Chapter 25

Strong(er) Concentration Bounds I: Martingales and Azuma's Inequality

We now go back to the topic of concentration bounds introduced in Chapter 2 and used throughout the book so far. This chapter and the next include more advanced concentration bounds and some of their applications.

25.1 Balls and Bins Revisited

We again consider a balls and bins experiment, but this time we are interested in the number of empty bins.

Problem 25.1. Consider throwing n balls into n bins independently and uniformly at random. We wish to asymptotically characterize the function $T \colon \mathbb{N} \to \mathbb{N}$ 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 $2/3$.

Define an indicator random variable X_i which is 1 iff the i-th bin is empty at the end of the experiment and define $X := \sum_{i=1}^{n} X_i$. We have that

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

which converges to n/e as $n \to \infty$. Our goal however is to (asymptotically) find $T(n)$ such that

$$\Pr(|X - \mathbb{E}[X]| \geqslant T(n)) \leqslant \frac{1}{3}.$$

This is certainly a concentration bound problem and we might be tempted to already use our standard toolkit from Chapter 2 here. In fact, going into "autopilot" and applying the additive Chernoff bound (Proposition 2.9), we conclude

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

Choosing $t = \sqrt{\ln(6)/2 \cdot n}$, we obtain

$$\Pr(|X - \mathbb{E}[X]| \geqslant t) \leqslant \frac{1}{3}.$$

However, our application of the Chernoff bound has a quite severe problem—the indicator random variables we considered are *not* independent! If we know that some $X_i = 1$, then we know that the n balls are now thrown uniformly into $n-1$ bins, making the probability of any other bin becoming empty lower.

As a result, our "proof" above is fundamentally flawed. On the other hand, it can be empirically verified that X is concentrated around its expected value (see the histogram in Figure 25.1), suggesting that we might be lacking mathematical machinery to prove what we want, rather than the entire statement being false.

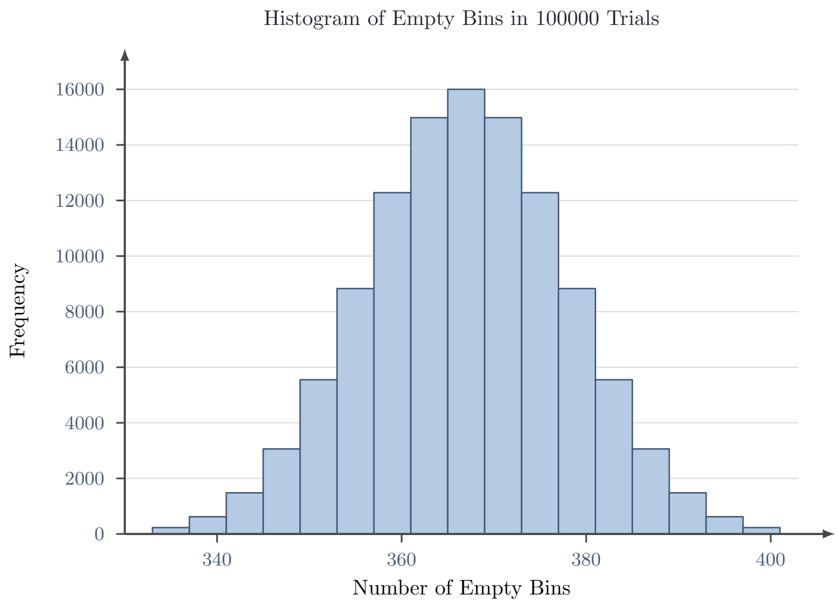

Figure 25.1: A histogram of frequency of empty bins in 100,000 trials of throwing 1000 balls into 1000 bins uniformly and independently (here, the number of empty bins X satisfies $\mathbb{E}[X] \approx 368$).

25.2 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 that may be correlated, but whose correlation is, in some sense, low. To continue, we need to take some detours.

25.2.1 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 25.2. Suppose we are interested in the following two random variables A and B:

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

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

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

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

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

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

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

where $\text{supp}(Y)$ is the set of points y such that $\Pr\left(Y = y\right) > 0$. We again emphasize that while $\mathbb{E}\left[X\right]$ is a number, $\mathbb{E}\left[X \mid Y\right]$ is a random variable (it becomes a number $\mathbb{E}\left[X \mid Y = y\right]$ for any fixed choice of y).

