MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436/15.085 Lecture 25

Lecturer: Yury Polyanskiy | Scribe: MIT Class Participants

MARTINGALES I

Content.

0. Background
1. Martingales: definition, examples
2. Azuma-Hoeffding Inequality
3. Optional stopping theorem

0 BACKGROUND

- Recall we defined conditional expectation $V = \mathbb{E}[A|X]$ as follows:

  $$\forall B = f(X), \mathbb{E}[AB] = \mathbb{E}[VB]$$

  We also learned that one computes conditional expectations, usually, by integrating

  $$\mathbb{E}[A|X = x] = \int_{\mathbb{R}} a P_{A|X}(da|x)$$

- We can also define conditional expectation with respect to a sigma-algebra $\mathcal{F}$:

  $$V = \mathbb{E}[A|\mathcal{F}]$$

  Namely, random variable V is a conditional expectation of A given $\mathcal{F}$ if a) $V \in \mathcal{F}$ and b) $\forall B \in \mathcal{F}$, $\mathbb{E}[AB] = \mathbb{E}[VB]$. Here we used common abuse of notation $V \in \mathcal{F}$ meaning "V is $\mathcal{F}$-measurable" (which, recall, means $\{V \leq v\} \in \mathcal{F}$ for every $v \in \mathbb{R}$.

- Recall $\sigma(X_0, X_1, ..., X_k)$ $\mathcal{F}_k$ where $\mathcal{F}_k$ is the smallest σ-algebra containing all events $\{X_i \leq a\}$. Recall also that

$$A \in \mathcal{F}_{\|} \iff \exists f : A = f(X_0, ..., X_k) \tag{1}$$

  Given a stochastic process $X_0, X_1, ...$

$$\mathcal{F}_k = \sigma(X_0, X_1, ..., X_k), \mathcal{F}_0 \subset \mathcal{F}_1 \subset ... \subset \mathcal{F}_\infty$$

  where $\mathcal{F}_\infty = \sigma(X_i, i \in \mathbb{Z}_+)$. The $\mathcal{F}_k$ we have defined here is known as the **standard filtration** generated by the stochastic process. We can think about each $\mathcal{F}_k$ as the valid questions you can ask (and answer) if you only know realization of the stochastic process up to time k.

- Before we were talking about a stochastic process in isolation. Now we will talk about stochastic process being adapted to some filtration $\mathcal{F}_k$. For simplicity, you can always think of a standard filtration generated by a mother (complicated) random process $\{X_k\}$. We say that Y_k is a stochastic process adapted to filtration $\mathcal{F}_k$ if $Y_k \in \mathcal{F}_k$ holds $\forall k \geq 0$.

  As an example, we can look at a simple process $Y_k = \text{sign}(X_0 + ... + X_k)$. Note that $Y_k \in \mathcal{F}_k$, i.e. Y_k is $\mathcal{F}_k$-measurable, because it is a function of $X_0, ...X_k$. However, knowledge of $Y_0, \ldots, Y_k$ is insufficient to reconstruct trajectory $X_0, \ldots, X_k$. So while Y_k is adapted to $\mathcal{F}_k$, the filtration $\mathcal{F}_k$ is much richer. This is a common situation in applications (since we are interested in functions of the mother process), and that's why we need the concept of filtration.

1 MARTINGALES

1.1 Definition

We introduce our main definition of a Martingale:

> **Definition 1.** *A process $(M_t, t = 0, ..., \infty)$ is a **martingale** with respect to filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset ...$ if:*
>
> 1. $M_t \in \mathcal{F}_t \; \forall t \geq 0$
> 2. $\mathbb{E}|M_t| < \infty$, *i.e. M_t is integrable*
> 3. $\mathbb{E}[M_t | \mathcal{F}_{t-1}] = M_{t-1}$

In the special case if $\mathcal{F}_t = \sigma(M_0, ..., M_t)$ we simply say "M_t is a martingale" (without mentioning filtration). In this case, the property 1) is automatic and being a martingale becomes essentially just the requirement that $\mathbb{E}[M_t|M_0, ..., M_{t-1}] = M_{t-1}$, for all $t \geq 1$.

1.2 History of Martingales

The word *martingale* comes from gambling. It describes a strategy in which a gambler makes a series of bets. For each bet, he wins if a coin lands on heads and loses if the coin lands on tails. For each successive loss, he doubles his bet, starting with \$1 on the first flip. At the time of winning (i.e. first time the coin lands on heads), the gambler will receive a net gain of \$1 $2^{t+1} - (1 + 2 + ... + 2^t) = 1$; however the expected loss at the time of winning is ∞.

