CSE525: Randomized Algorithms and Probabilistic Analysis

Lecture 18

Lecturer: Anna Karlin *Scribe: Alex Polozov*

1 Doob's martingale process

Let $Y_1, \ldots, Y_n$ be an arbitrary sequence of random variables. Let X be some random variable with finite expectation: $\mathbb{E}[X] < \infty$. We define *Doob's process* as follows:

$$X_0 \stackrel{\text{df}}{=} \mathbb{E}[X]$$
$$X_n \stackrel{\text{df}}{=} \mathbb{E}[X \mid Y_1, \ldots, Y_n]$$

Theorem 1. *Doob's process is a martingale.*

Proof By the law of total expectation ($\mathbb{E}[V \mid W] = \mathbb{E}[\mathbb{E}[V \mid U, W] \mid W]$),

$$\mathbb{E}[X_{n+1} \mid Y_1, \ldots, Y_n] = \mathbb{E}[\mathbb{E}[X \mid Y_1, \ldots, Y_{n+1}] \mid Y_1, \ldots, Y_n] = \mathbb{E}[X \mid Y_1, \ldots, Y_n] = X_n$$

■

1.1 Edge exposure martingale

An example of a Doob's process is an edge exposure martingale, which helps to calcuate an expectation of some graph-theoretic function of a random graph.

Let $G(n, p)$ be a random graph on n vertices, where each of the $m = \binom{n}{2}$ edges is present with probability p. Fix some particular ordering of the edges $e_1, \ldots, e_m$. Let $f(G)$ be any graph-theoretic function, such as chromatic number, maximum independent set size, maximum clique size etc.

We define m independent random variables $Y_1, \ldots, Y_m$:

$$Y_i \stackrel{\text{df}}{=} \begin{cases} 1 & \text{if edge } e_i \text{ is present with probability } p \\ 0 & \text{otherwise} \end{cases}$$

Then $X_k \stackrel{\text{df}}{=} \mathbb{E}[f(G) \mid Y_1, \ldots, Y_k]$ is a Doob's martingale. This is a conditional expectation of a function f, given a partial information about fixed states of k edges in this graph.

An illustration for this martingale type is given in figure 1. Here $X_i = \mathbb{E}[\text{max clique size} \mid e_1, \ldots, e_i]$. Figure 1 shows a tree of possibilities, branching on the state of every exposed edge e_i. Without any prior knowledge $X_0 = \mathbb{E}[\text{max clique size}] = 2$ (it can be verified by averaging clique sizes for all the leaves of the tree). The numbers in tree vertices show our max clique size expectation, conditioned upon the state of exposed edges – the value of random variable X_i. The tree also illustrates the main martingale property: at each vertex, the average of child numbers (the conditional expectation of X_{i+1}) is equal to the value of X_i.

2 Applications of the Optional Sampling Theorem

In the previous lecture we introduced the Optional Sampling Theorem:

Theorem 2. *Let $\{Z_t\}$ be a martingale with respect to a sequence $\{X_t\}$. If T is a stopping time for $\{X_t\}$, then $\mathbb{E}[Z_t] = \mathbb{E}[Z_0]$ wherever any of the following conditions holds:*

