CS 761: Randomized Algorithms **University of Waterloo: Winter 2025**

Lecture 9

February 11, 2025

Instructor: Sepehr Assadi *Scribes: Vihan Shah and Dinis Vitorino*

Topics of this Lecture

1 Balls and Bins Revisited

We again consider a balls and experiment. This time n balls are thrown into n bins independently, uniformly at random. We are interested in the number of empty bins after all n balls have been thrown. More precisely, we wish to asymptotically characterize the functions $T\colon \mathbb{N} \to \mathbb{N}$ such that the number of empty bins at the conclusion of the experiment, differs from its expected value by at most $T(n)$ with probability at least some big constant, e.g., 2/3.

Since we are studying the concentration of random variables, one feels tempted to use standard concentration inequalities like those presented in Lecture 3 (Markov, Chebyshev, or Chernoff). For any $i \in [n]$, define an indicator random variable where $X_i = 1$ iff the i-th bin is empty at the end of the experiment, and

$$X \coloneqq \sum_{i=1}^{n} X_i \ ,$$

so X counts the number of empty bins.

We can immediately conclude, using linearity of expectation, that

$$\mathbb{E}\left[X\right] = \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = n \cdot \mathbb{E}\left[X_1\right] = n \cdot \left(1 - \frac{1}{n}\right)^n \ ,$$

since the probability that any fixed bin remains empty after n balls have been thrown is $\left(1 - \frac{1}{n}\right)^n$ (the exponent comes from the n balls that are thrown independently, the $1 - \frac{1}{n}$ from the fact that each of these has to choose some bin which is not the one we are considering).

Going into auto-pilot and applying the additive Chernoff bound, we conclude

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{n}\right) .$$

Choosing $t = \sqrt{\ln(3) \cdot n}$, we obtain

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t\right) \leqslant \frac{2}{3} .$$

However, our application of the Chernoff bound had one, quite severe, problem – the indicator random variables we considered are *not* independent, turning our use of this bound into a fundamentally flawed approach. On the other hand, it can be empirically verified that X is concentrated around its expected value (see the histogram below), suggesting that we might be lacking mathematical machinery to prove what we want, rather than the entire statement being false.

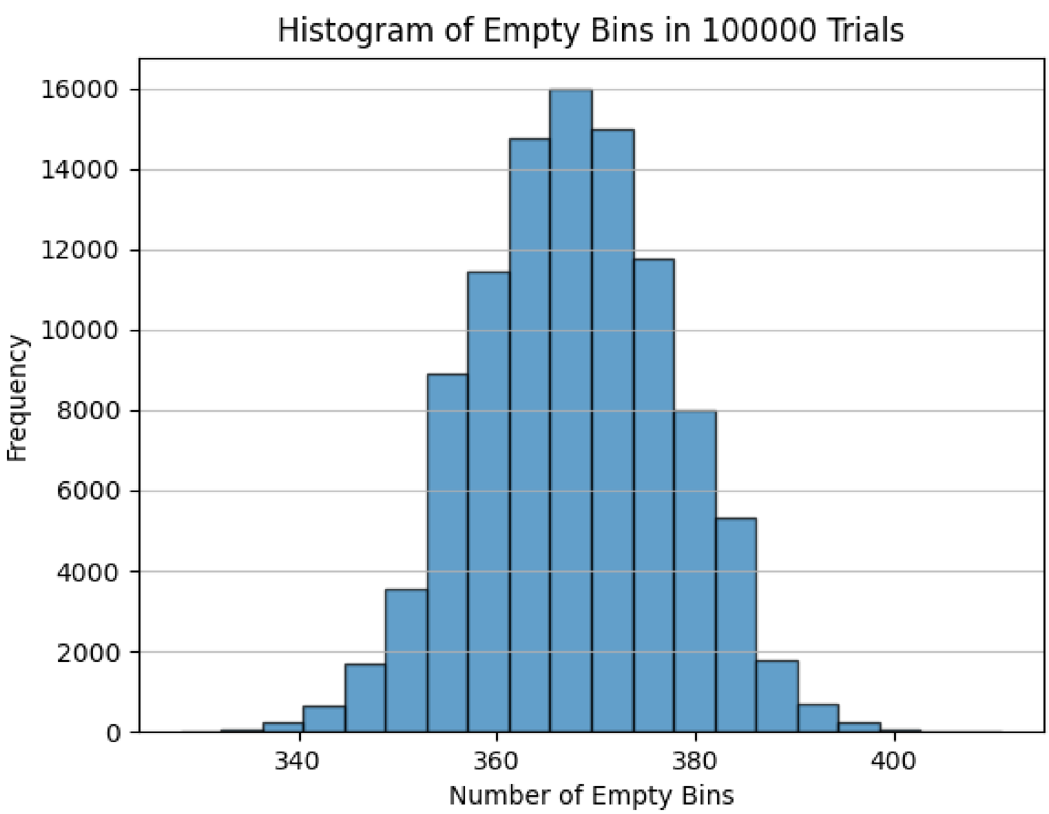

Figure 1: A histogram of frequency of empty bins in 100,000 trials of throwing 1000 balls into 1000 bins uniformly and independently. Note that in this case, the number of empty bins X satisfies $\mathbb{E}[X] \sim 367$.

