Plan Martingales

1. Basic Definitions
2. Examples
3. Overview of Results

Reading: G&S Section 12.1-12.4
Next Time: More Martingales

Midterm Exam: Tuesday 28 March in class
Sample exam problems ("Homework 5") available tomorrow at the latest

A Peculiar Etymology 1.

1. *Horse Riding.* A strap or arrangement of straps fastened at one end to the noseband, bit, or reins of a horse and at the other to its girth, in order to prevent it from rearing or throwing its head back, or to strengthen the action of the bit.
2. *Nautical.* A stay which holds down the jib-boom of a square-rigged ship, running from the boom to the dolphin-striker.
3. *Gambling.* Any of various gambling systems in which a losing player repeatedly doubles or otherwise increases a stake such that any win would cover losses accrued from preceding bets.
4. *Probability and Statistics.* stay tuned...

The Basic Definition 2. (from which the more general definition below will spring, and which will be the most common case, by why by specific when we can be general, eh?)

A stochastic process $(X_n)_{n \ge 0}$ is a *martingale* if

1. $\mathsf{E}|X_n| < \infty$ for all $n \ge 0$, and
2. $\mathsf{E}(X_{n+1} \mid X_0, \ldots, X_n) = X_n$.

This definition will be superseded below but is the base case.

Useful metaphor: Let X_n be a gambler's wealth at after the nth bet. This definition of a martingale captures a notion of a "fair" game. How?

The Original Martingale 3. Suppose you start with a large fortune w_0. (How large it needs to be, we will see.) You are offered a series of bets with a probability of winning equal to 1/2. (Yeah, right.)

The original martingale strategy is to bet \$$d$, say, on the first bet. If you win, stop. If you lose, bet \$$2d$ on second bet. Continue this way, stopping play as soon as you win and betting \$$2^n d$ on the $n+1$st bet.

You will win eventually, say at bet $t \ge 1$, at which point you will have won

$$2^{t-1}d - (d + 2d + 2^2 d + \cdots 2^{t-2}d) = d \tag{1}$$

dollars, where the second term is zero when $t = 1$. Seems like a sure thing, eh?

Let W_n denote your wealth after the nth bet in a series of bets like the above but without the stopping rule. Then, $W_0 = w_0$ and

$$W_n = w_0 + \sum_{k=1}^{n} \Xi_k d 2^{k-1}, \tag{2}$$

where Ξ_k are IID Bernoulli$\langle 1/2 \rangle$.

Now, (W_n) is a random walk, with $\mathsf{E}|W_n| < \infty$ and

$$\mathsf{E}(W_{n+1} \mid W_0, \ldots, W_n) = \mathsf{E}(W_{n+1} \mid \Xi_0, \ldots, \Xi_n) = W_n + \mathsf{E}(d2^n \Xi_{n+1}) = W_n. \tag{3}$$

Hence, $(W_n)_{n\geq 0}$ is a martingale by the above definition. Still, it's not quite what we want.

Let $T = \min n \geq 1 \colon \Xi_n = 1$ be the first bet that you win. What can we say about T? Well first off, it's Geometric$\langle 1/2 \rangle$, that is, $\mathsf{P}\{T = n\} = 2^{-n}$ for $n \geq 1$. But wait, is there more? There's something familiar abou this time....

$X_0 = W_0 = w_0$. This is your wealth. Define $X_n = W_{T\wedge n}$ for $n \geq 1$. This is your wealth on the the martingale system. Any questions?

Loosely, if $n \geq T$, $X_{n+1} = X_n$. If $n < T$, then $X_{n+1} = W_{n+1}$ and $X_n = W_n$. So, (X_n) should be a martingale. Formally, we need to show that $\mathsf{E}|X_n| < \infty$. And that $\mathsf{E}(X_{n+1} \mid X_0, \ldots, X_n) = X_n$.

Note that

$$X_{n+1} 1\{T \leq n\} = X_n 1\{T \leq n\} \tag{4}$$

$$X_{n+1} 1\{T > n\} = (X_n + \Xi_{n+1} d2^n)\, 1\{T > n\}. \tag{5}$$

That is,

$$X_{n+1} = X_n 1\{T \leq n\} + (X_n + \Xi_{n+1} d2^n)\, 1\{T > n\}. \tag{6}$$

What do we need to bring this home?