1.3 Examples

For the following examples of martingales, we introduce the notation

$$\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_t]$$

Example 1. *$S_n = X_0 + ... + X_n$ for X_i independent and $\mathbb{E}[X_i] = 0$. $\mathcal{F}_k = \sigma(X_0, ..., X_k)$. Then, we have $\mathbb{E}_{n-1} S_n = S_{n-1}$.*

Example 2. *$Y_n = X_0 \cdot X_1 \cdot ... \cdot X_n$ for X_i independent and $\mathbb{E}[X_i] = 1$. Then, we have $\mathbb{E}_{n-1} Y_n = Y_{n-1}$.*

Example 3 (Doob Martingale)**.** *Let Z be any random variable with finite expectation ($\mathbb{E}|Z| < \infty$) and $\mathcal{F}_t$ be any filtration. We define a Doob martingale:*

$$M_t = \mathbb{E}[Z|\mathcal{F}_t] \tag{2}$$

This make look like a rather special case, but it will turn out that many martingales we work with will turn out to be of this type. Think of it as if we have a "secret" Z and we are observing its average given the known information at time n. Over time, we learn more and more about this mother random variable, and approach knowing Z itself. A Doob martingale has the martingale property with respect to the given filtration:

$$\begin{aligned}
\mathbb{E}_{t-1} M_t &= \mathbb{E}[M_t|X_0, ..., X_{t-1}] \\
&= \mathbb{E}[\mathbb{E}[Z|X_0, ..., X_t]|X_0, ...X_{t-1}] \\
&= \mathbb{E}[Z|X_0, ..., X_{t-1}] \\
&= M_{t-1}
\end{aligned}$$

where in the second line, we use the tower property of conditional expectation.

As we will see below, Examples 1 and 2 are not Doob martingales unless they converge. We can, however, modify examples 1 and 2 so that they are Doob martingales. Suppose we restrict the martingales to within a certain window, for instance S_n for n such that $-100 \leq S_n \leq 100$, and freeze the process once it exceeds the boundary. Then, the martingales are Doob martingales (since they ar bounded!).

2 AZUMA'S INEQUALITY

By performing a simple computation and induction, we can see that $\mathbb{E}M_t = \mathbb{E}M_0$:

$$\begin{aligned}\mathbb{E}[M_t] &= \mathbb{E}[\mathbb{E}_{t-1}[M_t]] \\ &= \mathbb{E}[M_{t-1}]\end{aligned}$$

To compute the variance of M_t, assume without loss of generality that $\mathbb{E}[M] = 0$.

$$\begin{aligned}\mathrm{var}(M_t) &= \mathbb{E}M_t^2 && (3)\\ &= \mathbb{E}[M_t - M_{t-1} + M_{t-1}]^2 && (4)\\ &= \mathbb{E}[\mathbb{E}_{t-1}[(M_t - M_{t-1})^2 + M_{t-1}^2 + 2M_{t-1}(M_t - M_{t-1})]] && (5)\\ &= \mathbb{E}[M_{t-1}^2] + \mathbb{E}[M_t - M_{t-1}]^2 && (6)\\ &= \sum_{s=1}^{t} \mathbb{E}[M_s - M_{s-1}]^2 + \mathrm{var}(M_0) && (7)\end{aligned}$$

We obtain (5) by using the tower property of conditional expectation and we use the following simplification to obtain (6):

$$\mathbb{E}_{t-1}[M_{t-1}(M_t - M_{t-1})] = M_{t-1}(\mathbb{E}[M_t - M_{t-1}]) = 0$$

Finally, we obtain (7) by induction.

From this derivation, we can see that $|M_s - M_{s-1}| \leq c \Rightarrow \mathrm{var}M_t \leq c^2 t$. Martingales with bounded increments (within a constant c) grow with speed $\sim \sqrt{t}$. This leads us to the Azuma-Hoeffding inequality.

Theorem 1 (Azuma-Hoeffding Inequality). *If M_t is a martingale with $|M_t - M_{t-1}| \leq c_t$ a.s. $\forall t$, then*

$$\mathbb{P}(M_t - \mathbb{E}[M_t] > h) \leq \exp\left(\frac{-h^2}{2 \sum_{s=1}^{t} c_s^2}\right)$$

Proof. From Chernoff bound, we have $\mathbb{P}[\cdot] \leq e^{-\lambda h + \psi_t(\lambda)}\ \forall \lambda > 0$ where $\psi_t(\lambda) = \ln \mathbb{E}[e^{\lambda M_t}]$ is the log MGF. Without loss of generality, we are assuming $\mathbb{E}M = 0$.