2 Strengthening the Chernoff Bound

It turns out that one way to show concentration of the random variable X considered in the previous section is to use **Azuma's inequality**, a generalization of the additive Chernoff bound to random variables which may be correlated, but whose correlation is, in some sense, low. To continue, we need to take some detours.

2.1 Conditional Expectations

Azuma's inequality relies on *martingales*, which in turn rely on conditional expectations. We begin by studying the latter, starting with a simple example:

Example 1. Suppose we are interested in studying two random variables A and B, defined as:

$$B := \begin{cases} 0 & \text{with probability } \frac{1}{3} \\ 1 & \text{with probability } \frac{2}{3} \end{cases} \qquad A := \begin{cases} \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases} & \text{if } B = 0 \\ 1 & \text{if } B = 1 \end{cases}$$

It is natural to wonder about the expected value of A given that B has taken on specific values. If we do so, it is easy to conclude that

$$\mathbb{E}[A \mid B = 0] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2} \quad \text{and} \quad \mathbb{E}[A \mid B = 1] = 1 \ .$$

From this, we can additionally see that

$$\mathbb{E}[A] = \Pr(B = 0) \cdot \mathbb{E}[A \mid B = 0] + \Pr(B = 1) \cdot \mathbb{E}[A \mid B = 1] = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot 1 = \frac{5}{6} \ .$$

It is equally natural to define a random variable $\mathbb{E}[A \mid B]$ which is distributed "like the expected value of A given that B takes on particular values", but "inherits its randomness from B":

$$\mathbb{E}[A \mid B] := \begin{cases} \frac{1}{2} & \text{with probability } \frac{1}{3} \\ 1 & \text{with probability } \frac{2}{3} \end{cases} \ .$$

Armed with this intuition, we define the conditional expectation of a random variable X given a discrete random variable Y as

$$\mathbb{E}[X \mid Y] := \mathbb{E}[X \mid Y = y] \ \text{ with probability } \ \Pr(Y = y) \ \text{ for all } y \in \operatorname{supp}(Y) \ ,$$

where $\operatorname{supp}(Y)$ is the set of points y such that $\Pr(Y = y) > 0$. We again emphasize that while $\mathbb{E}[X]$ is a number, $\mathbb{E}[X \mid Y]$ is a random variable.

Some basic facts follow from this definition:

Proposition 2 (Law of Total Expectation)**.** *Let A be a random variable and B be a discrete random variable. Then*

$$\mathbb{E}[\mathbb{E}[A \mid B]] = \mathbb{E}[A] \ .$$

Proof. We have

$$[\mathbb{E}[\mathbb{E}[A \mid B]] = \sum_b \Pr(B = b) \cdot \mathbb{E}[A \mid B = b] = \sum_b \sum_a \Pr(B = b) \cdot a \cdot \Pr(A = a \mid B = b)$$
$$= \sum_b \sum_a a \cdot \Pr(A = a \wedge B = b) = \sum_a a \cdot \sum_b \Pr(A = a \wedge B = b) = \sum_a a \cdot \Pr(A = a) = \mathbb{E}[A] \,,$$

concluding the proof. □

Proposition 3 (Tower Property)**.** *Let A be a random variable and B, C be discrete random variables. Then*

$$\mathbb{E}[\mathbb{E}[A \mid B, C] \mid C] = \mathbb{E}[A \mid C] \ .$$

Proof. For any $c \in \operatorname{supp}(C)$, by applying Proposition 2 to the distribution of $A, B \mid C = c$, denoted by A', B', we have,

$$\mathbb{E}[\mathbb{E}[A \mid B, C = c] \mid C = c] = \mathbb{E}[\mathbb{E}[A' \mid B']] = \mathbb{E}[A'] = \mathbb{E}[A \mid C = c] \,.$$

By the equality in the statement now follows from the definition of $\mathbb{E}[\,\cdot \mid C]$. □

2.2 Martingales

With this new notion, we can define martingales:

Definition 4. Let $S_0, \ldots, S_n$ be a sequence of real-valued random variables and $\mathcal{F}_0, \ldots, \mathcal{F}_n$ be a sequence of multiset-valued random variables with $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n$. We say random variables $S_0, \ldots, S_n$ form a **martingale** with respect to $\mathcal{F}_0, \ldots, \mathcal{F}_n$ iff:

- for all $i \in [n] \cup \{0\}$, $\mathbb{E}[S_i] < \infty$;
- for all $i \in [n]$, $\mathbb{E}[S_i \mid \mathcal{F}_{i-1}] = S_{i-1}$.

