1 IEOR 4701: Martingales I: Discrete time

Martingales are stochastic processes that are meant to capture the notion of a fair game in the context of gambling. In a fair game, each gamble on average, independent of the past, yields no profit or loss. But the reader should not think that martingales are used just for gambling; they pop up naturally in numerous applications of stochastic modeling. They have enough structure to allow for strong general results while also allowing for dependencies among variables. Thus they deserve the kind of attention that Markov chains do. Gambling, however, supplies us with insight and intuition through which a great deal of the theory can be understood.

1.1 Basic definitions and examples

Definition 1.1 *A stochastic process* $\mathbf{X} = \{X_n : n \geq 0\}$ *is called a martingale (MG) if*

C1: $E(|X_n|) < \infty,\ n \geq 0$, *and*

C2: $E(X_{n+1}|X_0, \ldots, X_n) = X_n,\ n \geq 0.$

Notice that property C2 can equivalently be stated as

$$E(X_{n+1} - X_n|X_0, \ldots, X_n) = 0,\ n \geq 0. \tag{1}$$

In the context of gambling, by letting X_n denote your total fortune after the n^{th} gamble, this then captures the notion of a fair game in that on each gamble, independent of the past, your expected change in fortune is 0; on average you neither win or lose any money.

Taking expected values in C2 yields $E(X_{n+1}) = E(X_n),\ n \geq 0$, and we conclude that

$$E(X_n) = E(X_0),\ n \geq 0,\ \text{for any MG;}$$

At any time n, your expected fortune is the same as it was initially.

For notational simplicity, we shall let $\mathcal{G}_n = \sigma\{X_0, \ldots, X_n\}$ denote all the events determined by the r.v.s. $X_0, \ldots, X_n$, and refer to it as the *information* determined by $\mathbf{X}$ up to and including time n. Note that $\mathcal{G}_n \subset \mathcal{G}_{n+1},\ n \geq 0$; information increases as time n increases.

Then the martingale property C2 can be expressed nicely as

$$E(X_{n+1}|\mathcal{G}_n) = X_n,\ n \geq 0.$$

A very important fact is the following which we will make great use of throughout our study of martingales:

Proposition 1.1 *Suppose that* $\mathbf{X}$ *is a stochastic process satisfying C1.*
Let $\mathcal{G}_n = \sigma\{X_0, \ldots, X_n\},\ n \geq 0$. *Suppose that* $\mathcal{F}_n = \sigma\{U_0, \ldots, U_n\},\ n \geq 0$ *is information for some other stochastic process such that it contains the information of* $\mathbf{X}$*:*
$\mathcal{G}_n \subset \mathcal{F}_n,\ n \geq 0$. *Then if* $E(X_{n+1}|\mathcal{F}_n) = X_n,\ n \geq 0$, *then in fact* $E(X_{n+1}|\mathcal{G}_n) = X_n,\ n \geq 0$, *so* $\mathbf{X}$ *is a MG.*

$\mathcal{G}_n \subset \mathcal{F}_n$ implies that $\mathcal{F}_n$ also determines $X_0, \ldots, X_n$, but may also determine other things as well. So the above Proposition allows us to verify condition C2 by using more information than is necessary. In many instances, this helps us verify C2 in a much simpler way than would be the case if we directly used $\mathcal{G}_n$.
Proof :

$$\begin{aligned} E(X_{n+1}|\mathcal{G}_n) &= E(E(X_{n+1}|\mathcal{F}_n))|\mathcal{G}_n) \\ &= E(X_n|\mathcal{G}_n) \\ &= X_n. \end{aligned}$$

The first equality follows since $\mathcal{G}_n \subset \mathcal{F}_n$; we can always condition first on more information. The second equality follows from the assumption that $E(X_{n+1}|\mathcal{F}_n) = X_n$, and the third from the fact that X_n is determined by $\mathcal{G}_n$. ■

Because of the above, we sometimes speak of *a MG* $\mathbf{X}$ *with respect to* $\mathcal{F}_n,\ n \geq 0$, where $\mathcal{F}_n$ determines $X_0, \ldots, X_n$ but might be larger.

Examples

In what follows, we typically will define $\mathcal{F}_n$ from the start to be perhaps larger than is needed inorder to check C2.

1. *Symetric random walks.* Let $R_n = \Delta_1 + \cdots + \Delta_n,\ n \geq 1,\ \ R_0 = 0$ where $\{\Delta_n : n \geq 1\}$ is i.i.d. with $E(\Delta) = 0$, and $E(|\Delta|) < \infty$. That $\mathbf{R}$ is a MG is easily verified. C1: $E(|R_n|) \leq nE(|\Delta|) < \infty$. C2: We choose $\mathcal{F}_n = \sigma\{\Delta_1, \ldots, \Delta_n\}$, which clearly determines all that we need. $R_{n+1} = R_n + \Delta_{n+1}$ yielding $E(R_{n+1}|\mathcal{F}_n) = R_n + E(\Delta_{n+1}|\mathcal{F}_n)$
   $= R_n + E(\Delta_{n+1})$
   $= R_n + 0$
   $= R_n$

   For simplicity we chose $R_0 = 0$, and so $E(R_n) = E(R_0) = 0,\ n \geq 0$; but any initial condition, $R_0 = x$, will do in which case $\{R_n\}$ is still a MG, and $E(R_n) = E(R_0) = x,\ n \geq 0$.

2. (*Continuation.*) Assume further that $\sigma^2 = Var(\Delta) < \infty$. Then

   $X_n = R_n^2 - n\sigma^2,\ n \geq 0$ forms a MG.
   C1: $E(|X_n|) \leq E(R_n^2) + n\sigma^2$
   $= Var(R_n) + n\sigma_2 = n\sigma^2 + n\sigma^2 = 2n\sigma^2 < \infty$.