It is sufficient to prove that $\psi_t(\lambda) \leq \psi_{t-1}(\lambda) + \frac{\lambda^2 c_t^2}{2}$.

We can rewrite the following expression:

$$\mathbb{E}_{t-1} e^{\lambda M_t} = \mathbb{E}_{t-1} e^{\lambda(M_t - M_{t-1})} e^{\lambda M_{t-1}}$$

$\forall |x| \leq c_t$:

$$e^{\lambda x} \leq e^{-\lambda c_t} + (x + c_t)\frac{(e^{\lambda c_t} - e^{-\lambda c_t})}{2c_t}$$

Plugging in $x = M_t - M_{t-1}$.

$$\mathbb{E}_{t-1} e^{\lambda(M_t - M_{t-1})} \leq \frac{e^{-\lambda c_t} + e^{\lambda c_t}}{2}$$

because $\mathbb{E}_{t-1}(M_t - M_{t-1}) = 0$. Finally, using the fact that

$$\frac{e^{-p} + e^{p}}{2} \leq e^{\frac{p^2}{2}},$$

which can be checked using Matlab/Python, and substituting $p = \lambda c_t$, we get

$$\mathbb{E}_{t-1} e^{\lambda(M_t - M_{t-1})} \leq \frac{e^{-\lambda c_t} + e^{\lambda c_t}}{2} \leq e^{\frac{\lambda^2 c_t^2}{2}}$$

□

Example 4. *Suppose we throw M balls into n bins. Let V be the number of occupied bins. Let us define a process:*

$$M_t \triangleq \mathbb{E}[V | X_1, ..., X_t] \tag{8}$$

where X_i is the bin selected by the i^{th} ball. We can see that this process is a Doob martingale becase we are conditioning on increasing σ-algebras. Intuitively, we can see that at any step of the process, the conditional expectation *will not change by more than 1: $|M_t - M_{t-1}| \leq 1$. Thus, we have by Azuma-Hoeffding:*

$$\mathbb{P}[|V - \mathbb{E}V| > h] \leq e^{\frac{-h^2}{2M}}$$

This example demonstrates that, even if the exact mean is unknown, we can already guarantee that the distribution of V concentrates sharply around its mean. So all complexity of understanding V boils down to computing its expectation (which, in turn, can be done by sampling a few realizations, thanks to the concentration phenomenon).

3 OPTIONAL STOPPING THEOREM

Recall the following definition of the stopping time of a filtration:

Definition 2. $\tau : \Omega \to \mathbb{Z}_+ \cup \{+\infty\}$ *is a* ***stopping time of filtration*** $\mathcal{F}_0 \subset \mathcal{F}_1 \subset ...$ *if*

$$\{\tau \leq n\} \in \mathcal{F}_n \ \forall n \Longleftrightarrow \{\tau = n\} \in \mathcal{F}_n \ \forall n$$

Now, using the notation $a \wedge b = \min(a, b)$, let us define $Y_t = M_{\tau \wedge t}$. Think of this as a process that has values M_t until time τ and then has constant value M_τ. We are essentially defining a new process that follows the trajectory of M_t but then freezes once it reaches τ. For instance, we could define $\tau = \inf\{t : |M_t| \geq 100\}$.

Theorem 2. $Y_t = M_{\tau \wedge t}$ *is a martingale for any martingale* M_t *and stopping time* τ*.*

Proof.

$$Y_t = \sum_{r=0}^{t-1} M_r \mathbb{1}\{\tau = r\} + M_t \mathbb{1}\{\tau \geq t\}$$

Note that $\{\tau \geq t\} = \{\tau \leq t - 1\}^c \in \mathcal{F}_{t-1}$, so $\mathbb{E}_{t-1} \sum_{r=0}^{t-1} M_t \mathbb{1}\{\tau = r\} = \sum_{r=0}^{t-1} M_t \mathbb{1}\{\tau = r\}$. Substituting this, we get

$$
\begin{aligned}
\mathbb{E}_{t-1}Y_t &= \sum_{r=0}^{t-1} M_t \mathbb{1}\{\tau = r\} + (\mathbb{E}_{t-1}M_t)\mathbb{1}\{\tau \geq t\} \\
&= \sum_{r=0}^{t-1} M_t \mathbb{1}\{\tau = r\} + M_{t-1}\mathbb{1}\{\tau \geq t\} \\
&= M_{(t-1)\wedge\tau}
\end{aligned}
$$

where the equation from the first to the second line follows from $\mathbb{E}_{t-1}M_t = M_{t-1}$. We are done with the proof. □

This relates to the efficient market hypothesis: the price of a stock should be a martingale (with respect to filtration generated by all public information). Indeed, in this case defining a smart stopping time one is unable to improve the average price still.