To illustrate the intuition behind this definition, we present a **simple martingale process** revolving around a random walk:

Consider an n-step random walk on the number line starting at 0. At each time step, we choose with equal probability to take a step forwards, or to take a step backwards. More precisely, we consider $\{\xi_k\}_{k=1}^n$, a sequence of i.i.d. random variables taking value $+1$ or -1 each with probability $1/2$. Formally:

$$\mathbb{P}(\xi_k = 1) \;=\; \tfrac{1}{2}, \quad \mathbb{P}(\xi_k = -1) \;=\; \tfrac{1}{2}, \quad \text{and all } \xi_k \text{ are independent.}$$

The random walk process is then defined as a discrete-time stochastic process $\{M_i\}_{i=0}^n$ with

$$M_i \;=\; \sum_{k=1}^{i} \xi_k, \quad \text{and} \quad M_0 = 0.$$

Intuitively, M_i indicates the position of the random walker after i steps. We now introduce (multi-)sets $\mathcal{F}_i$ describing the first i ξ_k's:

$$\mathcal{F}_i = \{\xi_1, \xi_2, \ldots, \xi_i\} \;.$$

Clearly, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$. We now show that $\{M_i\}$ is a martingale (w.r.t. $\{\mathcal{F}_i\}$). Firstly, note that for all $i \in [n]$ we have $\mathbb{E}[M_i] \leqslant i \leqslant n$, implying that $\mathbb{E}[M_i]$ is finite. We now have to show that:

Claim 5. $\mathbb{E}[M_{i+1} \mid \mathcal{F}_i] \;=\; M_i$ *for all* $i \in [n-1]$.

Proof. We verify this for our random walk:

1. **Decompose** M_{i+1}:
$$M_{i+1} \;=\; \sum_{k=1}^{i+1} \xi_k \;=\; \underbrace{\sum_{k=1}^{i} \xi_k}_{= M_i} \;+\; \xi_{i+1}.$$

2. **Conditional expectation** given $\mathcal{F}_i$:
$$\mathbb{E}[M_{i+1} \mid \mathcal{F}_i] \;=\; \mathbb{E}\Big[M_i + \xi_{i+1} \;\Big|\; \mathcal{F}_i\Big].$$

3. $\mathcal{F}_i$ **completely captures** M_i and ξ_{i+1} is **independent** of $\mathcal{F}_i$:
$$\mathbb{E}[M_i \mid \mathcal{F}_i] \;=\; M_i, \quad \mathbb{E}[\xi_{i+1} \mid \mathcal{F}_i] \;=\; \mathbb{E}[\xi_{i+1}] \;=\; 0,$$
since ξ_{i+1} takes values ± 1 with equal probability.

4. **Use linearity of expectation**:

$$\mathbb{E}[M_{i+1} \mid \mathcal{F}_i] \;=\; \mathbb{E}[M_i \mid \mathcal{F}_i] \;+\; \mathbb{E}[\xi_{i+1} \mid \mathcal{F}_i] \;=\; M_i + 0 \;=\; M_i.$$

□

Hence, the sequence $\{M_i\}_{i=0}^n$ is a martingale.

2.3 Azuma's Inequality

Equipped with the understanding of martingales, we introduce a "Chernoff-like" inequality on them:

Proposition 6 (Azuma's Inequality [Hoe63, Azu67])**.** *Let $S_0, \ldots, S_n$ be a martingale (w.r.t. some sequence of random variables $\mathcal{F}_0, \ldots, \mathcal{F}_n$). If $|S_i - S_{i-1}| \leqslant c$ for all $i \in [n]$ and some constant $c \in \mathbb{R}_{>0}$, then*

$$\Pr\left(|S_n - S_0| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{n \cdot c^2}\right) \;.$$

3 Applications

3.1 Application 1: Balls and Bins

After this slight detour, we return to the balls and bins experiment from Section 1. We would like to use Azuma's inequality to show concentration of the number of empty bins around its expectation. Thus, our goal becomes to identify an appropriate martingale $S_0, \ldots, S_n$.