... the class steps in ...

...and thus we get that $(X_n)_{n\geq 0}$ is a martingale as defined above.

But would you want to use this strategy. Let D be the biggest debt you owed prior to winning. That is, $D = |X_{T-1}|$. Then,

$$\mathsf{E}D = \sum_{n=1}^{\infty} 2^{-n} d \sum_{k=0}^{n-2} 2^k = d \sum_{n=1}^{\infty} 2^{-n}(2^{n-1} - 1). \tag{7}$$

Even Bill Gates should think twice about using this approach.

Example 4. Another approach to Gambler's Ruin

Let S_n be a simple random walk with $S_0 = w \in \{0, \ldots, N\}$. Suppose we stop the walk when it first hits either 0 or N. Which will it hit first?

Write $S_n = w + \sum_{k=1}^n \Xi_k$ where the Ξ_K are IID Bernoulli$\langle p \rangle$. Let $q = 1 - p$.

Define $Y_n = \left(\frac{q}{p}\right)^{S_n}$. Let T be the first hitting time of $\{0, N\}$. We want to understand the limiting behavior of $X_n = Y_{T \wedge n}$.

As before,

$$X_{n+1} = X_n 1\{T \geq n\} + \left(\frac{q}{p}\right)^{S_n + \Xi_{n+1}} 1\{T < n\}. \tag{8}$$

Note that T is a stopping time with respect to $\Xi_1, \ldots, \Xi_n$. Hence,

$$\begin{aligned}
\mathsf{E}(X_{n+1} \mid \Xi_1, \ldots, \Xi_n) &= \mathsf{E}(X_n 1\{T \geq n\} \mid \Xi_1, \ldots, \Xi_n) + \mathsf{E}\left(\left(\frac{q}{p}\right)^{S_n + \Xi_{n+1}} 1\{T < n\} \,\middle|\, \Xi_1, \ldots, \Xi_n\right) && (9)\\
&= X_n 1\{T \geq n\} + \left(\frac{q}{p}\right)^{S_n} \mathsf{E}\left(\left(\frac{q}{p}\right)^{\Xi_{n+1}} \,\middle|\, \Xi_1, \ldots, \Xi_n\right) 1\{T < n\} && (10)\\
&= X_n 1\{T \geq n\} + \left(\frac{q}{p}\right)^{S_n} \mathsf{E}\left(\frac{q}{p}\right)^{\Xi_{n+1}} 1\{T < n\} && (11)\\
&= X_n 1\{T \geq n\} + \left(\frac{q}{p}\right)^{S_n} (p(q/p) + q(p/q)) 1\{T < n\} && (12)\\
&= X_n 1\{T \geq n\} + X_n 1\{T < n\} && (13)\\
&= X_n. && (14)
\end{aligned}$$

This isn't quite the definition, though. Perhaps we would work it out that the information in $\Xi_1, \ldots, \Xi_n$ is the same as the information in $X_0, \ldots, X_n$. A good strategy. But what a bother. Take this as motivation for the general definition below.

Now, $\mathsf{E}Y_n = (q/p)^w$ for all n, so doesn't it make sense that $\mathsf{E}Y_T$ would be the same? And thus, $\mathsf{E}X_n = (q/p)^w = \mathsf{E}X_T$ for all n. Hmmm...let's suppose it's true. Take this as motivation for one of the main results later.

If true, then

$$\mathsf{E}X_T = \left(\frac{q}{p}\right)^0 r_w + \left(\frac{q}{p}\right)^N (1 - r_w) = \left(\frac{q}{p}\right)^w, \tag{15}$$

where r_w is the ruin probability with initial wealth w. Then,

$$r_w = \frac{\left(\frac{q}{p}\right)^w - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N} n \tag{16}$$

as long as $p \neq \frac{1}{2}$. That's what we got before. Nice.

We've seen a martingale argument for the $p = \frac{1}{2}$ case as well. A lot of power in that simple assumption.

Intermediate Definition 5. A sequence of random variables $(Y_n)_{n\geq 0}$ is a martingale with respect to a sequence $(X_n)_{n\geq 0}$ if for all $n \geq 0$,

1. $\mathsf{E}|Y_n| < \infty$
2. $\mathsf{E}(Y_{n+1} \mid X_0, \ldots, X_n) = Y_n$.

