CSE 632 (Fall 2019): Analysis of Algorithms II : Randomized Algorithms

Lecture 15-16 (10/16,18/2019): Martingales and Azuma's Inequality

Lecturer: Shi Li *Scriber: John Rivera, Sen Lin*

1 Introduction

This lecture is about martingales. Let us start with an motivating example. Consider the following two games that a gambler can play:

Algorithm 1 Gambler's Game 1

for $i \leftarrow 1$ to n **do**
  the gambler gets a score of $\begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$

The gambler wins the game if the total score he get in $0.1n$. Using Chernoff bounds we learned from the last lecture, we can prove that the probability that the gambler wins the game is exponentially small, i,e, at most e^{-cn} for some constant $c > 0$. Now let us consider a different game, where the gambler has some control on what he can choose in each turn:

Algorithm 2 Gambler's Game 2

for $i \leftarrow 1$ to n **do**
  the gambler chooses any distribution for X_i over $[-1, 1]$ s.t. $\mathbb{E}[X_i] = 0$; the distribution may depend on the realizations of $X_1, X_2, \cdots, X_{i-1}$
  the gambler gets a score of X_i, randomly selected from the distribution he chooses

Again, the gambler wins the game if the total score he gets is at least $0.1n$. Now since the gambler can choose the distributions, can he win the game with much high probability than he can in the first game? In this lecture, we show that the answer is no. The accumulative scores he get in the game form a *martingale* sequence, which is the topic of the lecture.

2 Martingale Sequence

Definition 1. *A martingale sequence is a sequence $Y_1, Y_2, \cdots, Y_n$ of random variables such that for every $i \in \{2, 3, \cdots, n\}$, we have $\mathbb{E}[Y_i|Y_1, Y_2, Y_3, \cdots, Y_{i-1}] = Y_{i-1}$.*

A sequence $Y_1, Y_2, \cdots, Y_n$ of random variables is called a martingale sequence with respect to another sequence $X_1, X_2, \cdots, X_n$ if the following conditions hold.

1. *For every $i \in [n]$, Y_i is completely determined by $X_1, X_2, \cdots, X_i$.*
2. *For every $i \in \{2, 3, \cdots, n\}$, we have $\mathbb{E}[Y_i|X_1, X_2, \cdots, X_{i-1}] = Y_{i-1}$.*

Referring to the gambler's game 2 above, X_i is the score we obtain in the i-th iteration. Let $Y_i = \sum_{j=1}^{i} X_i$ is the total score we obtain from iteration 1 to iteration i. Then $Y_1, Y_2, \cdots, Y_n$ is a martingale sequence w.r.t $X_1, X_2, \cdots, X_n$ since $\mathbb{E}[Y_i|X_1, X_2, \cdots, X_{i-1}] = X_1 + X_2 + \cdots + X_{i-1} + \mathbb{E}[X_i|X_1, X_2, \cdots, X_{i-1}] = Y_{i-1}$ since $\mathbb{E}[X_i|X_1, X_2, \cdots, X_{i-1}] = 0$. This can also be used to prove that $Y_1, Y_2, \cdots, Y_n$ is a martingale sequence by itself.

3 Azuma's Inequality

Theorem 2 (Azuma's Inequality)**.** *Let $Y_1, Y_2, \cdots, Y_n$ be a martingale sequence with respect to another sequence $X_1, X_2, \cdots, X_n$. Let $Y_0 = \mathbb{E}[Y_1]$. Assume for $i \in [n]$, we always have $|Y_i - Y_{i-1}| \leq c_i$.*

Then $\forall t \geq 0$*, we have:*

$$\Pr[Y_n - Y_0 \geq t] \leq \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right) \text{ and } \Pr[Y_n - Y_0 \leq -t] \leq \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

Proof. Let $\lambda > 0$ be any number, then

$$\Pr[Y_n - Y_0 \geq t] = \Pr\left[e^{\lambda(Y_n - Y_0)} \geq e^{\lambda \cdot t}\right] \leq \frac{\mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right]}{e^{\lambda \cdot t}} \qquad \text{by Markov's Inequality.}$$