To do this, we introduce a martingale called a Doob martingale [Doo40] where $S_0 = \mathbb{E}[X]$ (i.e., S_0 is the expected number of empty bins) and $S_n = X$ (i.e., S_n is the random variable counting the number of empty bins at the end of the experiment). The S_i's in the middle show how our knowledge of the final number of empty bins evolves as randomness is gradually revealed step by step. We define multiset random variables $\mathcal{F}_0, \ldots, \mathcal{F}_n$, where $\mathcal{F}_i$ represents the multiset of bins that the first i balls landed in. The random variables $S_0, \ldots, S_n$ are defined as $S_i \coloneqq \mathbb{E}[X \mid \mathcal{F}_i]$, namely, S_i is the expected value of X given the information accumulated after seeing the outcome of first i balls.

Examples. Let us first familiarize ourselves with these variables. Firstly,

$$S_0 = \mathbb{E}[X \mid \mathcal{F}_0] = \mathbb{E}[X \mid \emptyset] = \mathbb{E}[X]\,.$$

Additionally,

$$\begin{aligned} S_1 = \mathbb{E}[X \mid \mathcal{F}_1] &= \mathbb{E}\left[\sum_{i=1}^n \begin{cases} \left(1-\frac{1}{n}\right)^{n-1} & \text{if } \mathcal{F}_1 \neq \{i\} \\ 0 & \text{otherwise} \end{cases} \middle|\, \mathcal{F}_1\right] \\ &= \mathbb{E}\left[(n-1)\cdot\left(1-\frac{1}{n}\right)^{n-1} \middle|\, \mathcal{F}_1\right] \\ &= \left\{(n-1)\cdot\left(1-\tfrac{1}{n}\right)^{n-1} \text{ with probability } 1\right. \end{aligned}$$

Note that $(n-1)\cdot\left(1-\frac{1}{n}\right)^{n-1} = n\cdot\left(1-\frac{1}{n}\right)^{n} = \mathbb{E}[X]$.

Similarly,

$$S_2 = \begin{cases} (n-1)\cdot\left(1-\frac{1}{n}\right)^{n-2} & \text{with probability } \Pr\left(\mathcal{F}_2 = \{i,i\}\right) \text{ for some } i \in [n] \\ (n-2)\cdot\left(1-\frac{1}{n}\right)^{n-2} & \text{with probability } \Pr\left(\mathcal{F}_2 = \{i,j\}\right) \text{ for some } i,j \in [n] \text{ with } i \neq j \end{cases}$$

Another sanity check is the following:

$$\begin{aligned}\mathbb{E}[S_2] &= \frac{1}{n}\cdot(n-1)\cdot\left(1-\frac{1}{n}\right)^{n-2} + \left(1-\frac{1}{n}\right)\cdot(n-2)\cdot\left(1-\frac{1}{n}\right)^{n-2}\\ &= \left(1-\frac{1}{n}\right)^{n-1} + (n-2)\cdot\left(1-\frac{1}{n}\right)^{n-1}\\ &= (n-1)\cdot\left(1-\frac{1}{n}\right)^{n-1}\\ &= \mathbb{E}[X].\end{aligned}$$

Iterating this process, it will eventually become clear that we are adding one support point when we move from S_i to S_{i+1}, which eventually results in $S_n = X$. In other words, the S_i's represent the best running estimate of the final number of empty bins as the experiment progresses. Initially, the expectation serves as our starting prediction, which continuously refines itself as more randomness is revealed, ultimately converging on the actual outcome.

Martingale properties. We now show that $\{S_i\}_{i=0}^{n}$ is a martingale (w.r.t. $\{\mathcal{F}_i\}$). Firstly, we have for all $i \in [0, n]$, $\mathbb{E}[S_i] \leqslant n$, because there are at most n bins, implying that $\mathbb{E}[S_i]$ is finite.

The tower property in Proposition 3 guarantees that for all $i \in [n]$,

$$\mathbb{E}[S_i \mid \mathcal{F}_{i-1}] = \mathbb{E}[\mathbb{E}[X \mid F_i] \mid \mathcal{F}_{i-1}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{i-1}, \mathcal{F}'_i] \mid \mathcal{F}_{i-1}] = \mathbb{E}[X \mid \mathcal{F}_{i-1}] = S_{i-1},$$

where $\mathcal{F}'_i$ denotes the random variable taking on the index of the bin the i-th ball landed in. Combining these two facts, we conclude that $S_0, \ldots, S_n$ is a martingale with respect to $\mathcal{F}_0, \ldots, \mathcal{F}_n$.

