Advanced Algorithms VIII (Fall 2020)

Instructor: Chihao Zhang
Scribed by: Ruihan Guo & Linsong Guo

Last modified on Nov 12, 2020

1 Martingale

Definition 1. *In a probability space* $(\Omega, \mathcal{F}, \mathbf{Pr})$*, a sequence of finite variables* $\{Z_n\}_{n\geq 0}$ *is a martingale w.r.t another sequence* $\{X_n\}_{n\geq 0}$ *if*

$$\forall n \geq 1, \mathbf{E}\left[Z_n \mid X_1, \ldots, X_{n-1}\right] = Z_{n-1}.$$

More formally, if we use $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ *to denote the* σ*-algebra generated by* $X_1, \cdots, X_n$*, then* $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$ *forms a filtration. Then we call* $\{Z_n\}_{n\geq 0}$ *a martingale if*

$$\forall n \geq 1, \mathbf{E}\left[Z_n \mid \mathcal{F}_{n-1}\right] = Z_{n-1}.$$

Simlarly, we say $\{Z_n\}_{n\geq 0}$ *a supermartingale if*

$$\forall n \geq 1, \mathbf{E}\left[Z_n \mid \mathcal{F}_{n-1}\right] \leq Z_{n-1},$$

and a submartingale if

$$\forall n \geq 1, \mathbf{E}\left[Z_n \mid \mathcal{F}_{n-1}\right] \geq Z_{n-1}.$$

1.1 Examples

1.1.1 The sum of independent random variables is a martingale

Claim 2. *Assume that* $X_1, \ldots, X_n$ *are n independent random variables with* $E[X_i] = 0$*. Let* $S_k = \sum_{i=1}^{k} X_i$ *and* $\overline{X}_k = (X_1, \ldots, X_k)$*. Then* $\{S_i\}_{i\geq 1}$ *is a martingale w.r.t.* $\{X_i\}_{i\geq 1}$*.*

Proof.

$$\mathbf{E}\left[S_i \,\middle|\, \overline{X}_{i-1}\right] = \mathbf{E}\left[S_{i-1} + X_i \,\middle|\, \overline{X}_{i-1}\right] = S_{i-1} + \mathbf{E}\left[X_i \,\middle|\, \overline{X}_{i-1}\right] = S_{i-1}.$$

□

1.1.2 The product of independent random variables is a martingale

Claim 3. *Assume that* $X_1, \ldots, X_n$ *are n independent random variables with* $\mathbf{E}[X_i] = 1$*. Let* $P_k = \prod_{i=1}^{k} X_i$*. Then* $\{P_i\}_{i\geq 1}$ *is a martingale w.r.t.* $\{X_i\}_{i\geq 1}$*.*

Proof.

$$\mathbf{E}\left[P_i \,\middle|\, \overline{X}_{i-1}\right] = \mathbf{E}\left[P_{i-1} \cdot X_i \,\middle|\, \overline{X}_{i-1}\right] = P_{i-1} \cdot \mathbf{E}\left[X_i \,\middle|\, \overline{X}_{i-1}\right] = P_{i-1}$$

□

1.1.3 Doob Sequence

Let $X_1, \ldots, X_n$ be a sequence of (unnecessarily independent) random variables and $f(\overline{X}_n) = f(X_1, \ldots, X_n) \in \mathbb{R}$ be a function. For $i \geq 0$, we define

$$Z_i = \mathbf{E}\left[f(\overline{X}_n) \,\middle|\, \overline{X}_i\right]$$

We can see that

$$Z_0 = \mathbf{E}\left[f(\overline{X}_n)\right];$$
$$Z_n = f(\overline{X}_n).$$

Therefore, Z_n is the value of the function given the input $\overline{X}_n$ and Z_0 is the average of the function value without any knowledge about the input. The sequence $\{Z_i\}_{i\geq 0}$ can be viewed as our estimation of the function value provided more and more information as i increases.

Lemma 4. *$\{Z_n\}_{n\geq 0}$ is a martingale w.r.t. $\{X_n\}_{n\geq 0}$.*

Proof.

$$\mathbf{E}\left[Z_i \,\middle|\, \overline{X}_{i-1}\right] = \mathbf{E}\left[\mathbf{E}[f(\overline{X}_n)] \,\middle|\, \overline{X}_{i-1}\right] = \mathbf{E}\left[f(\overline{X}_n) \,\middle|\, \overline{X}_{i-1}\right] = Z_{i-1}$$

□

2 Azuma-Hoeffding Inequality

Suppose we have a series of random variables $\{X_n\}_{n\geq 1}$, which satisfy $X_i \in [a_i, b_i]$. Without loss of generality, we assume $\mathbf{E}(X_i) = 0$. Otherwise, we can replace X_i with $X_i - \mathbf{E}(X_i)$. Let $S_k = \sum_{i=1}^{k} X_i$. The Azuma-Hoeffding inequality is:

Theorem 5. *If $\{S_n\}_{n\geq 1}$ is a martingale w.r.t. $\{X_n\}_{n\geq 1}$, then*

$$\mathbf{Pr}\left[S_n \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Now we will sketch a proof of the Azuma-Hoeffding, which is quite similar to our proof of the Hoeffding inequality. Recall when we were trying to prove the Hoeffding inequality, the most difficult part is to estimate the moment generating function, or namely the term

$$\mathbf{E}\left[e^{\alpha S_n}\right] = \mathbf{E}\left[\prod_{i=1}^{n} e^{\alpha X_i}\right].$$

We applied the independent properties of random variables and obtain

$$\mathbf{E}\left[\prod_{i=1}^{n} e^{\alpha X_i}\right] = \prod_{i=1}^{n} \mathbf{E}\left[e^{\alpha X_i}\right]$$

Then we use the Hoeffding lemma

$$\mathbf{E}\left[e^{\alpha X_i}\right] \le e^{-\frac{\alpha(b_i-a_i)^2}{8}}.$$

In the case of Azuma-Hoeffding, we can use the property of martingales instead of independence to obtain a similar bound. To see this, we have

$$\begin{aligned}\mathbf{E}\left[\prod_{i=1}^{n} e^{\alpha X_i}\right] &= \mathbf{E}\left[\mathbf{E}[\prod_{i=1}^{n} e^{\alpha X_i} | \overline{X}_{n-1}]\right] \\ &= \mathbf{E}\left[\prod_{i=1}^{n-1} e^{\alpha X_i} \mathbf{E}[e^{\alpha X_n} | \overline{X}_{n-1}]\right]\end{aligned}$$

The bounds then follows by an induction argument and a conditional expectation version of Hoeffding lemma:

$$\mathbf{E}\left[e^{\alpha X_n} \,\middle|\, \overline{X}_{n-1}\right] \le e^{-\frac{\alpha(b_i-a_i)^2}{8}}. \tag{1}$$

The proof the (1) is the same as our proof of Hoeffding lemma last time.

3 McDiarmid's Inequality

The Doob sequence we introduced in the first section is a class of important martingales. Recall that $Z_n = f(\overline{X}_n)$ and $Z_0 = \mathbf{E}\left[f(\overline{X}_n)\right]$. We would like to apply Azuma-Hoeffding to bound $|Z_n - Z_0|$. Sometimes it is more convenient to apply the following McDiarmid inequality to obtain concentration bounds. It is a consequence of Azuma-Hoeffding for Doob martingales.

Informally, given a Doob sequence $\{Z_i\}_{i\ge 0}$, in order to apply Azume-Hoeffding, we need to construct $\{X_i\}_{i\ge 0}$ and $\{S_i\}_{i\ge 0}$. We let $X_i = Z_i - Z_{i-1}$ and

$$S_i = X_1 + \cdots + X_i = (Z_1 - Z_0) + \cdots + (Z_i - Z_{i-1}) = Z_i - Z_0.$$

It remains to determine the "width" of each X_i. This is captured by the notion of c-Lipschitz.

Definition 6. *A function $f(x_1, \cdots, x_n)$ satisfies c-Lipschitz condition if*

$$\forall i \in [n], \forall x_i, \cdots, x_n, \forall y_i : \quad |f(x_1, \cdots, x_i, \cdots, x_n) - f(x_1, \cdots, y_i, \cdots, x_n)| \le c.$$

We are now ready to state and prove McDiarmid's inequality.

Theorem 7 (McDiarmid's Inequality). *Let f be a function on n variables satisfying c-Lipschitz condition and $X_1, \cdots, X_n$ be n independent variables. Then we have*

$$\mathbf{Pr}\left[|f(X_1, \cdots, X_n) - \mathbf{E}[f(X_1, \cdots, X_n)]| \ge t\right] \le 2e^{-\frac{2t^2}{nc^2}}.$$

Proof. We use f and $\{X_i\}_{i\ge 1}$ to define a Doob martingale $\{Z_i\}_{i\ge 1}$:

$$\forall i : Z_i = \mathbf{E}\left[f(\overline{X}_n) \,\middle|\, \overline{X}_i\right].$$

Let

$$X_i = Z_i - Z_{i-1} = \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_i\right] - \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}\right].$$

Next we try to determine the "width" of $Z_i - Z_{i-1}$. We first set a lower bound B_i:

$$Z_i - Z_{i-1} \geq \inf_x \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}, X_i = x\right] - \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}\right] \triangleq B_i.$$