For simplicity, we let $\Delta_n = Y_n - Y_{n-1}$. Then we always have $|\Delta_n| \leq c_n$. We then bound $\mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right]$ as follows

$$\mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right] = \mathbb{E}\left[e^{\lambda(Y_{n-1} - Y_0 + \Delta_n)}\right] = \mathop{\mathbb{E}}_{X_1, X_2, \cdots, X_{n-1}}\left[e^{\lambda(Y_{n-1} - Y_0)}\,\mathbb{E}\left[e^{\lambda \Delta_n} | X_1, X_2, \cdots, X_{n-1}\right]\right].$$

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda \Delta_n} | X_1, X_2, \cdots, X_{n-1}\right] &\leq \mathbb{E}\left[\frac{(c_n + \Delta_n)e^{\lambda c_n} + (c_n - \Delta_n)e^{-\lambda c_n}}{2c_n}\Big| X_1, X_2, \cdots, X_{n-1}\right] \\
&= \frac{c_n e^{\lambda c_n} + c_n e^{-\lambda c_n}}{2c_n} = \frac{e^{\lambda c_n} + e^{-\lambda c_n}}{2}
\end{aligned}$$

The inequality is by convexity of the exponential function and that $\Delta_n \in [-c_i, c_i]$. The first equality used that $\mathbb{E}[\Delta_n | X_1, X_2, \cdots, X_{n-1}] = \mathbb{E}[Y_n | X_1, X_2, \cdots, X_{n-1}] - Y_{n-1} = 0$.

Thus, we have

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda \Delta_n} | X_1, X_2, \cdots, X_{n-1}\right] &\leq \frac{e^{\lambda c_n} + e^{-\lambda c_n}}{2} \mathop{\mathbb{E}}_{X_1, X_2, \cdots, X_{n-1}}\left[e^{\lambda(Y_{n-1} - Y_0)} | X_1, X_2, \cdots, X_{n-1}\right] \\
&= \frac{e^{\lambda c_n} + e^{-\lambda c_n}}{2} \mathbb{E}\left[e^{\lambda(Y_{n-1} - Y_0)}\right].
\end{aligned}$$

We can prove the above inequality for $n-1, n-2, \cdots, 1$. Combining all the inequalities, we have

$$\mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right] \leq \prod_{i=1}^n \left(\frac{1}{2}e^{-\lambda c_i} + e^{\lambda c_i}\right) \mathbb{E}[e^{\lambda(Y_0 - Y_0)}] \leq \prod_{i=1}^n e^{(\lambda c_i)^2/2} = \exp\left(\prod_{i=1}^n (\lambda c_i)^2/2\right).$$

In the above sequence, we used that $\frac{e^x + e^{-x}}{2} \leq e^{x^2/2}$ for every $x \in \mathbb{R}$. This can be seen from the Taylor series:

$$\begin{aligned}
e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} \cdots \\
e^{-x} &= 1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} \cdots \\
\frac{e^x + e^{-x}}{2} &= 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \cdots \\
e^{x^2/2} &= 1 + \frac{x^2}{2} + \frac{x^4}{2 \cdot 2!} + \frac{x^6}{2 \cdot 3!} + \cdots
\end{aligned}$$

If we choose $\lambda = \dfrac{t}{\sum_{i=1}^n c_i^2}$, then

$$\Pr\left[Y_n - Y_n \geq t\right] \leq \frac{\mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right]}{e^{\lambda t}} \leq \frac{\exp\left(\prod_{i=1}^n (\lambda c_i)^2/2\right)}{e^{\lambda t}} = \exp\left(\sum_{i=1}^n (\lambda c_i)^2/2 - \lambda t\right).$$

Let $A = \sum_{i=1}^n c_i^2$ and $\lambda = \frac{t}{A}$, and the quantity on the right side is

$$\exp\left(\frac{\lambda^2 \cdot A}{2} - \lambda t\right) = \exp\left(\frac{t^2}{2A} - \frac{t^2}{A}\right) = \exp\left(-\frac{t^2}{2A}\right) = \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right).$$

This finishes the proof of the theorem. □

Then we go back to the gambler's game 2. Then we can define $c_1 = c_2 = \cdots = c_n = 1$ and $t = 0.1n$, then we have (notice that $Y_0 = 0$)

$$\Pr[Y_n \geq 0.1n] \leq \exp\left(-\frac{0.01n^2}{2n}\right) = e^{-n/200}.$$

So the probability is still exponentially small in n.

