[CS3958: Lecture 3] Martinalge(cont'd), Stopping time

Instructor: Chihao Zhang, Scribed by Yulin Wang

October 3, 2022

1 Martingale(cont'd)

Here is an example to apply the Azuma-Hoeffding inequality we learnt last time.

Example 1 (Balls-in-a-bag) *There are g green balls and r red balls in a bag and we want to estimate the ratio $\frac{r}{r+g}$ by drawing balls. There are two scenarios.*

- *Draw balls with replacement. Let $X_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$. Let $X = \sum_{i=1}^{n} X_i$. Then it is clear that each $X_i \sim \mathrm{Ber}\left(\frac{r}{r+g}\right)$ and $\mathbf{E}[X] = n \cdot \frac{r}{r+b}$.*

  Since all X_is are independent, we can directly apply Hoeffding's inequality and obtain

  $$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{n}\right).$$

- *Draw balls without replacement. Again let $X_i = \mathbf{1}[\text{the } i\text{-th ball is red}]$, then unlike the case of drawing with replacement, variables in $\{X_i\}$ are dependent. Let $X = \sum_{i=1}^{n} X_i$. We first calculate $\mathbf{E}[X]$.*

  For every $i \geq 1$, $\mathbf{E}[X_i]$ is the probability that the i-th draw is a red ball. Note that drawing without replacement is equivalent to first drawing a uniform permutation of $r + g$ balls and drawing each ball one by one in that order. Therefore, the probability of $X_i = 1$ is $\frac{r \cdot (r+g-1)!}{(r+g)!} = \frac{r}{r+g}$. So we have $\mathbf{E}[X] = n \cdot \frac{r}{r+g}$.

  Next, we consider the concentration of X. We apply Azuma-Hoeffding for a certain martingale. Consider the n-ary function $f(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i$ and the Doob sequence of f. That is, let $Z_i = \mathbf{E}\left[f(\overline{X_n}) \,\middle|\, \overline{X_i}\right]$, and then we know $\{Z_i\}_{0 \leq i \leq n}$ is a martingale. In order to satisfy the condition of Azuma-Hoeffding, note that

  $$Z_i = (Z_i - Z_{i-1}) + (Z_{i-1} - Z_{i-2}) + \cdots + (Z_1 - Z_0) + Z_0.$$

  Let $Y_i \triangleq Z_i - Z_{i-1}$ for $1 \leq i \leq n$, and thus

  $$Z_n - Z_0 = Z_n - \mathbf{E}[f] = \sum_{i=1}^{n} Y_i.$$

  In order to apply Azuma-Hoeffding, we need to bound $|Y_i| = |Z_i - Z_{i-1}|$. By definition,

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

If we use S_i to denote the number of 1s among $\overline{X_i}$, namely $S_i = \sum_{j=1}^{i} X_j$, then

$$\mathbf{E}\left[f(\overline{X_n}) \;\middle|\; \overline{X_i}\right] = \mathbf{E}\left[f(\overline{X_n}) \;\middle|\; S_i\right] = S_i + (n-i) \cdot \frac{r - S_i}{g + r - i}.$$

Therefore, $S_i = S_{i-1} + X_i$ and

$$\begin{aligned} Z_i - Z_{i-1} &= \left(S_i + (n-i) \cdot \frac{r - S_i}{g + r - i}\right) - \left(S_{i-1} + (n-i+1) \cdot \frac{r - S_{i-1}}{g + r - i + 1}\right) \\ &= \frac{g + r - n}{g + r - i}\left(Y_i + \frac{S_{i-1} - r}{g + r - i + 1}\right). \end{aligned}$$

Note that $r \geq S_{i-1}$ and $g \geq (i-1) - S_{i-1}$, so we have

$$\begin{aligned} Z_i - Z_{i-1} &\leq \frac{g + r - n}{g + r - i}\left(1 + \frac{S_{i-1} - r}{g + r - i + 1}\right) \leq \frac{g + r - n}{g + r - i} \leq 1, \\ Z_i - Z_{i-1} &\geq \frac{g + r - n}{g + r - i}\left(\frac{S_{i-1} - r}{g + r - i + 1}\right) \geq -\frac{g + r - n}{g + r - i} \geq -1. \end{aligned}$$

Therefore, $-1 \leq Y_i \leq 1$. And we can apply Azuma-Hoeffding to $Z_n - Z_0$ to obtain

$$\mathbf{Pr}\left[|X - \mathbf{E}[X]| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2n}\right).$$

1.1 McDiarmid's Inequality

The Doob sequence we used in the Balls-into-Bags example is a very powerful and general tool to obtain concentration bounds. For a model defined by n random variables $X_1, \ldots, X_n$ and any quantity $f(X_1, \ldots, X_n)$ that we want to estimate, we can apply the Azuma-Hoeffding inequality to the Doob sequence of f. As shown in the previous example, the quality of the bound relies on the *width* of the martingale.

