Chapter 8

Martingales

By Sariel Har-Peled, December 30, 2015[1]

'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

8.1. Martingales

8.1.1. Preliminaries

Let X and Y be two random variables. Let $\rho(x, y) = \mathbf{Pr}[(X = x) \cap (Y = y)]$. Then,

$$\mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] = \frac{\rho(x, y)}{\mathbf{Pr}[Y = y]} = \frac{\rho(x, y)}{\sum_z \rho(z, y)}$$

and $\mathbf{E}\left[X \,\middle|\, Y = y\right] = \sum_x x\,\mathbf{Pr}\left[X = x \,\middle|\, Y = y\right] = \frac{\sum_x x\rho(x, y)}{\sum_z \rho(z, y)} = \frac{\sum_x x\rho(x, y)}{\mathbf{Pr}[Y = y]}$.

Definition 8.1.1. The *conditional expectation* of X given Y, is the random variable $\mathbf{E}\left[X \,\middle|\, Y\right]$ is the random variable $f(y) = \mathbf{E}\left[X \,\middle|\, Y = y\right]$.

Lemma 8.1.2. *For any two random variables X and Y, we have* $\mathbf{E}\left[\mathbf{E}\left[X \,\middle|\, Y\right]\right] = \mathbf{E}\left[X\right]$.

Proof: $\mathbf{E}\Big[\mathbf{E}\big[X \mid Y\big]\Big] = \mathbf{E}_Y\Big[\mathbf{E}\big[X \mid Y = y\big]\Big] = \sum_y \mathbf{Pr}[Y = y]\,\mathbf{E}\big[X \mid Y = y\big]$

$$
\begin{aligned}
&= \sum_y \mathbf{Pr}\big[Y = y\big] \frac{\sum_x x\,\mathbf{Pr}[X = x \cap Y = y]}{\mathbf{Pr}[Y = y]} \\
&= \sum_y \sum_x x\,\mathbf{Pr}\big[X = x \cap Y = y\big] = \sum_x x \sum_y \mathbf{Pr}\big[X = x \cap Y = y\big] \\
&= \sum_x x\,\mathbf{Pr}\big[X = x\big] = \mathbf{E}\big[X\big]. \qquad \blacksquare
\end{aligned}
$$

Lemma 8.1.3. *For any two random variables X and Y, we have* $\mathbf{E}\Big[Y \cdot \mathbf{E}\big[X \,\big|\, Y\big]\Big] = \mathbf{E}\big[XY\big]$.

Proof: We have that $\mathbf{E}\Big[Y \cdot \mathbf{E}\big[X \mid Y\big]\Big] = \sum_y \mathbf{Pr}[Y = y] \cdot y \cdot \mathbf{E}\big[X \mid Y = y\big]$

$$
= \sum_y \mathbf{Pr}[Y = y] \cdot y \cdot \frac{\sum_x x\,\mathbf{Pr}[X = x \cap Y = y]}{\mathbf{Pr}[Y = y]} = \sum_x \sum_y xy \cdot \mathbf{Pr}\big[X = x \cap Y = y\big] = \mathbf{E}\big[XY\big]. \qquad \blacksquare
$$

8.1.2. Martingales

Intuitively, martingales are a sequence of random variables describing a process, where the only thing that matters at the beginning of the ith 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 8.1.4. A sequence of random variables $X_0, X_1, \ldots,$ is said to be a *martingale sequence* if for all $i > 0$, we have $\mathbf{E}\big[X_i \,\big|\, X_0, \ldots, X_{i-1}\big] = X_{i-1}$.

Lemma 8.1.5. *Let $X_0, X_1, \ldots,$ be a martingale sequence. Then, for all $i \geq 0$, we have* $\mathbf{E}[X_i] = \mathbf{E}[X_0]$.

8.1.2.1. Examples of martingales

Example 8.1.6. An example of martingales is 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 it either gains one dollar, or loses one dollar. Clearly, we have $\mathbf{E}\big[X_i \,\big|\, X_0, \ldots, X_{i-1}\big] = \mathbf{E}\big[X_i \,\big|\, X_{i-1}\big] = X_i$.

Example 8.1.7. Let $Y_i = X_i^2 - i$, where X_i is as defined in the above example. We claim that $Y_0, Y_1, \ldots$ is a martingale. Let us verify that this is true. Given Y_{i-1}, we have $Y_{i-1} = X_{i-1}^2 - (i-1)$. We have that

$$
\begin{aligned}
\mathbf{E}\big[Y_i \,\big|\, Y_{i-1}\big] &= \mathbf{E}\big[X_i^2 - i \,\big|\, X_{i-1}^2 - (i-1)\big] = \frac{1}{2}\Big((X_{i-1}+1)^2 - i\Big) + \frac{1}{2}\Big((X_{i-1}-1)^2 - i\Big) \\
&= X_{i-1}^2 + 1 - i = X_{i-1}^2 - (i-1) = Y_{i-1},
\end{aligned}
$$

which implies that indeed it is a martingale.

Example 8.1.8. Let U be a 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 $X_0, X_1, \ldots$ is a martingale.

