

Martingales

‘After that he always chose out a “dog command” and sent them ahead. It had the task of informing the inhabitants in the village where we were going to stay overnight that no dog must be allowed to bark in the night otherwise it would be liquidated. I was also on one of those commands and when we came to a village in the region of Milevsko I got mixed up and told the mayor that every dog-owner whose dog barked in the night would be liquidated for strategic reasons. The mayor got frightened, immediately harnessed his horses and rode to headquarters to beg mercy for the whole village. They didn’t let him in, the sentries nearly shot him and so he returned home, but before we got to the village everybody on his advice had tied rags round the dogs muzzles with the result that three of them went mad.’

The good soldier Svejk, Jaroslav Hasek

21.1. Martingales

21.1.1. Preliminaries

Let X and Y be two random variables. Let $\rho(x, y) = \mathbb{P}[(X = x) \cap (Y = y)]$. Observe that

$$\mathbb{P}[X = x \mid Y = y] = \frac{\rho(x, y)}{\mathbb{P}[Y = y]} = \frac{\rho(x, y)}{\sum_z \rho(z, y)}$$

and

$$\mathbb{E}[X \mid Y = y] = \sum_x x \mathbb{P}[X = x \mid Y = y] = \frac{\sum_x x \rho(x, y)}{\sum_z \rho(z, y)} = \frac{\sum_x x \rho(x, y)}{\mathbb{P}[Y = y]}.$$

The *conditional expectation* of X given Y , denoted by $\mathbb{E}[X \mid Y]$, is the random variable $f(Y)$, where $f(y) = \mathbb{E}[X \mid Y = y]$.

As a reminder, for any two random variables X and Y , we have

- (I) **Lemma 21.3.1:** $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.
- (II) **Lemma 21.3.2:** $\mathbb{E}[Y \cdot \mathbb{E}[X \mid Y]] = \mathbb{E}[XY]$.

21.1.2. Martingales

Intuitively, a martingale is a sequence of random variables describing a “fair game”: given the entire history up to the end of the $(i - 1)$ th step, the conditional expectation of the process at the i th step is equal to its current value. That is, the process has zero expected drift, regardless of how it arrived at its current state.

Intuitively (but not strictly correctly), the only thing that matters at the beginning of the i th step is where the process was in the end of the $(i - 1)$ th step. That is, it does not matter how the process arrived to a certain state, only that it is currently at this state.

Definition 21.1.1. A sequence of random variables $X_0, X_1, \dots$, is said to be a ***martingale sequence*** if for all $i > 0$, we have $\mathbb{E}[X_i \mid X_0, \dots, X_{i-1}] = X_{i-1}$.

In particular, note that for a martingale, we have $\mathbb{E}[X_i \mid X_{i-1}] = X_{i-1}$ as well (by the tower property). However, satisfying $\mathbb{E}[X_i \mid X_{i-1}] = X_{i-1}$ alone is strictly weaker than being a martingale, as the expectation must remain unbiased conditioned on the entire history $X_0, \dots, X_{i-1}$.

Lemma 21.1.2. Let $X_0, X_1, \dots$, be a martingale sequence. Then, for all $i \geq 0$, we have $\mathbb{E}[X_i] = \mathbb{E}[X_0]$.

Proof: By (I), and the martingale property, we have

$$\mathbb{E}[X_i] = \mathbb{E}[\mathbb{E}[X_i \mid X_{i-1}]] = \mathbb{E}[X_{i-1}] = \mathbb{E}[X_{i-2}] = \dots = \mathbb{E}[X_0]. \quad \color{red}{\text{🐼}}$$

21.1.2.1. Examples of martingales

Example 21.1.3. Consider the sum of money after participating in a sequence of fair bets. That is, let X_i be the amount of money a gambler has after playing i rounds. In each round the gambler either gains one dollar, or loses one dollar (with equal probability). Clearly, we have

$$\mathbb{E}[X_i \mid X_0, \dots, X_{i-1}] = \mathbb{E}[X_i \mid X_{i-1}] = X_{i-1} + \frac{1}{2} \cdot (+1) + \frac{1}{2} \cdot (-1) = X_{i-1}.$$

Example 21.1.4. Let $Y_i = X_i^2 - i$, where X_i is as defined in the above example. We claim that $Y_0, Y_1, \dots$ is a martingale. Let us verify that this is true. Conditioned on the history $X_0, \dots, X_{i-1}$ (which determines $Y_0, \dots, Y_{i-1}$), we have that

$$\begin{aligned}\mathbb{E}[Y_i \mid X_0, \dots, X_{i-1}] &= \mathbb{E}[X_i^2 - i \mid X_0, \dots, X_{i-1}] \\ &= \frac{1}{2}((X_{i-1} + 1)^2 - i) + \frac{1}{2}((X_{i-1} - 1)^2 - i) \\ &= X_{i-1}^2 + 1 - i = X_{i-1}^2 - (i - 1) = Y_{i-1},\end{aligned}$$

which implies that indeed it is a martingale.

