional Stopping of a Martingale

6.3.1 Stopping times and stopping strategies. Can one win a fair game?

Let $X = \{X_n\}_{n \geq 0}$ denote the history of a gambler's fortune in a game of chance. Assume that the size of the bet on each play is fixed by the house and so the gambler's only strategy option is do decide when to stop playing. Let T denote the time (or play) at which the gambler chooses to stop; this may depend on the outcome of the plays and so can be a random variable. The player leaves the game with a fortune of X_T, the value of the game at the random time T. (If we were to put in the dependence on ω explicitly, we would write $X_{T(\omega)}(\omega)$ for this value.)

Let the filtration $\{\mathcal{F}_n\}$ represent the flow of information available to the gambler during the game. The time T chosen by the gambler to end the game cannot be an arbitrary random variable, but must be consistent with this filtration. At a given time n, the gambler cannot look into the future, say, to see that he will suffer a devasting loss in he next play, and so decide to stop. Rather, at time n, he can only use the information available up to time n.

Definition A random variable T with values in $\bar{I\!\!N} := \{0, 1, 2, \ldots\} \cup \{\infty\}$ is called an $\{\mathcal{F}_n\}$-stopping time if

(1) $$\{T = n\} \in \mathcal{F}_n, \quad \text{for all non-negative integers } n.$$

Note: The condition (1) is equivalent to

(2) $$\{T \leq n\} \in \mathcal{F}_n, \quad \text{for all non-negative integers } n.$$

By the definition, if T is a stopping time, then $\{T > n\}$ is also $\mathcal{F}_n$-measurable for every n. Note that the possibility that $T = \infty$ (never stopping) is allowed. Moreover, if $\mathcal{F}_\infty := \sigma(\cup_0^\infty \mathcal{F}_n)$, then $\{T = \infty\}$ is $\mathcal{F}_\infty$-measurable. This is an immediate consequence of the identity $\{T = \infty\} = \cap_n \{T > n\}$ and the fact, just pointed out, that $\{T > n\}$ is $\mathcal{F}_n$-measurable for every n.

Now let $\{X_n\}_{n\geq 0}$ be a martingale and T a stopping time, both with respect to a filtration $\{\mathcal{F}_n\}$. Assume that $T < \infty$ almost surely, so that X_T makes sense almost surely. Since T is a non-clairvoyant, we should not be able to render a fair game favorable by using T to stop; mathematically, this means we should expect that

$$E\left[X_T\right] = E\left[X_0\right]. \tag{3}$$

Optional stopping theory for martingales concerns verifying when (3) is true. It is not an entirely trivial problem, especially since it is not always true! Here is a very basic example. Take the standard, symmetric, Bernoulli random walk. Thus $S_0 = 0$ and $S_n = \sum_1^n \xi_k$, where $\{\xi_k\}$ is a sequence of independent Bernoulli random variables with the common distribution, $\mathbb{P}(\xi = -1) = \mathbb{P}(\xi = 1) = 1/2$. Since the ξ_k have zero mean, $\{S_n\}$ is a martingale. Now let

$$T = \min\{n\,;\, S_n = 1\},$$

the first time that the random walk hits 1. This is clearly a stopping time:

$$\{T = n\} = \{S_1 < 1, S_2 < 1, \ldots, S_{n-1} < 1, S_n = 1\} \in \mathcal{F}_n^S.$$

Moreover, it will be shown later that $\mathbb{P}(T < \infty) = 1$. Since $S_T = 1$ by definition of T, it follows that

$$E\left[S_T\right] = 1 > E\left[S_0\right] = 0.$$

Thus T is a sure strategy for beating a fair game!

It would seem that we have achieved the holy grail of every gambler, a sure-fire way to beat the odds. However, before rashly boarding the bus for Atlantic City, a little optional stopping theory will help put this example into perspective. Notice that no matter how large K is, $\mathbb{P}(T > K) > 0$, because every possible path of the process S_n on an interval of length K has a positive probability and there are certainly many paths of length K that do not hit 1. Could one still beat the fair game, if instead, one had to stop by a certain fixed time? Notice also, that the strategy T requires that one be able to absorb arbitrarily large debts before leaving the game; in waiting for S_n to hit 1 there is not limit to how negative S_n might become. Is it possible to still beat the game if instead one has a limited amount of capital? Optional stopping theory will provide answers to these questions. It will also provide methods for computing probabilities of winning and losing simple games based on random walk.

6.3.2 Simple properties of stopping times; hitting times

It was shown in the last section that the first time for a random walk to hit 1 is a stopping time. This time is a particular example of a class of commonly encountered stopping times called hitting times. These are random times of the form

$$T_A := \begin{cases} \min\{k \geq 0; Y_k \in A\}, & \text{if } \{k \geq 0; Y_k \in A\} \neq \emptyset; \\ \infty, & \text{otherwise,} \end{cases}$$

where $\{Y_n\}$ is a stochastic process. Since

$$\{T = n\} = \{Y_1 \notin A, \ldots, Y_{n-1} \notin A, Y_n \in A\} \in \mathcal{F}_n^Y,$$

it is clear that T_A is an $\{\mathcal{F}_n^Y\}$-stopping times, and *a fortiori* a stopping time with respect to any filtration with respect to which $\{Y_n\}$ is adapted.

The next result is a simple lemma about closure of the stopping time property under simple operations. The proof is left to as an exercise.

Lemma 3.1 Let S and T be $\{\mathcal{F}_n\}$-stopping times. Then $S \wedge T := \min\{S, T\}$, $S \vee T := \max\{S, T\}$, and $S + T$ are also stopping times.

6.3.3 Optional Stopping Theorems

An optional stopping theorem is a result that extends martingale properties from deterministic times to stopping times. The aim of this section is to state the basic theorems and to develop some important applications. All results shall be stated for martingales or submartingales, and can be easily translated into facts about supermartingales.

Optional stopping theory is based on the following simple observations. First, if T is a random time and if one lets $V_k := \mathbf{1}_{T \geq k}$, $k \geq 0$, then

$$X_{T \wedge n} = X_0 + \sum_{j=0} \mathbf{1}_{T \geq j}(X_j - X_{j-1} = X_0 + (V.X)_n .$$

That is, stopped processes are stochastic integrals, in the sense defined in section 2.

The second observation is that if T is a stopping time then the process $V_k := \mathbf{1}_{\{T \geq k\}}$, $k \geq 0$, is $\{\mathcal{F}_n\}$-predictable. Indeed,

$$\{T \geq k\} = \{T \leq k-1\}^c \quad \text{is } \mathcal{F}_{k-1}\text{-measurable for every } k \geq 0.$$

The following is then an immediate consequence of Theorem 2.1 and Proposition 1.1.

Theorem 3.1 Let $X = \{X_n\}_{n \geq 0}$ be an $\{\mathcal{F}_n\}$-martingale (respectively, submartingale), and let T be an $\{\mathcal{F}_n\}$-stopping time. Then $\{X_{T \wedge n}\}$ is also an $\{\mathcal{F}_n\}$-martingale (respectively, submartingale). Thus, for every n,

$$E[X_{T \wedge n}] = E[X_0] \quad (\text{respectively} \quad E[X_{T \wedge n}] \geq E[X_0]) . \tag{2}$$

This theorem already answers the question raised in section 2 about whether it is possible to beat a fair game with a stopping strategy bounded in time. If X is a martingale, if T is a stopping time, and if K is a constant such that $I\!P(T \le K) = 1$, then by takin $n \ge K$ in (2), one finds $E[X_T] = E[X_0]$, so no bounded strategy can increases the expected payoff.

To go beyond bounded stopping times for optional stopping, it is necessary to seek conditions under which the limit as $n \to \infty$ can be taken in (2).

Theorem 3.2 (Optional Stopping Theorem Let $\{X_n\}_{n\ge 0}$ be an $\{\mathcal{F}_n\}$-martingale (respectively, submartingale), and let T be an $\{\mathcal{F}_n\}$-stopping time. Suppose

(i) $I\!P\,(T < \infty) = 1$;

(ii) $E\,[|X_T|] < \infty$; and

(iii) $\liminf_n E\left[|X_n|\mathbf{1}_{\{T>n\}}\right] = 0$.

Then

$$E\,[X_T] = E\,[X_0] \quad \text{(respectively,} \quad E\,[X_T] \ge E\,[X_0]\,.) \tag{3}$$

Proof: Assume $\{X_n\}_{n\ge 0}$ is a martingale. The submartingale case has a similar proof. Assumption (i) implies that X_T is well-defined, almost surely. From equation (2) of Theorem 3.1,

$$E\,[X_0] = E\,[X_{T\wedge n}] = E\left[X_T\mathbf{1}_{\{T\le n\}} + X_n\mathbf{1}_{\{T>n\}}\right]. \tag{4}$$

Now, assumption (ii) allows an application of the dominated convergence theorem to obtain:

$$\lim_n E\left[X_T\mathbf{1}_{\{T\le n\}}\right] = E\,[X_T]\,. \tag{5}$$

On the other hand, by assumption (iii), there exists a subsequence $\{n_k\}$ for which

$$\lim_k E\left[X_{n_k}\mathbf{1}_{\{T>n_k\}}\right] = 0. \tag{6}$$

Take limits on the right-hand side of (4) along the subsequence $\{n_k\}$ using (5) and (6). The result is

$$E\,[X_0] = \lim_k E\,[X_{T\wedge n_k}] = E\,[X_T]\,. \qquad \diamond$$

The proof of Theorem 3.2 contains the following result, which is worth stating separately for later use.

Corollary 3.1 Let the process $\{X_n\}_{n\ge 0}$ and the random time T (which need not be a stopping time) satisfy conditions (i)—(iii) of Theorem 3.2. Then, there is a subsequence $\{n_k\}$ for which $\lim_k E\,[X_{T\wedge n_k}] = E[X_T]$.

