Lecture 13: Martingale Difference Sequence & Azuma-Hoeffding Inequality

Martingale Difference Sequence

- Let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$ be a filtration
- Let $(\mathbb{F}_0, \mathbb{F}_1, \ldots, \mathbb{F}_n)$ be a martingale sequence with respect to the filtration above
- Let $\mathbb{Y}_0 = \mathbb{F}_0$, and $\mathbb{Y}_{t+1} = \mathbb{F}_{t+1} - \mathbb{F}_t$, for $0 \leqslant t < n$
- Intuition: $\mathbb{Y}_{t+1}$ measures the increase in $\mathbb{Y}_{t+1}$ from $\mathbb{Y}_t$. If $\mathbb{Y}_{t+1}$ is negative then it implies that $\mathbb{Y}_{t+1}$ is smaller than $\mathbb{Y}_t$
- Note that $\mathbb{E}\left[\mathbb{Y}_{t+1} | \mathcal{F}_t\right] = 0$, because we have $\mathbb{E}\left[\mathbb{F}_{t+1} | \mathcal{F}_t\right] = \mathbb{F}_t$

Theorem (Azuma's Inequality)

Suppose $(\mathbb{Y}_0, \dots, \mathbb{Y}_n)$ *be a martingale difference sequence with respect to the filtration* $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$. *Assume that the following condition holds for all* $x \in \Omega$ *and* $0 \leqslant t < n$.

$$\max_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y) - \min_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y) \leqslant c_{t+1}$$

Then the following large deviation bound holds

$$\mathbb{P}\left[\sum_{i=1}^{n} \mathbb{Y}_i \geqslant E\right] \leqslant \exp\left(-2E^2 / \sum_{i=1}^{n} c_i^2\right)$$

Subtlety. Fix t. For different $x \in \Omega$, it is possible that $\max_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y)$ is different from $\min_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y)$. All that matters is that their <u>difference</u> is bounded by c_{t+1}.

Proof Outline

- The proof outline is identical to the Hoeffding bound proof.
- If we prove the following bound, then we are done. For any $h > 0$, we have

$$\mathbb{E}\left[\exp\left(h\sum_{i=1}^{n}\mathbb{Y}_i\right)\right] \leqslant \exp\left(\frac{h^2}{8}\sum_{i=1}^{n}c_i^2\right)$$

This form of the inequality should remind us that we should be using the Hoeffding's Lemma in our proof.

Differences from Hoeffding's Bound

- The distribution $\mathbb{Y}_{t+1}$ can depend on the previous outcomes $(\omega_1, \ldots, \omega_t)$
- For different $x \in \Omega$, it is possible that $\max_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y)$ is different from $\min_{y \in \mathcal{F}(x)} \mathbb{Y}_{t+1}(y)$. All that matters is that their difference is bounded by c_{t+1}

Experiment.

- There are R red balls and B blue balls in an urn at time $t = 0$
- At any time, we sample a random ball from the urn (and we do not replace the ball back into the urn)
- We are interested in understanding the behavior of the random variable $\mathbb{S}_n$ that counts the total number of red balls at the end of time $t = n$ (that is, n balls are sampled without replacement from the urn)
- We assume that $R + B \geqslant n$, i.e., the bin never runs out of balls in our experiment

Formalization of the Problem I

- The variables $(\mathbb{X}_1, \ldots, \mathbb{X}_n)$ represent the balls we sample at time $1, \ldots, n$, respectively
- We are interest in understanding the concentration of the random variable

$$\mathbb{S}_n := \sum_{i=1}^{n} \mathbf{1}_{\{\mathbb{X}_i = R\}}$$

  Note that the probability of $\mathbb{X}_i = R$ depends on the sum $\mathbb{S}_{i-1}$
- Let us first calculate the expected value of this random value. Prove by mathematical induction that the following result is true for $n \geqslant 0$.

Lemma

$$\mathbb{E}\left[\mathbb{S}_n\right] = n \frac{R}{R+B}$$

Formalization of the Problem II

In this lecture, all results will be mentioned. No proofs shall be provided. Students are encouraged to prove these results on their own.