C2: $X_{n+1} = (R_n + \Delta_{n+1})^2 - (n+1)\sigma^2$
$= R_n^2 + \Delta_{n+1}^2 + 2R_n\Delta_{n+1} - (n+1)\sigma^2.$

$E(X_{n+1}|\mathcal{F}_n) = R_n^2 + E(\Delta_{n+1}^2) + 2R_nE(\Delta_{n+1}) - (n+1)\sigma^2$
$= R_n^2 + \sigma^2 - (n+1)\sigma^2$
$= R_n^2 - n\sigma^2 = X_n.$

3. *More general symetric random walks.* The random walk from Example 1 can be generalized by allowing each increment Δ_n to have its own mean 0 distribution; the MG property C2 still holds: Letting $\{\Delta_n : n \geq 1\}$ be an independent sequence of r.v.s. (that is not necessarily identically distributed) with $E(\Delta_n) = 0$ and $E(|\Delta_n|) < \infty,\ n \geq 1$, again yields a MG. If in addition, each Δ_n has the same variance $\sigma^2 < \infty$, then $R_n^2 - n\sigma^2$ from Example 2 also remains a MG.

4. Let $X_n = Y_1 \times \cdots \times Y_n,\ X_0 = 1$ where $\{Y_n : n \geq 1\}$ is i.i.d. with $E(Y) = 1$ and $E(|Y|) < \infty$. Then $\mathbf{X}$ is a MG. C1: $E(|X_n|) \leq (E(|Y|))^n < \infty$.

   C2: We use $\mathcal{F}_n = \sigma\{Y_1, \ldots, Y_n\}$. $X_{n+1} = Y_{n+1}X_n$. $E(X_{n+1}|\mathcal{F}_n) = X_nE(Y_{n+1}|\mathcal{F}_n)$
   $= X_nE(Y_{n+1}) = X_n \times 1 = X_n$.

5. *Doob's Martingale:*

   Let X be any r.v. such that $E(|X|) < \infty$. Let $\{Y_n : n \geq 0\}$ be any stochastic process (on the same probability space as X), and let $\mathcal{F}_n = \sigma\{Y_0, Y_1, \ldots, Y_n\}$. Then $X_n \stackrel{\text{def}}{=} E(X|\mathcal{F}_n)$ defines a MG called a Doob's MG. C1: $E(|X_n|) = E(|E(X|\mathcal{F}_n)|) \leq E(E(|X|\ |\mathcal{F}_n)) = E(|X|) < \infty$. (Note that $E(X_n) = E(X),\ n \geq 0$.)

   C2: $E(X_{n+1}|\mathcal{F}_n) = E(E(X|\mathcal{F}_{n+1})|\mathcal{F}_n)$
   $= E(X|\mathcal{F}_n)$ (because $\mathcal{F}_n \subset \mathcal{F}_{n+1}$)
   $= X_n$.

   In essence, X_n approximates X, and as n increases the approximation becomes more refined because more information has been gathered and included in the conditioning. For example, if X is completely determined by

   $$\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n \stackrel{\text{def}}{=} \cup_{n=0}^{\infty}\mathcal{F}_n = \sigma\{Y_0, Y_1, Y_2, \ldots\}$$

   then it seems reasonable that $X_n \to X,\ n \to \infty$, w.p.1. This is so, and in fact, it can be shown that in general, $X_n \to E(X|\mathcal{F}_\infty),\ n \to \infty$ w.p.1. The idea here is that $\mathcal{F}_\infty$ is the most information available, and so $E(X|\mathcal{F}_\infty)$ is the best approximation to X possible, given all we know.

1.2 Uniform integrability

Given a sequence of r.v.s. $\{X_n\}$ for which it is known apriori that $X_n \to X,\ n \to \infty$, w.p.1. for some r.v. X, it is of great importance in probability to determine conditions ensuring that $E(X_n) \to E(X),\ n \to \infty$. To quickly dispense with the notion that conditions might not be needed, we present a simple counterexample: Let U be uniformly distributed over $(0,1)$, and define $X_n = nI\{U \leq 1/n\},\ n \geq 1$. Then, w.p.1., $U > 0$ and for n sufficiently large, $U > 1/n$. Thus $X_n \to X = 0,\ n \to \infty$, w.p.1. But $E(X_n) = 1 \neq 0,\ n \geq 1$.

The following basic integration result yields some sufficient conditions:

Theorem 1.1 (Monotone convergence theorem) *If $0 \leq X_n \uparrow X$ wp1 as $n \to \infty$, then $E(X_n) \uparrow E(X)$.*

The point here is that $E(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} E(X_n)$ when the rvs are non-negative and monotone $0 \leq X_n \leq X_{n+1},\ n \geq 1$. (This theorem remains valid even if the limiting rv X has mass at ∞, $P(X = \infty) > 0$ in which case $E(X_n) \uparrow \infty$.)

Note that for a fixed non-negative rv X with $E(X) < \infty$ we have that $0 \leq XI\{X \leq x\}$ is monotone increasing in x to X, and hence using the monotone convergence theorem yields $E(XI\{X > x\}) \to E(X)$ as $x \to \infty$. But since $E(X) = E(XI\{X > x\}) + E(XI\{X \leq x\})$ we conclude that $E(XI\{X > x\}) \to 0$ yielding

$$E(|X|I\{|X| > x\}) \to 0, \textit{ as } x \to \infty \textit{ if } E(|X|) < \infty.$$

In general, the needed conditions ensuring that $E(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} E(X_n)$ involve what is called uniform integrability:

Definition 1.2 *A collection of r.v.s. $\{X_t : t \in T\}$ is said to be uniformly integrable (UI), if $\sup_{t\in T} E(|X_t|I\{|X_t| > x\}) \to 0$, as $x \to \infty$.*

Choosing $x > 0$ such that $\sup_{t\in T} E(|X_t|I\{|X_t| > x\}) \leq 1$, and noting that then regardless of $t \in T$,

$$E(|X_t|) = E(|X_t|I\{|X_t| \leq x\}) + E(|X_t|I\{|X_t| > x\}) \leq x + 1 < \infty$$

we conclude that if $\{X_t : t \in T\}$ is UI, then $\sup_{t\in T} E(|X_t|) < \infty$.