If $X_n = Y_n$ for all n, this reduces to the previous definition.

Definition 6. A *filtration* in a σ-field $\mathcal{F}$ is a sequence $(\mathcal{F}_n)_{n\geq 0}$ of sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ for all $n \geq 0$.

Write $\mathcal{F}_\infty = \lim_{n\to\infty} \mathcal{F}_n$ for the smallest σ-field containing all the $\mathcal{F}_n$s.

A sequence of random variables $(X_n)_{n\geq 0}$ is *adapted* to the filtration if X_n is $\mathcal{F}_n$-measurable for all $n \geq 0$. That is, events $\{X_n \in A\} \in \mathcal{F}_n$.

Intuitive unpacking follows.

General Definition 7. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration in $\mathcal{F}$ and let $(Y_n)_{n\geq 0}$ be a sequence of random variables adapted to that filtration.

Then, (Y_n) is a *martingale (with respect to* $(\mathcal{F}_n)$*)* if

1. $\mathsf{E}|Y_n| < \infty$
2. $\mathsf{E}(Y_{n+1} \mid \mathcal{F}_n) = Y_n$.

If $\mathcal{F}_n = \sigma(Y_0, \ldots, Y_n)$, we get the basic definition. If $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$, we get the intermediate definition. More intuitive unpacking ....

Definition 8. Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration in $\mathcal{F}$ and let $(Y_n)_{n\geq 0}$ be a sequence of random variables adapted to that filtration.

Then, (Y_n) is a *sub-martingale (with respect to* $(\mathcal{F}_n)$*)* if

1. $\mathsf{E}\max(Y_n, 0) < \infty$
2. $Y_n \leq \mathsf{E}(Y_{n+1} \mid \mathcal{F}_n)$.

And (Y_n) is a *super-martingale (with respect to* $(\mathcal{F}_n)$*)* if

1. $\mathsf{E}\max(-Y_n, 0) < \infty$
2. $Y_n \geq \mathsf{E}(Y_{n+1} \mid \mathcal{F}_n)$.

Questions 9.

a. Show that Y_n is a martingale if and only if it is a sub-martingale and a super-martingale.
b. Suppose that Y_n is a sub-martingale, what can you say about $-Y_n$?
c. Find three examples of a sub- or super-martingale.

Example 10. Simple Random Walk

Let S_n be a simple random walk as we have defined so often. Then, $\mathsf{E}|S_n| < \infty$ for all n because $|S_n| \le n$ and

$$\mathsf{E}(S_{n+1} \mid S_0, \ldots, S_n) = S_n + (p - q), \tag{17}$$

Note that $\sigma(S_0, \ldots, S_n) = \sigma(S_0, \Xi_1, \ldots, \Xi_n)$. Define $X_n = S_n - n(p - q)$. Then X_n is a martingale. (Why?)

Example 11. Sums of Random Variables

The same trick works with more general sums. Suppose that $(X_n)_{n\ge 0}$ are independent random variables with $\mathsf{E}|X_n| < \infty$. Let $Y_n = X_1 + \cdots + X_n$ for $n \ge 0$. Then, (Y_n) is a martingale by the same argument. (Make it.)

We can generalize this. Let $(X_n)_{n\ge 0}$ be a sequence of real-valued random variables. Let g_k and h_k be functions on $\mathbb{R}^k$ with the $|h_k| \le H_k < \infty$ for some constants H_k. Let f be a function so that $\mathsf{E}|f(g_{k+1}(X_0, \ldots, X_k))| < \infty$. Let $Z_k = f(g_{k+1}(X_0, \ldots, X_k))$.

Define Y_n by

$$Y_n = \sum_{k=0}^{n} (Z_k - \mathsf{E}(Z_k \mid X_0, \ldots, X_{k-1}))\, h_k(X_0, \ldots, X_{k-1}), \tag{18}$$

Then, $(Y_n)_{n\ge 0}$ is a martingale. Why?

This is example is not so interesting by itself, but it is quite a general mechanism for constructing martingales. And it helps you unpack complicated expressions.

Example 12. Variance of a Sum

Let $(X_n)_{n\geq 0}$ be IID random variables with $X_0 = 0$ and $\mathsf{E}X_n = 0$ and $\mathsf{E}X_n^2 = \sigma^2$ for $n \geq 1$. Define

$$Y_n = \left(\sum_{k=1}^{n} X_k\right)^2 - n\sigma^2. \tag{19}$$

Then (Y_n) is a martingale. Show this.

