Stats 150, Fall 2019
Lecture 23, Thursday, 11/21/2019

1 Warm-up Examples of Martingales

As warm-up, let's consider these examples. Suppose that $\mathbb{E}(Y|X) = \Psi(X)$. We'll look at two things:
1) Suppose that we don't know X, but we know $\Psi(X)$. Then what is

$$\mathbb{E}(Y|\Psi(X))?$$

2) If b is a bijection, then what is

$$\mathbb{E}(Y|b(X))?$$

These may be on our final exam. Pitman suggests that we think of this in terms of heights of students in rows or some other simulation or way of visualizing these problems.

2 Stopping Times for Martingales

There's quite a nice discussion of this in Durett. The setup here is to take $\Omega, ?, \mathbb{P})$ as our probability space, and $X_0, X_1, X_2, \dots$ is some background process for each $X_n = (X_n(\omega), \omega \in \Omega)$. This may or may not be numerical, as this is unimportant (e.g. X_n can be a vector in $\mathbb{R}^p$).
We'll say that $(M_n, n = 0, 1, 2, \dots)$ is a martingale, relative to $X_0, X_1, \dots)$ **if and only if**

$$\mathbb{E}(M_n \mid M_n, X_n, M_{n-1}, X_{n-1}, \dots) = M_n,$$

or equivalently

$$\mathbb{E}(M_{n-1} - M_n \mid M_n, X_n, \dots) = 0.$$

Recall that we need to assume absolute convergence, which is that $\mathbb{E}|M_n| < \infty, \forall_n$, where M_n is a function of $M_0, X0, X_1, \dots, X_n$.

Recall that our **stopping time** T, relative to $(X_0, X_1, \dots)$, takes on values in $\{0, 1, 2, \dots\}$ (we don't have a discussion if we allow T to take on value 0). This definition is equivalent to any of the following:
1) $(T = n)$ is determined by $(X_0, X_1, \dots, X_n)$ for all $n = 0, 1, 2, \dots$. This is an event but we can think of it as a Boolean function on the inputs X_i.
2) $(T \leq n)$ is determined by $(X_0, X_1, \dots, X_n)$ for all $n = 0, 1, 2, \dots$.
3) $(T > n)$ is determined by $(X_0, X_1, \dots, X_n)$ for all $n = 0, 1, 2, \dots$.
We need only subtract one from another to derive between the above. We decide values of T sequentially in a way that depends on $X_0, X_1, \dots$. It is no problem to allow M_0 to play a role, which is to replace $X_n \mapsto (X_n, M_n)$.

Let's recall something we know about stopping times, in this fairly large level of generality. Special to the structure of Markov chains, we have the Strong Markov Property. There's another special rule that applies to iid sequences $X_1, X_2, X_3, \dots$.

2.1 Case: IID Sequence

Assume that X_0 is anything, X_1, X_2, X_3 are iid real-valued random variables like X, with $\mathbb{E}|X| < \infty$.
Then we have a few key points:
1) $S_n - n\mathbb{E}X$ (after centering) is a martingale relative to $(X_0, X_1, X_2, \dots)$.
2) If T is a stopping time and $\mathbb{E}(T) < \infty$, then

$$\mathbb{E}S_T = (\mathbb{E}T)(\mathbb{E}X),$$

which is Wald's identity. We've proved in this the course in Lecture 12.
Notice that if we set $M_n := S_n - n\mathbb{E}X$, which is our martingale of the moment, then Wald's identity gives

$$\mathbb{E}M_T = \mathbb{E}M_0 = \mathbb{E}0 = 0.$$

Unpacking this a bit, assuming that $\mathbb{E}T < \infty$, we have $M_T = S_T - T\mathbb{E}(X)$, and so $\mathbb{E}M_T = 0$. This implies

$$\mathbb{E}S_T = (\mathbb{E}T)\mathbb{E}(X).$$

We have preservation of the expectation that works at the stopping time.
Notice that $S_0 = 0$ is to make empty sums ($\emptyset$). For example, we woudl like to include the case where T is independent of $(X_1, X_2, \dots)$. To see this, we need to go back to the definition of stopping times and realize that X_0 is included to allow that T is not only a function of $(X_1, X_2, \dots)$, and hence can be independent. Alternatively, we can even let $X_0 := T$ as an artificial way. This is why we are deliberately not including X_0 in our sum, including this only to allow for generality and ad hoc randomness.