Definition 1.3 *For a fixed $0 < p < \infty$, a collection of r.v.s. $\{X_t : t \in T\}$ is said to be bounded in L^p if $\sup_{t\in T} E(|X_t|^p) < \infty$.*

Thus we have shown above that

A uniformly integrable collection of r.v.s. is always bounded in L^1.

Noting that

$$E(|X_t|I\{|X_t| > x\}) = xP(|X_t| > x) + \int_x^\infty P(|X_t| > y)dy,$$

What UI means is that both $xP(|X_t| > x)$ and the *remainders*, $\int_x^\infty P(|X_t| > y)dy$, of the finite integrals $E(|X_t|)$ must converge to 0, as $x \to \infty$, uniformally over all $t \in T$.

It is immediate from the discussion right after the monotone convergence theorem that *every r.v. X such that $E(|X|) < \infty$ is itself is UI.*

Proposition 1.2 *Suppose $X_n \to X$, w.p.1., as $n \to \infty$, and $E(|X_n|) < \infty$, $n \geq 0$. Then the following are equivalent:*

1. *$\{X_n : n \geq 0\}$ is UI;*
2. *$E(|X|) < \infty$ and $E(|X_n - X|) \to 0$, as $n \to \infty$;*
3. *$E(|X|) < \infty$ and $E(|X_n|) \to E(|X|)$, as $n \to \infty$.*

The following result offers us a solution to our original problem.

Corollary 1.1 *Suppose $X_n \to X$, w.p.1., as $n \to \infty$, and $E(|X_n|) < \infty$, $n \geq 0$. If $\{X_n : n \geq 0\}$ is UI, then $E(|X|) < \infty$ and $E(X_n) \to E(X)$, as $n \to \infty$.*

Proof : From the equivalence of (1) and (2) of Proposition 1.2, we conclude that $E(|X|) < \infty$ and $|E(X_n) - E(X)| \leq E(|X_n - X|) \to 0$, as $n \to \infty$. ■

Proposition 1.3 *For a collection $\{X_t : t \in T\}$ of r.v.s. to be UI, it is sufficient that*

$$\sup_{t \in T} |X_t| \leq Y,$$

w.p.1., for a r.v. Y satisfying $E(Y) < \infty$. (A special case of interest is when Y is bounded.)

Proof : Under the stated hypothesis, we conclude that
$\sup_{t \in T} |X_t| I\{|X_t| > x\} \leq Y I\{Y > x\}$ and the result follows since $E(Y) < \infty$ and hence $E(Y I\{Y > x\}) \to 0$ as $x \to \infty$ (e.g., any rv X is UI if $E(|X|) < \infty$).

■

Corollary 1.2 *Any finite collection of r.v.s. $X_1, \ldots, X_k$ such that $E(|X_i|) < \infty$, $1 \leq i \leq k$ is UI.*

Proof : Choose $Y = |X_1| + \cdots + |X_k|$ and apply Proposition 1.3. ■

The following is one of the most useful results in probability.

Theorem 1.2 (Dominated Convergence Theorem) *Suppose $X_n \to X$, w.p.1., as $n \to \infty$, and $E(|X_n|) < \infty$, $n \geq 0$. If*

$$\sup_{n \geq 0} |X_n| \leq Y,$$

w.p.1., for a r.v. Y satisfying $E(Y) < \infty$, then $E|X| < \infty$ and $E(X_n) \to E(X)$, as $n \to \infty$.

Note: The dominated convergence theorem in the special case when Y is bounded is called the *bounded convergence theorem.*

Proof : Follows directly from Corollary 1.1 and Proposition 1.3. ■

Lemma 1.1 *Suppose $X \geq 0$ and $E(X) < \infty$. Then for any (proper) r.v. Z it holds that*

$$E(XI\{Z > x\}) \to 0, \ x \to \infty.$$

Proof : $XI\{Z > x\} \to 0, \ x \to \infty$, w.p.1., and $XI\{Z > x\} \leq X$. Since $E(X) < \infty$ the result follows via the dominated convergence theorem. ■

1.3 Optional Stopping

Stopped Martingales

Recall that a *stopping time* τ with respect to a stochastic process $\{X_n : n \geq 0\}$ is a discrete r.v. with values in $\{0, 1, 2, \ldots\}$ such that for each $n \geq 0$, the event $\{\tau = n\}$ is determined by (at most) $\{X_0, \ldots, X_n\}$, the information up to and including time n. It is easily seen that the property can be restated equivalently as the event $\{\tau > n\}$ is determined by (at most) $\{X_0, \ldots, X_n\}$. ($\{\tau > n\}$ denotes the event that you have not stopped by time n; and this can only depend on the information up to and including time n.) Unless otherwise stated, we shall always assume that all stopping times in question are proper, $P(\tau < \infty) = 1$. (Of course, in some examples, we must first prove that is so.)

Let $a \wedge b = \min\{a, b\}$.

