CSE 525: Randomized Algorithms **Spring 2023**

Lecture 9: Intro to Martingales

Lecturer: Shayan Oveis Gharan *04/25/2023*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

We have seen that if $X = X_1 + \cdots + X_n$ is a sum of independent $\{0,1\}$ random variables, then X is tightly concentrated around its expected value $\mathbb{E}[X]$. The fact that the random variables were $\{0,1\}$-valued was not essential; similar concentration results hold if we simply assume that they are in some bounded range $[-L, L]$.

We further see that if the independence condition is relaxed and the random variables are assumed to be negatively correlated still the same concentration bound holds. In this lecture we will discuss another generalization of the independence assumption.

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$. We say the sequence is $\{X_i\}$ is a martingale with respect to another sequence of random variables $\{Y_i\}$ if for every i, it holds that

$$\mathbb{E}[X_{i+1}|Y_0, Y_1, \ldots, Y_i] = X_i.$$

Note that this is equivalent to

$$\mathbb{E}[X_{i+1} - X_i|Y_0, Y_1, ..., Y_i] = 0.$$

If one thinks of $\{Y_0, Y_1, \ldots, Y_i\}$ as all the "information" up to time i, then this says that the difference $X_{i+1} - X_i$ is unbiased conditioned on the past information up to time i.

We say $X_0, \ldots,$ is a martingale w.r.t itself, if for any $i \geq 1$,

$$\mathbb{E}[X_{i+1}|X_0, X_1, \ldots, X_i] = X_i$$

for every $i = 0, 1, 2, \ldots$. In such a case, we can write for any i, we have

$$\mathbb{E}[X_i] = \mathbb{E}[\mathbb{E}[X_i|X_0, \ldots, X_{i-1}]] = \mathbb{E}[X_{i-1}] = \cdots = \mathbb{E}[X_0].$$

where we used the law of conditional expectations,

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

Example 9.1. *Consider a gambler that goes to a casino and bets every day. Let Y_i be a random variable indicating the amount of money the gambler wins or loses on day i and let X_i be the gamblers total win/loss up to day i. Then,*

$$\mathbb{E}[X_{i+1}|Y_0, \ldots, Y_{i-1}] = X_{i-1} + \mathbb{E}[Y_i] = X_{i-1}.$$

So, it is a martingale. Note that this is also a martingale even if the amount of bet on day i is dependent on the total win or loss up to day i. Here we assumed $\mathbb{E}[Y_i] = 0$ that is the casino is fair.

Remark 9.2. *The correct level of generality at which to define martingales involves a filtration. Formally, this is an increasing sequence of σ-algebras on our measure space $(\Omega, \mu, F) : F_0 \subseteq F_1 \subseteq \cdots \subseteq F$. Then, a sequence of random variables $\{X_i\}$ is a martingale with respect to the filtration $\{F_i\}$ if $\mathbb{E}[X_{i+1}|F_i] = X_i$ for every $i \geq 0$.*

9.1 Doob Martingales

One reason martingales are so powerful is that they model a situation where one gains progressively more information over time. Suppose that U is a set of objects, and $f : U \to \mathbb{R}$. Let X be a random variable taking values in U, and let $\{Y_i\}$ be another sequence of random variables. The associated Doob martingale is given by

$$X_i = \mathbb{E}\left[f(X)|Y_0, Y_1, ..., Y_i\right].$$

In words, this is our "estimate" for the value of $f(X)$ given the information contained in $\{Y_0, \ldots, Y_i\}$. To see that this is always a martingale with respect to $\{Y_i\}$, observe that

$$\mathbb{E}\left[X_{i+1}|Y_0, \ldots, Y_i\right] = \mathbb{E}\left[\mathbb{E}\left[f(X)|Y_0, \ldots, Y_{i+1}\right]|Y_0, \ldots, Y_i\right] = \mathbb{E}\left[f(X)|Y_0, \ldots, Y_i\right] = X_i,$$

where we have used the tower rule of conditional expectations.

Example 1: Balls in bins Suppose we throw m balls into n bins one at a time. At step i, we place ball i in a uniformly random bin. Let $Y_1, Y_2, \ldots, Y_m$ be the sequence of (random) choices, and let Y denote the final configuration of the system, i.e. exactly which balls end up in which bins. Now, we can consider a functional like $f(Y) = \#$ of empty bins. Then, if

$$X_i = \mathbb{E}\left[f(Y)|Y_1, \ldots, Y_i\right],$$

then $\{X_i\}$ is a (Doob) martingale. It is straightforward to calculate that

$$\mathbb{E}\left[X_m\right] = \mathbb{E}\left[X_0\right] = \mathbb{E}\left[f(Y)\right] = n \cdot (1 - 1/n)^m.$$

Suppose we are interested the concentration of $X_m = f(Y)$ around its mean value. Of course, we can write $X_m = Z_1 + \cdots + Z_n$ where Z_i is the indicator of whether the i-th been is empty after all the balls have been thrown. But note, unfortunately, that the $\{Z_i\}$ variables are not independent, in fact they are negatively correlated. In particular, if we know that $Z_1 = 1$ (bin 1 is empty), it decreases slightly the likelihood that other bins are empty. So, we can use concentration inequalities for Negatively correlated random variables, but we will see a different technique to prove that $\sum_i Z_i$ is concentrated.