4 Doob Martingales

Theorem 3. *Let $A, X_1, X_2, \ldots, X_n$ be random variables in a common probability space. For every $i \in \{0, 1, 2, 3, \cdots, n\}$, let $Y_i = \mathbb{E}[A|X_1, X_2, \ldots, X_i]$. Then $Y_1, Y_2, \ldots, Y_n$ is a martingale sequence w.r.t $X_1, X_2, \ldots, X_n$. The martingale is called is the Doob martingale for $Y_1, Y_2, \cdots, Y_n$.*

A special but very common case is when A is completely decided by $X_1, X_2, \cdots, X_n$. That is, there is a function f such that $A = f(X_1, X_2, \cdots, X_n)$.

Proof of Theorem 3. First, for every $i \in [n]$, Y_i is completed determined by $X_1, \ldots, X_i$ by the definition. It suffices to check if $\mathbb{E}[Y_i|X_1, X_2, \ldots, X_{i-1}] = Y_{i-1}$ for every $i \in \{2, 3, \cdots, n\}$.

$$\begin{aligned}
&\mathbb{E}[Y_i|X_1, X_2, \ldots, X_{i-1}] \\
&= \mathop{\mathbb{E}}_{X_i|X_1,X_2,\ldots,X_{i-1}} \mathbb{E}[Y_i|X_1, X_2, \cdots, X_{i-1}, X_i] \quad Y_i \text{ is deterministic conditioned on } X_1, \cdots, X_i \\
&= \mathop{\mathbb{E}}_{X_i|X_1,X_2,\ldots,X_{i-1}} \mathbb{E}[A|X_1, X_2, \ldots, X_{i-1}, X_i] \\
&= \mathbb{E}[A|X_1, X_2, \ldots, X_{i-1}] \\
&= Y_{i-1}
\end{aligned}$$

□

5 Application of Azuma's Inequality to Lipschitz Functions

5.1 Lipschitz Property

Definition 4. *A function $f : (x_1, x_2, \ldots, x_n) \to \mathbb{R}$ is said to have Lipschitz constant $c_i \geq 0$ on the i-th coordinate, if for every $x_1, x_2, \cdots, x_n$ and x_i' we have*

$$|f(x_1, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i$$

Lemma 5. *Let $f : (x_1, x_2, \cdots, x_n) \to \mathbb{R}$ be a function with Lipschitz constants $c_1, c_2, \cdots, c_n$ on the n coordinates. Assume $X_1, X_2, \ldots, X_n$ are **independent** random variables. Let $\mu = \mathbb{E}[f(X_1, X_2, \ldots, X_n)]$. Then $\forall t \geq 0$, we have*

$$\Pr[f(X_1, X_2, \ldots, X_n) - \mu \geq t] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right), \text{ and}$$
$$\Pr[f(X_1, X_2, \ldots, X_n) - \mu \leq -t] \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right).$$

Notice that the Chernoff bounds can be applied when $f = \sum_{i=1}^n X_i$. Using Azuma's inequality, we can prove concentration bounds for any Lipschitz function f.

Proof. For every $i \in [n]$, we define $Y_i = \mathbb{E}[f(X_1, X_2, \ldots, X_n)|X_1, X_2, \ldots, X_i]$ (Doob martingale). What needs to be proved is that for every i,

$$|Y_i - Y_{i-1}| \leq c_i.$$

Fix $X_1, X_2, \cdots, X_i$, we have from the definition of Y-variables and the independence of X variables

$$\begin{aligned}
Y_i &= \mathop{\mathbb{E}}_{X_{i+1},X_{i+2},\ldots,X_n} [f(X_1, X_2, \ldots, X_n)] \\
Y_{i-1} &= \mathop{\mathbb{E}}_{X_i',X_{i+1},X_{i+2},\ldots,X_n} [f(X_1, X_2, X_{i-1}, X_i', X_{i+1}, X_n)]
\end{aligned}$$

