CS 761 : Randomized Algorithms , Spring 2023, Waterloo

Lecture 8 : Martingales

We study the concept of martingales and see some examples.

Then we prove the Azuma-Hoeffding inequality and derive some consequences.

Martingales

Definition A sequence of random variables $X_1, X_2, \ldots$ is a martingale with respect to the sequence of random variables $Y_1, Y_2, \ldots$ if for all $i \geq 1$, the following conditions hold:

- X_i is a function of $Y_1, Y_2, \ldots, Y_i$;
- $\mathbb{E}[|X_i|] < \infty$
- $\mathbb{E}[X_{i+1} \mid Y_1, Y_2, \ldots, Y_i] = X_i$.

A sequence of random variables $X_1, X_2, \ldots$ is called a martingale if it is a martingale with respect to itself.

A standard example is when Y_i denotes the amount a gambler wins on the i-th game, and X_i denotes the total winning at the end of game i.

Assuming the games are fair, i.e. $E[Y_i]=0$, then it is clear that $X_1, X_2, \ldots$ is a martingale wrt $Y_1, Y_2, \ldots$.

Note that we don't assume Y_i are independent. e.g. the bet size of the i-th game could depend on previous outcomes.

Another example is from random walks on an infinite 2-dimensional grid.

Let Y_i be the current position where Y_1 is the origin. and X_i be the ℓ_1-distance from the origin.

Assume that we move up/down/left/right with equal probability at each step. then $X_1, X_2, \ldots$ is a martingale wrt $Y_1, Y_2, \ldots$.

Note again that we don't assume the Y_i are independent, e.g. the step size at each step could depend on current position.

We will see an interesting example of this type of random walk process next time.

Conditional Expectations

Before we see more properties and more examples, let's review some basic identities about conditional expectations.

For two random variables X, Y, the conditional expectation of X wrt $Y=y$ is $\mathbb{E}[X \mid Y=y]$,

and so $\mathbb{E}[X|Y]$ is a random variable which is a function of Y, and thus

$$\mathbb{E}[\mathbb{E}[X|Y]] = \sum_y \Pr(Y=y) \cdot \mathbb{E}[X|Y=y] = \sum_y \Pr(Y=y) \cdot \sum_x x \cdot \Pr(X=x|Y=y) = \sum_x x \cdot \sum_y \Pr(X=x \cap Y=y) = \mathbb{E}[X].$$

This should be understood as a way to compute $\mathbb{E}[X]$, by first conditioned on the value of Y, and then take the weighted average depending on the probability that $Y=y$.

The same identity holds even when we condition on another random variable Z, i.e.

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

While the notation may look complicated, it is easier to understand if we fix the value of $Z=z$.

We will use the following facts. Let X be a random variable, and $\vec{Y},\vec{Z}$ be sequences of random variables.

- For any arbitrary functions f and g, $\mathbb{E}[\,\mathbb{E}[f(X)\,g(X,\vec{Y})\,|\,X]\,] = \mathbb{E}[\,f(X)\cdot\mathbb{E}[g(X,\vec{Y})|X]\,]$.
- $\mathbb{E}[X] = \mathbb{E}[\,\mathbb{E}[X|\vec{Y}]\,]$
- $\mathbb{E}[X|\vec{Z}] = \mathbb{E}[\,\mathbb{E}[X|\vec{Y},\vec{Z}]\,|\,\vec{Z}]$.

One basic property of martingale is $\mathbb{E}[X_i] = \mathbb{E}[\mathbb{E}[X_i|Y_1,\dots Y_{i-1}]] = \mathbb{E}[X_{i-1}] = \mathbb{E}[X_{i-2}] = \dots = \mathbb{E}[X_0]$,

which just says that no matter what information we conditioned on it holds that $\mathbb{E}[X_i]=\mathbb{E}[X_{i-1}]$.

<u>Doob Martingale</u>

A useful way to construct a martingale is by estimating a quantity with progressively more information.