The vertex exposure filtration Recall that $G_{n,p}$ denotes the random graph model where an undirected graph on n vertices is chosen by including every edge independently with probability p. Suppose the vertices are numbered $\{1, 2, ..., n\}$ Let $G \sim G_{n,p}$ and denote by G_i the induced subgraph on the vertices $\{1, \ldots, i\}$. G_0 denotes the empty graph. Let $\chi(G)$ denote the chromatic number of G, i.e., the minimum number of colors we need to obtain a proper coloring of the vertices of G and consider the Doob martingale $X_i = \mathbb{E}\left[\chi(G)|G_0, ..., G_i\right]$. If we wanted to understand concentration properties of $X_n = chi(G)$, this seems even more daunting. The chromatic number is a very complicated parameter of a graph, it is not even polynomial time computable (assuming $P \neq NP$)! Nevertheless, we will see that martingale concentration inequalities allow us to achieve tight concentration using very limited information about a sequence of random variables.

9.2 Azuma-Hoeffding Inequality

Say that a martingale $\{X_i\}$ has is $(c_1, c_2, \ldots,)$-bounded increments if $|X_i - X_{i-1}| \leq c_i$ for all $i \geq 1$ **with probability 1**.

Theorem 9.3. *If $\{X_i\}$ is $(c_1, c_2, \dots)$-bounded a martingale then, for every $\lambda > 0$, and $n \geq 0$, we have*

$$\mathbb{P}[X_n \geq X_0 + \lambda] \leq e^{-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}}, \quad \mathbb{P}[X_n \leq X_0 - \lambda] \leq e^{-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}}$$

Lemma 9.4. *Let X be a random variable with $\mathbb{E}[X] = 0$ and $|X| \leq 1$ w.p. 1. Then,*

$$\mathbb{E}\left[e^{aX}\right] \leq e^{a^2/2}.$$

Proof. The first observation is that e^{aX} is a convex function (in the range $[-1, +1]$). Therefore, it lies below the linear function which connects, e^{-a} to e^{+a}. In particular, for any $x \in [-1, +1]$, writing $x = \frac{1+x}{2}(+1) + \frac{1-x}{2}(-1)$ we have

$$e^{ax} \leq \frac{1+x}{2}e^{+a} + \frac{1-x}{2}e^{-a} = \frac{e^a + e^{-a}}{2} + x\frac{e^a - e^{-a}}{2}$$

Using $X \in [-1, 1]$, taking expectation from both sides we obtain

$$\begin{aligned}
\mathbb{E}\left[e^{aX}\right] &\leq \frac{e^a + e^{-a}}{2} \\
&\underset{\text{Taylor Expansion}}{=} 1 + \frac{a^2}{2} + \frac{a^4}{4!} + \cdots = \sum_{k=0}^{\infty} \frac{a^{2k}}{(2k)!} \\
&\leq \sum_{k=0}^{\infty} \frac{a^{2k}}{2^k k!} = e^{a^2/2}.
\end{aligned}$$

□

Now, we use the above lemma to prove the Azuma-Hoeffding inequality:

Proof. Let t be a parameter that we choose later. We write

$$\begin{aligned}
\mathbb{E}\left[e^{tX_n} | X_0, \dots, X_{n-1}\right] &= \mathbb{E}\left[e^{tX_{n-1}} e^{t(X_n - X_{n-1})} | X_0, \dots, X_{n-1}\right] \\
&= e^{tX_{n-1}} \mathbb{E}\left[e^{t(X_n - X_{n-1})} | X_0, \dots, X_{n-1}\right] \\
&\underset{\text{Lemma 9.4}}{\leq} e^{tX_{n-1}} e^{t^2 c_n^2/2}
\end{aligned}$$

where in the last inequality we used that $\mathbb{E}[X_n - X_{n-1} | X_0, \dots, X_{n-1}] = 0$ and that $|X_n - X_{n-1}| \leq c_n$ w.p. 1. Now, taking expectations from both sides, we write

$$\mathbb{E}\left[e^{tX_n}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{tX_n} | X_0, \dots, X_{n-1}\right]\right] \leq \mathbb{E}\left[e^{tX_{n-1}}\right] e^{t^2 c_n^2/2}.$$

So, by induction,

$$\mathbb{E}\left[e^{tX_n}\right] \leq \exp\left(t^2 \sum_{i=1}^n c_i^2/2\right)$$

Finally (assume $X_0 = 0$ for simplicity), we write,

$$\mathbb{P}[X_n \geq \lambda] = \mathbb{P}\left[e^{tX_n} \geq e^{t\lambda}\right] \leq \frac{\mathbb{E}\left[e^{tX_n}\right]}{e^{t\lambda}} \leq \exp\left(t^2 \sum_{i=1}^n c_i^2/2 - t\lambda\right)$$