Applying Azuma's inequality. Finally, to apply Azuma's inequality, we show that for all $i \in [n]$,

$$|S_i - S_{i-1}| \leqslant 1.$$

Let f'_i, f_{i-1} be events such that $\Pr(F_i = f_{i-1}, f'_i) \neq 0$ and $\Pr(\mathcal{F}_{i-1} = f_{i-1}) \neq 0$ (that is f_{i-1} is a possible "assignment of bins to balls $1, \ldots, i$" and $f'_i \in [n]$ represents the bin the i-th ball was thrown into). Thus,

$$S_i = \mathbb{E}[X \mid \mathcal{F}_i = (f_{i-1}, f'_i)] = \mathrm{empty}(f_{i-1}, f'_i)\cdot\left(1-\frac{1}{n}\right)^{n-i}$$

and

$$S_{i-1} = \mathbb{E}[X \mid \mathcal{F}_{i-1} = f_{i-1}] = \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)^{n-(i-1)}$$

where $\mathrm{empty}(f)$ denotes the number of bins not present in the multiset f (or equivalently, the number of bins left empty if balls were thrown into bins "as dictated by the multiset f"). With this definition of $\mathrm{empty}(f)$, we can immediately conclude that $|\mathrm{empty}(f_{i-1}, f'_i) - \mathrm{empty}(f_{i-1})| \leqslant 1$. Hence, for all $i \in [n]$:

$$\begin{aligned}|S_i - S_{i-1}| &= \left|\mathrm{empty}(f_{i-1}, f'_i)\cdot\left(1-\frac{1}{n}\right)^{n-i} - \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)^{n-(i-1)}\right|\\ &= \left(1-\frac{1}{n}\right)^{n-i}\cdot\left|\mathrm{empty}(f_{i-1}, f'_i) - \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)\right|\\ &\leqslant \left|\mathrm{empty}(f_{i-1}, f'_i) - \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)\right| \qquad \text{(since } (1-1/n)\leqslant 1 \text{ and } n-i\geqslant 0)\\ &= \begin{cases}\left|\mathrm{empty}(f_{i-1}) - \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)\right| & \text{if } \mathrm{empty}(f_{i-1}, f'_i) = \mathrm{empty}(f_{i-1})\\ \left|(\mathrm{empty}(f_{i-1}) - 1) - \mathrm{empty}(f_{i-1})\cdot\left(1-\frac{1}{n}\right)\right| & \text{otherwise}\end{cases}\\ &\leqslant \max\left(\frac{\mathrm{empty}(f_{i-1})}{n}, 1 - \frac{\mathrm{empty}(f_{i-1})}{n}\right)\\ &\leqslant 1.\end{aligned}$$

This is because the number of empty bins empty$(\cdot)$ can change by at most 1 when one more ball is thrown. Thus, Azuma's inequality in Proposition 6 applies with $c = 1$. We can conclude that

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{n}\right)$$

and, as before, choosing $t = \sqrt{\ln(3) \cdot n}$

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t\right) \leqslant \frac{2}{3}\ .$$

Thus, we indeed managed to prove precisely the bound we "claimed" using Chernoff bound, but this time properly by using Martingales and Azuma's inequality.

3.2 Application 2: Chromatic Number

We now consider the problem of finding the chromatic number of a graph G sampled from $\mathbb{G}(n, 1/2)$, denoted by $X = \chi(G)$ (a random graph where an edge is present between each pair of vertices with probability $1/2$). Recall that the chromatic number $\chi(G)$ of G is the minimum number of colors we can color the vertices of the graph with such that no edge is monochromatic i.e., has the same color on both vertex endpoints.

It is known that the expected chromatic number of $G \sim \mathbb{G}(n, 1/2)$ is $\Theta(n/\log n)$ [Bol88], but to our knowledge, the constants here are not well-understood yet. We will use Azuma's inequality to show that the chromatic number is within $\pm\sqrt{\ln(3) \cdot n}$ of this expectation with constant probability. Note that this provides a sharper concentration result than the $\pm\Theta(n/\log n)$ bound around the expected value.

The proof goes as before by defining a *proper* martingale and then applying Azuma's inequality.

3.2.1 A First Attempt – Edge Exposure Martingales

Let us start with the perhaps most obvious choice for defining our martingales. We first define the multisets $\{\mathcal{F}_k\}$ which "reveal" information over time. This is achieved by ordering the edges, e.g., enumerating all potential edges in lexicographical order as $e_1, e_2, \ldots, e_{\binom{n}{2}}$ (any fixed ordering of the potential edges would work just as well). Now let $\{\xi_k\}_{i=1}^{\binom{n}{2}}$ be a sequence of i.i.d. random variables taking values 0 or 1 with equal probability $1/2$. Intuitively, ξ_i is a random variable indicating whether the i^{th} edge is present in G. We reveal edges incrementally, i.e., we define the multi-sets $\mathcal{F}_i$ for all $i \in [\binom{n}{2}]$ as $\mathcal{F}_i = \{\xi_1, \xi_2, \ldots, \xi_i\}$. In other words, after i steps, we only know whether each of the first i edges in our list is present or not; we do *not* yet know about edges $e_{i+1}, \ldots, e_{\binom{n}{2}}$. Clearly, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_{\binom{n}{2}}$.