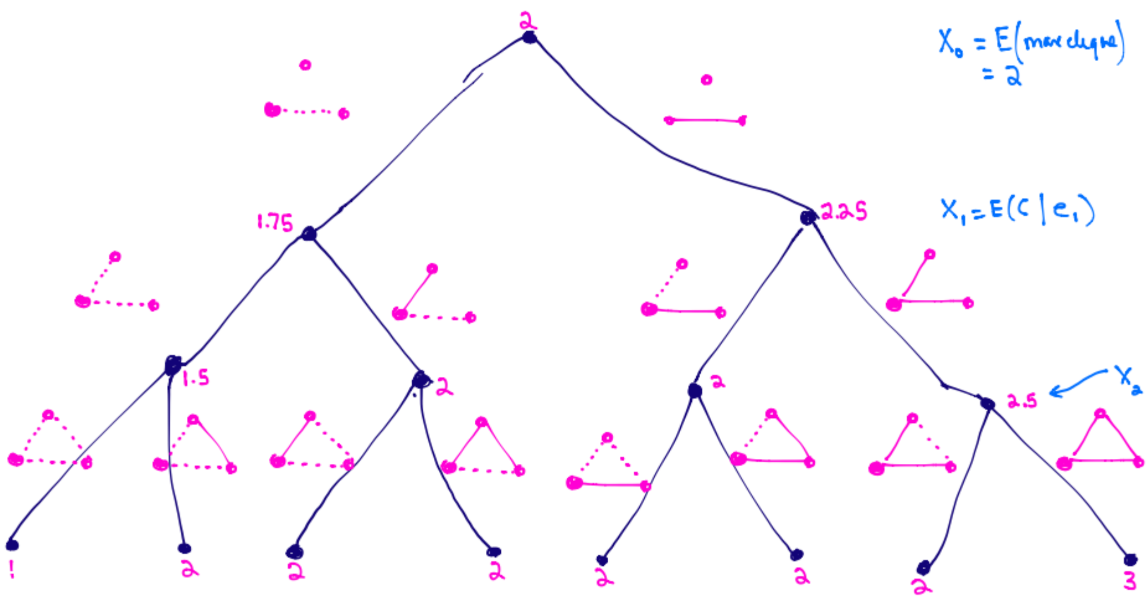

Figure 1: A tree of possibilities for the edge exposure martingale $X_i = \mathbb{E}[\text{max clique size} \mid e_1, \ldots, e_i]$.

- *Z_i are bounded: $\exists c$ s.t. $\forall i \, |Z_i| \le c$*
- *T is bounded*
- *$\mathbb{E}[T] < \infty$ and $\exists c$ s.t. $\mathbb{E}[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] \le c$ (i.e. Z_i changes are bounded on each step)*

We show several examples of its applications here.

2.1 Unbiased random walk on a line

Consider a random walk on a line starting at 0. On each step the probability of moving in either direction (right or left) is 1/2. Let's say we are interested in two particular points on a line: $-a$ and b. What's the probability of reaching one of them (say, $-a$) before the other?

This problem is closely related to fair gambling: if a gambler walks in a casino with a dollars and a goal of winning b dollars, and makes a sequence of fair bets (each bet either gives him a dollar or takes a dollar with equal probability) then the probability of ruining is exactly the probability of our random walk reaching $-a$ before reaching b.

Let Y_i be a step direction at time i:

$$Y_i \stackrel{\text{df}}{=} \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

Let $X_n = \sum_{i=1}^{n} X_i$, a position of the random walk at time n. As shown earlier, X_n is a martingale.

The time of the walk reaching $-a$ or b is a stopping time: it is completely determined by the current value of X_n.

$$T \stackrel{\text{df}}{=} \min\{n \mid X_n = -a \text{ or } X_n = b\}$$

Let $v_a \stackrel{\text{df}}{=} \Pr(X_n \text{reaches } -a \text{ before reaching } b)$.

By the Optional Sampling Theorem, $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$ (third condition holds). On the other hand, $\mathbb{E}[X_T] = v_a \cdot (-1) + (1 - v_a) \cdot b$. Therefore $v_a = \frac{b}{a+b}$.

2.2 Unbiased random walk on a line: stopping time

Under the same conditions of an unbiased random walk on a line we are interested in calculating $\mathbb{E}[T]$.

Let's define a different martingale: $Z_n \stackrel{\text{df}}{=} X_n^2 - n$. It is a martingale because $\mathbb{E}[Y_n] = 0$ and $\text{Var}[Y_n] = 1$, as shown earlier.

By the Optional Sampling Theorem, $\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0$. On the other hand,

$$\mathbb{E}[Z_T] = (v_a a^2 + (1-b)b^2) - \underbrace{\mathbb{E}[T]}_{n=T}$$

Hence, $\mathbb{E}[T] = \frac{b}{a+b}a^2 + \frac{a}{a+b}b^2 = ab$.