Proposition 1.4 *If $\mathbf{X} = \{X_n : n \geq 0\}$ is a MG, and τ is a stopping time w.r.t. $\mathbf{X}$, then the stopped process $\overline{\mathbf{X}} = \{\overline{X}_n : n \geq 0\}$ is a MG, where*

$$\begin{aligned} \overline{X}_n &\stackrel{\text{def}}{=} \begin{cases} X_n, & \textit{if } \tau > n\ ; \\ X_\tau, & \textit{if } \tau \leq n \end{cases} \\ &= X_{n \wedge \tau}. \end{aligned}$$

Since $\overline{X}_0 = X_0$, we conclude that $E(\overline{X}_n) = E(X_0), \ n \geq 0$: Using any stopping time as a gambling strategy yields at each fixed time n, on average, no benefit; the game is still fair.

Proof : (C1:) Since

$$|\overline{X}_n| \leq \max_{0 \leq k \leq n} |X_k| \leq |X_0| + \cdots + |X_n|,$$

we conclude that $E|\overline{X}_n| \leq E(|X_0|) + \cdots + E(|X_n|) < \infty$, from C1 for $\mathbf{X}$.

(C2:) It is sufficient to use $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ since $\sigma\{\overline{X}_0, \ldots, \overline{X}_n\} \subset \mathcal{F}_n$ by the stopping time property that $\{\tau > n\}$ is determined by $\{X_0, \ldots, X_n\}$. Noting that both $\overline{X}_n = X_n$ and $\overline{X}_{n+1} = X_{n+1}$ if $\tau > n$, and $\overline{X}_{n+1} = \overline{X}_n$ if $\tau \leq n$ yields

$$\overline{X}_{n+1} = \overline{X}_n + (X_{n+1} - X_n)I\{\tau > n\}.$$

Thus

$$\begin{aligned}
E(\overline{X}_{n+1}|\mathcal{F}_n) &= \overline{X}_n + E((X_{n+1}-X_n)I\{\tau>n\}|\mathcal{F}_n) \\
&= \overline{X}_n + I\{\tau>n\}E((X_{n+1}-X_n)|\mathcal{F}_n) \\
&= \overline{X}_n + I\{\tau>n\}\cdot 0 \\
&= \overline{X}_n.
\end{aligned}$$

■

Since $\lim_{n\to\infty} n\wedge\tau = \tau$, w.p.1., we conclude that $\lim_{n\to\infty}\overline{X}_n = X_\tau$, w.p.1. It is therefore of interest to know when we can interchange the limit with expected value:

$$\textit{When does} \quad \lim_{n\to\infty} E(\overline{X}_n) = E(X_\tau)\ ? \tag{2}$$

For if (2) holds, then since $E(\overline{X}_n) = E(X_0),\ n\geq 0$, we would conclude that

$$E(X_\tau) = E(X_0). \tag{3}$$

From Corollary 1.1, we know that uniform integrability (UI) of $\{\overline{X}_n\}$ is the needed condition for the desired interchange, so we at first state this important result, and then give some reasonable sufficient conditions (useful in many applications) ensuring the UI condition.

Theorem 1.3 (Martingale Optional Stopping Theorem) *If $\mathbf{X} = \{X_n : n\geq 0\}$ is a MG and τ is a stopping time w.r.t. $\mathbf{X}$ such that the stopped process $\overline{\mathbf{X}}$ is UI, then (3) holds: Your expected fortune when stopping is the same as when you started; the stopping strategy does not help to increase your expected fortune.*

Proposition 1.5 *If $\mathbf{X}$ is a MG and τ is a stopping time w.r.t. $\mathbf{X}$, then each of the following conditions alone ensures that $\overline{\mathbf{X}}$ is (UI) and hence that (3) holds:*

1. $\sup_{n\geq 0}|\overline{X}_n| \leq Y$, w.p.1., *where Y is a r.v. such that $E(Y)<\infty$.*
2. *The stopping time τ is bounded: $P(\tau\leq k)=1$ for some $k\geq 1$.*
3. *$E(|X_\tau|)<\infty$ and $E(|X_n|;\tau>n)\to 0,\ n\to\infty$.*
4. *$E(\tau)<\infty$ and $\sup_{n\geq 0}E(|X_{n+1}-X_n|\ |\mathcal{F}_n)\leq B$, some $B<\infty$.*
5. *There exists a $\delta>0$ and a $B>0$ such that $\sup_{n\geq 0}E(|\overline{X}_n|^{1+\delta})\leq B$.*

Proof :

1. The dominated convergence theorem.
2. $|\overline{X}_n| \leq Y \stackrel{\text{def}}{=} \max\{|X_1|,\ldots,|X_k|\}$, and thus the dominated convergence theorem applies since $E(Y)\leq E(|X_1|)+\cdots+E(|X_k|)<\infty$.

3. $$|\overline{X}_n| = |X_\tau| I\{\tau \le n\} + |X_n| I\{\tau > n\}.$$

   Thus if $E(|X_n|; \tau > n) \to 0$, then $\lim_{n\to\infty} E(|\overline{X}_n|) = \lim_{n\to\infty} E(|X_\tau| I\{\tau \le n\})$. But $|X_\tau| I\{\tau \le n\} \to |X_\tau|$ and $|X_\tau| I\{\tau \le n\} \le |X_\tau|$ with $E(|X_\tau|) < \infty$ by assumption. Thus from the dominated convergence theorem we obtain $\lim_{n\to\infty} E(|\overline{X}_n|) = E(|X_\tau|) = E(\lim_{n\to\infty} |\overline{X}_n|)$ which is equivalent to UI (via 3 in Proposition 1.2).