Now we return to the idea of uniform integrability and introduce the crucial concept of a **uniformly integrable martingale (uim)**. First we consider a very useful and simple criterion for getting a wealth of uims.

Proposition 1. *Let M_t be a martingale, τ be a stopping time such that $\mathbb{E}\tau < \infty$, and $\mathbb{E}_{t-1}|M_t - M_{t-1}| \leq c$ a.s., then $Y_t \triangleq M_{t\wedge\tau}$ is a* ***uniformly integrable Martingale****.*

Proof.

$$
\begin{aligned}
|Y_t - Y_0| &\leq \sum_{s=1}^{t} |Y_s - Y_{s-1}| \\
&= \sum_{s=1}^{t} |M_s - M_{s-1}|\mathbb{1}\{\tau \geq s\} \\
&\leq \sum_{s=1}^{\infty} |M_s - M_{s-1}|\mathbb{1}\{\tau \geq s\} =: W
\end{aligned}
$$

The conditions imply

$$
\mathbb{E}W < \infty \Rightarrow |Y_t| \leq W + |Y_0| \ \forall t
$$

□

We will see in the next lecture that every u.i.m. $\iff$ Doob Martingale. In particular, every bounded martingale is Doob. For now we state the crown jewel of martingale theory:

Theorem 3 (Optional stopping theorem)**.** *Let M_t be a uniformly integrable Martingale and let τ be a stopping time such that $\mathbb{P}[\tau < \infty] = 1$. Then:*

$$\mathbb{E}M_\tau = \mathbb{E}M_0$$

Proof. We first prove a special case: Suppose $\tau \leq L$ a.s. where L is some constant. Then:

$$M_\tau = \sum_{t=0}^{L} M_t \mathbb{1}\{\tau = t\} \tag{9}$$

$$= \sum_{t=0}^{L} (\mathbb{E}_t M_L) \mathbb{1}\{\tau = t\} \tag{10}$$

$$= \sum_{t=0}^{L} \mathbb{E}_t M_L \mathbb{1}\{\tau = t\} \tag{11}$$

The key insight to obtain (10) was to use the property of martingales from part 3 of the definition. Now, we can take the expected value of both sides of (11):

$$\mathbb{E}M_\tau = \sum_t \mathbb{E}M_L \mathbb{1}\{\tau = t\} \tag{12}$$

$$= \mathbb{E}M_L \tag{13}$$

$$= \mathbb{E}M_0 \tag{14}$$

Note that (12) forms a partition because $\sum_t \mathbb{1}\{\tau = t\} = 1$ a.s.

Note that a similar argument shows

$$\mathbb{E}|M_\tau| \leq \mathbb{E}|M_L|$$

Indeed, one only needs to notice that $|\mathbb{E}_t M_L| \leq \mathbb{E}_t |M_L|$.

The general case follows in two steps. First define $\tau_L = \tau \wedge L$, then by the previous argument we have

$$\mathbb{E}|M_{\tau_L}| \leq \sup_L \mathbb{E}|M_L| < \infty\,,$$

where the last inequality follows from uniform integrability (which implies uniform boundedness). So since $M_{\tau_L} \to M_\tau$ as $L \to \infty$ almost surely, we have via Fatou's lemma

$$\mathbb{E}|M_\tau| < \infty\,.$$

Finally,

$$M_\tau = M_{\tau_L} + (M_\tau - M_{\tau_L})1\{\tau \geq L\}$$

By the first part of the proof $\mathbb{E}[M_{\tau_L}] = \mathbb{E}[M_0]$. So we only need to show that as $L \to \infty$ the expectation of the second term vanishes.

Note that as $L \to \infty$ we have $\mathbb{P}[\tau \geq L] \to 0$. Thus $\mathbb{E}[|M_\tau|1\{\tau \geq L\}] \to 0$. Similarly, from uniform integrability of $\{M_L\}$ we have $\mathbb{E}[|M_{\tau_L}|1\{\tau \geq L\}] = \mathbb{E}[|M_L|1\{\tau \geq L\}] \to 0$. This completes the proof. □