As with the balls and bins experiment, we define the *Doob martingale* with respect to $\{\mathcal{F}_i\}$ as

$$M_i \;=\; \mathbb{E}\left[X \,\middle|\, \mathcal{F}_i\right].$$

We now show that $\{M_i\}$ is a martingale (w.r.t. $\{\mathcal{F}_i\}$). Firstly, we have $\mathbb{E}[M_i] \leqslant \binom{n}{2}$, implying that $\mathbb{E}[M_i]$ is finite. We now have to show that:

$$\mathbb{E}[\, M_{i+1} \mid \mathcal{F}_i \,] \;=\; M_i.$$

Using the *tower property* of Proposition 3:

$$\mathbb{E}[\, M_{i+1} \mid \mathcal{F}_i \,] \;=\; \mathbb{E}\Big[\mathbb{E}[\, X \mid \mathcal{F}_{i+1} \,] \;\Big|\; \mathcal{F}_i\Big] \;=\; \mathbb{E}[\, X \mid \mathcal{F}_i \,] \;=\; M_i.$$

Hence, $\{M_i\}_{i=0}^{\binom{n}{2}}$ is a martingale with respect to $\{\mathcal{F}_i\}$.

Remark. We should remark that we can define a *Doob martingale* for any stochastic process as follows. Let Y be a bounded random variable and think of $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_m$ (for some finite m), as providing some "filtration", where $\mathcal{F}_0 = \emptyset$ and $\mathcal{F}_m$ deterministically determines Y, and each step provides outcomes of more randomness used in defining Y, until in the last step fully determines Y. Define for $t \in [m]$,

$$Z_t := \mathbb{E}\left[Y \mid \mathcal{F}_t\right];$$

then, $Z_1, \ldots, Z_m$ form a martingale with respect to $\mathcal{F}_0, \ldots, \mathcal{F}_m$. To see, why this is a martingale, we have that for every $t > 1$, every $|Z_t| < \infty$ since Y is bounded and moreover,

$$\mathbb{E}\left[Z_t \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[\mathbb{E}\left[Y \mid \mathcal{F}_t\right] \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[Y \mid \mathcal{F}_{t-1}\right] = Z_{t-1},$$

where the first and last equality are by the definition of Z_t and Z_{t-1}, respectively, and the middle equality is by the tower property of Proposition 3.

We can now apply Azuma's inequality to it provided that $|M_i - M_{i-1}| \leqslant 1$ holds for all $i \in [\binom{n}{2}]$.

Claim 7. $|M_i - M_{i-1}| \leqslant 1$ *for all* $i \in [\binom{n}{2}]$

Proof. We start with the following fact. Consider two graphs G and G' on the same vertex set, differing in the status of exactly *one* potential edge $e = \{u, v\}$. Then

$$\left|\chi(G) - \chi(G')\right| \leqslant 1.$$

We show this by assuming, without loss of generality, that G is the graph with the extra edge. Clearly, any valid coloring of G is a valid coloring of G'. Hence, $\chi(G') \leqslant \chi(G)$. Now take an optimal coloring of G' and color G with it. If the extra edge e "violates this coloring", then color $\max\{u, v\}$ with a new color. This clearly forms a valid coloring of G using only $\chi(G') + 1$ colors. It follows that $\chi(G) \leqslant \chi(G') + 1$. Thus, $\chi(G') \leqslant \chi(G) \leqslant \chi(G') + 1$.

Additionally, recall that:

$$M_i = \mathbb{E}\big[X \,\big|\, \mathcal{F}_i\big] \quad \text{and} \quad M_{i-1} = \mathbb{E}\big[X \,\big|\, \mathcal{F}_{i-1}\big].$$

The only difference between $\mathcal{F}_{i-1}$ and $\mathcal{F}_i$ is that (exactly) one new edge e_i has been revealed (i.e., we now know whether it is present or absent). *Conditioned on* $\mathcal{F}_{i-1}$, there are two possible worlds for G as far as edge e_i is concerned: e_i present, or e_i absent. The above shows that $\chi(\cdot)$ can differ by at most 1 between the two possibilities. Since the two random variables differ by at most 1 pointwise for every outcome, the conditional expectations of X under "e_i present" versus "e_i absent" can differ by at most 1. Formally:

$$\left|\,\mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 1] - \mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 0]\right| \leqslant 1.$$