Example 21.1.5. Let U be an urn with b black balls, and w white balls. We repeatedly select a ball and replace it by c balls having the same color. Let X_i be the fraction of black balls after the first i trials. We claim that the sequence $X_0, X_1, \dots$ is a martingale.

Indeed, let $n_i = b + w + i(c - 1)$ be the number of balls in the urn after the i th trial. Clearly,

$$\begin{aligned}\mathbb{E}[X_i \mid X_{i-1}, \dots, X_0] &= X_{i-1} \cdot \frac{(c - 1) + X_{i-1}n_{i-1}}{n_i} \\ &\quad + (1 - X_{i-1}) \cdot \frac{X_{i-1}n_{i-1}}{n_i} \\ &= \frac{X_{i-1}(c - 1) + X_{i-1}n_{i-1}}{n_i} \\ &= X_{i-1} \frac{c - 1 + n_{i-1}}{n_i} \\ &= X_{i-1} \frac{n_i}{n_i} = X_{i-1}.\end{aligned}$$

Example 21.1.6. Let G be a random graph on the vertex set $V = \{1, \dots, n\}$ obtained by independently choosing to include each possible edge with probability p . The underlying probability space over *random graphs* is denoted by $\mathbf{G}_{n,p}$. Arbitrarily label the $m = n(n - 1)/2$ possible edges with the sequence $1, \dots, m$. For $1 \leq j \leq m$, define the indicator random variable I_j , which takes values 1 if the edge j is present in G , and has value 0 otherwise. These indicator variables are independent and each takes value 1 with probability p .

Consider any real valued function f defined over the space of all graphs, e.g., the clique number, which is defined as being the size of the largest complete subgraph. The *edge exposure martingale* is the sequence of random variables $X_0, \dots, X_m$ such that

$$X_i = \mathbb{E}[f(G) \mid I_1, \dots, I_i],$$

while $X_0 = \mathbb{E}[f(G)]$ and $X_m = f(G)$. That this sequence of random variables is a martingale follows immediately from a theorem that will be described in the next lecture.

One can define similarly a *vertex exposure martingale*, where $X_i = \mathbb{E}[f(G) \mid G_i]$, and G_i is the graph induced on the first i vertices of the random graph G .

Example 21.1.7 (The sheep of Mabinogion). The following is taken from a medieval Welsh manuscript based on Celtic mythology:

“And he came towards a valley, through which ran a river; and the borders of the valley were wooded, and on each side of the river were level meadows. And on one side of the river he saw a flock of white sheep, and on the other a flock of black sheep. And whenever one of the white sheep bleated, one of the black sheep would cross over and become white; and when one of the black sheep bleated, one of the white sheep would cross over and become black.” – *Peredur the son of Evrawk*, from the *Mabinogion*.

More concretely, we start at time 0 with w_0 white sheep, and b_0 black sheep. At every iteration, a random sheep is picked, it bleats, and a sheep of the other color turns to this color. The game stops as soon as all the sheep have the same color. No sheep dies or is born during the game. Let X_i be the expected number of black sheep in the end of the game, after the i th iteration. For reasons that we will see later on, this sequence is a martingale.

The original question is somewhat more interesting – if we are allowed to take away a sheep in the end of each iteration, what is the optimal strategy to maximize X_i ?

21.1.2.2. Azuma's inequality

A sequence of random variables $X_0, X_1, \dots$ has *bounded differences* if $|X_i - X_{i-1}| \leq \Delta$, for some Δ .

Theorem 21.1.8 (Azuma's Inequality). *Let $X_0, \dots, X_m$ be a martingale with $X_0 = 0$, and*

$$|X_{i+1} - X_i| \leq 1, \quad \text{for } i = 0, \dots, m-1.$$

For any $\lambda > 0$, we have

$$\mathbb{P}[X_m > \lambda\sqrt{m}] < \exp(-\lambda^2/2).$$

Proof: Let $\alpha = \lambda/\sqrt{m}$. Let $Y_i = X_i - X_{i-1}$, so that $|Y_i| \leq 1$ and $\mathbb{E}[Y_i | X_0, \dots, X_{i-1}] = 0$.