4. Follows from (3):

$$\begin{aligned}
E(|X_\tau|) &\le E(|X_0|) + \sum_{n=1}^{\infty} E(|X_n - X_{n-1}| I\{\tau > n-1\}) \\
&= E(|X_0|) + \sum_{n=1}^{\infty} E\{E(|X_n - X_{n-1}| I\{\tau > n-1\} | \mathcal{F}_{n-1})\} \\
&= E(|X_0|) + \sum_{n=1}^{\infty} E\{I\{\tau > n-1\} E(|X_n - X_{n-1}| \,|\mathcal{F}_{n-1})\} \\
&\le E(|X_0|) + B \sum_{n=1}^{\infty} P(\tau > n-1) \\
&= E(|X_0|) + BE(\tau) \\
&< \infty.
\end{aligned}$$

   Similarly, $E(|X_n|; \tau > n) \le E(|X_0| + Bn; \tau > n) \le E(|X_0| + B\tau; \tau > n) \to 0$.

5. From Markov's inequality,

$$P(|\overline{X}_n| > x) \le x^{-(1+\delta)} E(|\overline{X}_n|^{1+\delta}) \le Bx^{-(1+\delta)},$$

   yielding

$$\begin{aligned}
\sup_{n\ge 0} E(|X_n| I\{|X_n| > x\}) &= \sup_{n\ge 0} x P(|X_n| > x) + P(\int_x^\infty P(|X_n| > y) dy \\
&\le Bx^{-\delta} + \int_x^\infty By^{-(1+\delta)} dy = 2Bx^{-\delta},
\end{aligned}$$

   which tends to 0, as $x \to \infty$, uniformly in n.

■

1.4 Applications

1. *Wald's equation.* Let $\{Y_n : n \geq 1\}$ be i.i.d. with finite mean $\mu = E(Y)$. Let $\Delta_n = Y_n - \mu$, so that the Δ_n are i.i.d. with mean 0. Now let $R_n = \Delta_1 + \cdots + \Delta_n,\ n \geq 1$, $R_0 = 0$, denote the associated *symetric random walk*, which we know is a MG. Let τ be any stopping time w.r.t. $\{Y_n\}$ such that $E(\tau) < \infty$. If the required UI condition is met, then we conclude from Theorem 1.3 that $E(R_\tau) = 0 = E(R_0)$. Since
$$R_\tau = -\tau\mu + \sum_{j=1}^{\tau} Y_j,$$
taking expected values then yields the well-known Wald's equation,
$$E\Big\{\sum_{j=1}^{\tau} Y_j\Big\} = E(\tau)E(Y).$$
The UI condition is met via (4) of Proposition 1.5: $E(\tau) < \infty$ is assumed, and
$$E(|R_{n+1} - R_n|\ |\mathcal{F}_n) = E(|\Delta|) < \infty;\ B = E(|\Delta|).$$

2. *Hitting times for the simple symetric random walk.* Let $\{R_n\}$ denote the simple symetric random walk, with $R_0 = 0$; the increment distribution is $P(\Delta = 1) = P(\Delta = -1) = 0.5$. For fixed integers $a > 0$ and $b > 0$, let
$$\tau = \min\{n \geq 0 : R_n \in \{a, -b\}\}, \tag{4}$$
the first passage time of the random walk to level a or $-b$. (We already know from basic random walk theory that $P(\tau < \infty) = 1$.) If the required UI condition is met, then $E(R_\tau) = 0$. But by definition of τ, $R_\tau = a$ or $R_\tau = -b$.
Letting $p_a = P(R_\tau = a)$ and $p_{-b} = P(R_\tau = -b) = 1 - p_a$, we conclude that $0 = E(R_\tau) = ap_a - b(1 - p_a)$, or
$$p_a = \frac{b}{a+b}, \tag{5}$$
and we have computed the probability that the random walk goes up by a units before dropping down by b units. This gives the solution to the *gambler's ruin problem* when $p = 0.5$.

   UI: Noting that up to time τ, the random walk is restricted within the bounded interval $[-b, a]$, the UI condition is obtained via (1) of Proposition 1.5:

   $\sup_{n\geq 0} |\overline{R}_n| \leq \max\{a, b\}$.

3. *Continuation.* $X_n = R_n^2 - n$ defines yet another MG since $\sigma^2 = Var(\Delta) = 1$. Let τ be as in (4). If UI holds, then $0 = E(X_0) = E(X_\tau) = E(R_\tau^2) - \tau$, or $E(\tau) = E(R_\tau^2)$. Using p_a and p_{-b} from (5), we conclude that
$$E(\tau) = \frac{b}{a+b}a^2 + \frac{a}{a+b}b^2 = ab. \tag{6}$$

UI:

$$\begin{aligned}|X_{n\wedge\tau}| &= |R^2_{n\wedge\tau} - n\wedge\tau| \\ &\le R^2_{n\wedge\tau} + n\wedge\tau \\ &\le Y \stackrel{\text{def}}{=} (\max\{a,b\})^2 + \tau.\end{aligned}$$

Thus the UI condition can obtained via (1) of Proposition 1.5 if we can show that $E(\tau) < \infty$. To this end, consider the random walk with $R_0 = 0$ but restricted to the states $-b, \ldots, a$, where the transition probabilities for the boundaries are changed to $P(X_1 = -b|X_0 = -b) = 1 = P(X_1 = a|X_0 = a)$; a and $-b$ are now absorbing states. Then τ can be interpreted for the Markov chain as the time until absorption when initially starting at the origin; thus from finite state space Markov chain theory, $E(\tau) < \infty$.