The remainder of this subsection treats applications of optional stopping to random walks. First, some general theorems are proved; then, the optional stopping method is used to calculate hitting probabilities and expected exit times for Bernoulli random walks. Throughout the discussion, $S := \{S_n\}$ will denote a random walk with independent, identically distributed steps; i.e.,

$$S_n = S_0 + \sum_1^n \xi_k$$

where $\xi_1, \xi_2, \dots$ is a sequence of i.i.d. random variables independent of S_0. Because $S_n - S_{n-1} = \xi_n$, the random variables ξ_k, $k \geq 1$, are called the *increments* of S. When the increments are integrable, μ shall denote their common mean ($\mu = E[\xi_k]$) and, when the increments are square integrable, σ^2 shall denote their common variance. It was seen in section 1 that if S_0 is integrable and if ξ is integrable with mean 0, then S_n is a martingale with respect to the filtration $\{\mathcal{F}_n^S\}$, where

$$\mathcal{F}_n^S = \sigma\left\{S_0, \xi_1, \dots, X_n\right\}, \qquad n \geq 1.$$

To reiterate, the martinale property follows simply from the facts that $\mu = 0$ and that ξ_{n+1} is independent of $\mathcal{F}_n^S$ for every $n \geq 1$. More generally,

$$S_n - n\mu - S_0 = \sum_1^n (\xi_k - \mu)$$

is a martingale, since it is a sum of mean zero increments $\xi_k - \mu$. The following observation is also useful.

Lemma 3.1 Assume that $\{\xi_k\}$ is a sequence of i.i.d. random variables with mean $\mu = 0$ and finite variance σ^2. Then $\{(S_n - S_0)^2 - n\sigma^2\}$ is an $\{\mathcal{F}_n^S\}$-martingale.

Proof: Let $S_n' = S_n - S_0 = \sum_1^n \xi_k$. Then,

$$\left(S_{n+1}'\right)^2 - (n+1)\sigma^2 = \left(S_n'\right)^2 - n\sigma^2 + 2S_n'\xi_{n+1} + \xi_{n+1}^2 - \sigma^2. \tag{7}$$

Since, S_n' is $\mathcal{F}_n^S$-measurable, and since ξ_{n+1} is independent of $\mathcal{F}_n^S$ and has mean 0,

$$E\left[S_n'\xi_{n+1}/\mathcal{F}_n^S\right] = S_n' E\left[\xi_{n+1}/\mathcal{F}_n^S\right] = S_n' E[\xi_{n+1}] = 0.$$

Also, $E\left[\xi_{n+1}^2/\mathcal{F}_n^S\right] = E\left[\xi_{n+1}^2\right] = \sigma^2$. Thus, taking conditional expectations on both sides of (7),

$$E\left[\left(S_{n+1}'\right)^2 - (n+1)\sigma^2/\mathcal{F}_n^S\right] = \left(S_n'\right)^2 - n\sigma^2. \qquad \diamond$$

The next result gives simple conditions under which the optional stopping theorem holds for random walks.

Proposition 3.1 Let $\{S_n\}$ be a random walk with i.i.d. increments of finite mean and variance. Let T be an $\{\mathcal{F}_n^S\}$-stopping time. If

$$E[T] < \infty \tag{8}$$

and if $E[|S_0|] < \infty$, then

$$E\left[S_T\right] = E\left[S_0\right] + \mu E[T]. \tag{9}$$

Moreover, if (8) holds, then $E\left[(\sum_1^T \xi_k)^2\right] < \infty$.

Remarks Observe that for a nonrandom time N,

$$E[S_N] = E[S_0] + \mu N.$$

Equation (9) is a generalization of this identity to stopping times. The last statement of the theorem is an auxiliary fact that goes into the proof of (9).

Proof: It suffices to consider the case $\mu = 0$; The general case is derived from the mean 0 case by applying the mean 0 result to the martingale $\{S_n - n\mu - S_0\}$. Thus, assume $\mu = 0$.

Let $\{S'_n\}$ denote the zero mean martingale $\{S_n - S_0 = \sum_1^n \xi_k\}$. To prove (8), we apply the Optional Stopping Theorem to $\{S'_n\}$ and T. For this, it is only necessary to verify the conditions (i)—(iii) of the Optional Stopping Theorem. Condition (i), $I\!\!P(T < \infty) = 1$, is immediate from the assumption (8) that $E[T] < \infty$.

To prove condition (iii), apply the Cauchy-Schwartz inequality to find

$$E\left[\mid S'_n \mid \mathbf{1}_{\{T>n\}}\right] \leq \sqrt{E[S'_n)^2] I\!\!P(T > n)} = \sqrt{\sigma^2 n I\!\!P(T > n)}.$$

However,

$$n I\!\!P(T > n) \leq E\left[T \mathbf{1}_{\{T>n\}}\right],$$

and so, by application of dominated convergence, $\lim_n n I\!\!P(T > n) = 0$. Conditon (iii) follows.

To prove (ii), apply Theorem 3.1 to $(S'_n)^2 - n\sigma^2$, which is a martingale by Lemma 3.1. One finds that $0 = E\left[(S'_{T\wedge n})^2 - (T \wedge n)\sigma^2\right] = 0$ for all n. Thus,

$$\sigma^2 E\left[T \wedge n\right] = E\left[(S'_{T\wedge n})^2\right] \geq E\left[(S'_T)^2 \mathbf{1}_{\{T\leq n\}}\right]. \tag{10}$$

for $n \geq 0$. Let $n \to \infty$ and use monotone convergence on each side. Then $E\left[(S'_T)^2\right] \leq \sigma^2 E[T] < \infty$. Therefore $E\left[|S'_T|\right] < \infty$; (for a r.v. X, $(E[|X|])^2 \leq E[X^2]$ by Cauchy-Schwartz.) $\diamond$

Given a Borel subset A of the real line, recall that the time to hit A^c is

$$T_{A^c} := \min\{n\,;\, S_n \notin A\}.$$

This may also thought of as the first time to exit A. Our aim is to apply optional stopping to such exit times, and fortunately there is a simple sufficient condition for the integrability criterion (8) for these times.

Lemma 3.2 Assume that the i.i.d. random variables ξ_n, $n \geq 1$, are non-degenerate, that is $\mathbb{P}(\xi_k = 0) < 1$. For any bounded set A, $E\left[T_{A^c}\right] < \infty$.

Proof: Let $A \subset (a, b)$ where a and b are finite, and set $T = T_{(a,b)^c}$. Since $T_{A^c} \leq T$, it will suffice to prove $E[T] < \infty$.

For any real x, let $E_x[T]$ denote the expected value of T given $S_0 = x$. Because the steps ξ_n, $n \geq 1$ are independent of S_0,

$$E[T] = E\left[E\left[T/S_0\right]\right] = E\left[E_{S_0}\left[T\right]\right].$$

Therefore, to complete the proof it suffices to show

$$\sup_x E_x\left[T\right] < \infty.$$

Because of the assumed non-degeneracy of the steps, (and by replacing ξ_n by $-\xi_n$, $n \geq 1$, if necessary), it may be assumed that there is a $c > 0$ such that $\eta := \mathbb{P}(\xi_k \geq c) > 0$. Choose a positive integer N so large that $Nc \geq (a + b)$. We shall partition time into blocks of length N and build an event on each block that guarantees that the walk must leave (a, b) in that time block. This event, for block k, is

$$B_k := \left\{\xi_{kN+1} \geq c, \xi_{kN+2} \geq c, \ldots, \xi_{(k+1)N} \geq c\right\}.$$

Indeed, $B_k \subset \left\{S_{(k+1)N} - S_{kN} \geq A{+}b\right\}$, and so

$$B_k \subset \left\{kN \leq T \leq (k{+}1)N\right\}.$$

By independence of the increments, the events $B_1, B_2, \ldots$ are independent, and $\mathbb{P}(B_k) = \eta^N$ for each k. Thus

$$\mathbb{P}\left(T > (k{+}1)N\right) \leq \mathbb{P}\left(\bigcap_1^k B_j^c\right) = \left(1 - \eta^N\right)^k. \tag{11}$$

Recall that $E[Z] = \int_0^\infty \mathbb{P}(Z > z)\,dz \leq \sum_0^\infty \mathbb{P}(Z > k)$ for positive random variables Z. Thus, from (11),

$$N^{-1}E[T] \leq \sum_0^\infty \left(1 - \eta^N\right)^k < \infty. \qquad \diamond$$

Notice that no assumption on the mean or variance of ξ was made in Lemma 3.2.

In the proof of Theorem 3.2, optional stopping at $T \wedge n$ was used to obtain a bound on $E[S_T^2]$. When T is the exit time from a bounded set, this argument leads to an identity for the expected value of T, which is useful later in deriving explicit formulae for Bernoulli random walks.

Proposition 3.2 Let $\{S_n\}$ be a random walk with independent increments, let $S_0 = 0$, and suppose A is a bounded set. Then

$$E\left[S_{T_{A^c}}^2\right] = \sigma^2 E\left[T_{A^c}\right].$$

Proof: For convenience, let $\sigma^2 = 1$. From equation (9),

$$E\left[n \wedge T_{A^c}\right] = E\left[S_{T_{A^c}}^2 \mathbf{1}_{\{T_{A^c} \le n\}}\right] + E\left[S_n^2 \mathbf{1}_{\{T_{A^c} > n\}}\right]. \tag{12}$$

Since, $|S_n| \le K := \sup\{|x|\,;\, x \in A\} < \infty$ if $T_{A^c} > n$,

$$\lim_{n\to\infty} E\left[S_n^2 \mathbf{1}_{T_{A^c}}\right] \le K^2 \lim I\!\!P\left(T_{A^c} > n\right) = 0.$$

Therefore, one can let $n \to \infty$ on both sides of (12) to finish the proof. ◇

Application 1. The symmetric Bernoulli random walk. The simple, symmetric Bernoulli walk is the process $S_n = \sum_1^n \xi_k$, where $\{\xi_k\}$ is a sequence of independent, symmetric, Bernoulli random variables ($I\!\!P\left(\xi = -1\right) = I\!\!P\left(\xi = 1\right) = 1/2$.) Optional stopping is applied to solve the classical problems of determining ruin probabilites and expected exit times.

Let $-\infty < -a < 0 < b < \infty$, where a and b are positive integers, and let T_{ab} denote the first time that the walk hits either $-a$ or b. Because the walk takes unit steps, T_{ab} is the same as $T_{(-a,b)^c}$. Moreover, the event that the walk hits $-a$ before hitting b is the same as $\{S_{T_{ab}} = -a\}$.

We are interested in computing

$$r := I\!\!P\left(S_{T_{ab}} = -a\right) \qquad \text{and} \qquad E[T_{ab}].$$

Calculating r is the classical problem of gambler's ruin. Imagine that gambler A (Alice) has a dollars and gambler B (Bob) has b dollars. They play a fair game in which each player bets 1 dollar on every play and they agree to play until one of them is broke. If S_n represents Alice's accumulated gain or loss from n plays, then S_n is a sum of n i.i.d. symmetric, Bernoulli random variables, and, at time n, Alice has a total fortune of $S_n + a$ dollars. Thus $r = I\!\!P\left(S_{T_{ab}}\right) = -a$ is the probability that Alice loses, or the probability of her ruin.