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

$$\begin{aligned}\mathbf{E}\Big[X_i \,\Big|\, X_{i-1}, \ldots, X_0\Big] &= X_{i-1} \cdot \frac{(c-1) + X_{i-1} n_{i-1}}{n_i} + (1 - X_{i-1}) \cdot \frac{X_{i-1} n_{i-1}}{n_i} \\ &= \frac{X_{i-1}(c-1) + X_{i-1} n_{i-1}}{n_i} = X_{i-1} \frac{c - 1 + n_{i-1}}{n_i} = X_{i-1} \frac{n_i}{n_i} = X_{i-1}.\end{aligned}$$

Example 8.1.9. Let G be a random graph on the vertex set $V = \{1, \ldots, n\}$ obtained by independently choosing to include each possible edge with probability p. The underlying probability space is called $\mathbf{G}_{n,p}$. Arbitrarily label the $m = n(n-1)/2$ possible edges with the sequence $1, \ldots, m$. 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 defined to be the sequence of random variables $X_0, \ldots, X_m$ such that

$$X_i = \mathbf{E}\Big[f(G) \,\Big|\, I_1, \ldots, I_i\Big],$$

while $X_0 = \mathbf{E}[f(G)]$ and $X_m = f(G)$. This sequence of random variable begin a martingale follows immediately from a theorem that would be described in the next lecture.

One can define similarly a ***vertex exposure martingale***, where the graph G_i is the graph induced on the first i vertices of the random graph G.

Example 8.1.10 (The sheep of Mabinogion). The following is taken from 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 get born during the game. Let X_i be the expected number of black sheep in the end of the game, after the ith iteration. For reasons that we would see later on, this sequence is a martingale.

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

8.1.2.2. Azuma's inequality

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

Theorem 8.1.11 (Azuma's Inequality.). *Let $X_0, \ldots, X_m$ be a martingale with $X_0 = 0$, and $|X_{i+1} - X_i| \leq 1$ for all $0 \leq i < m$. Let $\lambda > 0$ be arbitrary. Then* $\mathbf{Pr}\Big[X_m > \lambda\sqrt{m}\Big] < \exp\Big(-\lambda^2/2\Big)$.

Proof: Let $\alpha = \lambda/\sqrt{m}$. Let $Y_i = X_i - X_{i-1}$, so that $|Y_i| \le 1$ and $\mathbf{E}\left[Y_i \mid X_0, \ldots, X_{i-1}\right] = 0$.

We are interested in bounding $\mathbf{E}\left[e^{\alpha Y_i} \mid X_0, \ldots, X_{i-1}\right]$. Note that, for $-1 \le x \le 1$, we have

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

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

$$\begin{aligned}
\mathbf{E}\left[e^{\alpha Y_i} \mid X_0, \ldots, X_{i-1}\right] &\le \mathbf{E}\left[h(Y_i) \mid X_0, \ldots, X_{i-1}\right] = h\left(\mathbf{E}\left[Y_i \mid X_0, \ldots, X_{i-1}\right]\right) \\
&= h\left(0\right) = \frac{e^{\alpha} + e^{-\alpha}}{2} \\
&= \frac{\left(1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} + \cdots\right) + \left(1 - \alpha + \frac{\alpha^2}{2!} - \frac{\alpha^3}{3!} + \cdots\right)}{2} \\
&= 1 + \frac{\alpha^2}{2} + \frac{\alpha^4}{4!} + \frac{\alpha^6}{6!} + \cdots \\
&\le 1 + \frac{1}{1!}\left(\frac{\alpha^2}{2}\right) + \frac{1}{2!}\left(\frac{\alpha^2}{2}\right)^2 + \frac{1}{3!}\left(\frac{\alpha^2}{2}\right)^3 + \cdots = \exp\left(\alpha^2/2\right),
\end{aligned}$$

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

Hence, by Lemma 8.1.3, we have that

$$\begin{aligned}
\mathbf{E}\left[e^{\alpha X_m}\right] &= \mathbf{E}\left[\prod_{i=1}^{m} e^{\alpha Y_i}\right] = \mathbf{E}\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) e^{\alpha Y_m}\right] \\
&= \mathbf{E}\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right) \mathbf{E}\left[e^{\alpha Y_m} \mid X_0, \ldots, X_{m-1}\right]\right] \le e^{\alpha^2/2}\, \mathbf{E}\left[\prod_{i=1}^{m-1} e^{\alpha Y_i}\right] \\
&\le \exp\left(m\alpha^2/2\right).
\end{aligned}$$

Therefore, by Markov's inequality, we have

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

implying the result. ■

Here is an alternative form.

Theorem 8.1.12 (Azuma's Inequality). *Let $X_0, \ldots, X_m$ be a martingale sequence such that and $|X_{i+1} - X_i| \le 1$ for all $0 \le i < m$. Let $\lambda > 0$ be arbitrary. Then $\mathbf{Pr}\left[|X_m - X_0| > \lambda\sqrt{m}\right] < 2\exp\left(-\lambda^2/2\right)$.*

Example 8.1.13. Let $\chi(H)$ be the chromatic number of a graph H. What is chromatic number of a random graph? How does this random variable behaves?

Consider the vertex exposure martingale, and let $X_i = \mathbf{E}\left[\chi(G) \mid G_i\right]$. Again, without proving it, we claim that $X_0, \ldots, X_n = X$ is a martingale, and as such, we have: $\mathbf{Pr}\left[|X_n - X_0| > \lambda\sqrt{n}\right] \le e^{-\lambda^2/2}$. However, $X_0 = \mathbf{E}[\chi(G)]$, and $X_n = \mathbf{E}\left[\chi(G) \mid G_n\right] = \chi(G)$. Thus,

$$\mathbf{Pr}\left[\left|\chi(G) - \mathbf{E}[\chi(G)]\right| > \lambda\sqrt{n}\right] \le e^{-\lambda^2/2}.$$

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