Some basic facts follow from this definition:

Fact 25.3 (Law of Total Expectation)**.** *Let A be a random variable and B be a discrete random variable. Then*

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

Proof. We have

$$\begin{aligned} \mathbb{E}\left[\mathbb{E}\left[A \mid B\right]\right] &= \sum_b \Pr\left(B = b\right) \cdot \mathbb{E}\left[A \mid B = b\right] = \sum_b \sum_a \Pr\left(B = b\right) \cdot a \cdot \Pr\left(A = a \mid B = b\right) \\ &= \sum_b \sum_a a \cdot \Pr\left(A = a \wedge B = b\right) = \sum_a a \cdot \sum_b \Pr\left(A = a \wedge B = b\right) \\ &= \sum_a a \cdot \Pr\left(A = a\right) = \mathbb{E}\left[A\right], \end{aligned}$$

concluding the proof. □

Fact 25.4 (Tower Property)**.** *Let A be a random variable and B, C be discrete random variables. Then*

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

Proof. For any $c \in \text{supp}(C)$, by applying Fact 25.3 to the distribution of $A, B \mid C = c$, denoted by A', B', we have

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

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

25.2.2 Martingales

With the definition of conditional expectation as a random variable, we can now define martingales:

Definition 25.5. 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 $i \in [n] \cup \{0\}$, S_i is determined by $\mathcal{F}_i$;
- 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 (this is also called the *gambler's (ruin) random walk*). 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:

$$\Pr(\xi_k = 1) = \frac{1}{2}, \quad \Pr(\xi_k = -1) = \frac{1}{2}, \quad \text{and all } \xi_k \text{ are independent.}$$

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 multisets $\mathcal{F}_i$ 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. The main part is to show the following.

Claim 25.6. $\mathbb{E}[M_{i+1} \mid \mathcal{F}_i] = M_i$ *for all* $i \in \{0\} \cup [n-1]$.

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.

25.2.3 Azuma's Inequality

We now get to a more advanced concentration bound: a "Chernoff-like" bound for martingales.

Proposition 25.7 (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*

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

The bounded differences on their own can only confine the martingale to $|S_n - S_0| \leqslant c \cdot n$; Azuma's inequality sharpens this to show that with high probability, $|S_n - S_0| \lesssim c\sqrt{n}$ instead.

25.3 Applications

25.3.1 Application 1: Balls and Bins

After this slight detour, we return to the balls and bins experiment from Problem 25.1. We would like to use Azuma's inequality to show concentration of the number of empty bins around its expectation. Thus, our goal is 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 (see Figure 25.2). 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 the first i balls.

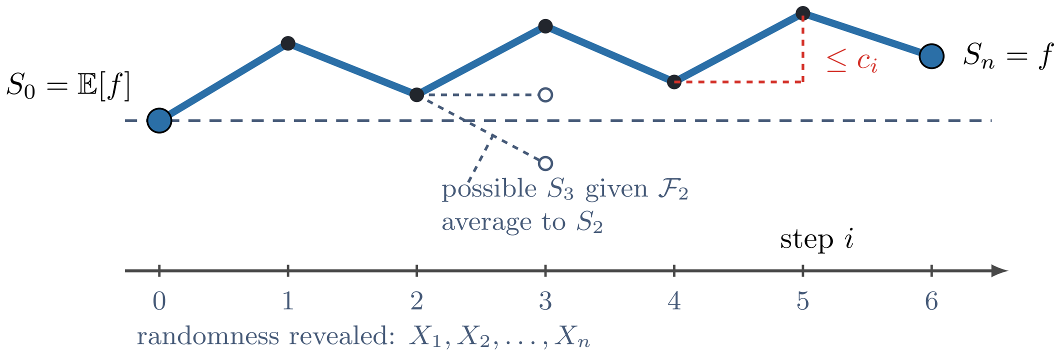

Figure 25.2: An illustration of a Doob (exposure) martingale. The picture is schematic, with one fair bit revealed per step.

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,

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

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,

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

Another sanity check is the following:

$$\begin{aligned} \mathbb{E}[S_2] =& \underbrace{\frac{1}{n}}_{\Pr(\mathcal{F}_2=\{i,i\})} \cdot (n-1)\cdot\left(1-\frac{1}{n}\right)^{n-2} + \underbrace{\left(1-\frac{1}{n}\right)}_{\Pr(\mathcal{F}_2=\{i,j\})} \cdot (n-2)\cdot\left(1-\frac{1}{n}\right)^{n-2} \\ =& \left(1-\frac{1}{n}\right)^{n-1} + (n-2)\cdot\left(1-\frac{1}{n}\right)^{n-1} \\ =& (n-1)\cdot\left(1-\frac{1}{n}\right)^{n-1} = \mathbb{E}[X]. \end{aligned}$$

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 $i \in [n] \cup \{0\}$, $\mathbb{E}[S_i] \leqslant n$, because there are at most n bins, so $\mathbb{E}[S_i]$ is finite.