2.3 Biased random walk

Consider now a biased random walk:

$$Y_i \stackrel{\text{df}}{=} \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } q = 1-p < p \end{cases}$$

Our new martingale measures the drift of the random walk:

$$X'_0 \stackrel{\text{df}}{=} 1$$
$$X'_n \stackrel{\text{df}}{=} \left(\frac{q}{p}\right)^{\sum_{i=1}^{n} Y_i}$$

Again, we are interested in the probability v_a of the random walk reaching $-a$ before reaching b. Let T be a stopping time for any of these events, as above.

By the Optional Sampling Theorem, $\mathbb{E}[X'_T] = \mathbb{E}[X'_0] = 1$. On the other hand,

$$\mathbb{E}[X'_T] = v_a \left(\frac{q}{p}\right)^{-a} + (1-v_a)\left(\frac{q}{p}\right)^{b}$$

Hence

$$v_a = \frac{1 - \left(\frac{q}{p}\right)^b}{\left(\frac{q}{p}\right)^{-a} - \left(\frac{q}{p}\right)^b}$$

3 Tail inequalities

Theorem 3 (Azuma-Hoeffding Inequality)**.** *If $\{X_i\}$ is a martingale such that $\forall k\ |X_k - X_{k+1}| \le c_k < \infty$ (i.e. martingale changes are bounded, possibly with different bounds on each step), then*

$$\forall t \ge 0,\ R > 0 \qquad \Pr(|X_t - X_0| > R) \le 2\exp\left\{-\frac{R^2}{2\sum_{k=1}^{t} c_k^2}\right\}$$

Proof The proof is by induction on t. The base case $t = 0$ is trivial.

In the induction step, let $\Pr(|X_t - X_0| > R) \le 2\exp\left\{-R^2 \Big/ 2\sum_{k=1}^{t} c_k^2\right\}$.

By the convexity of $f(x) = e^{\lambda x}$, we have

$$\forall x \in [-c, c] \qquad e^{\lambda x} \leq \frac{\left(1 - \frac{x}{c}\right) e^{-\lambda c} + \left(1 + \frac{x}{c}\right) e^{\lambda c}}{2} \stackrel{\text{df}}{=} \ell(x)$$

after rewriting $x = -c\frac{1-\frac{x}{c}}{2} + c\frac{1+\frac{x}{c}}{2}$.

Thus if X has $\mathbb{E}[X] = 0$ and $|X| \leq c$, then

$$\mathbb{E}\left[e^{\lambda X}\right] \leq \mathbb{E}[\ell(X)] = \frac{e^{\lambda c} + e^{-\lambda c}}{2} = \sum_{k=0}^{\infty} \frac{(\lambda c)^{2k}}{(2k)!} \leq \sum_{k=0}^{\infty} \frac{(\lambda c)^{2k}}{2^k k!} = e^{\frac{(\lambda c)^2}{2}}$$

Therefore $\mathbb{E}\left[e^{\lambda(X_{t+1} - X_t)} \mid X_0, \ldots, X_{t-1}\right] \leq e^{(\lambda c_t)^2/2}$.

$$\mathbb{E}\left[e^{\lambda X_{t+1}} \mid X_0, \ldots, X_{t-1}\right] = e^{\lambda X_t}\, \mathbb{E}\left[e^{\lambda(X_{t+1} - X_t)} \mid X_0, \ldots, X_{t-1}\right] \leq e^{\lambda X_t} e^{(\lambda c_t)^2/2}$$

Taking expectations and applying inductive assumption:

$$\mathbb{E}\left[e^{\lambda X_{t+1}}\right] \leq e^{(\lambda c_t)^2/2}\, \mathbb{E}\left[e^{\lambda X_t}\right] \leq \exp\left\{\lambda^2 \sum_{i=1}^{t+1} \frac{c_i^2}{2}\right\}$$

Finally, $\Pr(X_t \geq R) = \Pr\left(e^{\lambda X_t} \geq e^{\lambda R}\right) \leq e^{-\lambda R} e^{\lambda^2 \sum c_i^2/2}$.

Optimization gives us $\lambda \stackrel{\text{df}}{=} R / \sum_{i=1}^{t} c_i^2$, hence $\Pr(X_t \geq R) \leq \exp\left\{-R^2 \Big/ 2 \sum_{i=1}^{t} c_i^2\right\}$.