The following preliminary remarks are required for the computation. By Lemma 3.2, $E[T_{ab}] < \infty$, and, in particular, $I\!\!P(T_{ab} < \infty) = 1$, so that the random walk cannot stay strictly between $-a$ and b forever. Thus, S_n must eventually hit either $-a$ or b, and hence, if does not hit $-a$ first, it must hit b at T_{ab}; that is

$$I\!\!P\left(S_{T_{ab}} = b\right) = 1 - I\!\!P\left(S_{T_{ab}} = -a\right) = 1 - r.$$

To find r, apply Theorem 3.3. Since $\mu = 0$,

$$0 = E\left[S_{T_{ab}}\right] = -a\mathbb{P}\left(S_{T_{ab}} = -a\right) + b\mathbb{P}\left(S_{T_{ab}} = b\right) = -r(a+b) + b.$$

Therefore,

$$r = \frac{b}{a+b}.$$

Now apply Theorem 3.4. Since $\sigma^2 = 1$, the expected length of the game is

$$E[T_{ab}] = E[S_{T_{ab}}^2] = a^2 r + b^2(1-r) = ab.$$

Application 2. Asymmetric Bernoulli random walk. In this example, we compute classical probabilities of ruin in the case of unfair games. Thus, S_n will now be a sum of i.i.d. Bernoulli random variables with $p := \mathbb{P}(\xi_k = 1)$, $q = 1 - p = \mathbb{P}(\xi_k = -1)$, and $p \neq q$. From example 6 of section 1, $Z_n := (q/p)^{S_n}$, $n \geq 0$ is a martingale. As in application 1, it is assumed that $S_0 = 0$.

Let a and b be positive integers and define T_{ab}, as before, to be the first time $\{S_n\}$ hits either $-a$ or b. Lemma 3.2 still applies and says that $E[T_{ab}] < \infty$. Since $|Z_n|$ is bounded by $\max\{(q/p)^a, (q/p)^b\}$ as long as $n \leq T_{ab}$,

$$E\left[|Z_{T_{ab}}|\right] < \infty \quad \text{and} \quad \lim_n E\left[Z_n \mathbf{1}_{\{T_{ab} > n\}}\right] = 0.$$

Therefore the conditions of the optional stopping theorem are valid for $\{Z_n\}$ and T_{ab}. Hence, noting that $Z_0 = (q/p)^{S_0} = 1$,

$$1 = E\left[\left(\frac{q}{p}\right)^{S_n}\right] = (q/p)^{-a}\mathbb{P}(S_{T_{ab}} = -a) + (q/p)^b \mathbb{P}(S_{T_{ab}} = b).$$

Since, as before, $\mathbb{P}(S_{T_{ab}} = b) = 1 - \mathbb{P}(S_{T_{ab}} = -a)$, we can solve to find the probability of ruin:

$$\mathbb{P}\left(S_{T_{ab}} = -a\right) = \frac{(q/p)^a - (q/p)^{a+b}}{1 - (q/p)^{a+b}}.$$

We can also apply Theorem 3.2 since $E[T_{ab}] < \infty$. Here $\mu = p - q$. Thus,

$$\begin{aligned}
E[T_{ab}] &= \frac{1}{p-q} E[S_{T_{ab}}] \\
&= \frac{1}{p-q}\left(b - (a+b)\mathbb{P}\left(S_{T_{ab}} = -a\right)\right) \\
&= \frac{1}{p-q}\left(\frac{b(1-(q/p)^a) - a(q/p)^a(1-(q/p)^b)}{1-(q/p)^{a+b}}\right).
\end{aligned}$$

Hitting times for Bernoulli random walks, continued. For a random walk $\{S_n\}$ on the integer lattice and an integer b, let $T_b := \min\{n\,;\, S_n = b\}$ with the understanding that $T_b = \infty$ if the random walk never hits b. Assume all random walks start from 0. We are now in a position to study these hitting times for the symmetric Bernoulli walk.

Proposition 3.3 With probability one, the symmetric Bernoulli random walk eventually hits every integer. That is,

$$\mathbb{P}\left(\bigcap_{-\infty}^{\infty}\{T_k < \infty\}\right) = 1.$$

Proof: It suffices to prove that $\mathbb{P}(T_b < \infty) = 1$ for each integer b, and, by symmetry, it may be assumed that b is positive. However, $T_b < \infty$, if and only if there exist some positive integer a such that $\{S_n\}$ hits b before it hits $-a$; in other words,

$$\{T_b < \infty\} = \bigcup_{a=1} 1^{\infty}\{S_{T_{ab}} = b\}$$

The events in the union increase with increasing a, because if the walk hits b before it hits $-a$, then it certainly hits b before hitting $-(a+1)$. Thus

$$\mathbb{P}(T_b < \infty) = \lim_{a\to\infty} \mathbb{P}(S_{T_{ab}} = b) = 1. \tag{13}$$

But we know from computing the probability of ruin above that $\mathbb{P}(S_{T_{ab}} = b) = a/(a+b)$. Taking the limit in (13) as $a \to \infty$ then shows that $\mathbb{P}(T_b < \infty) = 1$. ◇

Consider the stopping time T_1, the first time to hit 1. Since we now know this to be finite a.s., T_1 provides a strategy to beat a fair game, as discussed in section 1. However, because the optional stopping identity, $E[S_T] = E[S_0]$ is *not* true for T_1, it follows from Theorem 2 that,

$$E[T_1] = \infty;$$

for the symmetric Bernoulli random walk; that is, the expected amount of time one needs to get to 1 is infinite.

Consider now the case of the asymmetric Bernoulli walk.

Proposition 3.5 Let $\{S_n = \sum_1^n \xi_k\}$ be the asymmetric random walk with $p = \mathbb{P}(\xi=1) < q = \mathbb{P}(\xi=-1)$. Then for positive integers b,

$$\mathbb{P}(T_b < \infty) = \left(\frac{p}{q}\right)^b < 1.$$

Proof: From the calculation of the ruin probability in asymmetric random walk, one finds after a bit of algebra that

$$\mathbb{P}(S_{T_{ab}} = b) = \frac{(p/q)^b - (p/q)^{a+b}}{1-(p/q)^{a+b}}.$$

Since $p < q$, this tends to $(p/q)^b$ as $a \to \infty$. Hence, using (13) again,

$$\mathbb{P}(T_b < \infty) = \lim_{a\to\infty} \mathbb{P}(S_{T_{ab}} = b) = (p/q)^b. \qquad \diamond$$

Application 3. Probability of absorbtion for a Markov chain.

The martingale techniques applied to random walk extend to more general Markov chains. To illustrate, consider the Markov chain defined at the end of section 1. In this chain, modelling a generalized random walk on the non-negative integers, the transition probabilities out of state i for $i \geq 1$ are given by

$$p_{ij} = \begin{cases} p_i & , \text{ if } j = i+1; \\ q_i & , \text{ if } j = i-1; \\ 0 & , \text{ otherwise,} \end{cases}$$

As with Bernoulli random walk, these transition probabilities allow only one step up or down at each time, but the step probabilities now depend on the state and so the steps are no longer identically distributed. If any one of the probabilities p_i or q_i is zero it creates a barrier to either upward or downward movement from state i. To avoid this *it is assumed* that p_i and q_i are strictly positive for all $i \geq 1$.

Since the state space in this model is the non-negative integers, the chain cannot take a step in the negative direction from state 0, and other transition probabilities out of state 0 need to be specified. In this discussion, we will be concerned only with computing the probability that the chain eventually hits state 0 given that it starts in state i. For any $i \geq 1$ this probability will not depend on what the transition probabilities out of state 0 are, and so it will be convenient to assume that state 0 is absorbing, that is, that $p_{00} = 1$ and $p_{0j} = 0$ for every $j \geq 1$. When this is the case equation (12) of section 1 identifies all P-harmonic functions for this chain.

Let $\{Z_n\}$ denote a Markov chain with transition matrix P. The notation $\mathbb{P}_i(A)$ indicates the probability of A when $Z_0 = i$ almost surely. Let T_0 be the first time that the chain hits the state 0. Then $\mathbb{P}_i(T_0 < \infty)$ is the probability that the chain hits state 0 if it starts from i.

Propostion 3.6 If $\displaystyle\sum_{j=1}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_1} = \infty$ then

$$\mathbb{P}_i(T_0 < \infty) = 1 \qquad \text{for every } i \geq 1. \tag{14}$$

But if $\displaystyle\sum_{j=1}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_1} < \infty$, then

$$\mathbb{P}_i(T_0 < \infty) = \frac{\sum_{j=i}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_1}}{\sum_{j=1}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_1} = \infty} \qquad i \geq 1. \tag{15}$$

Proof: Assume that $Z_0 = i$. Let T_N be the first time that the chain hits state N. Then if $0 < i < N$, the stopping time $T_0 \wedge T_N$ is the first time that the chain hits either 0 or N, that is, the first time it exits $[1, N-1]$. Because p_i and $q_i[S1]$ are assumed strictly positive for all i, it can be shown much as in Lemma 3.2, that $E_i[T_0 \wedge T_N] < \infty$ and hence $\mathbb{P}_i(T_0 \wedge T_N < \infty) = 1$ for all $N > i$; proof is left to the reader. As a consequence,

$$\mathbb{P}_i(T_N < T_0) = 1 - \mathbb{P}_i(T_0 < T_N). \tag{16}$$

Since only unit steps of the process are allowed, it is also true that

$$\{T_0 < \infty\} = \bigcup_{N>i} \{T_0 < T_N\}.$$

Thus,

$$\mathbb{P}_i\left(T_0 < \infty\right) = \lim_{N\to\infty} \mathbb{P}_i\left(T_0 < T_N\right). \tag{17}$$

Now proceed much as in the calculation of ruin probabilities for simple random walk. By setting $\alpha = 1$ and $\beta = 0$ in equation (12) of section 1, observe that

$$f(0) = 1, \quad f(1) = 0, \quad f(i) = 1 - \sum_{j=1}^{i-1} \frac{q_j \cdots q_1}{p_j \cdots p_i}$$

defines a P-harmonic function. Therefore $\{f(Z_n)\}$ is a martingale and so, since $Z_0 = i$, the optional stopping theorem implies

$$f(i) = E\left[f(Z_{T_0 \wedge T_N})\right]. \tag{18}$$

Of course, the conditions of the optimal stopping theorem must be verified, but this is easy once it is known that $\mathbb{P}_i(T_0 \wedge T_N < \infty) = 1$; the proof is similar in spirit to that used above in treating asymmetric random walk.