The tower property in Fact 25.4 guarantees that for all $i \in [n]$,

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

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

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

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

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

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

and

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

where 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 empty(f), we can conclude that $|\text{empty}(f_{i-1}, f'_i) - \text{empty}(f_{i-1})| \leqslant 1$. Hence, for all $i \in [n]$:

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

Thus, Azuma's inequality in Proposition 25.7 applies with $c = 1$. We can conclude that

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

and, as before, choosing $t = \sqrt{2\ln(6) \cdot n}$

$$\Pr\left(|X - \mathbb{E}[X]| \geqslant t\right) \leqslant \frac{1}{3}\ .$$

Thus, we indeed managed to prove the bound we claimed using Chernoff bound with a wrong proof (up to a factor 2 loss in the range of deviation), but this time properly, by using martingales and Azuma's inequality.

25.3.2 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 $X = \chi(G)$ (to be precise, $\mathbb{G}(n, 1/2)$ samples 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 endpoints.

It is known that the expected chromatic number of $G \sim \mathbb{G}(n, 1/2)$ is $\sim n/(2\log n)$ [Bol88]. We will use Azuma's inequality to show that the chromatic number is within $\pm\,\Theta(\sqrt{n})$ of this expectation with constant probability. Note that this provides a sharper concentration result than the $\pm\,o(n/\log n)$ bound around the expected value.

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

A First Attempt: Edge Exposure Martingales

Let us start with the perhaps most obvious choice for defining our martingale. We first define the multisets $\{\mathcal{F}_i\}$ 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 $\{\xi_i\}_{i=1}^{\binom{n}{2}}$ 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 multisets $\mathcal{F}_i$ 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* w.r.t. $\{\mathcal{F}_i\}$ 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 Fact 25.4:

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

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. Define for $t \in [m] \cup \{0\}$,

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

then, $Z_0, \ldots, Z_m$ 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 $t \in [m]$, $|Z_t| < \infty$ 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 Fact 25.4.

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

Claim 25.8. *$|M_i - M_{i-1}| \leqslant 1$ 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

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

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 one of the endpoints of e 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:

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

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:

$$\big|\,\mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 1] \;-\; \mathbb{E}[X \mid \mathcal{F}_{i-1}, e_i = 0]\big| \;\leq\; 1.$$

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:

$$\begin{aligned}
M_{i-1} &= \mathbb{E}[\,X \mid \mathcal{F}_{i-1}] \\
&= \mathbb{E}\big[\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i\,] \bigm| \mathcal{F}_{i-1}\big] && \text{(Tower property of Fact 25.4)} \\
&= \frac{1}{2}\,\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i = 1\,] \;+\; \frac{1}{2}\,\mathbb{E}[\,X \mid \mathcal{F}_{i-1}, e_i = 0\,]. && (\Pr e_i = 1 \mid \mathcal{F}_{i-1} = 1/2)
\end{aligned}$$

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:

$$\Pr|M_{\binom{n}{2}} - M_0| \geqslant t \leqslant 2 \cdot \exp\left(-\frac{t^2}{2 \cdot \binom{n}{2}}\right).$$

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

The takeaway is that we need to define the $\mathcal{F}_i$'s in a cleverer way: Azuma's inequality is a *dimension-dependent* concentration bound and therefore 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.

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 $\{\xi_i\}_{i=1}^{\binom{n}{2}}$ 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 vertex 1 has no lower-ranked neighbors; $N^+(2) = \{\xi_1\}$, 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, $N^+(i)$ has $i-1$ 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 union of the first $N^+(i)$'s:

$$\mathcal{F}_i = N^+(1) \cup N^+(2) \cup \cdots \cup N^+(i).$$

What we have done is create a grouping of the random variables $\{\xi_i\}$, 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, but 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-difference condition". We first make a simple observation.

Claim 25.9. *Consider two graphs G and G' on the same vertex set, differing in the neighborhood of exactly* one *vertex. 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, say, 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 claim:

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

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

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

Claim 25.10. $|M_i - M_{i-1}| \leqslant 1$ *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 edges from vertex i to the 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 25.9). Since any two of these 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:

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

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:

$$\begin{aligned}
M_{i-1} &= \mathbb{E}[\, X \mid \mathcal{F}_{i-1}] \\
&= \mathbb{E}\big[\mathbb{E}[\, X \mid \mathcal{F}_{i-1}, N^+(i)\,] \bigm| \mathcal{F}_{i-1}\big] && \text{(Tower property of Fact 25.4)} \\
&= \sum_{\ell=1}^{2^{i-1}} \frac{1}{2^{i-1}}\, \mathbb{E}[\, X \mid \mathcal{F}_{i-1}, N^+(i) = N_\ell^+(i)\,]. && \text{(as } \Pr\left(N^+(i) = N_\ell^+(i) \mid \mathcal{F}_{i-1}\right) = 1/2^{i-1})
\end{aligned}$$

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 $t = \sqrt{2\ln(6)\cdot n}$:

$$\Pr\left(|M_n - M_0| \geqslant t\right) \leqslant 2 \cdot \exp\left(-\frac{t^2}{2n}\right) \leqslant 1/3.$$

Thus, we get failure with small constant probability. This tells us that X is within $\pm\sqrt{2\ln(6)\cdot n}$ of $\mathbb{E}\left[X\right]$ with constant probability.

We conclude this chapter with a remark. We just know the value of $\mathbb{E}\left[X\right]$ to within $\pm\Theta(n/\log n)$, and yet managed to prove that whatever the value of $\mathbb{E}\left[X\right]$ is, X is within $\pm\Theta(\sqrt{n})$ of it with constant probability; in other words, the concentration we have proved is even sharper than our current estimates for the range $\mathbb{E}\left[X\right]$ itself belongs to.

Exercises

Exercise 25.1. In this exercise, we use Azuma's inequality to reprove the additive Chernoff bound (Proposition 2.9) with a slightly weaker constant in the exponent. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [0,1]$, let $S := \sum_{i=1}^n X_i$, and write $\mu := \mathbb{E}\left[S\right]$. Set $\mathcal{F}_i := \{X_1, \ldots, X_i\}$ and define the Doob martingale $Z_i := \mathbb{E}\left[S \mid \mathcal{F}_i\right]$ for $i \in [n] \cup \{0\}$.

1. Show that $Z_i = \sum_{j\leqslant i} X_j + \sum_{j>i} \mathbb{E}\left[X_j\right]$, so that $Z_0 = \mu$ and $Z_n = S$, and conclude that the increment is $Z_i - Z_{i-1} = X_i - \mathbb{E}\left[X_i\right]$.

2. Deduce that $|Z_i - Z_{i-1}| \leqslant 1$ for all $i \in [n]$, and apply Azuma's inequality (Proposition 25.7) to obtain
$$\Pr\left(|S - \mu| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{2n}\right).$$

Exercise 25.2 (♦). Azuma's inequality has a widely used corollary for functions of independent inputs, called the **bounded-differences (or McDiarmid's) inequality**. Let $X_1, \ldots, X_m$ be independent random variables, with X_i chosen from $\mathcal{X}_i$ and let $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_m \to \mathbb{R}$ be a function. We are interested in proving a concentration bound for $f(X_1, \ldots, X_m)$.

We say that f is *c-Lipschitz* for some constant $c > 0$ iff the following holds: for every coordinate $i \in [m]$ and every pair of inputs differing only in index i,

$$|f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m) - f(x_1, \ldots, x_{i-1}, x'_i, x_{i+1}, \ldots, x_m)| \leq c \ .$$

Write $\mu := \mathbb{E}\left[f(X_1, \ldots, X_m)\right]$ and $\mathcal{F}_i := \{X_1, \ldots, X_i\}$.

1. Show that $S_i := \mathbb{E}\left[f(X_1, \ldots, X_m) \mid \mathcal{F}_i\right]$ is a martingale with $S_0 = \mu$ and $S_m = f(X_1, \ldots, X_m)$.
2. Prove the bounded-difference condition $|S_i - S_{i-1}| \leqslant c$ for all $i \in [m]$.
3. Conclude McDiarmid's inequality:

$$\Pr\left(|f(X_1, \ldots, X_m) - \mu| \geqslant t\right) \leqslant 2 \exp\left(-\frac{t^2}{2m \cdot c^2}\right) .$$

4. Recover the empty-bins bound of Section 25.3.1 in a single line by exhibiting the appropriate f and c.
5. Explain why the same argument, applied to $f = \chi(G)$ with the neighborhoods of the n vertices as the independent inputs, reproduces the $\pm\sqrt{2\ln(3) \cdot n}$ concentration of the chromatic number obtained in this chapter by vertex exposure martingales.

Note: In both exercises above, the obtained exponents for additive Chernoff and McDiarmid's inequalities are weaker than the standard bounds by a factor of 4. This is simply because we have stated a simpler form of Azuma's inequality for clarity. Using the slightly stronger bounds in Azuma's inequality allows one to recover the precise bounds in these exercises as well.