- Now, we shall prove a concentration bound around this expected value

The Filtration and the Martingale I

- Let
$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n$$
represent the natural ball-exposure filtration for this problem.
- This statement, in short, states that $\Omega = \{R, B\}^n$ and, for any $x \in \Omega$ and $0 \leqslant i \leqslant n$, we have
$$\mathcal{F}_i(x) = \{x_1 x_2 \ldots x_i\} \times \{R, B\}^{n-i}$$
That is, $\mathcal{F}_i(x)$ is the set of all $y \in \Omega$ such that $x_1 = y_1, \ldots,$ $x_i = y_i$

The Filtration and the Martingale II

- Now, we need to define the random functions $\mathbb{F}_0, \ldots, \mathbb{F}_n$ that are $\Omega \to \mathbb{R}$.

  $$\mathbb{F}_i(x) := \mathbb{E}\left[\mathbb{S}_n | \mathcal{F}_i\right](x)$$

  Let us parse this statement. Recall that $\mathcal{F}_i(x)$ denotes the set of all $y \in \Omega$ that agree at the first i entries with x, i.e., the subset $\{x_1 x_2 \ldots x_i\} \times \{R, B\}^{n-i}$. Now, $\mathbb{F}_i(x)$ represents the conditional expectation of $\mathbb{S}_n$ restricted to x in the subset $\mathcal{F}_i(x)$.
- Observe that $\mathbb{F}_0 = \mathbb{E}[\mathbb{S}_n]$, i.e., the expected value of $\mathbb{S}_n$ in this experiment. We have already computed this quantity previously, i.e., we have $\mathbb{F}_0 = n\frac{R}{R+B}$.
- Observe that $\mathbb{F}_i$ is $\mathcal{F}_i$-measurable, for $0 \leqslant i \leqslant n$
- Now, we need to prove that the martingale property holds. That is, we need to prove (the functional identity) $\mathbb{E}\left[\mathbb{F}_{i+1} | \mathcal{F}_i\right] = (\mathbb{F}_i | \mathcal{F}_i)$, for all $0 \leqslant i < n$

The Filtration and the Martingale III

- Note that $(\mathbb{F}_0, \dots, \mathbb{F}_n)$ is Doob's martingale for the function $\mathbb{S}_n$. So, it is a martingale. Nevertheless, let us prove that $(\mathbb{F}_0, \dots, \mathbb{F}_n)$ is a martingale with respect to the ball-exposure filtration $(\mathcal{F}_0, \dots, \mathcal{F}_n)$ using elementary techniques. Towards this, we need to compute the following quantity

$$(\mathbb{F}_i | \mathcal{F}_i)(x) = ?$$

Prove the following result.

Lemma

Let $0 \leqslant i \leqslant n$. Let $\mathbb{S}_i(x)$ represent the number of red balls in the first i samples of $x \in \{R, B\}^n$. Then, we have

$$(\mathbb{F}_i | \mathcal{F}_i)(x) = \mathbb{S}_i(x) + (n - i)\frac{R - \mathbb{S}_i(x)}{R + B - i}$$

The Filtration and the Martingale IV

Intuitively, we have seen $\mathbb{S}_i(x)$ until time $t = i$. In the future, we expect to see $(n-i)\frac{R-\mathbb{S}_i(x)}{R+B-i}$ red balls (there are $R - \mathbb{S}_i(x)$ red balls left in the urn among $R + B - i$ balls).

At time time $t = i + 1$, the probability that we see a red ball is $p = \frac{R-\mathbb{S}_i(x)}{R+B-i}$. So, we have

$$\mathbb{E}\left[\mathbb{F}_{i+1}|\mathcal{F}_i\right](x) = p\left(\mathbb{S}_i(x) + 1 + (n-i-1)\frac{R-\mathbb{S}_i(x)-1}{R+B-i-1}\right)$$
$$(1-p)\left(\mathbb{S}_i(x) + (n-i-1)\frac{R-\mathbb{S}_i(x)}{R+B-i-1}\right)$$

We need to prove that the RHS is equal to $\mathbb{S}_i(x) + (n-i)\frac{R-\mathbb{S}_i(x)}{R+B-i}$. This step is left as an exercise. (Think: You have already proved this result earlier!)

The Filtration and the Martingale V

- Let us calculate the value of c_{i+1}, for $0 \leqslant i < n$.

$$
\begin{aligned}
&= \max_{y \in \mathcal{F}_i(x)} \mathbb{F}_{i+1}(y) - \min_{y \in \mathcal{F}_i(x)} \mathbb{F}_{i+1}(y) \\
&= \left(\mathbb{S}_i(x) + 1 + (n-i-1) \frac{R - \mathbb{S}_i(x) - 1}{R + B - i - 1} \right) \\
&\quad - \left(\mathbb{S}_i(x) + (n-i-1) \frac{R - \mathbb{S}_i(x)}{R + B - i - 1} \right) \\
&= 1 - \frac{n-i-1}{R+B-i-1} \\
&< 1 =: c_{i+1}
\end{aligned}
$$

The Filtration and the Martingale VI

- By Azuma's inequality, we have

$$\mathbb{P}\left[\mathbb{F}_n - \mathbb{F}_0 \geqslant E\right] \leqslant \exp\left(-2E^2 / \sum_{i=1}^{n} c_i^2\right)$$

This inequality is equivalent to

$$\mathbb{P}\left[\mathbb{F}_n - n\frac{R}{R+B} \geqslant E\right] \leqslant \exp(-2E^2/n)$$