Note that $M_i = \mathbb{E}\left[X \mid \mathcal{F}_{i-1}, e_i\right]$ can take the values $a_0 = \mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 0]$ or $a_1 = \mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 1]$. Also, we have:

$$\begin{aligned}
M_{i-1} &= \mathbb{E}[\,X \mid \mathcal{F}_{i-1}] \\
&= \mathbb{E}\big[\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i\,] \,\big|\, \mathcal{F}_{i-1}\big] && \text{(Tower property of Proposition 3)} \\
&= \frac{1}{2}\,\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i = 1\,] + \frac{1}{2}\,\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i = 0\,]. && (\Pr(e_i = 1 \mid \mathcal{F}_{i-1}) = 1/2)
\end{aligned}$$

M_{i-1} is the average of a_0 and a_1. Thus, its value lies between the two, implying that:

$$\begin{aligned}
|M_i - M_{i-1}| &\leqslant \max\left(|a_0 - M_{i-1}|, |a_1 - M_{i-1}|\right) && (M_i \text{ is } a_0 \text{ or } a_1) \\
&\leqslant \max\left(|a_0 - a_1|, |a_1 - a_0|\right) && (M_{i-1} \text{ lies between } a_0 \text{ and } a_1) \\
&\leqslant 1.
\end{aligned}$$

This proves the claim for all $i \in [\binom{n}{2}]$. □

We can now apply Azuma's inequality:

$$\Pr\left(|S_n - S_0| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{\binom{n}{2}}\right).$$

We want failure with constant probability, so we need to set $t \approx n$. This tells us that X is within $\pm n$ of $\mathbb{E}[X]$. Unfortunately, this is an absolutely trivial statement, because $1 \leqslant \chi(G) \leqslant n$ for any graph G!

We seem to need to define the $\mathcal{F}_i$'s in a cleverer way: Azuma's inequality is a *dimension-dependent* concentration bound and as such is sensitive to the number of variables in its martingale; as such, to obtain a better bound, we need to reduce the number of variables from $\approx n^2$ to something much smaller.

3.2.2 Refining our Approach – Vertex Exposure Martingales

Earlier, we constructed the $\mathcal{F}_i$'s by revealing one edge at a time (Edge Exposure Martingale). Instead, one could reveal one vertex, or more precisely the neighborhood of one vertex, at a time (Vertex Exposure Martingale). Vertices are assumed to be numbered from 1 to n, so we reveal them in that order (any other order also works). We still use the indicator random variables for the edges $\{\xi_k\}_{i=1}^{\binom{n}{2}}$ that we used in the edge exposure martingale.

We define $N^+(u)$ as the set of edges incident on u that go to lower ranked vertices. So $N^+(1) = \emptyset$, because no other vertices have been revealed. $N^+(2) = \{\xi_1\}$, where ξ_1 is the random variable that corresponds to the edge $\{2, 1\}$. Similarly, $N^+(3) = \{\xi_2, \xi_3\}$, and so on. Basically, $N^+(i)$ will have $i - 1$ random variables corresponding to the edges $(i, 1), (i, 2), \ldots, (i, i-1)$. We can now define the multisets $\mathcal{F}_i$ for all $i \in [n]$ as the first i $N^+(k)$'s: $\mathcal{F}_i = \{N^+(1), N^+(2), \ldots, N^+(i)\}$.

What we have done is create a grouping of the random variables ξ_k, and reveal them in groups instead of one at a time, reducing the number of $\mathcal{F}_i$'s. In other words, after i steps we know the graph induced on the first i vertices, we do *not* yet know about edges incident on vertices $i+1$ to n. Clearly, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$.

Similar to the previous analysis, we define the *Doob martingale* with respect to $\{\mathcal{F}_i\}$ as

$$M_i \;=\; \mathbb{E}\left[X \,\middle|\, \mathcal{F}_i\right],$$

and as before, we have that this is indeed a martingale with respect to $\{\mathcal{F}_i\}$.

We are almost ready to apply Azuma's inequality to it, but before doing so, we have to establish the "bounded-ness condition". We first make a simple observation.

Claim 8. *Consider two graphs G and G' on the same vertex set, differing in the neighborhood of exactly* one *vertex i. Then $\left|\chi(G) - \chi(G')\right| \;\leqslant\; 1$.*

Proof. The two graphs G and G' are on the same vertex set, differing in the neighborhood of exactly *one* vertex i. This means that both graphs induced on $V \setminus \{i\}$ are identical, but they have potentially different edges incident on vertex i. For such graphs we want to show that:

$$\left|\chi(G) - \chi(G')\right| \;\leqslant\; 1.$$

Consider graph G and color it using the optimal coloring of G'. Now just color i with a completely new color. Any edge not incident on i is not monochromatic, because it also exists in G'. Any edge incident on i is not monochromatic, since i has a completely new color. This implies: $\chi(G) \leqslant \chi(G') + 1$. We can similarly show $\chi(G') \leqslant \chi(G) + 1$, which implies $\left|\chi(G) - \chi(G')\right| \;\leqslant\; 1$. □

We now show that $|M_i - M_{i-1}| \leqslant 1$ for all $i \in [n]$.