We are interested in bounding $\mathbb{E}[e^{\alpha Y_i} | X_0, \dots, X_{i-1}]$. Note that, for $-1 \leq x \leq 1$, we have

$$f(x) = e^{\alpha x} \leq h(x) = \frac{e^\alpha + e^{-\alpha}}{2} + \frac{e^\alpha - e^{-\alpha}}{2}x,$$

as $f(x) = e^{\alpha x}$ is a convex function, $h(-1) = e^{-\alpha} = f(-1)$, $h(1) = e^\alpha = f(1)$, and $h(x)$ is a linear function. Thus,

$$\begin{aligned} \mathbb{E}[e^{\alpha Y_i} | X_0, \dots, X_{i-1}] &\leq \mathbb{E}[h(Y_i) | X_0, \dots, X_{i-1}] \\ &= h(\mathbb{E}[Y_i | X_0, \dots, X_{i-1}]) \\ &= h(0) = \frac{e^\alpha + e^{-\alpha}}{2} \\ &= \frac{1}{2} \sum_{k=0}^{\infty} \frac{\alpha^k + (-\alpha)^k}{k!} = \sum_{k=0}^{\infty} \frac{\alpha^{2k}}{(2k)!} \\ &\leq \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\alpha^2}{2}\right)^k = \exp(\alpha^2/2), \end{aligned}$$

as $(2i)! \geq 2^i i!$.

Let $\tau = \mathbb{E}[e^{\alpha X_m}] = \mathbb{E}[\prod_{i=1}^m e^{\alpha Y_i}]$. By the law of total expectation, conditioning on $X_0, \dots, X_{m-1}$ yields

$$\tau = \mathbb{E} \left[\mathbb{E} \left[\prod_{i=1}^m e^{\alpha Y_i} \mid X_0, \dots, X_{m-1} \right] \right]$$

$$= \mathbb{E} \left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i} \right) \mathbb{E}[e^{\alpha Y_m} \mid X_0, \dots, X_{m-1}] \right].$$

Since $\mathbb{E}[Y_m \mid X_0, \dots, X_{m-1}] = 0$, our earlier bound applies to step m :

$$\mathbb{E}[e^{\alpha Y_m} \mid X_0, \dots, X_{m-1}] \leq \exp(\alpha^2/2).$$

Therefore,

$$\tau \leq e^{\alpha^2/2} \mathbb{E} \left[\prod_{i=1}^{m-1} e^{\alpha Y_i} \right].$$

Repeating this conditioning inductively for each factor $i = m-1, \dots, 1$, we obtain

$$\tau \leq \exp(m\alpha^2/2).$$

Therefore, by Markov's inequality, we have

$$\begin{aligned} \mathbb{P}[X_m > \lambda\sqrt{m}] &= \mathbb{P}[e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}] \\ &\leq \frac{\mathbb{E}[e^{\alpha X_m}]}{e^{\alpha\lambda\sqrt{m}}} \leq \exp(m\alpha^2/2 - \alpha\lambda\sqrt{m}) \\ &= \exp(m(\lambda/\sqrt{m})^2/2 - (\lambda/\sqrt{m})\lambda\sqrt{m}) = e^{-\lambda^2/2}, \end{aligned}$$

implying the result. 

Here is an alternative form.

Theorem 21.1.9 (Azuma's Inequality). *Let $X_0, \dots, X_m$ be a martingale sequence such that $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then*

$$\mathbb{P}[|X_m - X_0| > \lambda\sqrt{m}] < 2\exp(-\lambda^2/2).$$

Example 21.1.10. Let $\chi(H)$ be the chromatic number of a graph H . What is the chromatic number of a random graph? How does this random variable behave?

Consider the vertex exposure martingale, and let $X_i = \mathbb{E}[\chi(G) \mid G_i]$. Note that exposing the edges incident to vertex i changes the chromatic

number by at most 1, so $|X_i - X_{i-1}| \leq 1$. Again, without proving it, we claim that $X_0, \dots, X_n$ is a martingale, and as such, by Azuma's inequality,

$$\mathbb{P}[|X_n - X_0| > \lambda\sqrt{n}] < 2\exp(-\lambda^2/2).$$

However, $X_0 = \mathbb{E}[\chi(G)]$, and $X_n = \mathbb{E}[\chi(G) \mid G_n] = \chi(G)$. Thus,

$$\mathbb{P}[|\chi(G) - \mathbb{E}[\chi(G)]| > \lambda\sqrt{n}] < 2\exp(-\lambda^2/2).$$

Namely, the chromatic number of a random graph is highly concentrated! And we do not even (need to) know what is the expectation of this variable!

21.2. Bibliographical notes

Our presentation follows [MR95].

21.3. From previous lectures

Lemma 21.3.1. *For any two random variables X and Y , we have*

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X].$$

Lemma 21.3.2. *For any two random variables X and Y , we have $\mathbb{E}[Y \cdot \mathbb{E}[X \mid Y]] = \mathbb{E}[XY]$.*

References

- [MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge, UK: Cambridge University Press, 1995.