Analyzing lower tail $\Pr(X_t < -\lambda)$, we get a similar bound, which gives us a factor of 2 in the expression.

■

3.1 Random walk on a line

The first example of Azuma-Hoeffding inequality application is a random walk on line. As before, let Y_i denote a step direction taken at time i:

$$Y_i \stackrel{\text{df}}{=} \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$

and let X_n denote a position of the random walk at time n:

$$X_n = \sum_{i=1}^{n} Y_i$$

We are interested in estimating the likelihood of the random walk diverging far from origin. Since every step is bounded by 1 ($|X_k - X_{k+1}| \leq 1$), in the notation of theorem 3 we have $\sum_{k=1}^{t} c_k^2 = t$. Consequently,

$$\Pr(|X_t - X_0| \geq \lambda) \leq 2e^{-\lambda^2/2t}$$

Here Azuma-Hoeffding inequality tells us that the random walk in t steps is likely to stay within an area of $\sqrt{t}$ around origin. If $\lambda \gg \sqrt{t}$, then $\Pr(|X_t - X_0| \geq \lambda) = \mathcal{O}(e^{-t^{\epsilon}})$.

A similar result can be established for a biased random walk ($p \neq q$). The corresponding martingale is

$$X_t = \sum_{i=1}^{t} Y_i - t(p - q)$$

3.2 Chromatic number

Consider a *vertex exposure martingale* $X_i \stackrel{\text{df}}{=} \mathbb{E}\left[\chi(G) \mid G_1, \ldots, G_i\right]$ in a random graph $G(n, 1/2)$, where G_i is a subgraph of G induced by first i vertices, and $\chi(G)$ is a chromatic number.

The gap between X_i and X_{i+1} is at most 1, because a vertex uses no more than one new color. Consequently, we can apply Azuma-Hoeffding inequality for $X_n = \chi(G)$ and $X_0 = \mathbb{E}\left[\chi(G)\right]$:

$$\Pr\left(\left|\chi(G) - \mathbb{E}\left[\chi(G)\right]\right| \geq \lambda\sqrt{n}\right) \leq 2e^{-2\lambda^2}$$

3.3 Pattern matching

Consider a random string of characters $X = (X_1, \ldots, X_n)$ – for example, a DNA sequence. Each character is chosen independently and uniformly at random from a fixed alphabet Σ of size s. We are interested in the number of occurrences of a particular pattern $B = (b_1, \ldots, b_k)$ (say, "ACCTA") in the sequence X. Formally, let F be the number of occurrences of the pattern B in the sequence X. Our goal is to find $\mathbb{E}\left[F\right]$ and estimate the concentration of F around its mean.

The mean can be easily calculated combinatorically:

$$\mathbb{E}\left[F\right] = (n - k + 1)\left(\frac{1}{s}\right)^k$$

To estimate the concentration, we define the following Doob's martingle:

$$Z_0 \stackrel{\text{df}}{=} \mathbb{E}\left[F\right]$$
$$Z_i \stackrel{\text{df}}{=} \mathbb{E}\left[F \mid X_1, \ldots, X_i\right]$$

Z_i defines the expected number of occurrences of the pattern *in the entire sequence*, given only the first i characters. Clearly, $Z_n = F$.

Notice that, when a new character X_{i+1} is exposed, it adds at most k new occurrences of B in expectation: from the leftmost one with $b_k = X_{i+1}$ to the rightmost one with $b_1 = X_{i+1}$. Hence $|Z_{i+1} - Z_i| \leq k$.

Now, by Azuma-Hoeffding inequality,

$$\Pr\left(\left|F - \mathbb{E}\left[F\right]\right| \geq \lambda\right) \leq 2\exp\left\{-\frac{\lambda^2}{2nk^2}\right\}$$

For $\lambda = ck\sqrt{n}$ we get $\Pr\left(\left|F - \mathbb{E}\left[F\right]\right| \geq ck\sqrt{n}\right) \leq 2e^{-c^2/2}$.