Let us first repeat the argument in the previous example. The Doob sequence is $Z_i = \mathbf{E}\left[f(\overline{X_n}) \;\middle|\; \overline{X_i}\right]$ for every $0 \leq i \leq n$. For every $0 \leq i \leq n$, we let

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

where $X_j = Z_j - Z_{j-1}$. Then we apply Azuma-Hoeffding to $S_n = Z_n - Z_0 = f(\overline{X_n}) - \mathbf{E}\left[f(\overline{X_n})\right]$.

We need to determine the width of each X_i. This is relatively easy if the function f and the variables $\{X_i\}_{1 \leq i \leq n}$ have certain nice properties.

Definition 1 (c-Lipschitz function) *A function $f(x_1, \cdots, x_n)$ satisfies c-Lipschitz condition if*

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

The McDiarmid's inequality is the application of Azuma-Hoeffding inequality to Lipschitz f and independent $\{X_i\}$.

Theorem 2 (McDiarmid's Inequality) *Let f be a c-Lipschitz function on n variables and $X_1, \cdots, X_n$ be n independent variables. Then we have*

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

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

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

Let

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

Next we try to determine the width of Y_i. Clearly

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

and

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

The gap between the upper bound and the lower bound is

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

For every x, y and $\sigma_1, \ldots, \sigma_{i-1}$,

$$\begin{aligned}
&\mathbf{E}\left[f(\overline{X}) \,\middle|\, \bigwedge_{1\leq j\leq i-1} X_j = \sigma_j, X_i = y\right] - \mathbf{E}\left[f(\overline{X}) \,\middle|\, \bigwedge_{1\leq j\leq i-1} X_j = \sigma_j, X_i = x\right] \\
&= \sum_{\sigma_{i+1},\ldots,\sigma_n} \Bigg(\mathbf{Pr}\left[\bigwedge_{i+1\leq j\leq n} X_j = \sigma_j \,\middle|\, \bigwedge_{1\leq j\leq i-1} X_j = \sigma_j, X_i = y\right] \cdot f(\sigma_1,\ldots,\sigma_{i-1}, y, \sigma_{i+1},\ldots,\sigma_n) \\
&\quad - \mathbf{Pr}\left[\bigwedge_{i+1\leq j\leq n} X_j = \sigma_j \,\middle|\, \bigwedge_{1\leq j\leq i-1} X_j = \sigma_j, X_i = x\right] \cdot f(\sigma_1,\ldots,\sigma_{i-1}, x, \sigma_{i+1},\ldots,\sigma_n)\Bigg) \\
&\overset{(\heartsuit)}{=} \sum_{\sigma_{i+1},\ldots,\sigma_n} \mathbf{Pr}\left[\bigwedge_{i+1\leq j\leq n} X_j = \sigma_j\right] \cdot \left(f(\sigma_1,\ldots,\sigma_{i-1}, y, \sigma_{i+1},\ldots,\sigma_n) - f(\sigma_1,\ldots,\sigma_{i-1}, x, \sigma_{i+1},\ldots,\sigma_n)\right) \\
&\overset{(\clubsuit)}{\leq} c.
\end{aligned}$$

where ($\heartsuit$) uses independence of $\{X_i\}$ and ($\clubsuit$) uses the c-Lipsichitz property of f.

Applying Azuma-Hoeffding, we have

$$\mathbf{Pr}\left[|Z_n - Z_0| \geq t\right] = \mathbf{Pr}\left[\left|f(X_1,\cdots,X_n) - \mathbf{E}\left[f(X_1,\cdots,X_n)\right]\right| \geq t\right] \leq 2e^{-\frac{2t^2}{nc^2}}.$$

□

Let us examine two applications of McDiarmid's inequality.