What can you say about $M_n = \left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)^2 - \frac{\sigma^2}{n}$?

Example 13. The Doob Process

Let X be a random variable with $\mathsf{E}|X| < \infty$. Let $(Z_n)_{n\geq 0}$ be an arbitrary sequence of random variables.

Define $X_n = \mathsf{E}(X \mid Z_0, \ldots, Z_n)$.

Note that $\mathsf{E}|X_n| \leq \mathsf{E}|X|$. (Why?)

And

$$\begin{aligned}
\mathsf{E}(X_{n+1} \mid Z_0, \ldots, Z_n) &= \mathsf{E}(\mathsf{E}(X \mid Z_0, \ldots, Z_{n+1}) \mid Z_0, \ldots, Z_n) && (20)\\
&= \mathsf{E}(X \mid Z_0, \ldots, Z_n) && (21)\\
&= X_n, && (22)
\end{aligned}$$

by the Mighty Conditioning Identity.

Example 14. Harmonic Functions on Markov Chains

Let X be a countable-state Markov chain on $\mathcal{S}$ with transition probabilities P. Let Δ be the drift operator: $\Delta V = PV - V$.

Recall that any V for which $\Delta V = 0$ we called *harmonic*. This is a function for which

$$\mathsf{E}(V(X_{n+1}) \mid X_n) = V(X_n). \tag{23}$$

Hmmm...

Define $Y_n = V(X_n)$ for a harmonic V. As long as $\mathsf{E}|V(X_n)|$ is finite, we've got a martingale. This will hold, for example, if we choose a bounded harmonic function. This is a useful mechanism for discovering martingales in Markov Chains.

Example 15. Eigenvector Induced Martingales for Markov Chains

A slight generalization. Now, let V be an eigenfunction (eigenvector) of P with eigenvalue λ. That is, for all $s \in \mathcal{S}$,

$$\sum_{s' \in \mathcal{S}} P(s, s')V(s') = \lambda V(s), \tag{24}$$

or equivalently,

$$\mathsf{E}(V(X_{n+1}) \mid X_n) = \lambda V(X_n). \tag{25}$$

So, define

$$Y_n = \lambda^{-n} V(X_n) \tag{26}$$

for such a V and $n \geq 0$. If $\mathsf{E}|V(X_n)| < \infty$, we have

$$\begin{aligned} \mathsf{E}(Y_{n+1} \mid X_0, \dots X_n) &= \mathsf{E}\left(\lambda^{-(n+1)} V(X_{n+1}) given X_0, \dots, X_n\right) & (27) \\ &= \lambda^{-n}\lambda^{-1}\mathsf{E}(V(X_{n+1}) \mid X_n) & (28) \\ &= \lambda^{-n}\lambda^{-1}\lambda V(X_n) & (29) \\ &= Y_n, & (30) \end{aligned}$$

then (Y_n) is a martingale with respect to X.

This has a direct application to Branching Processes which we'll see in the near future.

Example 16. Discretization and Derivatives

Let U be a Uniform$\langle 0, 1\rangle$ random variable. Define $X_n = k2^{-n}$ for the unique k such that $k2^{-n} \leq U < (k+1)2^{-n}$. As n increases, X_n gives finer and finer information about U.

Let f be a bounded function on $[0, 1]$ and define

$$Y_n = 2^n \left(f(X_n + 2^{-n}) - f(X_n)\right). \tag{31}$$

What is Y_n approximating here as $n \to \infty$?

What is the distribution of U given $X_0, \dots, X_n$?

Show that Y_n is a martingale wrt X_n.

Past Example 17. Likelihood Ratios
Example from homework. Very important in some applications such as sequential analysis.

Future Example 18. False Discovery Rates
Next time will show how a martingale argument proves the result of Benjamini and Hochberg (1995).

Outline 19. Main Results for (sub and super) martingales

1. Decomposition into martingale plus predictable process
2. Strong convergence theorems
3. Upcrossing Inequalities
4. Large Deviation Bounds
5. Maximal Inequalities
6. Optional Sampling of Process at Stopping Times
7. Optional Stopping of Process at Stopping Times