Above, X_i' is a random variable that has the same distribution as X_i. Notice that since we already fixed X_i, we need to use X_i' in the definition of Y_{i-1}. Thus,

$$\begin{aligned}Y_i - Y_{i-1} &= \mathop{\mathbb{E}}_{X_{i+1},X_{i+2},\ldots,X_n}[f(X_1, X_2, \ldots, X_n)] \\ &- \mathop{\mathbb{E}}_{X_i',X_{i+1},X_{i+2},\ldots,X_n}[f(X_1, X_2, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)] \\ &= \mathop{\mathbb{E}}_{X_i',X_{i+1},\ldots,X_n}[f(X_1, X_2, \ldots, X_n) - f(X_1, \ldots, X_{i-1}, X_{i'}, X_{i+1}, \ldots, X_n)]\end{aligned}$$

Notice that the quantity inside $\mathbb{E}[]$ is always between $-c_i$ and c_i, we have $|Y_i - Y_{i-1}| \le c_i$.

Additionally,

- $Y_0 = \mathbb{E}[f(X_1, X_2, \ldots, X_n)] = \mu$
- $Y_n = \mathbb{E}[f(X_1, X_2, \ldots, X_n)|X_1, X_2, \ldots, X_n] = f(X_1, X_2, \ldots, X_n)$

The difference between Y_n and μ is $f(X_1, X_2, \ldots, X_n) - \mu$, which is a function of $X_1, X_2, \ldots, X_n$. Then, applying the Azuma's inequality, Lemma 5 is proved. □

6 Examples

6.1 Bin Packing of Items of Random Sizes

Let $X_1, X_2, \ldots, X_n$ be independent random variables from $[0, 1]$ (they do not necessarily have the same distribution). Let $f(X_1, X_2, \ldots, X_n)$ be the minimum number of boxes of capacity 1 to hold the n items of sizes $X_1, X_2, \ldots, X_n$ respectively.

Observation 6. *For every $i \in [n]$, $x_1, x_2, \cdots, x_n$ and x_i', we have*

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \le 1.$$

Let $\mu = \mathbb{E}[f(X_1, X_2, \ldots, X_n)]$. Applying Lemma 5, we have

Lemma 7. *$\forall t \ge 0$, $\Pr[f(X_1, X_2, \ldots, X_n) - \mu > t] \le \exp(-\frac{t^2}{2n})$.*

The problem of computing $f(x_1, x_2, \ldots, x_n)$ is actually a NP-hard problem, but Lemma 7 always holds.

6.2 Chromatic Number of Random Graphs

Suppose we have a graph $G \sim \mathcal{G}(n, p)$ which means that we have n vertices in G and each edge is present with probability p independently. Indeed, the analysis in the section works even if edges have different probabilities of being present in G, as long as the events are independent.

Let $f(G)$ be the minimum number of colors to color the vertices of G, s.t. there are not two adjacent vertices having the same color. (Vertex Coloring Problem).

First we define X_i to be the set of edges in G between i and $\{1, 2, \ldots, i-1\}$

- X_1: ϕ
- X_2: the set of edges between 2 and $\{1\}$
- X_3: the set of edges between 3 and $\{1, 2\}$
- ...

Then, $G = (V, X_1 \cup X_2 \cup X_3 \cdots X_n)$, so G and $f(G)$ are determined by $X_1, X_2, \cdots, X_n$. We define $Y_i = \mathbb{E}[f(G)|X_1, X_2, \ldots, X_i]$. i.e., given $X_1, X_2, \ldots, X_i$, then Y_i is computed by the expectation of $f(G)$ with sampling $X_{i+1}, X_{i+2}, \ldots, X_n$.

And we have:

Observation 8. $|f(G(X_1, X_2, \ldots, X_n)) - f(G(X_1, X_2, \ldots, X_{i-1}, X_i', X_{i+1} \ldots, X_n))| \le 1$.

I.e, any changes to X_i will at most increase one color, because given a coloring for $G(X_1, X_2, \ldots, X_n)$, can always use a new color for i ot obtain a coloring for $G(X_1, X_2, \cdots, X_{i-1}, X_i', X_{i+1}, \cdots, X_n)$. Then as the Section 6.1, we can have the bound that

$$\Pr[f(G) - \mu \ge t] \le \exp\left(-\frac{t^2}{2n}\right).$$