University of Washington Math 523A Lecture 1
Martingales: definitions and examples

Lecturer: Eyal Lubetzky

March 30, 2009

Basic definitions

Let $(\Omega, \mathcal{F})$ be a measurable space. A *filtration* $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \ldots \subset \mathcal{F}$ is an increasing sequences of sub-σ-algebras of $\mathcal{F}$. A sequence of random variables (X_t) is *adapted* to a filtration $(\mathcal{F}_t)$ if X_t is $\mathcal{F}_t$-measurable for all t.

Given a stochastic process, one can think of $(\mathcal{F}_t)$ as the "history so far". In many cases, it will be useful to consider the natural filtration generated by X_t, i.e., $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$ is the smallest σ-algebra in which $\{X_i : i \le t\}$ are measurable. In other cases, the stochastic process will include some extra randomness beyond that which is observed in the variables X_i.

Throughout the course, we will use the fact that $\mathbb{E}[\mathbb{E}[X|A]] = \mathbb{E}X$ (the so-called "tower of expectations" property). In particular, if we are conditioning on a sequence of random variables and have

$$\mathbb{E}\left[X \mid Y_1, Y_2, \ldots, Y_k\right] = f(Y_1)$$

for some function f, then

$$\mathbb{E}\left[X \mid Y_1\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid Y_1, Y_2, \ldots, Y_k\right] \mid Y_1\right] = \mathbb{E}\left[f(Y_1) \mid Y_1\right] = f(Y_1) \ .$$

In general, whenever we have a sub-σ-algebra, $\mathcal{F}_1 \subset \mathcal{F}_2$, and have $\mathbb{E}\left[X \mid \mathcal{F}_2\right] = Y$ for some variable $Y \in \mathcal{F}_1$, then

$$\mathbb{E}\left[X \mid \mathcal{F}_1\right] = \mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{F}_2\right] \mid \mathcal{F}_1\right] = \mathbb{E}\left[Y \mid \mathcal{F}_1\right] = Y \ .$$

We now move on to the definition of a martingale.

Definition 1. *A sequence of random variables (X_t) adapted to a filtration $(\mathcal{F}_t)$ is a* martingale *(with respect to $(\mathcal{F}_t)$) if the following holds for all t:*

(i) $\mathbb{E}|X_t| < \infty$

(ii) $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = X_t$

If instead of condition (ii) *we have $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \ge X_t$ for all t, we then say that (X_t) is a* submartingale *with respect to $(\mathcal{F}_t)$.*

If instead of condition (ii) *we have $\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] \le X_t$ for all t, we then say that (X_t) is a* supermartingale *with respect to $(\mathcal{F}_t)$.*

When $\mathcal{F}_t$ is the natural filtration generated by (X_t), the above condition (ii) can be rewritten as $\mathbb{E}[X_{t+1} \mid X_t, \ldots, X_1] = X_t$.

The distinction between the terms submartingale and supermartingale is best remembered by the following: in a supermartingale, the current variable X_t is an overestimate for the upcoming X_{t+1}, whereas in a submartingale X_t is an underestimate for this value.

Examples

1. Sums of i.i.d. variables

Let $Y_1, Y_2, \ldots$ be i.i.d. variables with mean 0:

$$Y_i \sim Y_1 \quad , \quad \mathbb{E}Y_1 = 0 \ ,$$

and let

$$\mathcal{F}_t = \sigma(Y_1, \ldots, Y_t) \quad , \quad S_n = \sum_{i=1}^{n} Y_i \ .$$

Clearly,

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[Y_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}Y_{n+1} = S_n \ ,$$

where in the second inequality we used the fact that the Y_i's are independent.

Now, consider a slightly modified setting, where for some $p > 0$ fixed and $q \triangleq 1 - p$,

$$Y_1 \sim \begin{cases} 1 & p \ , \\ -1 & q \ . \end{cases}$$

In this case, setting $\mu = \mathbb{E}Y_1 = p - q$ we have that $X_n = S_n - n\mu$ is martingale, since

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + \mu \ ,$$

and so

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = S_n + \mu - (n+1)\mu = X_n \ .$$

Remark. As an exercise, show that if $\mathbb{E}Y_1 = 0$ and $\mathbb{E}Y_1^2 = \sigma^2 < \infty$, then S_n^2 is a submartingale.

2. The "double or nothing" martingale

A casino runs a game in the form of independent trials Y_i, where $Y_i \in \{\pm 1\}$ with probability $\frac{1}{2}$ each, and the payoff (whenever $Y_i = 1$) is \$1 for each \$1 wagered. Note that by the previous example, $\sum Y_i$ is a martingale. We will later see that if $\sum Y_i$ was instead a supermartingale, then no gambling strategy can "beat the house"; meanwhile, however, the sequence Y_i represents fair coin tosses, and we consider the following gambling scheme:

- Gary the gambler sets \$1 as the initial bet.

- If he loses, Gary doubles the existing bet.
- Upon winning, Gary leaves the game (equivalently, he wagers \$0 in each subsequent round).