2.2 Generalizing Our Example

Now the question arises: Suppose $(M_n, n = 0, 1, 2, \dots)$ is a martingale, and T is a stopping time with $\mathbb{E}T < \infty$. We may ask: Is it true that $\mathbb{E}M_t = \mathbb{E}M_0$? This worked in the iid case, but does this work more generally?

Pitman answers that **no**, we do not have this unnecessarily if T is **unbounded.** However, we do have this property if T is bounded. Let's define what this means.

Definition: Bounded -

We say that T is bounded if there is some **fixed** number b where

$$\mathbb{P}(T \leq b) = 1.$$

That is, it is certain that T is less than or equal to the bound.

2.3 Issues When T is Unbounded

Example: From the previous class, we consider an exponential or multiplicative martingale. Let $X_1, X_2, \dots$ be iid Bernoulli $\left(\frac{1}{2}\right)$ random variables. We are simply doing fair coin tosses to decide our X_i, so that X has value 0 with probability $\frac{1}{2}$, and value 1 with probability $\frac{1}{2}$. In formulas, we'll define our game:

$$M_0 := 1 \quad (\emptyset \text{ product } = 0)$$
$$M_n := 2^n X_1 X_2 \cdots X_n = \prod_{k=1}^{n} (2X_k),$$

which we know commonly as the “double-or-nothing” gambling game. This is more or less historically the first major use of the martingale term in a gambling context.
Notice that 0 is an absorbing state in this game. We should check that this is a martingale by a prior discussion. A very simple check would be to see that

$$\begin{aligned}\mathbb{E}M_n &= \mathbb{E}\left(\prod_{k=1}^{n} 2X_k\right)\\ &= \prod_{k=1}^{n} \underbrace{(\mathbb{E}2X_k)}_{2\cdot\frac{1}{2}=1}.\end{aligned}$$

Also, we should double-check (Pitman jokes as this is a double-or-nothing game) that:

$$\mathbb{P}(M_n = 2^n) = \left(\frac{1}{2}\right)^n = 2^{-n}.$$

So there is a very small chance that we end up with a very large fortune. On the other hand, we have the other path:

$$\mathbb{P}(M_n = 0) = 1 - 2^{-n}.$$

To verify this, see that:

$$\mathbb{E}M_n = 2^{-n} \cdot 2^n + (1 - 2^{-n}) \cdot 0 = 1.$$

We looked at this example to see that there can be issues (trouble) at particular stopping times. Recall that we've defined that $T := \min\{n : M_n = 0\}$. Notice that $\mathbb{P}(T > n) = 2^{-n}$ so that T is unbounded. Then:

$$\begin{aligned}\mathbb{E}T &= \sum_{n=0}^{\infty} \mathbb{P}(T > n)\\ &= \sum_{n=0}^{\infty} 2^{-n}\\ &= 2 < \infty.\end{aligned}$$

and so $T \sim \text{Geometric}\left(\frac{1}{2}\right)$ on $\{1, 2, 3, \dots\}$. This is good because

$$\mathbb{P}(T < \infty) = 1.$$

However, $\mathbb{P}(M_T = 0) = 1$. So in summary, we found:

$$1 = \mathbb{E}M_0 = \mathbb{E}M_1 = \mathbb{E}M_2 = \cdots, \qquad M_n \to M_T = 0,$$

but it turns out $\boxed{\mathbb{E}M_T \neq 1}$.

Remark: If we have unbounded stopping times, the preservation of expectations can fail.

Let's look at a generic picture of a CDF, and assume that $X \geq 0$ for simplicity. The shaded region above the CDF is $\mathbb{E}(X)$. One way to understand this is to think of creating a uniform random variable U on $[0, 1]$ and pick the value $X = X(U)$ to be the inverse CDF.
Now for our present martingale, we have a sequence of pictures.
Notice that every one of these areas is 1, but the shaded region (rectangle) is getting much longer. In the limit, we are guaranteed to lose our money (with probability 1) even in this fair gamble.

Break time.