Claim 9. $|M_i - M_{i-1}| \leqslant 1$ *for all* $i \in [n]$

Proof. Recall that:

$$M_i \;=\; \mathbb{E}\big[X \,\big|\, \mathcal{F}_i\big] \quad \text{and} \quad M_{i-1} \;=\; \mathbb{E}\big[X \,\big|\, \mathcal{F}_{i-1}\big].$$

When going from $\mathcal{F}_{i-1}$ to $\mathcal{F}_i$, we revealed the neighborhood of exactly one vertex i to vertices in $[i-1]$. *Conditioned on* $\mathcal{F}_{i-1}$, there are 2^{i-1} possible worlds for G as far as vertex i is concerned: for all vertices $k \in [i-1]$, the edge (i,k) could be present or absent. This gives us 2^{i-1} possible neighborhoods (to previous vertices) for vertex i: $N_1^+(i), N_2^+(i), \ldots, N_{2^{i-1}}^+(i)$. Now fix any possibility for the remaining graph, i.e., for vertices from $i+1$ to n fix any choices of edges incident on them.

We hereby obtain 2^{i-1} different graphs $G_1, G_2, \ldots, G_{2^{i-1}}$. $\chi(\cdot)$ can differ by at most 1 between any two such graphs, because they differ in the neighborhood of only one vertex (Claim 8). Since the two random variables differ by at most 1 pointwise for every fixed value they may assume, the conditional expectations of X under any two neighborhoods $N_j^+(i)$ and $N_k^+(i)$ can differ by at most 1. Formally:

$$\big|\,\mathbb{E}[X \mid \mathcal{F}_{i-1},\, N^+(i) = N_j^+(i)] \;-\; \mathbb{E}[X \mid \mathcal{F}_{i-1},\, N^+(i) = N_k^+(i)]\big| \;\leqslant\; 1.$$

Note that $M_i = \mathbb{E}\left[X \mid \mathcal{F}_{i-1}, N^+(i)\right]$ can take 2^{i-1} values depending on the value of $N^+(i)$. Let a_0 be the smallest of the 2^{i-1} values which occurs for $N^+(i) = N_j^+(i)$. Let a_1 be the largest of the 2^{i-1} values which occurs for $N^+(i) = N_k^+(i)$. Also, we have:

$$\begin{aligned}
M_{i-1} \;&=\; \mathbb{E}[\,X \mid \mathcal{F}_{i-1}] \\
&=\; \mathbb{E}\big[\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, N^+(i)\,] \,\big|\, \mathcal{F}_{i-1}\big] && \text{(Tower property of expectation)} \\
&=\; \sum_{\ell=1}^{2^{i-1}} \frac{1}{2^{i-1}}\, \mathbb{E}[\,X \mid \mathcal{F}_{i-1}, N^+(i) = N_\ell^+(i)\,]. && \left(\Pr\left(N^+(i) = N_\ell^+(i) \mid \mathcal{F}_{i-1}\right) = 1/2^{i-1}\right)
\end{aligned}$$

M_{i-1} is the average of 2^{i-1} numbers between a_0 and a_1. Thus, its value lies between the two implying that:

$$\begin{aligned}
|M_i - M_{i-1}| &\leqslant \max_\ell \left(|M_{i-1} - \mathbb{E}[\,X \mid \mathcal{F}_{i-1}, N^+(i) = N_\ell^+(i)\,]|\right) && (M_i \text{ is one of these } 2^{i-1} \text{ values}) \\
&\leqslant |a_0 - a_1| && (\text{all the values lie between } a_0 \text{ and } a_1) \\
&\leqslant 1.
\end{aligned}$$

This proves the claim for all $i \in [n]$. □

We can now apply Azuma's inequality with $t = \sqrt{\ln(3) \cdot n}$:

$$\Pr\left(|S_n - S_0| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{n}\right) \leqslant 2/3.$$

Thus, we get failure with small constant probability. This tells that X is within $\pm\sqrt{\ln(3) \cdot n}$ of $\mathbb{E}[X]$ with constant probability. Note that we just know the value of $\mathbb{E}[X]$ to within $\pm\Theta(n/\log n)$. What we have shown here is that whatever the value of $\mathbb{E}[X]$ is, X is within $\pm\sqrt{\ln(3) \cdot n}$ of it with constant probability. We are able to show this, because these additive concentration bounds do not depend on the expectation of the random variable we are dealing with.

References

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357 – 367, 1967. 5

[Bol88] B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988. 7

[Doo40] J. L. Doob. Regularity properties of certain families of chance variables. Jan 1940. 5

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 5