Let X be a random variable with $\mathbb{E}[X]<\infty$, and $Y_1,Y_2,\dots,Y_n$ be a sequence of random variables.

The associated Doob martingale is $Z_i = \mathbb{E}[X \mid Y_1,Y_2,\dots,Y_i]$ for $0\le i\le n$.

To see that $Z_1,\dots,Z_n$ is a martingale wrt $Y_1,\dots,Y_n$, note that

$$\mathbb{E}[Z_{i+1}|Y_1,\dots,Y_i] = \mathbb{E}[\,\mathbb{E}[X|Y_1,\dots,Y_i,Y_{i+1}]\,|\,Y_1,\dots,Y_i] = \mathbb{E}[X|Y_1,\dots,Y_i] = Z_i.$$

Informally, it just says that the <u>average</u> of the estimated value of X with the additional information Y_{i+1}

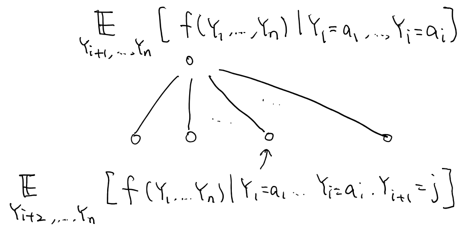
is just equal to the estimated value of X before the additional information is revealed.

To get a better intuition, consider the case where $X=f(Y_1,Y_2,\dots Y_n)$,

and now the values of $Y_1,Y_2,\dots,Y_n$ are revealed one at a time.

Then, the weighted average of the new estimations, by definition,

is equal to the previous estimation, i.e.



$$\mathbb{E}[f(\vec{Y})|Y_1=a_1,\dots,Y_i=a_i] = \sum_j \Pr(Y_{i+1}=j)\cdot\mathbb{E}[f(\vec{Y})|Y_1=a_1,\dots,Y_i=a_i,Y_{i+1}=j] = \mathbb{E}_{Y_{i+1}}[\,\mathbb{E}[f(\vec{Y})|Y_1=a_1,\dots,Y_i=a_i,Y_{i+1}]\,]$$

<u>Balls and Bins</u>: It is a simple random process where we throw m balls into n bins one at a time,

where ball i is placed in a uniform random bin at step i.

Let $C_1,\dots,C_m$ be the sequence of random variables where C_i is the location of ball i.

Let X be the number of empty bins after all m balls are thrown.

Then $Z_i = \mathbb{E}[X|C_1,\dots,C_i]$ is a (Doob) martingale.

We will prove a concentration bound for X using this martingale.

<u>Vertex Exposure Martingale</u> Let $G\sim G_{n,p}$ and G_i be the induced subgraph on the vertices $\{1,2,\dots,i\}$.

Let $\chi(G)$ be the chromatic number of G, and consider the Doob martingale $Z_i = \mathbb{E}[\chi(G) \mid G_1, \ldots, G_i]$.

It is not easy to compute $\mathbb{E}[\chi(G)]$, but we will use this martingale to prove a concentration bound for $\chi(G)$.

Azuma-Hoeffding Inequality

The following theorem is a generalization of the Chernoff-Hoeffding theorem in L03, where the independence assumption is replaced by the martingale property, and the bounded variable assumption is replaced by bounded differences.

Theorem Let $X_1, X_2, \ldots$ be a martingale with respect to $Y_1, Y_2, \ldots$. Suppose the X_i's satisfy the bounded difference condition that $|X_i - X_{i-1}| \le c_i$ for some real number c_i for $i > 1$ always (with probability 1).

Then, $\Pr(|X_n - X_1| > \lambda) \le 2 \cdot \exp\left(\frac{-\lambda^2}{2\sum_i c_i^2}\right)$.

The proof is similar to that of Chernoff bound.

We start by bounding $\Pr(X_n > \lambda) = \Pr(e^{tX_n} > e^{t\lambda}) \le \frac{\mathbb{E}[e^{tX_n}]}{e^{t\lambda}}$ for any $t > 0$.