Let τ denote the first round in which Gary wins:

$$\tau \triangleq \min\{t : Y_t = 1\} \ .$$

If we let X_t denote the total earnings (positive or negative) of Gary after playing t rounds (i.e., at the beginning of round $t+1$), the following holds:

$$\begin{aligned}
&X_0 = 0 \ , \\
&\text{If } t > \tau \text{ then } X_{t+1} = X_t \text{ (no bet), and otherwise:} \\
&X_{t+1} = X_t + Y_t 2^t \sim \begin{cases} X_t - 2^t & \frac{1}{2} \\ X_t + 2^t & \frac{1}{2} \end{cases} \ .
\end{aligned}$$

Indeed, by definition (X_t) is a martingale with respect to the natural filtration it defines:

$$\mathbb{E}[X_{t+1} \mid \mathcal{F}_t] = X_t \ .$$

The advantage in this gambling strategy is obvious: τ is a geometric random variable $\text{Geo}(\frac{1}{2})$, and so we are guaranteed to win: $\mathbb{P}(\tau < \infty) = 1$. Moreover, on the event that $\tau = k$ ($k \geq 1$) we have

$$X_{k+1} = 2^k - \sum_{i=0}^{k-1} 2^i = 1 \ ,$$

hence, once we win a round, its gain will cover our entire cumulative loss up to that point, plus an extra \$1. That is to say, this method guarantees profit with probability 1.

The drawback is clear as well — if one uses this method, he or she must be prepared to suffer a substantial cumulative loss before reaching the round that would yield this mentioned profit. Indeed, on the event that $\tau = k$ ($k \geq 2$) we have lost rounds $1, 2, \ldots, k-1$, in which we wagered $1, 2, \ldots, 2^{k-2}$ dollars resp., and so

$$X_k = -\sum_{i=0}^{k-2} 2^i = -(2^{k-1} - 1) \ ,$$

and so, since $\tau \sim \text{Geo}(\frac{1}{2})$, and if $\tau = 1$ we immediately win \$1,

$$\mathbb{E}X_\tau = \frac{1}{2} - \sum_{k=2}^{\infty} 2^{-k}(2^{k-1} - 1) = -\infty \ .$$

Remark. In the above example, clearly for any fixed m we have that $\mathbb{E}X_m = X_0 = 0$. Compare this to the random stopping time τ, where $\mathbb{E}X_\tau = -\infty$. In the future, we will learn to distinguish between cases where one can consider a random stopping time for the martingale such that it would still maintain its properties.

3. Branching processes

Let $X_i^{(t)}$ be i.i.d. random variables, $X_i^{(t)} \sim X$ for some X (a typical example one may consider is when $X \sim \text{Geo}(p)$ for some $0 < p < 1$). The branching process is the stochastic process describing the evolution of a population, as follows:

$$Z_0 = 1 \ , Z_{t+1} = \sum_{i=1}^{Z_t} X_i^{(t+1)} \ .$$

That is, each of the elements from the previous round spawns new siblings independently according to the same law (the law of X, or equivalently, of Z_1). The key observation is that, on the event $Z_t = k$, we have

$$Z_{t+1} \sim \sum_{i=1}^{k} X_i^{(t+1)} \ ,$$

that is, Z_{t+1} is the sum of k independent copies of X. In particular, (Z_t) is a Markov chain: the distribution of Z_{t+1} is completely determined by the value of Z_t (rather than the entire history $\mathcal{F}_t$).

Let $m \triangleq \mathbb{E}X$. By the above discussion,

$$\mathbb{E}[Z_{t+1} \mid Z_t] = mZ_t \ ,$$

and iterating we have that for any integer $t \geq 1$,

$$\mathbb{E}Z_t = m^t \ .$$

Therefore, letting $W_t \triangleq m^{-t} Z_t$ we get

$$\mathbb{E}[W_{t+1} \mid \mathcal{F}_t] = m^{-(t+1)}(mZ_t) = W_t \ ,$$

and so (W_t) is a martingale. This is no surprise: Since each element spawns new elements with expectation m, one anticipates that the population size should be roughly m^k after k rounds.

The following martingale is less obvious. Let $f : [0,1] \to \mathbb{R}^+$ be the probability generating function of X:

$$f(s) \triangleq \mathbb{E}\left[s^X\right] \ ,$$

and let q be the smallest non-negative solution to the equation $f(s) = s$. It then follows that $Y_t = q^{Z_t}$ is a martingale, since, recalling that the distribution of Z_t is determined by Z_{t-1},

$$\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] = \mathbb{E}\left[q^{\sum_{i=1}^{Z_t} X_i^{(t+1)}} \mid Z_t\right] = \prod_{i=1}^{Z_t} \mathbb{E}\left[q^{X_i^{(t+1)}}\right] = (f(q))^{Z_t} = q^{Z_t} = Y_t \ ,$$

where in the second inequality we used the fact that the variables X_i are independent, and after that we used the definitions of $f(s)$ and q.

Remark. Though this will not be used in the present example, in the future we will see that q is in fact the extinction probability for the branching process, that is:

$$q = \mathbb{P}(Z_t = 0 \text{ for some } t) \ .$$

Remark. The martingale (W_t) is in fact a special case of the following: if P is the transition matrix of a Markov chain (X_t), and g is a non-negative eigenfunction of P corresponding to an eigenvalue λ, then $W_t \triangleq \lambda^{-t} g(X_t)$ is a martingale. In the example above, $g(x) = x$ and $X_t = Z_t$.