Example 2 (Pattern Matching) *Let $B \in \{0,1\}^k$ be a fixed string. For a uniformly at random string $X \in \{0,1\}^n$, what is the expected number of occurrences of B in X?*

For example, **1001** occurs 2 times in **1001001**.

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, and we can enumerate the first position of each occurrence. Let $Y_i \triangleq \mathbf{1}[(X_i, X_{i+1}, \ldots, X_{i+k-1}) = B]$. Then by the linearity of expectation, we have

$\mathbf{E}[Y_i] = \frac{1}{2^k}$ for any $1 \le i \le n-k+1$ since $X_i, X_{i+1}, \ldots X_{i+k-1}$ are independent.

$$\mathbf{E}[f] = \sum_{i=1}^{n-k+1} \mathbf{E}[Y_i] = \frac{n-k+1}{2^k}.$$

We can then use McDiarmid's inequality to show that f is well-concentrated. To see this, note that variables in $\{X_i\}$ are independent and the function f is k-Lipschitz: If we change one bit of X, the number of occurrences changes at most k. Therefore,

$$\mathbf{Pr}[|Z_n - Z_0| \ge t] = \mathbf{Pr}[|f - \mathbf{E}[f]| \ge t] \le 2e^{-\frac{2t^2}{nk^2}}.$$

Example 3 (Chromatic Number of $\mathcal{G}(n,p)$) *Another application of McDiarmid's Inequality is to establish the concentration of chromatic number for Erdős-Rényi random graphs $\mathcal{G}(n,p)$. For a graph $G \sim \mathcal{G}(n,p)$, we use $\chi(G)$ to denote its chromatic number, i.e. the minimum number q so that G can be properly colored using q colors. There are different ways to represent G using random variables.*

Recall the notation $\mathcal{G}(n,p)$ specifies a distribution over all undirected simple graphs with n vertices. In the model, each of the $\binom{n}{2}$ possible edges exists with probability p independently.

- *The most natural way is to introduce a variable X_e for every pair of vertices $e = \{u, v\} \in \binom{V}{2}$ where $X_e = \mathbf{1}[\text{the edge } e \text{ exists in } G]$. Then $\{X_e\}$ are independent and the chromatic number can be written as a function $\chi(G) = f(X_{e_1}, X_{e_2}, \ldots, X_{e_{\binom{n}{2}}})$. It is easy to see that χ is 1-Lipschitz as removing or adding one edge can only change the chromatic number by one at most. So by McDiarmid's inequality, we obtain that*

  $\binom{V}{2}$ here denotes all subset of V of size 2.

  $$\mathbf{Pr}[|f - \mathbf{E}[f]| \ge t] \le 2e^{-2t^2/\binom{n}{2}}.$$

  However, this bound is not satisfactory as we need to set $t = \Theta(n)$ in order to upper bound the RHS by a constant.

- *We can encode the graph G in a more efficient way while reserving the Lipschitz and the independence property. Suppose the vertex set of G is $\{v_1, \ldots, v_n\}$. We define n random variables $Y_1, \cdots, Y_n$, where Y_i encodes the edges between v_i and $\{v_1, \cdots, v_{i-1}\}$. Once $Y_1, \cdots, Y_n$ are given, the graph is known, so the chromatic number can be written as a function $\chi(G) = g(Y_1, \ldots, Y_n)$. Since Y_i only involves the connections between v_i and $v_1, \cdots, v_{i-1}$, $\{Y_i\}$ are independent.*

  It is also easy to see that g is 1-Lipschitz as well since if Y_i changes, the chromatic number changes by one at most. Applying McDiarmid's inequality, we obtain that

  $$\mathbf{Pr}[|g - \mathbf{E}[g]| \ge t] \le 2e^{-\frac{2t^2}{n}}.$$

2 Stopping Time

Suppose $Z_0, Z_1, \ldots, Z_n, \ldots$ is a martingale. We know that for any t, $\mathbf{E}[Z_t] = \mathbf{E}[Z_0]$. However, does $\mathbf{E}[Z_\tau] = \mathbf{E}[Z_0]$ still hold if τ is a random variable?

Consider the following gambling strategy in a fair game. At the first round, the gambler bet \$1. Then he simply double his stake until he wins

The stretagy was called martingale!

Let τ be the first time he wins. Then expected money he win at time τ is 1, which is not equal to 0, his initial money. In order to understand the phenomenon, let us first formally introduce *stopping time*.