Consider the martingale difference sequence $D_i = X_i - X_{i-1}$ for $i > 1$. Note that $\mathbb{E}[D_i] = 0$.

We compute the numerator by conditional expectation so that

$$\mathbb{E}[e^{tX_n}] = \mathbb{E}\left[\mathbb{E}[e^{tX_{n-1}} e^{tD_n} \mid Y_1, \ldots, Y_{n-1}]\right] = \mathbb{E}\left[e^{tX_{n-1}}\, \mathbb{E}[e^{tD_n} \mid Y_1, \ldots, Y_{n-1}]\right]$$

We compute $\mathbb{E}[e^{tD_n} \mid Y_1, \ldots, Y_{i-1}]$ using the following claim.

Claim Let D be a random variable such that $|D| \le 1$ always and $\mathbb{E}[D] = 0$. Then $\mathbb{E}[e^{tD}] \le e^{t^2/2}$.

Proof The idea is the same as in Hoeffding's extension in L03.

Note that e^{tx} is convex in the interval $x \in [-1, +1]$ for any $t > 0$, and so its graph lies below the line joining $(-1, e^{-t})$ and $(1, e^t)$, which has the equation $\frac{1}{2}(e^t + e^{-t} + x(e^t - e^{-t}))$.

Therefore, $\mathbb{E}[e^{tD}] \le \mathbb{E}\left[\frac{1}{2}(e^t + e^{-t} + D(e^t - e^{-t}))\right] = \frac{1}{2}(e^t + e^{-t}) = \sum_{i \ge 0} \frac{t^{2i}}{(2i)!} \le \sum_{i \ge 0} \frac{t^{2i}}{2^i \cdot i!} = e^{t^2/2}$. □

Using the claim by scaling, $\mathbb{E}[e^{tD_n} \mid Y_1, \ldots, Y_{i-1}] \le e^{t^2 c_n^2/2}$.

It follows by induction that $\mathbb{E}[e^{tX_n}] = \mathbb{E}\left[e^{tX_{n-1}}\, \mathbb{E}[e^{tD_n} \mid Y_1, \ldots, Y_{n-1}]\right] \le e^{t^2 c_n^2/2} \cdot \mathbb{E}[e^{tX_{n-1}}] \le e^{t^2 \sum_i c_i^2/2}$.

Therefore, $\Pr(X_n > \lambda) \le \min_{t>0} \mathbb{E}[e^{tX_n}] / e^{t\lambda} \le \min_{t>0} e^{t^2 \sum_i c_i^2/2 - t\lambda} \le e^{-\lambda^2/2\sum_i c_i^2}$ by setting $t = \lambda / \sum_i c_i^2$.

This completes the proof for the upper tail, while the proof for the lower tail is symmetrical, and we are done.

Exercise: Derive the Chernoff-Hoeffding bound in L03 from the Azuma-Hoeffding inequality.

The Method of Bounded Differences

Applying Azuma-Hoeffding inequality on a Doob martingale gives the following corollary.

Theorem Let $Y_1, \dots, Y_n$ be an arbitrary set of random variables and let f be a function satisfying the property that for each $i \in [n]$, there is a real number c_i such that $\left| \mathbb{E}[f \mid Y_1, \dots, Y_i] - \mathbb{E}[f \mid Y_1, \dots, Y_{i-1}] \right| \le c_i$.

Then. $\Pr[\, |f - \mathbb{E}f| > \lambda \,] \le \exp(-\lambda^2 / 2\sum_i c_i^2)$.

The advantage of this theorem is that there is no assumption that $Y_1, \dots, Y_n$ are independent.
but it could be difficult to bound c_i in practice.

The following special case is easier to apply but requires independence.

We say a function $f(x_1, \dots, x_n)$ satisfies the Lipschitz property with constants c_i for $i \in [n]$ if $|f(a) - f(a')| \le c_i$ whenever a and a' differ in just the i-th coordinate for $i \in [n]$.