Optimizing for t we need to set $t = \frac{\lambda}{\sum_{i=1}^n c_i^2}$. And, we get

$$\mathbb{P}[X_n \geq \lambda] \leq \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}\right)$$

as desired. □

9.3 Concentration of Lipschitz Functions

Let $f : U_1 \times U_2 \times \cdots \times U_n \to \mathbb{R}$ be a function. We say f is $c_1, c_2, \ldots, c_n$ Lipschitz if for any $1 \leq i \leq n$ and for any $x_1 \in U_1, x_2 \in U_2, \ldots, x_{i-1} \in U_{i-1}, x_i, x_i' \in U_i, , x_{i+1} \in U_{i+1}, \ldots, x_n \in U_n$ we have

$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i.$$

Then,

$$\mathbb{P}_{x_1,\ldots,x_n \sim U_1 \times \cdots \times U_n} \left[f(x_1, \ldots, x_n) \geq \mathbb{E} f + \lambda \right] \leq \exp\left(- \frac{\lambda^2}{2 \sum_{i=1}^n c_i^2} \right),$$

where $x_1, \ldots, x_n$ are chosen independently from $U_1, \ldots, U_n$.

Proof. Consider the product distribution, where Y_i is chosen from U_i independently. Define a Doob martinagle as follows: $X_0 = \mathbb{E} f(Y_1, \ldots, Y_n)$. For every i,

$$X_i = \mathbb{E}\left[f(Y) | Y_1, \ldots, Y_i \right] = \mathbb{E}_{x_{i+1},\ldots,x_n \sim U_{i+1} \times \cdots \times U_n} \left[f(Y_1, \ldots, Y_i, x_{i+1}, \ldots, x_n) \right].$$

Then, the resulting martingale is $c_1, \ldots, c_n$ bounded. So, by Azuma-Hoeffding's inequality it is concentrated.l □

Remark 9.5. *The above theorem means that if we have any set of independent random variable $Z_1, \ldots, Z_n$ and some quantity $f(Z)$ that we care about does not depend too much on changing any single piece of information, then $f(Z)$ is tightly concentrated about its mean. This is a vast generalization of the fact that sums of independent, bounded random variables are highly concentrated*

Remark 9.6. *This fact also generalizes beyond independent. Pemantle and Peres proved that any Lipschitz function of strongly Rayleigh distributions is tightly concentrated around its expected value. Cryan, Guo and Mousa proved that any Lipschitz function on (discrete) log-concave probability distributions is tightly concentrated around its expected value.*

9.4 Applications

Balls in Bins. First let's apply the Azuma-Hoeffding inequality to the balls and bins process. Recall that for a sequence of choices $Y_1, \ldots, Y_m$ (where Y_i is the bin that the i-th ball is thrown into), we put $f(Y_1, \ldots, Y_m)$ to be the number of empty bins. Then, clearly f is 1-Lipschitz: Changing the fate of ball i can only change the number of empty bins by 1. Therefore the corresponding martingale $X_i = \mathbb{E}\left[f | Y_1, \ldots, Y_i \right]$ has 1-bounded increments, and Azuma's inequality implies that

$$\mathbb{P}\left[X \geq \mathbb{E}\left[X \right] + \lambda \right] \leq e^{-\lambda^2/2m}$$

Recall that $X_0 = \mathbb{E}\left[X_n \right] = n(1 - 1/m)^n$. Consider the situation where $m = n$ and thus $X_0 \approx n/e$. If we $\lambda = c\sqrt{n}$, then with probability $1 - e^{-c^2}$, the number of empty bins is in the interval $[n/e - c\sqrt{n}, n/e + c\sqrt{n}]$

The chromatic number. Similarly, consider the vertex exposure martingale. We have to be a little more careful here to describe a graph G by a sequence $(Z_1, \ldots, Z_n)$ of independent random variables. The key is to think about Z_i containing the information on edges from vertex i to the vertices $\{1, \ldots, i-1\}$ so that we have independence. Since we can identify a graph G with the vector $(Z_1, \ldots, Z_n)$, we can think of the chromatic number as a function $\chi(Z_1, \ldots, Z_n)$. The function χ satisfies the 1-Lipschitz property because changing the edges adjacent to some vertex i can only change the chromatic number by 1. The chromatic number cannot increase by more than one because we could always color i a new color; it cannot decrease

by more than one because if we could color the graph without vertex i with c colors, then we can color the whole graph with $c+1$ colors. So the martingale $X_i = \mathbb{E}\left[\chi(G)|Z_1, \ldots, Z_i\right]$ has 1-bounded increments and Azuma's inequality tells us that

$$\mathbb{P}\left[\chi(G) \geq \mathbb{E}\left[\chi(G)\right] + \lambda\right] \leq e^{-\lambda^2/2n}.$$

Even without having any idea how to compute $\mathbb{E}\left[[]\,\chi(G)\right]$, we are able to say something significant about its concentration properties.

Remark 9.7. *It turns out that if $G \sim G(n, 1/2)$, then $\mathbb{E}\left[[]\,\chi(G)\right] \approx n/(2\log_2 n)$, so the concentration window - which is $O(\sqrt{n})$ - is again quite small with respect to the expectation.*