- If $\tau = 1$, he wins 1 dollar.
- If $\tau = 2$, he wins $-1 + 2 = 1$ dollar.
- If $\tau = 3$, he wins $-1 - 2 + 4 = 1$ dollar.
- ...

Definition 3 (Stopping Time) *Let $\tau \in \mathbb{N} \cup \{\infty\}$ be a random variable. We say τ is a* stopping time *if for all $t \geq 0$, the event "$\tau \leq t$" is $\mathcal{F}_t$-measurable.*

For example, the first time that a gambler wins five games in a row is a stopping time, since for a given t, this can be determined by looking at the outcomes of all the previous games, and therefore the time is $\mathcal{F}_t$-measurable. However, the *last* time the gambler wins five games in a row is *not* a stopping time, since determining whether the time is t cannot be done without knowing $X_{t+1}, X_{t+2}, \ldots$

2.1 Optional Stopping Theorem(OST)

The optional stopping theorem provides sufficient condition for $\mathbf{E}[Z_\tau] = \mathbf{E}[Z_0]$ to hold.

Theorem 4 (Optional Stopping Theorem) *Let $\{X_t\}_{t\geq 0}$ be a martingale and τ be a stopping time with respect to $\{\mathcal{F}_t\}_{t\geq 0}$. Then $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$ if* at least one of *the following conditions holds: 1. τ is bounded almost surely, that is, $\exists n \in \mathbb{N}$ such that $\mathbf{Pr}[\tau \leq n] = 1$; 1. $\mathbf{Pr}[\tau < \infty] = 1$, and there is a finite M such that $|X_t| \leq M$ for all $t < \tau$; or 1. $\mathbf{E}[\tau] < \infty$, and there is a constant c such that $\mathbf{E}[|X_{t+1} - X_t| \mid \mathcal{F}_t] \leq c$ for all $t < \tau$.*

We will prove the theorem next time. Let us look back at the boy-or-girl example mentioned in the first class.

Example 4 (Boy or Girl) *Suppose there is a country in which people only want boys. What is the sex ratio of the country in the following three scenarios?*

- *Each family continues to have children until they have a boy.*
- *Each family continues to have children until there are more boys.*
- *Each family continues to have children until there are more boys or there are 10 children.*

We can model the problem as a random walk. Suppose there is a man walking randomly on a one-dimensional axis. Let $\{X_t\}_{t\geq 0}$ be the positions of the man *at each time where* X_t *stands for* the number of boys minus the number of girls *in the first* t *children of a family. Starting at* $X_0 = 0$*, at time* 0*, the man takes a step* $c_t \in_R \{-1, 1\}$ *and reach* X_{t+1}*, i.e.,* $X_{t+1} = X_t + c_t$*. It is easy to verify that* $\{X_t\}_{t\geq 0}$ *is a martingale. The three scenarios mentioned correspond to the following three different definitions of a stopping time* τ*. The identity* $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$ *means that the sex ratio is balanced. We will check respectively whether it is the case using OST.*

- *Let* τ *be the first time* t *such that* $c_t = 1$*. Then* $\mathbf{E}[\tau] < \infty$ *since by definition* $\tau \sim \mathrm{Geom}\left(\frac{1}{2}\right)$*, and* $|X_{t+1} - X_t| \leq 1$ *for all* $t < \tau$*. Therefore from the 3rd condition of OST we have* $\mathbf{E}[X_\tau] = \mathbf{E}[X_0] = 0$*. In other words, if the man stops at the first time of* $c_t = 1$*, then the expected final position is* 0*.*
- *Let* τ *be the first time* t *such that* $X_t = 1$*, then of course* $\mathbf{E}[X_\tau] = 1 \neq \mathbf{E}[X_0]$*. This process is called "1-d random walk with one absorbing barrier" and it is well-known that* $\mathbf{E}[\tau] = \infty$*. No condition in OST is satisfied.*
- *Let* τ *be the minimum between* 10 *and the first time* t *such that* $X_t = 1$*. In this case,* τ *is at most* 10*, which satisfies the first condition of OST. Therefore we have* $\mathbf{E}[X_\tau] = \mathbf{E}[X_0] = 0$*.*

The property $\mathbf{E}[\tau] = \infty$ of the random work is called "null recurrent". You can find more on this from my lecture on stochastic processes.