Theorem If $f(X_1, \dots, X_n)$ satisfies the Lipschitz property with constants c_i for $i \in [n]$ and $X_1, \dots, X_n$ are independent random variables, then $\Pr_{X_1 \dots X_n}[\, |f - \mathbb{E}f| > \lambda \,] \le 2 \cdot \exp(-\lambda^2 / 2\sum_i c_i^2)$.

Proof We just need to check $\left| \mathbb{E}[f \mid X_1, \dots, X_i] - \mathbb{E}[f \mid X_1, \dots, X_{i-1}] \right| \le c_i$ for $i \in [n]$ so that we can apply previous theorem.

To do so, it suffices to check that $\left| \mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = a] - \mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = b] \right| \le c_i \quad \forall a, b$ for $i \in [n]$.

Since the random variables $X_1, \dots, X_n$ are independent,

$$\mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = a] = \sum_{a_{i+1}, \dots, a_n} \mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = a, X_{i+1} = a_{i+1}, \dots, X_n = a_n] \cdot \Pr[X_{i+1} = a_{i+1}, \dots, X_n = a_n \mid X_1, \dots, X_{i-1}]$$

$$= \sum_{a_{i+1}, \dots, a_n} f(X_1, \dots, X_{i-1}, a, a_{i+1}, \dots, a_n) \cdot \Pr[X_{i+1} = a_{i+1}, \dots, X_n = a_n]$$

The same holds when $X_i = b$, and thus $\left| \mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = a] - \mathbb{E}[f \mid X_1, \dots, X_{i-1}, X_i = b] \right| \le c_i$ from Lipschitz property. $\square$

Note that even in this simplification, it is already a significant generalization that the sum of independent, bounded random variables is highly concentrated around its expected value.

Simple Applications

Let's consider the two examples of Doob martingales that we have seen above.

Balls and bins Let $C_1, \dots, C_m$ be the (random) bin of ball i, and $X = f(C_1, \dots, C_m)$ be the number of empty bins.

Then clearly f is 1-Lipschitz, as changing the location of one ball could only change # empty bins by 1.

Therefore, by the theorem, $\Pr\left[\, |\text{\#empty bins} - \mathbb{E}[\text{\# of empty bins}] > \lambda \,\right] \le 2e^{-\lambda^2/2m}$.

So, with high probability, # empty bins is concentrated within a window of $O(\sqrt{m})$ to the expected value.

It is a simple exercise that $\mathbb{E}[\text{\# empty bins}] = n(1 - \frac{1}{n})^n \approx \frac{n}{e}$ when $m = n$ where n is # bins.

So, the window of concentration is a lower-order term comparing to the expected value.

Note that this is an example where Chernoff bound does not apply, as whether bin i is empty is not independent on whether other bins are empty or not (although the dependency is not that strong).

Graph coloring Consider the expected chromatic number $\mathbb{E}[\chi(G)]$ of $G \sim G_{n,p}$.

We use the vertex exposure martingale, where each vertex i is revealed one by one, with the edges from i to $\{1,2,\dots,i-1\}$ are independent random variables from the edges that we have seen before.

So, let F_i be the set of edges from i to $\{1,2,\dots,i-1\}$ and $\chi(G) = f(F_1, F_2, \dots, F_n)$.

Then, observe that f is 1-Lipschitz, as changing the edges of one vertex could only change $\chi(G)$ by one.

Therefore, by the theorem, $\Pr[\,|\mathbb{E}[\chi(G)] - \chi(G)| > \lambda\,] \le 2e^{-\lambda^2/2n}$.

This is an example where we can prove concentration without knowing the expected value.

It turns out that $\mathbb{E}[\chi(G)] = \frac{n}{2\log_2 n}$ (see Alon-Spencer chapter 10), and so again the window of concentration (which is $O(\sqrt{n})$) is a lower-order term comparing to the expected value.

References
- Chapter 12 of Mitzenmacher-Upfal.
- Chapter 5 of "Concentration of measure for the analysis of randomized algorithms" by Dubhashi and Panconesi, in which there are generalizations of the method of bounded differences.