4. *Hitting times for the simple non-symetric random walk.* We now consider the simple random walk $\{R_n\}$ in which $P(\Delta = 1) = p$, $P(\Delta = -1) = q$ and $p \ne q$; $E(\Delta) = p - q$. Although $\{R_n\}$ is no longer a MG, the transformed process $X_n = (q/p)^{R_n}$ is readily verfied to be a MG, with $X_0 = 1$. (It is a special case of a MG of the form $X_n = Y_1 \times \cdots \times Y_n$ in which $\{Y_n\}$ is i.i.d. with $E(Y) = 1$; here $Y_n = (q/p)^{\Delta_n}$.) Let τ be defined as (4), and observe that $0 \le X_{n\wedge\tau} \le (q/p)^a$, if $p < q$, and $0 \le X_{n\wedge\tau} \le (p/q)^b$, if $p > q$; yielding the fact that $X_{n\wedge\tau}$ is bounded hence UI. Let p_a and p_{-b} defined as before. Thus we conclude that

$$1 = E\{(q/p)^{R_\tau}\},$$

yielding

$$1 = (q/p)^a p_a + (q/p)^{-b} p_{-b}$$

or

$$p_a = \frac{1 - (q/p)^{-b}}{(q/p)^a - (q/p)^{-b}}. \tag{7}$$

This gives the solution to the *gambler's ruin problem* when $p \ne 0.5$.

If $q > p$ then random walk $\{R_n\}$ is negative drift transient; $lim_{n\to\infty} R_n = -\infty$ w.p.1., and R_n reaches a finite maximum $M \stackrel{\text{def}}{=} \max_n R_n$ before drifting towards $-\infty$. Thus p_a in (7) increases to $P(M \ge a)$ as $b \to \infty$ yielding

$$P(M \ge a) = \lim_{b\to\infty} p_a = (p/q)^a,\ a \ge 0, \tag{8}$$

and we conclude that

The maximum, M, of the simple random walk with negative drift has a geometric distribution with parameter p/q.

Note that $P(M = 0) = 1 - P(M \ge 1) = 1 - p/q > 0$; there is a positive probability that the random walk will never go above the origin.

1.5 Sub and super martinagles

Relaxing the equality in C2 for the definition of a MG allows for superfair and subfair games, yielding the notions of submartingales and supermartingales:

Definition 1.4 *A stochastic process* $\mathbf{X} = \{X_n : n \geq 0\}$ *is called a submartingale (SUBMG) if*

C1: $E(|X_n|) < \infty,\ n \geq 0$*, and*

(SUB)C2: $E(X_{n+1}|X_0, \ldots, X_n) \geq X_n,\ n \geq 0.$

Similarly, $\mathbf{X}$ *is called a supermartingale (SUPMG) if C1 holds and the inequality in (SUB)C2 is replaced by* $\leq$*, referred to as Condition (SUP)C2.*

In what follows, we state results in terms of submartingales, the results for super martingales being analogous due to the fact that $\{-X_n\}$ is a SUBMG iff $\{X_n\}$ is a SUPMG.

Note that if $\{X_n\}$ is a SUBMG, then $E(X_{n+1}) \geq E(X_n),\ n \geq 0$, so in particular, $E(X_n) \geq E(X_0),\ n \geq 0$. Also note that as with martingales, we can use information $\mathcal{F}_n$ that is more than needed to verify (SUB)C2; Proposition 1.1 goes through for submartingales, and we can (and will) speak of a submartingale with respect to some information $\mathcal{F}_n,\ n \geq 0$.

Examples of submartingales are easily obtained by transforming martingales. The key connection is *Jensen's inequality*, which asserts that $E(f(X)) \geq f(E(X))$ for any convex function $f = f(x)$, and any r.v. X provided that both $E(|X|) < \infty$ and $E(|f(X)|) < \infty$.

Jensen's inequality generalizes to a conditional form which asserts that

$$E(f(X)|\mathcal{F}) \geq f(E(X|\mathcal{F})),$$

for any collection of events $\mathcal{F}$, and any r.v. X provided that both $E(|X|) < \infty$ and $E(|f(X)|) < \infty$.

Proposition 1.6 *If* $\mathbf{X}$ *is a MG, and if* $f = f(x)$ *is a convex function such that* $E(|f(X_n)|) < \infty,\ n \geq 0$*, then* $\{f(X_n) : n \geq 0\}$ *is a SUBMG.*

Proof : C1 holds by the hypothesis. (SUB)C2: It is sufficient to use $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$ since $\sigma\{f(X_0), \ldots, f(X_n)\} \subset \mathcal{F}_n$. Then

$E(f(X_{n+1})|\mathcal{F}_n) \geq f(E(X_{n+1}|\mathcal{F}_n)$ (Conditional Jensen's inequality)
$= f(X_n)$ (since $\mathbf{X}$ is a MG). ■

Common examples of convex functions include $f(x) = |x|$ and $f(x) = x^2$ yielding, for example, the two subMG's $\{|R_n|\}$ and $\{R_n^2\}$ for any symetric random walk $\{R_n\}$ (as long as for $|R_n|$, $E(|\Delta|) < \infty$ and for R_n^2, $Var(\Delta) < \infty$).

A perusal of the proof of Proposition 1.4 reveals that if $\mathbf{X}$ is a SUBMG, then the stopped process defined by $\overline{X}_n = X_{n\wedge\tau}$ is a SUBMG too, and the martingale optional stopping theorem (Theorem 1.3) extends to an analogous theorem for SUBMG's:

Theorem 1.4 (Submartingale Optional Stopping Theorem) *If* $\mathbf{X} = \{X_n : n \geq 0\}$ *is a SUBMG and* τ *is a stopping time w.r.t.* $\mathbf{X}$ *such that the stopped process* $\overline{\mathbf{X}}$ *is UI, then* $E(X_\tau) \geq E(X_0)$.