The upper bound of $Z_i - Z_{i-1}$ is

$$Z_i - Z_{i-1} \leq \sup_y \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}, X_i = y\right] - \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}\right].$$

The gap between the upper bound and the lower bound is

$$\begin{aligned}
&\sup_{x,y} \left(\mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}, X_i = y\right] - \mathbf{E}\left[f(\overline{X}) \,\middle|\, \overline{X}_{i-1}, X_i = x\right]\right) \\
&\overset{(\heartsuit)}{=} \sup_{x,y} \left(\mathbf{E}\left[f\left(x_1, \cdots, x_{i-1}, y, x_{i+1} \cdots, x_n\right) - f\left(x_1, \cdots, x_{i-1}, x, x_{i+1}, \cdots, x_n\right) \,\middle|\, \overline{X}_{i-1}\right]\right) \\
&\leq c,
\end{aligned}$$

where $(\heartsuit)$ uses the fact that X_i is independent of $\overline{X}_{i-1}$.

Hence we have $X_i = Z_i - Z_{i-1} \in [B_i, B_i + c]$. Applying Azuma-Hoeffding, we have

$$\mathbf{Pr}\left[|Z_n - Z_0| \geq t\right] \leq 2e^{-\frac{2t^2}{nc^2}}.$$

Note that $Z_0 = \mathbf{E}\left[f(X_1, \cdots, X_n)\right]$ and $Z_n = f(X_1, \cdots, X_n)$, we have

$$\mathbf{Pr}\left[\left|f\left(X_1, \cdots, X_n\right) - \mathbf{E}\left[f\left(X_1, \cdots, X_n\right)\right]\right| \geq t\right] \leq 2e^{-\frac{2t^2}{nc^2}}.$$

□

4 Applications of McDiarmid's Inequality

4.1 Pattern Matching

Problem 8. *Let $B \in \{0,1\}^k$ be a fixed string. For a random string $X \in \{0,1\}^n$, what is the expected number of occurrences of B in X?*

The expectation can be easily calculated using the linearity of expectations. We define n independent random variables $X_1, \cdots, X_n$, where X_i denotes i-th character of X. Let $F(X_1, \cdots, X_n)$ be the number of occurrences of B in X. Note that there are at most $n - k + 1$ occurrences of B in X, we can enumerate all the occurrences. By the linearity of expectation, we have

$$\mathbf{E}\left[F\right] = \frac{n - k + 1}{2^k}.$$

We can then use McDarmid's inequality to show that F is well-concentrated. For every i, we define $Z_i \triangleq \mathbf{E}\left[F(\overline{X}_n) \,\middle|\, \overline{X}_i\right]$. The sequence $\{Z_i\}$ is a Doob sequence and therefore it is a martingale.

If we change one bit of X, the number of occurrences changes at most k. Hence F satisfies k-Lipschitz condition. Applying McDiarmid's Inequality with $t = \delta k\sqrt{n}$, we have

$$\mathbf{Pr}\left[|Z_n - Z_0| \geq \delta k\sqrt{n}\right] = \mathbf{Pr}\left[|F - \mathbf{E}\left[F\right]| \geq \delta k\sqrt{n}\right] \leq 2e^{-2\delta^2}.$$

4.2 Chromatic Number

Another application of McDiarmid's Inequality is to establish the concentration of choromatic number for Erdős-Rényi random graphs $G(n, p)$. The notation $G(n, p)$ specifies a distribution over all undirected graphs with n vertices. In the model, each of the $\binom{n}{2}$ possible edges exists with probability p independently.

We define n random variables $X_1, \cdots, X_n$, where X_i denotes the edges between v_i and $\{v_1, \cdots, v_{i-1}\}$. Once $X_1, \cdots, X_n$ are given, the graph is known. Since X_i only involves the connections between v_i and $v_1, \cdots, v_{i-1}$, the n variables are independent.

Let $\chi(X_1, \cdots, X_n)$ be the chromatic number. For every i, we define $Z_i \triangleq \mathbf{E}\left[\chi(X_1, \ldots, X_n) \,\middle|\, \overline{X}_i\right]$. Then $\{Z_i\}$ is a Doob sequence.

If X_i changes, the chromatic number changes at most 1. Hence χ satisfies 1-Lipschitz conditions. Applying McDiarmid's Inequality with $t = \delta\sqrt{n}$, we have

$$\mathbf{Pr}\left[|Z_n - Z_0| \geq \delta\sqrt{n}\right] = \mathbf{Pr}\left[|\chi - \mathbf{E}[\chi]| \geq \delta\sqrt{n}\right] \leq 2e^{-2\delta^2}.$$