Next evaluate the right-hand side of (18), using (16), to find

$$f(i) = \mathbb{P}_i(T_0 < T_n)\,(1 - f(N)) + f(N),$$

and hence, from (17),

$$\mathbb{P}_i\left(T_0 < \infty\right) = \lim_{N\to\infty} \frac{f(N) - f(i)}{f(N) - 1}.$$

The identities (14) and (15) now follow from the definition of $f(i)$. ◇

6.3.4 Optional stopping: some refinements

The sequence of expected values $\{E[X_n]\}$ of a submartingale $\{X_n\}_{n\geq 0}$ is a non-decreasing sequence. It is natural to expect that this property extend to stopping times, as we shall now explain, beginning with an extension of Theorem 3.1. Let

$\{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-submartingale, and let S and T be $\{\mathcal{F}_n\}$-stopping times such that $S \leq T$, a.s. Define the process

$$V_k = \mathbf{1}_{\{S<k\leq T\}} \qquad k \geq 1.$$

It is easy to check that V is $\{\mathcal{F}_n\}$-predictable and that $(V.X)_n = X_{T\wedge n} - X_{S\wedge n}$. Therefore it follows from Theorem 6.5 that

$$(19) \qquad X_{T\wedge n} - X_{S\wedge n} \quad \text{is a submartingale and} \quad E[X_{S_n}] \leq E[X_{T\wedge n}] \ \ \forall n.$$

To remove n in the last inequality, it suffices to impose the conditions of Theorem 3.2 on T.

Theorem 3.4 (Optional stopping, part (ii)) Let $\{X_n\}_{n\geq 0}$ be a submartingale and let S and T be two stopping times with $S \leq T$, a.s. Assume that T satisfies conditions (i)—(iii) of Theorem 3.2. Then S satisfies (i)—(iii) and

$$(20) \qquad E[X_S] \leq E[X_T].$$

(When $\{X_n\}_{n\geq 0}$ is a martingale, then equality holds in (20).)

Proof: The main step is proving that S also satisfies (i)—(iii) of Theorem 3.2. Once this has been done, then, by Corollary 3.1, there exists a subsequence $\{n'\}$ such that $\lim_{n'} E[X_{T\wedge n'}] = E[X_T]$ and $\lim_{n'} E[X_{S\wedge n'}] = E[X_S]$. Equation (20) then follows by taking limits along the subsequence $\{n'\}$ in (19).

It remains to prove S satisfies (i)—(iii). Condition (i) is immediate from the assumption $I\!P(T < \infty) = 1$ since $S \leq T$ a.s. Condition (iii) for S is also easy. Because $\{S > n\} \subset \{T > n\}$,

$$E\left[|X_n|\mathbf{1}_{\{S>n\}}\right] \leq E\left[|X_n|\mathbf{1}_{\{T>n\}}\right],$$

and hence the assumption of property (iii) for T implies that for S.

The proof of (ii) is slightly more involved. First notice that since the function $x \to (x)^+ := 0 \wedge x$ is increasing and convex, Proposition 1.3 implies that $\{(X_n)^+\}$ is a submartingale. Thus, by (19),

$$E\left[(X_{S\wedge n})^+\right] \leq E\left[(X_{T\wedge n})^+\right].$$

for all n. Secondly, since $\{X_{S\wedge n}\}$ is a submartingale, $E[X_0] \leq E[(X_{S\wedge n})^+] - E[(X_{S\wedge n})^-]$. Thus $E[(X_{S\wedge n})^-] \leq E[(X_{S\wedge n})^+] - E[X_0]$, and hence

$$(21) \qquad E\left[X_{S\wedge n}|\right] = E[(X_{S\wedge n})^-] + E[(X_{S\wedge n})^+] \leq 2E\left[|X_{T\wedge n}|\right] - E[X_0].$$

By Corollary 3.1 applied to the submartingale $\{(X_n)^+\}$, there is a subsequence $\{n_k\}$ such that $\lim_k E[(X_{T\wedge n_k})+] = E[(X_T)^+]$. Therefore, by taking limits in (21) along the subsequence $\{n_k\}$ and applying Fatou's lemma,

$$E\left[|X_S|\right] \leq E\left[(X_T)^+\right] - E\left[X_0\right] \leq E[|X_T|] - E[X_0] < \infty. \qquad \textit{diamond}$$

It is also possible to extend the martingale property to stopping times. The idea is to define for any stopping time T, a σ-algebra $\mathcal{F}_T$ of events that occur by time T. Then if S and T are stopping time satisfying the conditions of Theorem 3.4, and if $\{X_n\}_{n\geq 0}$ is a martingale (submartingale) one proves

$$X_S = E\left[X_T/\mathcal{F}_S\right] \qquad \left(X_S \leq E\left[X_T/\mathcal{F}_S\right]\right). \tag{22}$$

This extension is very important in advanced martingale theory, but we shall not make further use of it, so the reader may skip this topic on a first reading.

How does one define the class of events of a filtration $\{\mathcal{F}_n\}$ that occur up to or before a stopping time? Let A be any event and consider the disjoint partition $A = \cup_{n=0}^{\infty} (A \cap \{T = n\})$. Then we should require that for each n, $A \cap \{T = n\}$ depends only on the information known by time n. The next definition gives this idea a more general looking, but equivalent, formulation.

Definition Given a filtration $\{\mathcal{F}_n\}$ and an $\{\mathcal{F}_n\}$-stopping time T,

$$\mathcal{F}_T := \{A \in \mathcal{F}\,;\, A \cap \{T \leq n\} \in \mathcal{F}_n \ \ \forall n \geq 0\}\,.$$

Here is a simple example. Let B be any set of non-negative integers. Then $\{T \in B\}$ is an event in $\mathcal{F}_T$. Doing the next exercise will help the reader understand $\mathcal{F}_T$ better.

Exercise. (a) $\mathcal{F}_T$ is a σ-algebra .
(b) An event $A \in \mathcal{F}_T$ if and only if $A \cap \{T = n\} \in \mathcal{F}_n$ for every nonnegative integer n.
(c) If $\{X_n\}_{n\geq 0}$ is $\{\mathcal{F}_n\}$-adapted and if T is a stopping time, then X_T is $\mathcal{F}_T$-measurable.
(d) Let $\{\mathcal{F}_n^X\}$ be the filtration generated by $\{X_n\}_{n\geq 0}$. Then $\mathcal{F}_T^X$ is the σ-algebra generated by the random variables $\{X_{T\wedge n}\}_{n\geq 0}$.

Theorem 3.5 Let $\{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-martingale (submartingale) and suppose that S and T are $\{\mathcal{F}_n\}$-stopping times that satisfy the assumptions of Theorem 3.4. Then (22) holds.

(Proof to be added.)

6.4 Martingale Moment Inequalities

The version of the Optional Stopping Theorem expressed in Theorem 3.4, which says that $E[X_S] \leq E[X_T]$ when X is a submartingale and the stopping time S is less than the stopping time T, looks innocent, but has profound and remarkable consequences for the probabilistic structure of martingales. In particular, one can bound the moments and probability tails of maximum of a submartingale up to time N by moments of the process at time N; in other words, the average behavior of the process over a time interval is constrained by its average behavior at the terminal time. The most important of these martingale inequalities are developed in this section.

Actually, we have already seen an example of a martingale inequality earlier in these notes, namely, Kolmogorov's inequality, which we proved in Chapter 3 as part of the proof of Kolmogorov's three series theorem. Kolmogorov's inequality shows how to bound tail probabilities for the *maximum* of a mean 0 random walk up to time N in terms of the second moment of the value of the walk *at* time N. If you examine the proof closely, you will see that it is really a stopping time argument. Kolmogorov's inequality is in fact a prototype of the martingale inequalities developed here, and it is worthwhile to reread its proof after studying the material of this section.

Throughout the discussion $\{X_n\}_{n\geq 0}$ will be a submartingale, unless otherwise mentioned. We shall study the auxiliary maximum process,

$$X_n^* := \max_{0\leq k\leq n} X_k.$$

Let T be a stopping time satisfying conditions (i)—(iii) of the optional stopping theorem. Our first goal is to bound the probability of the event

$$\{X_T^* \geq \lambda\}$$

for a positive level λ. The idea is to consider the first time before T, if any, at which the maximum process X^* exceeds λ, and accordingly, we define

$$S := T \wedge \min\{k\,;\, X_k \geq \lambda\}.$$

Then from Theorem 3.5,

$$E[X_S] \leq E[X_T]. \tag{1}$$

What are the consequences of (1)? Observe first that

$$\{S<T\} = \left\{X_{T-1}^* \geq \lambda\right\} \quad \text{and} \quad \{S{=}T\} = \{S{=}T, X_T \geq \lambda\} \cup \{X_T^* < \lambda\}.$$

Observe also that $X_S \geq \lambda$ on $\{S<T\} \cup \{S=T, X_T \geq \lambda\}$. Therefore,

$$\begin{aligned} E\left[S_X\right] &= E\left[X_S \mathbf{1}_{\{S<T\}} + X_T \mathbf{1}_{\{S=T, X_T\geq\lambda\}} + X_S \mathbf{1}_{\{X_T^*<\lambda\}}\right] \\ &\geq \lambda I\!P\left(X_T^* \geq \lambda\right) + E\left[X_T \mathbf{1}_{\{X_T^*<\lambda\}}\right]. \end{aligned}$$

Thus, from (1),

$$\lambda I\!P\left(X_T^* \geq \lambda\right) \leq E\left[X_T \mathbf{1}_{\{X_T^*\geq\lambda\}}\right]. \tag{2}$$

Note how much stronger this is than Markov's inequality! This inequality and its major consequences are summarized in the next theorem.