3 Expectations In Bounded Stopping Times

If we assume that (M_n) is a martingale relative to $(X_0, X_1, \dots)$, and T is a **bounded** stopping time, then

$$\mathbb{E}M_0 = \mathbb{E}M_T.$$

First, we'll want to see that is it true that $\mathbb{E}M_T$ can be found by conditioning on T. If $T \le b$, then:

$$\begin{aligned} M_T &= \underbrace{\sum_{n=0}^{b} \mathbb{1}(T=n)}_{1} M_T \\ &= \sum_{n=0}^{b} M_n \mathbb{1}(T=n). \end{aligned}$$

Then this tells us that we can certainly compute that

$$\begin{aligned} \mathbb{E}M_T &= \sum_{n=0}^{b} \mathbb{E}\left[M_n \mathbb{1}(T=n)\right] \\ &= \sum_{n=0}^{b} \underbrace{\mathbb{E}(M_n | T=n)} \cdot \mathbb{P}(T=n) \quad \text{(equivalently by conditioning)}. \end{aligned}$$

This underbraced part can be problematic. Now if we assume that T is independent of (M_n), then we can continue in the obvious way to get:

$$\begin{aligned} \mathbb{E}M_T &= \sum_{n=0}^{b} (\mathbb{E}M_n)\mathbb{P}(T=n) \\ &= \sum_{n=0}^{b} (\mathbb{E}M_0)\mathbb{P}(T=n) \\ &= \mathbb{E}(M_0) \sum_{n=0}^{b} \mathbb{P}(T=n) \\ &= (\mathbb{E}M_0) \cdot 1 \\ &= \mathbb{E}M_0. \end{aligned}$$

Pitman suggests that we formulate examples where $\mathbb{E}(M_n | T=n) \neq \mathbb{E}M_0$. This would mean that our approach fails as we have just done, which might lead us to think our result is incorrect. However, what happens is that if there are terms that are larger, then this must be compensated by some terms that are smaller.

Now we have a key idea for how to proceed without this large assumption that T is independent of (M_n). We'll take T to be a stopping time, even allowing the case where $\mathbb{P}(T=\infty) > 0$.
Notice that if (M_n) is a martingale relative to (M_n) and T is a stopping time relative to (X_n), then

$$(M_{n \wedge T}, n = 0, 1, 2, \dots)$$

is a martingale relative to $(X_0, X_1, \dots)$. Recall that we've defined:

$$\begin{aligned} n \wedge m &:= \min\{n, m\} \\ n \vee m &:= \max\{n, m\}. \end{aligned}$$

So we have:

$$M_{T\wedge n} := \begin{cases} M_T, & T < n \\ M_n = M_T, & T = n \\ M_n, T > n. \end{cases}$$

Intuitively, in the fair gambling game, we can choose to stop and walk away with our money at any (stopping) point in time. Now let's look at the mathematical proof.

As a matter of technique, we will check the following:

1) $M_{n\wedge T}$ is a function of $(M_0, X_0, M_1, X_1, \ldots, M_n, X_n)$.

$$M_{n\wedge T} = \sum_{k=0}^{n-1} \underbrace{\mathbb{1}(T = k)}_{f(M_0, X_0, \ldots, M_k, X_k)} M_k + \underbrace{\mathbb{1}(T \geq n)}_{f(M_0, X_0, \ldots, M_n, X_n)} M_n,$$

where we can evaluate these indicators by definition of stopping time.

2) We need to check the martingale property, and this is most often accessible via the difference form. That is, we check that

$$\mathbb{E}(M_{(n+1)\wedge T} - M_{n\wedge T} \mid X_0, M_0, \ldots, X_n, M_n) = 0.$$

Notice that after we've stopped this, the difference is trivial and there is nothing to check. It remains to check for the case when we have not yet stopped. Via the method of indicators, we have:

$$M_{(n+1)\wedge T} - M_{n\wedge T} = \mathbb{1}(T \leq n) \cdot 0 + \mathbb{1}(T > n) \cdot (M_{n+1} - M_n),$$

which directly implies that

$$\mathbb{E}\left[M_{(n+1)\wedge T} - M_{n\wedge T}\right] \mathbb{1}(A_v) = \mathbb{E}(M_{n+1} - M_n)\mathbb{1}(T > n)\mathbb{1}(A_v) = 0,$$

for all A_v functions of $(X_0, M_0, \ldots, X_n, M_n)$.

Pitman closes today's lecture with a life lesson.

Ultimate Rule for Cooking Perfect Toast:
Cook until you see the toast smokes, and cook for 10 seconds less.

This is one intricate example of what is invalid as a stopping time. As another example, we cannot play in the stock market and stop just before we lose (we cannot see into or condition on the future).

Lecture ends here.