More generally, we obtain

Theorem 1.5 *If* $\mathbf{X} = \{X_n : n \geq 0\}$ *is a SUBMG and both* τ_1 *and* τ_2 *are stopping times w.r.t.* $\mathbf{X}$ *such that* $P(\tau_1 < \tau_2) = 1$*, then*

$$E(X_{\tau_1 \wedge n}) \leq E(X_{\tau_2 \wedge n}),\ n \geq 0.$$

If in addition both stopped processes $\overline{\mathbf{X}}$ *are UI, then also*

$$E(X_0) \leq E(X_{\tau_1}) \leq E(X_{\tau_2}).$$

Proof : Representation (2) for stopped processes yields

$$\begin{aligned}
E(X_{\tau_1 \wedge (n+1)} - X_{\tau_1 \wedge n}) &= E((X_{n+1} - X_n) I\{\tau_1 > n\}) \\
&= E(E((X_{n+1} - X_n) I\{\tau_1 > n\} | \mathcal{F}_n)) \\
&= E(I\{\tau_1 > n\} E((X_{n+1} - X_n) | \mathcal{F}_n)) \\
&\leq E(I\{\tau_2 > n\} E(X_{n+1} - X_n) | \mathcal{F}_n)) \\
&= E(X_{\tau_2 \wedge (n+1)} - X_{\tau_2 \wedge n}),\ n \geq 0.
\end{aligned}$$

The inequality followed since $E(X_{n+1} - X_n) | \mathcal{F}_n)) \geq 0$ by the SUBMG property, and $P(\tau_1 < \tau_2) = 1$ by assumption. Thus

$$\begin{aligned}
E(X_{\tau_1 \wedge n}) &= E(X_0) + \sum_{k=0}^{n-1} E(X_{\tau_1 \wedge (k+1)} - X_{\tau_1 \wedge k}) \\
&\leq E(X_0) + \sum_{k=0}^{n-1} E(X_{\tau_2 \wedge (k+1)} - X_{\tau_2 \wedge k}) \\
&= E(X_{\tau_2 \wedge n}),
\end{aligned}$$

and we conclude that $E(X_0) \leq E(X_{\tau_1 \wedge n}) \leq E(X_{\tau_2 \wedge n}),\ n \geq 0$, where $E(X_0) \leq E(X_{\tau_1 \wedge n})$ follows since $X_{\tau_1 \wedge n}$ is a SUBMG.

Using the UI assumption we can take limits inside the expectations as $n \to \infty$ yielding $E(X_0) \leq E(X_{\tau_1}) \leq E(X_{\tau_2})$. ∎

1.6 Martingale convergence theorems

The Doob's martingale $X_n \stackrel{\text{def}}{=} E(X|\mathcal{F}_n)$ from Example 5 appears to converge, and it turns out that this martingale is the canonical example of a uniformly integrable (UI) martingale. But not all MG's are UI, and convergence is possible with the weaker condition, bounded in L^1:

Theorem 1.6 (Submartingale convergence theorem) *If* $\mathbf{X}$ *is a SUBMG which is bounded in* L^1*, that is,*

$$\sup_{n\geq 0} E(|X_n|) \leq B < \infty,$$

then

$$\lim_{n\to\infty} X_n = X, \text{ w.p.1. } \textit{for some r.v. } X, \textit{ and } E(|X|) \leq B < \infty. \tag{9}$$

Since a MG is a SUBMG, the above of course covers MG's too. Observing that for a non-negative MG, $E(|X_n|) = E(X_n) = E(X_0)$, we conclude that *every non-negative MG is bounded in* L^1, yielding

Corollary 1.3 *Every non-negative MG converges: (9) holds.*

Recalling that UI for any stochastic process implies bounded in L^1, we can apply Proposition 1.2 (and Corollary 1.1) to obtain

Corollary 1.4 (L^1 SUBMG convergence theorem) *If* $\mathbf{X}$ *is a uniformly integrable SUBMG, then (9) holds and in addition* $E(|X_n - X|) \to 0$*, and* $E(X) = \lim_{n\to\infty} E(X_n)$*.*

$E(|X_n - X|) \to 0$ is called *convergence in* L^1 of X_n to X. That is why we refer to the above result in this way. We usually denote this by $X_n \xrightarrow{L^1} X$.

1.7 Proof of the submartingale convergence theorem

The key to understanding why the Theorem is true, and how to prove it, is actually just a sample-path idea and is very intuitive: for any sequence of numbers $\{x_n : n \geq 0\}$, if the sequence does NOT converge to a constant (finite or infinite), then there must exist an interval $[a, b]$ with $a < b$ such that for infinitely many values of n it holds that $x_n < a$ and for infinitely many values of n it holds that $x_n > b$. To see this, recall that x_n does not converge if and only if $l^- \stackrel{\text{def}}{=} \underline{\lim}\, x_n < l^+ \stackrel{\text{def}}{=} \overline{\lim}\, x_n$; but if $l^- < l^+$, then we can choose any $a < b$ such that $l^- < a < b < l^+$, and visa versa.[1]

Each time the sequence passes below a and then later goes above b, we say that an *upcrossing* of $[a, b]$ has ocurred. Letting $U[a, b]$ denote the total number (over all time n) of upcrossings, we conclude that x_n does NOT converge if and only if there exists an interval $[a, b]$ with $a < b$ such that $U[a, b] = \infty$.

[1] l^- always exists since $\underline{\lim}\, x_n \stackrel{\text{def}}{=} \lim_{N\to\infty} \inf_{n\geq N} x_n$ is a non-decreasing convergence, $\inf_{n\geq N} x_n \uparrow l^-$, and similarly l^+ always exists since $\overline{\lim}\, x_n \stackrel{\text{def}}{=} \lim_{N\to\infty} \sup_{n\geq N} x_n$ is a non-increasing convergence, $\sup_{n\geq N} x_n \downarrow l^+$, where $\pm\infty$ is allowed and $l^- \leq l^+$. Convergence is the case when $l^- = l^+$ in which case the limit is $x = l^- = l^+$.