Theorem 4.1. (Doob's martingale inequalities) Let $\{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-submartingale and let T be a stopping time satisfying (i)—(iii) of the optional stopping theorem.
(a) For $\lambda > 0$,

$$\mathbb{P}\left(X_T^* \geq \lambda\right) \leq \frac{E\left[X_T \mathbf{1}_{\{X_T^* \geq \lambda\}}\right]}{\lambda} \leq \frac{E[X_T^+]}{\lambda}. \tag{3}$$

(b) If $\{X_n\}_{n\geq 0}$ is a martingale and $\lambda > 0$, then

$$\begin{aligned} \mathbb{P}\left(\max_{0\leq k\leq T} \mid X_k \mid \geq \lambda\right) &\leq \frac{E[X_T|]}{\lambda} \\ \mathbb{P}\left(\max_{0\leq k\leq T} X_k \leq -\lambda\right) &\leq \frac{E[X_T^-]}{\lambda} \end{aligned} \tag{4}$$

(c) If $\{X_n\}_{n\geq 0}$ is a *non-negative* submartingale and $p > 1$, then

$$E\left[(X_T^*)^p\right] \leq \left(\frac{p}{p-1}\right)^p E\left[X_T^p\right]. \tag{5}$$

Remark. If $\{X_n\}_{n\geq 0}$ is a martingale, then, by Proposition , $\{|X_n|\}$ is a submartingale, and hence by (5),

$$E\left[\max_{0\leq k\leq T} \mid X_k \mid^p\right] \leq \left(\frac{p}{p-1}\right)^p E\left[|X_T|^p\right]. \tag{6}$$

Proof: Part (a) follows directly from (2).

The first inequality of (b) is a consequence of (a) applied to the submartingale $\{|X_n|\}$. The second inequality of (b) is derived by application of (a) to the martingale $\{-X_n\}$.

To prove (c), recall the following formula for positive random variables Z,

$$E[Z^p] = \int_0^\infty \mathbb{P}(Z^p \geq z)\, dz = \int_0^\infty \mathbb{P}(Z \geq z) p z^{p-1}\, dz.$$

By using (a) in this formula,

$$\begin{aligned} E\left[(X_T^*)^p\right] &\leq \int_0^\infty E[X_T \mathbf{1}_{\{X_T^* \geq z\}}] p z^{p-2}\, dz = E[X_T \int_0^{X_T^*} p z^{p-2}\, dz] \\ &= (\frac{p}{p-1}) E[X_T (X_T^*)^{p-1}] \\ &\leq (\frac{p}{p-1}) (E[X_T^p])^{1/p} \left(E[(X_T^*)^p]\right)^{p/p-1}. \end{aligned}$$

Inequality (5) now follows by some simple algebra. ⋄

The following corollary generalizes Kolmogorov's inequality, which now becomes a special case.

Corollary 4.1. Let $\{X_n\}_{n\geq 0}$ be a square-integrable martingale. Then for any $N \geq 1$,

$$I\!\!P\left(\max_{0\leq k\leq N} \mid X_k \mid\geq \lambda\right) \leq \frac{E[X_n^2]}{\lambda^2}.$$

Proof: Apply (3) of Theorem 3.5 to the submartingale $\{X_n^2\}$. ⋄.

6.5 Doob's Upcrossing Inequality

A picturesque way to introduce the upcrossing inequality is by consideration of a "buy low, sell high" investment strategy in a stock. Imagine that a process $\{X_n\}_{n\geq 0}$ models the successive prices of a stock with several endearing characteristics. First, $\{X_n\}_{n\geq 0}$ is a submartingale—we expect to earn money. Second, the stock never trades below a level a—perhaps whenever it threatens to go below a a fairy godmother steps in and bids the price up to a. Finally, the stock trades without transaction costs. A broker enters the market with no shares and decides to trade this stock using the classic "buy low, sell high" strategy. He fixes a level b greater than a in advance, buys a share of the stock when it hits level a and holds onto it until it the next time it rises to level b or above, at which time he immediately sells. He continues to trade in this fashion, buying at a and selling at b, making a profit of at least $b - a$ every time, for the duration of trading.

This is a nice market for the broker—he can't lose money because the price never goes below a, and on the average he expects to gain—and so he might be feeling pretty smug about his strategy. However, we onlookers, as initiates into the secrets of martingale theory, are even more insufferably smug. We know that the favorability of a game of chance cannot be improved by a predictable strategy and "buy low, sell high" is predictable. Therefore we suspect that on the average our broker will do at least as well by just buying a share of stock at time zero and holding onto it! Let us see that this is true.

The first step is to represent the broker's earnings process as a stochastic integral (sum) in the sense of section 2. To this end define the process V by

$$V_k = \begin{cases} 1, & \text{if the broker holds a share during period from } k-1 \text{ to } k; \\ 0, & \text{otherwise.} \end{cases}$$

Then the total earnings of the broker by time n are given by $(V.X)_n$, in the notation of section 2. The process V is clearly predictable because V_k depends only on whether the stock was held in the previous period and what its price at time $k-1$ was. Hence $\{(V.X)_n\}$ is a submartingale. The following lemma will clinch our hunch about whether "buy low, sell high" beats "sit and wait."

Lemma 5.1 Let $\{X_n\}_{n\geq 0}$ be an $\{\mathcal{F}_n\}$-martingale and let $\{V_k\}$ be a bounded, $\{\mathcal{F}_n\}$-predictable process and K a positive constant such that $V_n \leq K$ a.s. for all n. Then

$$E\left[(V.X)_n\right] \leq KE\left[X_n - X_0\right]. \tag{1}$$

Proof: For each k, $1 \leq k \leq n$,

$$\begin{aligned} E\left[V_k(X_k - X_{k-1})\right] &= E\left[V_k E\left[X_k - X_{k-1}/\mathcal{F}_k\right]\right] \leq KE\left[E\left[X_k - X_{k-1}/\mathcal{F}_k\right]\right] \\ &= KE\left[X_k - X_{k-1}\right] \end{aligned}$$

The first equality is due to the predictability of V, and the middle inequality to the bound on V_k and the submartingale property of $\{X_n\}_{n\geq 0}$, which implies that $0 \leq E\left[E\left[X_k - X_{k-1}/\mathcal{F}_k\right]\right]$ a.s. Summing up gives

$$E\left[(V.X)_n\right] = E\left[\sum_1^n V_k(X_k - X_{k-1})\right] \leq KE\left[\sum_1^n (X_k - X_{k-1})\right] = KE\left[X_n - X_0\right]. \quad \diamond$$

Return now to our hypothetical broker. For him V_k is bounded by 1 always, and hence, by (1),

$$E[(V.X)_n] \leq E[X_n - X_0], \tag{2}$$

and the right-hand side is just the expected profit gained by buying one share at time 0 for price X_0 and selling it at time n for its price X_n.

Inequality (2) is simple, but contains hidden depths. From it one can derive an inequality that bounds the fluctations of the submartingale X. To explain this, we introduce the concept of an upcrossing. Let $\underline{x} = (x_0, x_1, \ldots)$ be a sequence of real numbers. The restriction $(x_k, x_{k+1}, \ldots, x_m)$ of $\underline{x}$ to the finite interval of indices $[k, m]$ is called an *upcrossing* of the interval $[a, b]$ if

$$x_k \leq a, \qquad a < x_j < b \quad \text{for } k < j < m, \qquad \text{and} \;\; x_m \geq b.$$

To count upcrossings, we use the notation

$$U_n^{a,b}(\underline{x}) := \text{ number of upcrossings of } \;[a, b]\; \text{ in } \;(x_0, \ldots, x_n).$$

Returning again to the broker, buying at a and selling when the price rises above b, notice that he makes a profit of at least $b - a$ dollars on every upcrossing of $[a, b]$ completed by the stock price process $\{X_n\}_{n\geq 0}$. In symbols,

$$(b - a)U_n^{a,b}(\{X_k\}) \leq (V.X)_n$$

By taking expectations on both sides and using (2) we have proved the following upcrossing inequality.

Lemma 5.2 Let $\{X_n\}_{n\geq 0}$ be a submartingale such that $X_n \geq a$ a.s. for all n. For any $b > a$,

$$E\left[U_n^{a,b}(\{X_k\})\right] \leq \frac{E[X_n - X_0]}{b-a}. \tag{3}$$

The general martingale upcrossing inequality follows from this lemma.

Theorem 5.1 Let $\{X_n\}_{n\geq 0}$ be a submartingale. Then for any $n \geq 0$ and $b > a$,

$$E\left[U_n^{a,b}(\{X_k\})\right] \leq \frac{E[(X_n-a)^+ - (X_0-a)^+]}{b-a} \tag{4}$$

Proof: Let $Z_n := (X_n - a)^+$, $n \geq 01$. The function $x \to (x-a)^+$ is non-decreasing and convex and hence, from Proposition 1.3, $\{Z_n\}$ is a submartingale. Moreover, $\{X_n\}$ makes an upcrossing of $[a, b]$ if and only if $\{Z_n\}$ makes an upcrossing of $[0, b–a]$. Inequality (32) then follows by applying Lemma 5.2 to $\{Z_n\}$ with a replaced by 0 and b replaced by $b-a$. $\diamond$

The upcrossing inequalities are due ultimately to Doob. They are important in deriving martingale convergence theorems, which are taken up in the next section.

6.6 Martingale Convergence Theorems

Martingale convergence theorems address the question of existence and properties of $\lim_n X_n$ when $\{X_n\}_{n\geq 0}$ is a (sub)(super)martingale. Examples 2, 3 and 4 of section 1 already show that this is an interesting issue. They concern processes of the form $\{E\,[Z/\mathcal{F}_n]\}$ in the cases

(i) $\{\mathcal{F}_n\}$ is a filtration and hence $\{E\,[Z/\mathcal{F}_n]\}$ is a martingale (examples 2 and 3), and;

(ii) $\{\mathcal{F}_n\}$ is a decreasing sequence of σ-algebras, and hence $\{E\,[Z/\mathcal{F}_n]\}$ is a reverse martingale (example 4).

How do these processes behave as $n \to \infty$? The answer in the first case is Lévy's upward theorem: $\lim_n E\,[Z/\mathcal{F}_n] = E\,[Z/\mathcal{F}_\infty]$, both a.s. and in the L^1-sense, where $\mathcal{F}_\infty$ is, in essence, the upward limit of the filtration, that is, the smallest σ-algebra containing $\mathcal{F}_n$ for all n. In the second case of a reverse martingale, the result is $\lim_n E\,[Z/\mathcal{F}_n] = E\,[Z/\bigcap F_n]$, both a.s. and in L^1, and this is similar in spirit to the martingale case, because $\bigcap_n \mathcal{F}_n$ may be thought of as the downward limit of a sequence of decreasing σ-algebras. These results will be proved later on as consequences of more general theorems that identify conditions under which martingales converge.

