Plan Martingales cont'd

0. Questions for Exam
2. More Examples
3. Overview of Results

Reading: study
Next Time: first exam

Midterm Exam: Tuesday 28 March in class
Sample exam problems ("Homework 5") and HW4 solutions on-line

Example 1. 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

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

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 2. 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

$$\mathsf{E}(X_{n+1} \mid Z_0, \ldots, Z_n) = \mathsf{E}(\mathsf{E}(X \mid Z_0, \ldots, Z_{n+1}) \mid Z_0, \ldots, Z_n) \tag{2}$$

$$= \mathsf{E}(X \mid Z_0, \ldots, Z_n) \tag{3}$$

$$= X_n, \tag{4}$$

by the Mighty Conditioning Identity.

Example 3. 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

$$\mathsf{E}(V(X_{n+1}) \mid X_n) = V(X_n). \tag{5}$$

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.

(Similarly, a V with $\Delta V \geq 0$ or $\Delta V \leq 0$ is called subharmonic or superharmonic respectively. What do you think we get in that case?)

Example 4. 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 $s \in \mathcal{S}$,

$$\sum_{s' \in \mathcal{S}} P(s, s') V(s') = \lambda V(s), \tag{6}$$

or equivalently,

$$\mathsf{E}(V(X_{n+1}) \mid X_n) = \lambda V(X_n). \tag{7}$$

So, define

$$Y_n = \lambda^{-n} V(X_n) \tag{8}$$

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

$$\mathsf{E}(Y_{n+1} \mid X_0, \ldots X_n) = \mathsf{E}\left(\lambda^{-(n+1)} V(X_{n+1}) \mid X_0, \ldots, X_n\right) \tag{9}$$

$$= \lambda^{-n} \lambda^{-1} \mathsf{E}(V(X_{n+1}) \mid X_n) \tag{10}$$

$$= \lambda^{-n} \lambda^{-1} \lambda V(X_n) \tag{11}$$

$$= Y_n, \tag{12}$$

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 5. 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 $k2^{-n} \le U < (k+1)2^{-n}$. As n increases, X_n gives finer and finer information about U.

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

$$Y_n = 2^n \left(f(X_n + 2^{-n}) - f(X_n) \right). \tag{13}$$

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

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

Show that Y_n is a martingale wrt X_n.

Past Example 6. Likelihood Ratios
This is an example from homework but still worth reminiscing about.

Let f_0 and f_1 be known PDFs. Let $Y_0, \ldots, Y_n$ be an IID sample from a PDF f. We know that f is either f_0 or f_1, but we don't know which. One way to infer the generating density is through a likelihood ratio test of

$$\begin{aligned} \mathrm{H}_0 &: f = f_0 \\ \mathrm{H}_1 &: f = f_1 \end{aligned}$$

Define

$$R_n = \frac{f_0(Y_0) \cdots f_0(Y_n)}{f_1(Y_0) \cdots f_1(Y_n)}. \tag{14}$$

for $n \geq 0$. R_n is the likelihood ratio statistic computed from data up to index n. (We can assume $f_0 > 0$ on the support of f_1 to make this well defined, but the notion extends beyond this.)

Suppose the null hypothesis (H_0) is true. Then expected values of the Ys are taken relative to f_0, so

$$\mathsf{E}\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} = \int \frac{f_1(y)}{f_0(y)} f_0(y)\,\mathrm{d}y = \int f_1(y)\,\mathrm{d}y = 1. \tag{15}$$

Thus,

$$\begin{aligned} \mathsf{E}(R_{n+1} \mid Y_0, \ldots, Y_n) &= \mathsf{E}\left(R_n \frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \,\middle|\, Y_0, \ldots, Y_n \right) && (16) \\ &= R_n \mathsf{E}\left(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \,\middle|\, Y_0, \ldots, Y_n \right) && (17) \\ &= R_n \mathsf{E}\left(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} \right) && (18) \\ &= R_n. && (19) \end{aligned}$$

Under the null hypothesis R_n is a martingale with respect to the Ys.

Under the alternative hypothesis (H_1), however, if there is a region with high f_1 probability but low f_0 probability, then the R_ns will tend to grow. If $\mathsf{E}\left(\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}\right) > 1$ in this case, the process will be a submartingale. For example, when f_0 is a Normal$\langle 0, 1\rangle$ and f_1 is a Normal$\langle 1, 1\rangle$, the expected ratio is about 2.7 under the alternative.

Future Example 7. False Discovery Rates
Suppose we have m hypothesis tests and corresponding p-values $P_1, \ldots, P_m$. Let $P_{(0)} \equiv 0$ and $P_{(1)} < \cdots P_{(m)}$ are the p-values in increasing order. Then, Benjamini and Hocbherg (1995) used the following threshold to determine which null hypotheses would be rejected.

$$P^* = \max\left\{P_{(i)} \colon P_{(i)} \le \alpha \frac{i}{m}\right\} \tag{20}$$

They showed that if we reject all null hypothesis for which $P_i \le P^*$, we will control the *false discovery rate* at level α. That is,

$$\mathsf{E}\frac{\text{falsely rejected nulls}}{\text{rejected nulls}} \le \alpha. \tag{21}$$

We will see later how to prove this using a cute martingal argument.

Outline 8. Main Results for (sub and super) martingales
Just to give you a flavor for what we can do with this.

1. Decomposition into martingale plus predictable process
   A submartingale can
2. Strong convergence theorems
   If Y is a submartingale with $\mathsf{E}Y_n^+ \le M < \infty$ for some M and all n, then there exists a random variable Y_∞ such that Y_n converges to Y_∞ with probability 1. Under some conditions, this can be made even stronger.
3. Crossing Inequalities
   Bounds on the expected number of crossings of any level.
4. Large Deviation Bounds
   Bounds on the probabilities of large excursions in the process.
5. Maximal Inequalities
   Example: if X is a submartingale, then

$$\mathsf{P}\left\{\max_{1 \le i \le n} X_i > u\right\} \le \frac{\mathsf{E}X_n^+}{u}. \tag{22}$$

6. Optional Sampling of Process at Stopping Times
   If $T_1 \le T_2 \le \cdots$ are stopping times satisfying some conditions (such as being bounded by a sequence of deterministic times), then a submartingale Y *sampled* at these times remains a submartingale.
7. Optional Stopping Theorem
   Let X be a martingale and T a stopping time. Then, $\mathsf{E}X_T = \mathsf{E}X_0$ if

   A. $\mathsf{P}\{T < \infty\} = 1$,
   B. $\mathsf{E}|X_T| < \infty$,
   C. $\mathsf{E}(X_n 1\{T > n\}) \to 0$ as $n \to \infty$.

   These conditions, fairly mild, can be simplified in some cases.