Sketch of the proof

What we will show is that for a SUBMG satisfying $\sup_{n\geq 0} E(|X_n|) < \infty$, it holds that for any interval $[a,b]$ with $a < b$, the expected number of upcrossings, $E(U[a,b])$, is finite, and hence wp1, the number of upcrossings is finite; $P((U[a,b] = \infty) = 0$. From here we will then conclude that wp1, X_n must converge; there exists a rv X such that $lim_{n\to\infty} X_n = X$, wp1. Finally we will argue that $E|X| < \infty$ (so that in particular $P(X < \infty) = 1$).

In the following, let $U_N[a,b]$ denote the number of upcrossings up to time N, and thus $U[a,b] = \lim_{N\to\infty} U_N[a,b]$.

Lemma 1.2 (Upcrossings inequality) *If $\{X_n\}$ is a SUBMG, then for any $a < b$*

$$E(U_N[a,b]) \leq \frac{E(X_N - a)^+}{b-a}.$$

Corollary 1.5 *If $\{X_n\}$ is a SUBMG that is bounded in L^1 ($\sup_{n\geq 0} E|X_n| \leq B < \infty$), then for any $a < b$*

$$E(U[a,b]) \leq \frac{|a| + B}{b-a} < \infty,$$

and so $P((U[a,b] = \infty) = 0$.

Proof : For all $N \geq 0$, $E(X_N - a)^+ \leq E|X_N| + |a| \leq B + |a|$. Moreover, by non-negativity and monotonicity (in N) of $U_N[a,b] \uparrow U[a,b]$, we have (via monotone convergence theorem) that $\lim_{N\to\infty} E(U_N[a,b]) = E(U[a,b])$. Thus the result follows from the upcrossings inequality. ■

Proof :[**Submartingale convergence theorem**]

Let $\mathcal{E} = \{X_n \text{ does not converge}\}$. Then

$$\mathcal{E} = \bigcup_{a<b} \{U[a,b] = \infty\}.$$

From Corollary 1.5, $P((U[a,b] = \infty) = 0$ for each pair $a < b$.

If we restrict our numbers $a < b$ always to be rational numbers a_i, b_i so that the number of such intervals is countable, then

$$\mathcal{E} = \bigcup_{a_i<b_i} \{U[a_i,b_i] = \infty\},$$

a countable union, and so from the basic laws of probability

$$P(\mathcal{E}) \leq \sum_{a_i<b_i} P(\{U[a_i,b_i] = \infty\}) = 0,$$

yielding $P(\mathcal{E}) = 0$. Thus there exists a rv X such that $lim_{n\to\infty} X_n = X$, wp1.

From Fatou's Lemma, we have

$$E|X| = E(|\varliminf X_n|) \leq \varliminf E(|X_n|) \leq \sup E(|X_n|) \leq B < \infty.$$

■

1.7.1 Examples

1. Recall MG's of the form $X_n = Y_1 \times Y_2 \times \cdots \times Y_n$ where $X_0 = 1$ and the Y_i are iid with mean $E(Y) = 1$. Let's take the special case when $P(Y = 0) = P(Y = 2) = 0.5$. Then since this forms a non-negative MG, Corollary 1.3 tells us that it must converge wp1 to a rv X such that $E|X| < \infty$. Clearly, as soon as one of the $Y_i = 0$, X_n remains at 0 forever after, so we see that $X = 0$ wp1.

2. *Continuation:* Let's choose Y to have a uniform distribution over the interval $(0, 2)$. Once again, X_n forms a non-negative MG hence converges to an X with $E|X| < \infty$. But what is X now?

   Here we take logarithms to obtain

   $$\ln(X_n) = \sum_{i=1}^{n} \ln(Y_i),$$

   and use the SLLN to conclude that $\ln(X_n)/n \to E(\ln(Y))$, wp1. But $E(\ln(Y)) = 0.5 \int_0^2 \ln(x)dx = \ln(2) - 1 < 0$; hence $\ln(X_n) \to -\infty$ wp1. Thus (taking exponentials), we conclude that $X_n \to e^{-\infty} = 0$, wp1, and thus $X = 0$ wp1.

3. Consider a Markov chain on $\{0, 1, 2, \ldots\}$ where $P_{0,0} = 1$ and for each $i \geq 1$,

   $$P_{i,j} = e^{-i} \frac{i^j}{j!}, \ j \geq 0.$$

   In other words, the i^{th} row forms a Poisson distribution with mean i, $i \geq 1$. In particular $E(X_{n+1}|X_n = i) = i$ yielding $E(X_{n+1}|X_n) = X_n$. But by the Markov property $E(X_{n+1}|X_n, X_{n-1}, \ldots, X_0) = E(X_{n+1}|X_n)$; thus we conclude that X_n is a non-negative MG. Hence it converges to an X with $E|X| < \infty$. It can be shown that $X = 0$ wp1: The communication classes of this chain are $C_1 = \{0\}$ and $C_2 = \{1, 2, 3 \ldots\}$. Communication classes have the nice property that all members within a class are either recurrent (all together) or transient (all together). Clearly C_2 can't be recurrent, for if it were, then if $X_0 = 1$, then the chain would visit state 1 infinitely many times and each time it did it would move next, independent of the past, into the absorbing state 0 wp $e^{-1} > 0$; so in fact (via Borel-Cantelli) it would with certainty eventually hit 0 and never come back, contradicting recurrence. Thus C_2 is a transient set, meaning that each state $i \geq 1$ has the property that it is visited only a finite number of times and then never again. The only way that can happen is if either the chain converges to ∞ or to 0. But it can't converge to ∞ since the martingale convergence theorem says that the limit X satisfies $E|X| < \infty$. Thus $X = 0$ wp1 as was to be shown. Note that this is so regardless of initial conditions, $X_0 = i$.