Of course, a martingale need not converge. Any mean 0 random walk with non-degenerate, i.i.d. increments is a martingale, but almost-surely it does not converge. (The proof is left as an exercise of medium difficulty).

6.6.1 Almost-sure convergence of martingales

The key to martingale convergence theorems is the upcrossing inequality. Recall from section 6.5 that if $\underline{x} = (x_0, x_1, \ldots)$ is a sequence of numbers, the number of upcrossings of $[a,b]$ contained in the finite portion $(x_0, x_1, \ldots, x_n)$ of $\underline{x}$ is denoted by $U_n^{[a,b]}(\underline{x})$. The number of upcrossings of $[a,b]$ made by the infinite sequence $\underline{x}$ is then

$$U_\infty^{[a,b]}(\underline{x}) := \lim_{n\to\infty} U_n^{[a,b]}(\underline{x}).$$

Upcrossings enter into convergence theory because of the following fact. Given a sequence $\underline{x}$, $\liminf_n x_n < \limsup_n x_n$ *if and only if* there exist finite numbers $a < b$, which we may assume to be rational, such that $U_\infty^{[a,b]}(\underline{x}) = \infty$. (Here, either or both of $\liminf_n x_n$ and $\limsup_n x_n$ are allowed to be ∞ or $-\infty$.) Therefore, if it is known of $\underline{x}$ that $U_\infty^{[a,b]}(\underline{x}) < \infty$ for all rational a and b such that $a < b$, then it must follow that $\lim x_n$ exists, although this limit might be ∞ or $-\infty$. Since the upcrossing inequality gives a way to control the upcrossings of a submartingale, it leads to a convergence criterion.

Theorem 6.1 (Basic martingale convergence theorem)
Let $\{X_n\}_{n\geq 0}$ be a submartingale. If

(1) $$\sup_n E\left[X_n^+\right] < \infty,$$

then $X_\infty = \lim_n X_n$ exists almost surely, and $E\left[|X_\infty|\right] < \sup_n E[|X_n|] < \infty$.

Proof: Let $a < b$ be arbitrary real numbers. By the upcrossing inequality in Theorem 5.1, the inequality $(x-a)^+ \leq x^+ + |a|$, and the monotone convergence theorem,

$$E\left[U_\infty^{[a,b]}(\{X_k\})\right] = \lim_n E\left[U_n^{[a,b]}(\{X_k\})\right] \leq \limsup_n \frac{E[X_n^+] + |a|}{b-a}$$
$$\leq \frac{\sup_n E[X_n^+] + |a|}{b-a} < \infty.$$

Thus, for every $a < b$, $I\!P\left(U_\infty^{[a,b]}(\{X_k\}) = \infty\right) = 0$, and so

$$I\!P\left(\bigcup_{a<b,\, a,b\in Q} \left\{U_\infty^{[a,b]}(\{X_k\}) = \infty\right\}\right) = 0.$$

It follows that $X_\infty = \lim_n X_n$ exists almost surely, although at this point one cannot exclude the possibility that $X_\infty = \lim_n X_n = \pm\infty$.

It remains to show that, in fact, X_∞ is integrable and hence is finite a.s. First observe that since $\{X_n\}$ is a submartingale, $E[X_0] \leq E[X_n] = E[X_n^+ - X_n^-]$ for every n. Therefore $E[|X_n|] = E[X_n^+] + E[X_n^-] \leq 2E[X_n^+] - E[X_0]$, and

$$\sup_n E\left[|X_n|\right] \leq 2\sup_n E\left[X_n^+\right] - E\left[X_0\right] < \infty.$$

By Fatou's lemma, $E\left[|X_\infty|\right] \le \liminf_n E\left[|X_n|\right] \le \sup_n E\left[|X_n|\right] < \infty.$ $\diamond$

Here is an interesting corollary that is often useful when dealing with martingales.

Corollary 6.1 If $\{X_n\}_{n\ge 0}$ is a non-negative supermartingale, then $X_\infty = \lim_n X_n$ exits almost surely and $E[X_\infty] \le E[X_0]$.

Proof: Note that $\{-X_n\}$ is a submartingale and, because $\{X_n\}$ is non-negative, $(-X_n)^+ = 0$ for all n. Thus the martingale convergence theorem applies to give the existence of X_∞. Since $0 \le E[X_n] \le E[X_0]$ for all n, a Fatou's lemma argument implies $E[X_\infty] \le \sup_n E[X_n] \le E[X_0]$. $\diamond$

For an example, consider the martingale $\{E\left[Z/\mathcal{F}_n\right]\}$, where Z is integrable and $\{\mathcal{F}_n\}$ is a filtration. Because $x \to x^+$ is a convex function, the conditional Jensen inequality implies $(E\left[Z/\mathcal{F}_n\right])^+ \le E\left[Z^+/\mathcal{F}_n\right]$. Hence,

$$\text{(2)} \qquad \sup_n E\left[(E\left[Z/\mathcal{F}_n\right])^+\right] \le \sup_n E\left[E\left[Z^+/\mathcal{F}_n\right]\right] = E[Z^+] < \infty.$$

Therefore Theorem 6.1 implies that $\lim_n E\left[Z/\mathcal{F}_n\right]$ exists almost surely. This goes part way towards proving Lévy's upward theorem. To finish one must identify the limit with $E\left[Z/\sigma(\cup_n \mathcal{F}_n)\right]$ and for this it turns out that one needs in addition to prove that convergence holds in the L^1-norm as well. The question of L^1 convergence is taken up in a subsequent section and Lévy's theorem will proved there as a consequence of a more general theorem.

Theorem 6.1 as it stands does not apply directly to reverse martingales. However, since a reverse martingale $\{X_n\}$ considered in reverse order from any n, that is $X_n, X_{n-1}, \ldots, X_0$, is a martingale, the upcrossing inequality for martingales translates immediately to

$$E\left[U_n^{[a,b]}\left(X_n, X_{n-1}, \ldots, X_0\right)\right] \le \frac{E[X_0^+] + |a|}{b-a}.$$

Therefore, letting $n \to \infty$, the total number of upcrossings $[a, b]$ by the reverse time sequence is also bounded by $(E[X_0^+]+|a|)/(b-a) < \infty$ and hence is a.s. bounded. Then, following the argument of Theorem 6.1, one has

Corollary 6.2 A reverse martingale converges almost surely.

Again, a statement about L^1-convergence is needed to properly identify the limit of a reverse martingale, and this will be developed later.

Meanwhile, there is a nice application of Theorem 6.1 and Corollary 6.1 to questions of recurrence and transience of Markov chains. To explain this requires a few definitions and then some facts about Markov chains that are intuitively clear, but that we present without rigorous proof.

Let a countable state space Λ and a stochastic transition matrix P on Λ be given. States i and j in Λ are said to *communicate* if a Markov chain with transition matrix P will reach j with positive probability if it starts from i and, conversely, will reach i with positive probability if it starts from j. It may be shown that a chain can reach j from i if and only if there is a path in Λ from i to j such that the transition probability from each state of this path to the next is positive. A chain is said to be communicating if each state communicates with every other. (This terminology is imprecise, but standard. Communcication of states is solely a propery of the transition probability matrix P. To say a chain is communicating means that all states communicate for the transition probability matrix of this chain.)

A state i for a chain is said to be *recurrent* if, given that the chain starts in i, it visits i again at some later date with probability 1. It is a fact that *if* i *is recurrent, then the chain visits* i *at an infinite number of times.* Indeed, since i is recurrent with probability one there is a finite first time T at which the chain returns to i. As a consequence of the Markov propery, it can be shown that the process $X_T, X_{T+1}, \ldots$, which is just $\{X_n\}$ restarted at T, is just a Markov chain starting from i with the same transition matrix P. This fact, known as the Strong Markov Property, requires a bit of proof which is omitted; like the Markov property it states that the future of the process after return to i depends on the past up to the return to i only through the present value i at the time of return. But i is recurrent and so the chain must return a second time to i. After the second time it has to return a third, and so on. Hence the chain will visit i at an infinite number of times. A state that is not recurrent is called *transient*. A Markov chain will visit any one of its transient states only a finite number of times.

It is another fact that *if* i *is recurrent and if the chain can reach* j *from* i*, then* j *must also be recurrent.* Here is the idea of the proof, again without rigor. Assume the chain starts in a recurrent state i and decompose its paths into successive excursions starting at and returning to i, with no visits to i inbetween. It has been seen that the number of such excursions is infinite. The Strong Markov property implies in addition that the excursions are independent and have identical statistical properties. Let p be the probability that the chain visits j during one of the excursions. This probability is the same for all excursions because they are identically distributed. By independence of excursions, the chain can reach j from i if and only if $p > 0$, and, in the case that $p > 0$, it will reach j eventually with probability one, because it has an infinite number of independent opportunities to do so. Thus if the chain can reach a state j with positive probability from a recurrent state i, it must in fact reach j with probability one. Now, once the chain visits j, it must return to i with probability one because i is recurrent. But after this return to i, it must again visit j. Therefore, the probability that a chain starting from j must return to j will be one, and so j is recurrent.

It is thus clear that in a communicating chain, either all the states are recurrent or they are all transient. A chain with communicating states is called recurrent or transient according as the states themselves are recurrent or transient. A communicating, recurrent chain will visit all of its states infinitely often, while a communicating, transient chain will visit every state only a finite (random) number of times. It is of interest to determine whether a given chain, i.e. a given probability transition matrix, for a communicating chain is recurrent or transient. The martingale convergence theorem gives a nice criterion.

Proposition 6.1 Let P be the probability transition matrix of a communicating chain with a countably infinite state space. If there is a non-constant P-superharmonic function which is bounded below, then the chain is transient.

Proof: We shall do a proof by contradiction. Let $\{X_n\}$ be a chain with transition matrix P, assume it is recurrent and assume f is a non-constant superharmonic function which is bounded below. Because constant functions are harmonic, $g(i) := f(i) - \inf_\Lambda f(k)$ is a superharmonic function which is non-negative. Therefore, without loss of generality it may be assumed that f is non-negative. Let i and j be states for which $f(i) \neq f(j)$. By Proposition 1.2, $\{f(X_n)\}$ is a supermartingale, and hence by Corollary 6.1 $\lim_n f(X_n)$ exists almost surely. But since the chain is recurrent, it visits i and j infinitely often with probability one, and since $f(i) \neq f(j)$, this means $\{f(X_n)\}$ cannot converge. Thus a contradiction has been obtained, and so the chain must be transient. $\diamond$

From this Proposition, one can derive the following.

Theorem 6.2 Let P be the probability transition matrix of a communicating chain with a countably infinite state space. Then P is recurrent if and only if the only non-negative P-superharmonic functions are constant functions.

Proof: The 'if' direction is an immediate consequence of the previous Proposition. The proof of the 'only if' direction is a consequence of the following observation. For any chain on the state space Λ and any fixed $j \in \Lambda$, let

$$f(i) = \mathbb{P}\left(\text{there exists } n \geq 0 \text{ s.t. } X_n = j \;/\; X_0 = i\right), \quad i \in \Lambda.$$

Then f is non-negative and superharmonic. Here is the proof. Clearly $f(j) = 1$ and $f(i) \leq 1$ for all i. Thus

$$f(j) = 1 = \sum_{k \in \Lambda} p_{jk} \geq \sum_{k \in \Lambda} p_{jk} f(k) = (Pf)(j).$$

If $i \neq j$ then by conditioning and the Markov property (see equation (3) in section 1),

$$\begin{aligned} f(i) &= \sum_{k \in \Lambda} \mathbb{P}\left(\text{there exists } n \geq 0 \text{ s.t. } X_n = j \;/\; X_1 = k, X_0 = i\right) p_{ik} \\ &= \sum_{k \in \Lambda} \mathbb{P}\left(\text{there exists } n \geq 0 \text{ s.t. } X_n = j \;/\; X_1 = k\right) p_{ik} = \sum_{k \in \Lambda} f(k) p_{ik} \\ &= (Pf)(i). \end{aligned}$$

These two calculations show that f is superharmonic.

Now, if the only non-negative superharmonic function are constant, then, since $f(j) = 1$, it follows that $f(i) = 1$ for all i. Hence, the chain reaches j from any i with probability one, and interchanging j and i, reaches i from j with probability one. Hence, if the chain starts in i, it will return to i with probability one, proving that the chain is recurrent. $\diamond$.

As an application, let us reconsider the chain on the non-negative integers defined by

$$p_{ij} = \begin{cases} p_i & \text{, if } j = i{+}1; \\ q_i & \text{, if } j = i{-}1; \\ 0 & \text{, otherwise,} \end{cases}$$

if $i \geq 1$. This was the chain studied in example 6 of section 1 and again in section 3, where probabilities to hit the 0 state were calculated. Let us assume here that $p_{01} = 1$ (the chain goes to state 1 from state 0 with probability one) and that $p_i > 0$ and $q_i > 0$ for all $i \geq 1$. Then it is easy to see that this chain is communicating. When is it recurrent and when is it transient? We can answer this directly from Proposition 3.6, because the chain will be recurrent if and only if the probability of reaching 0 from any other state is 1, and Proposition 3.6 says that this is the case if and only if $\sum_{j=1}^{\infty} q_j \cdots q_1 / p_j \cdots p_i = \infty$. Let us rederive this by direct use of Theorem 6.2. Although formally the two derivations are different, it is apparent they are closely related, because the 'only if' part of the the proof of Theorem 6.2 actually relied on computing the probability of hitting a given state from another.

Define $f(0) = 1$, $f(1) = 0$ and let

$$f(i) = 1 - \sum_{j=1}^{i-1} \frac{q_j \cdots q_1}{p_j \cdots p_i}, \quad i \geq 2.$$

Then we know from section 1 that f is superharmonic. If

$$\sum_{j=1}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_i} < \infty, \tag{3}$$

then f will be bounded below and hence the chain will be transient.

On the other hand, let g be any superharmonic function; then

$$g(i) \geq p_i g(i{+}1) + q_i g(i{-}1) \quad \text{or} \quad g(i{+}1) - g(i) \leq \frac{q_i}{p_i} \left(g(i) - g(i{-}1) \right), \quad i \geq 1.$$

If $g(i) = g(i{-}1)$ for $i \leq k$, and if $g(k{+}1) \neq g(k)$, it then follows that $g(k{+}1) < g(k)$ and, for $i > k{+}1$,

$$g(i) - g(k) \leq (g(k{+}1) - g(k)) \sum_{j=k+1}^{i-1} \frac{q_j \ldots q_k}{p_j \ldots p_k}.$$

From this one sees that

$$\sum_{j=1}^{\infty} \frac{q_j \cdots q_1}{p_j \cdots p_i} = \infty, \tag{4}$$

implies that any non-constant superharmonic function is not bounded below; thus, by Theorem 6.2, (4) implies that the chain is recurrent.

6.7 Uniform Integrability

To go further in martingale convergence theory requires the concept of a uniformly integrable family of random variables. We interrupt the discussion of martingales long enough to treat this topic, which, although ultimately a part of abstract integration theory, is presented here in the language of probability theory.

Uniform integrability is a condition that connects convergence in probability to convergence in L^1-norm. We know that convergence in probability of a sequence of random variables does not imply convergence in L^1. A simple example is the sequence $X_n = n\mathbf{1}_{A_n}$, where $\{A_n\}$ is a sequence of events such that $I\!P(A_n) = 1/n$ for every positive integer n. Then $X_n \to 0$ in probability, but $E[X_n] = 1$ for all n. The impediment to L^1 convergence here is that as n gets larger, the non-zero values of X_n become uniformly large as well. The condition of uniform integrability prevents this kind of behavior without imposing a uniform deterministic bound on the value of the X_n's.

Definition A collection of random variables $\{Y_\alpha\,;\,\alpha \in I\}$, where I is an arbitrary index set, is said to *uniformly integrable* (u.i.) if

$$\lim_{K\to\infty} \sup_{\alpha\in I} E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right] = 0. \tag{1}$$

Remarks. 1. For any $K > 0$,

$$\begin{aligned} \sup_\alpha E[|Y_\alpha|] &= \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|\le K\}} + |Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right] \\ &\le K + \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right]. \end{aligned} \tag{2}$$

Therefore, if $\{Y_\alpha\,;\,\alpha \in I\}$ is u.i., then it is bounded in L^1:

$$\sup_\alpha E[|Y_\alpha|] < \infty.$$

2. Inequality (2) is easily generalized. For any event A,

$$\begin{aligned} \sup_\alpha E[|Y_\alpha|\mathbf{1}_A] &= \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_A\mathbf{1}_{\{|Y_\alpha|\le K\}} + |Y_\alpha|\mathbf{1}_A\mathbf{1}_{\{|Y_\alpha|>K\}}\right] \\ &\le K I\!P(A) + \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right]. \end{aligned} \tag{3}$$

3. (Exercise) Any finite family of integrable random variables is uniformly integrable.

There is an alternate characterization of uniform integrability that is a consequence of inequality (3).

Lemma 7.1 $\{Y_\alpha\,;\,\alpha \in I\}$ is uniformly integrable if and only if

$$\sup_\alpha E\left[|Y_\alpha|\right] < \infty, \quad \text{and} \tag{4}$$

$$\lim_{\delta\downarrow 0} \sup_{\mathbb{P}(A)\le\delta} \sup_\alpha E\left[|X_\alpha|\mathbf{1}_A\right] = 0. \tag{5}$$

Proof: Assume that the family of r.v.'s is u.i. Then (4) follows by remark (1). To obtain (5) set $K = \delta^{-1/2}$ in (3). Then

$$\sup_{\mathbb{P}(A)\le\delta} \sup_\alpha E\left[|X_\alpha|\mathbf{1}_A\right] \le \delta^{1/2} + \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>\delta^{-1/2}\}}\right],$$

and this approaches 0 as $\delta \downarrow 0$ by u.i.

Now assume that (4) and (5) hold. By Markov's inequality

$$\sup_\alpha \mathbb{P}\left(|Y_\alpha| > K\right) \le \frac{\sup_\alpha E[|Y_\alpha|]}{K},$$

and by (4) this approaches 0 as $K \to \infty$. Thus from (5) one obtains

$$\lim_{K\to\infty} \sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right] = 0,$$

which means that the family is uniformly integrable. ◇

A theorem of de la Vallée-Poussin provides a necessary and sufficient condition for uniform integrability,

Lemma 7.2 A family $\{Y_\alpha\,;\,\alpha \in I\}$ is uniformly integrable if and only if there is a positive, increasing function ϕ on $[0,\infty)$ satisfying $\lim_{t\to\infty}\phi(t)/t = \infty$ for which

$$\sup_\alpha E\left[\phi(|Y_\alpha|)\right] < \infty. \tag{6}$$

Proof: Assume (6), where $\lim_{t\to\infty}\phi(t)/t = \infty$. Then

$$\begin{aligned}\sup_\alpha E\left[|Y_\alpha|\mathbf{1}_{\{|Y_\alpha|>K\}}\right] &= \sup_\alpha E\left[\frac{|Y_\alpha|}{\phi(|Y_\alpha|)}\phi(|Y_\alpha|)\mathbf{1}_{\{|Y_\alpha|>K\}}\right]\\ &\le \left(\sup_{t\ge K}\frac{t}{\phi(t)}\right)\sup_\alpha E[\phi(|Y_\alpha|)].\end{aligned}$$

By assumption this approaches 0 as $K \to \infty$.

The necessity of (6) when the family of random variables is u.i. is rarely used. We present a proof for the curious, but it can be skipped without harm. Assume that $\{Y_\alpha\,;\,\alpha \in I\}$ is u.i. Let g be a function on $[0, \infty)$ with the following properties

(i) $g(0) < 1$ and g is continuous, nonincreasing and positive;

(ii) $\lim_{x\to\infty} g(x) = 0$.

(iii) There is a number N such that

$$g(x) \geq \sup_\alpha E\left[|Y_\alpha| \mathbf{1}_{\{|Y_\alpha|>x\}}\right], \quad \text{for all } x \geq N. \tag{7}$$

Such a g exists because of the uniform integrability property. We shall prove (6) for $\phi(x) := -x\ln(g(x))$. For this it suffices to show

$$\sup_\alpha E\left[\phi(|Y_\alpha|)\mathbf{1}_{\{|Y|\geq N\}}\right] < \infty. \tag{8}$$

We work with this restricted expectation to more easily exploit (7).

Now let Y be in the family $\{Y_\alpha\,;\,\alpha \in I\}$, and define the measure

$$\mu_Y(A) := E\left[|Y|\mathbf{1}_A(|Y|)\right],$$

on the Borel sets A of $[0, \infty)$. Notice that $\mu([z, \infty)) \leq g(z)$ for $z \geq N$. Then

$$\begin{aligned}
E\left[|Y|(-\ln g(|Y|))\mathbf{1}_{\{|Y|\geq N\}}\right] &= \int_{[N,\infty)} (-\ln g(y))\,\mu(dy) \\
&= \int_{[N,\infty)} \left(-\int_{[N,y)} \frac{dg(z)}{g(z)} - \ln g(N)\right) \mu(dy) \\
&= \int_{[N,\infty)} \mu([z,\infty))\,\frac{-dg(z)}{g(z)} - \mu([Z,\infty))\ln g(Z) \\
&\leq g(1) - g(Z)\ln g(Z).
\end{aligned}$$

This proves (8) because the right hand side does not depend on Y in the family $\{Y_\alpha\,;\,\alpha \in I\}$. ◇

Remark: (For those who read the proof of the converse in lemma 7.2.) In Lemma 7.2 it may be assumed that the function ϕ is convex. Indeed, take the function $\phi(x) = -x\ln(g(x))$ constructed in the proof. It is increasing. Take the largest convex function dominated by ϕ—this can be defined as the function whose epigraph is the convex hull generated by the epigraph of ϕ. Call this function ψ. Then one can verify that ψ is non-negative and $\lim_{x\to\infty} \psi(x)/x = \infty$. Since ψ is dominated by ϕ, (6) holds for ψ also.

Families of random variables of the form $\{E\left[Z/\mathcal{F}_n\right]$ can appear as martingales or reverse martingales, if the sequence of σ-algebras $\{\mathcal{F}_n\}$ is increasing or decreasing. The following result says that such families are uniformly integrable if Z is integrable.

Lemma 7.3 If $\{\mathcal{F}_\alpha\,;\,\alpha\in I\}$ is a family of σ-algebras and if Z is integrable, then $\{E\,[Z/\mathcal{F}_\alpha]\,;\,\alpha\in I\}$ is a uniformly integrable family of random variables.

Proof: This is left as an exercise; the lemma can be verified by checking the criterion (1) directly. Notice also that one can use Lemma 7.2. A single integrable random variable Z constitutes a uniformly integrable family, and hence by the Remark following Lemma 7.2, there is a positive, convex ϕ with $\lim_{x\to\infty}\phi(x)/x=\infty$ such that $E[\phi(|Z|)]<\infty$. By the conditional Jensen inequality, $E[\phi(|E\,[Z/\mathcal{G}])|]\leq E[E\,[\phi(|Z|)/\mathcal{G}]]=E[\phi(|Z|)]$, for any σ-algebra $\mathcal{G}$. $\diamond$

Here is the main theorem of the section, relating uniform integrability with convergence of expectations.

Theorem 7.1 Let $\{X_n\}$ be a sequence of random variables which converges in probability to a random variable X. Then

$$E[|X|]<\infty,\quad E[X_n]<\infty\quad\text{for all } n,\quad\text{and}\quad \lim_{n\to\infty}E\,[|X-X_n|]=0 \tag{9}$$

$$\text{if and only if}$$

$$\{X_n\}\quad\text{is uniformly integrable.}$$

Proof: Assume uniform integrability. We prove (9). First, it was shown in (2) that $\sup_n E[X_n]<\infty$. Choose a subsequence $\{n_k\}$ such that $\{X_{n_k}$ converges a.s. to X. Then by Fatou's lemma

$$E[|X|]\leq\liminf_k E[X_{n_k}]\leq\sup_n E[|X_n|]<\infty.$$

It remains to show L^1-convergence. For any $\epsilon>0$, let $A_n^\epsilon:=\{|X-X_n|>\epsilon\}$.

$$\begin{aligned}E\,[|X-X_n|]&\leq\epsilon+E\left[|X-X_n|\mathbf{1}_{A_n^\epsilon}\right]\\&\leq\epsilon+E\left[|X_n|\mathbf{1}_{A_n^\epsilon}\right]+E\left[|X|\mathbf{1}_{A_n^\epsilon}\right]\\&\leq\epsilon+\sup_k E\left[|X_k|\mathbf{1}_{A_n^\epsilon}\right]+E\left[|X|\mathbf{1}_{A_n^\epsilon}\right]\end{aligned}\tag{10}$$

Since $X_n\to X$ in probability, $I\!\!P(A_n^\epsilon)\to 0$. Hence by Lemma 7.1 applied to (10),

$$\lim_{n\to\infty}E\,[|X-X_n|]\leq\epsilon.$$

Now take $\epsilon\downarrow 0$.

Conversely, suppose that (9) holds. We shall prove uniform integrability by verifying (5) of Lemma 7.1, Since any finite set of integrable random variables is uniformly integrable (see Remark 3 after the definition of uniform integrability),

$$\lim_{\delta\downarrow 0}\sup_{I\!\!P(A)\leq\delta}\sup_{1\leq n\leq N}E\,[|X_n|\mathbf{1}_A]=0. \tag{11}$$

Given $\epsilon > 0$, choose N so that $E[|X - X_n|] < \epsilon$ for $n \geq N$. Then

$$\begin{aligned}\sup_n E\left[|X_n|\mathbf{1}_A\right] &\leq \sup_{1\leq n\leq N} E\left[|X_n|\mathbf{1}_A\right] + \sup_{n\geq N} E\left[|X - X_n|\mathbf{1}_A\right] + E\left[|X|\mathbf{1}_A\right] \\ &\leq \sup_{1\leq n\leq N} E\left[|X_n|\mathbf{1}_A\right] + E\left[|X|\mathbf{1}_A\right] + \epsilon\end{aligned}$$

Thus, from (11),

$$\lim_{\delta\downarrow 0} \sup_{\boldsymbol{P}(A)\leq\delta} \sup_n E\left[|X_n|\mathbf{1}_A\right] \leq \epsilon.$$

Again, take $\epsilon \downarrow 0$ to complete the proof. ◇

6.8 More Martingale Convergence

We combine uniform integrability with Theorem 6.1 to obtain the following basic result. As you will see, it implies that uniform integrability is a necessary and sufficient condition for a martingale to be of the form $\{E\left[Z/\mathcal{F}_n\right]\}$.

Theorem 8.1 Let $\{X_n\}$ be a submartingale. Then

(12) $$X_\infty = \lim_n X_n \quad \text{exists and} \quad \lim_{n\to\infty} E[|X - X_n|] = 0$$

if and only if

(13) $$\{X_n\} \text{ is uniformly integrable.}$$

When (12), or equivalently (13), holds,

$$X_n \leq E\left[X_\infty/\mathcal{F}_n\right] \qquad \text{a.s. for all } n.$$

If $\{X_n\}$ is a martingale then (12), or equivalently (13), implies

$$X_n = E\left[X_\infty/\mathcal{F}_n\right] \qquad \text{a.s. for all } n.$$

Proof: Assume that $\{X_n\}$ is uniformly integrable. Then $\sup_n E[|X_n|] < \infty$, and hence Theorem 6.1 implies that X_∞ exists and is finite almost surely. Theorem 7.1 then implies that $\lim_{n\to\infty} E[|X - X_n|] = 0$.

Conversely, Theorem 7.1 proves that (12) implies (13).

Now assume (12), or equivalently (13). Then, for any n

$$E\left[\left|\, E\left[X_\infty/\mathcal{F}_n\right] - E\left[X_m\infty/\mathcal{F}_n\right] \,\right|\right] \leq E\left[E\left[|X_\infty - X_m|/\mathcal{F}_n\right]\right] \leq E\left[|X - X_m|\right].$$

Therefore, it follows that $E\left[X_m/\mathcal{F}_n\right]$ converges in L^1 to $E\left[X_\infty/\mathcal{F}_n\right]$ as $m \to \infty$. By the submartingale propery $E\left[X_m/\mathcal{F}_n\right] \geq X_n$ a.s. for all m, and hence its limit $E\left[X_\infty/\mathcal{F}_n\right] \geq X_n$ a.s. also. If $\{X_n\}$ is actually a martingale, then $E\left[X_m/\mathcal{F}_n\right] = X_n$ a.s. for all m, and equality passes to the limit also. ◇

Corollary 8.1 (Lévy's upward convergence theorem) Let $\{\mathcal{F}_n\}$ be a filtration, let $\mathcal{F}_\infty := \sigma(\cup_n \mathcal{F}_n)$, and let Z be an integrable random variable. Then

$$E[Z/\mathcal{F}_n] \to E[Z/\mathcal{F}_\infty] \quad \text{as } n \to \infty \text{, both almost surely and in } L^1.$$

Proof: The martingale $\{E[Z/\mathcal{F}_n]\}$ is uniformly integrable by Lemma 7.3. Theorem 8.1 then implies that it has a limit X_∞ and that

$$E[Z/\mathcal{F}_n] = E[X_\infty/\mathcal{F}_n], \quad n \geq 0. \tag{14}$$

This limit is clearly $\mathcal{F}_\infty$-measurable.

Let $\mathcal{C}$ be the class of all sets A satisfying

$$E[Z\mathbf{1}_A] = E[X_\infty \mathbf{1}_A]. \tag{15}$$

It is easily checked that this class $\mathcal{C}$ is a monotone class. We will show that $\mathcal{C}$ contains the algebra $\cup_n \mathcal{F}_n$. This will complete the proof, because it then follows by the monotone class theorem that $\mathcal{C}$ must contain $\mathcal{F}_\infty$ as well. Since (15) must thus be true for all sets in $\mathcal{F}_\infty$, and since X_∞ is $\mathcal{F}_\infty$-measurable, it follows that $X_\infty = E[Z/\mathcal{F}_\infty]$.

Thus it remains to show $\mathcal{C}$ contains $\cup_n \mathcal{F}_n$. But if A is aan event in the algebra $\cup_n \mathcal{F}_n$, there is an n such that $A \in \mathcal{F}_n$. Now apply (14): by conditioning inside the expectation on $\mathcal{F}_n$,

$$\begin{aligned} E[Z\mathbf{1}_A] &= E[\mathbf{1}_A E[Z/\mathcal{F}_n]] = E[\mathbf{1}_A E[X_\infty/\mathcal{F}_n]] \\ &= E[\mathbf{1}_A X_\infty] \end{aligned}$$

Thus (15) is verified. ◇

Corollary 8.2 Let $\{X_n\}_{n\geq 0}$ be a reverse martingale with respect to the decreasing sequence of σ-algebras $\{sF_n\}$. Then, as $n \to \infty$, X_n converges almost surely and in L^1 to $E[X_0/\cap_n \mathcal{F}_n]$

Proof: Exercise.