Chapter 1

A GUIDE TO CONCENTRATION BOUNDS

Josep Díaz, Jordi Petit and Maria Serna
Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica Catalunya
Jordi Girona Salgado 1–3
08034-Barcelona.
diaz,jpetit,mjserna@lsi.upc.es

Abstract In this chapter we survey some of the concentration bounds that are used in the field of combinatorics and theoretical computer science. We have tried to emphasize the intuition behind the use of the different bounds. Throughout the text, the reader can follow the application of the different bounds to the same examples, comparing each time with the previous results.

1. INTRODUCTION

Worst case analysis of an algorithm gives information on hard instances that may not be representative of its real behavior on the data of interest. In such situations, the analysis of its *average behavior* may give more useful information. Given a probability distribution on the set of inputs and a performance measure, we wish to analyze the expected value of this performance, seen as a random variable on the probability space defined by the inputs of a given size. Relevant performance measures may be running time, memory size, solution quality, number of executions of a type of instructions, etc. For instance, in a sorting algorithm, the uniform distribution on the symmetric group S_n is a natural distribution when the input to be sorted has n keys. As one of the basic operations used by many sorting algorithms is to compare values, the number of comparisons performed along the algorithm's execution is a possible measure to analyze. For the Quicksort algorithm, it is a classical result that the the expected number of comparisons is $O(n \log n)$ [5]. However, even the expected behavior may not give enough information: the dispersion of the random variable must be taken into account. A way to solve this problem is to show that the value of the random variable is concentrated around the expected value, so that the probability of getting an instance such that the performance deviates very much from the expected value, is very small. In the case of Quicksort, a recent result states that the probability that the real number of comparisons performed by the algorithm deviates more than ϵ times their expectation is $O(n^{-\epsilon \ln \ln n})$ [30].

Another important technique in computer science and discrete mathematics is the *probabilistic method*. The basic idea behind the probabilistic method is to show that a certain property holds for some combinatorial structure. To do so, one constructs an appropriate probability space, and shows that the property has non-zero probability on it. A lot of research has been devoted to the probabilistic method, see for example [31] for a survey and [1] for a reference book. Although the goal of every application of the probabilistic method is to show that a particular desirable event occurs with positive probability, frequently it is needed to show first that the probability of some other related events are very small. For example, in most of the applications of the *Lovász local lemma*, in order to show that the probability of the union of a set of events is less than one, there is a first step proving that each event has very small probability [1, Chapter 5]. Concentration bounds are a useful tool to prove that the probability of a sequence of events becomes very small.

We have mentioned two areas where showing concentration around the expected value is important. In general, *concentration bounds* are used to prove that the probability that a random variable deviates significatively from its expectation is sufficiently small.

Theorems like the *law of large numbers*, the *central limit theorem* or the *de Moivre–Laplace limit theorem* are some of the most basic theorems in probability theory (see e.g. [10, 14]). However, it turns out that these classical results are not the most adequate tools to study the problems associated with the analysis of algorithms and discrete mathematics. A first reason is that, in probability, classical limit theorems provide asymptotic results that only apply in the infinite limit, but for most of the applications in discrete mathematics and algorithmics we are interested in large but finite cases. A second reason is that we are interested in explicit rates of convergence for the concentration bounds, while the above mentioned theorems do not deal with rates of convergence.

The survey is organized as follows. In Section 2, we present basic notation and results that the reader should know. In Section 3, we present several basic concentration bounds for a random variable. In Section 4, we derive results for the sum of independent random variables. In Section 5, we extend the concentration results to the case of functions of independent random variables. In Section 6, we present the martingale concept and use it to derive bounds for the case of dependent random variables that form a martingale. Finally, in Section 7, we give a quick glance to the topic of isoperimetric inequalities.

Some of the basic results shown here are included as parts of text-books, such as Chapter 7 and Appendix A of [1] or Chapter 4 of [32]. There are also two encyclopedic and excellent surveys [28, 29]; however the goal of this survey is to give a "basic introduction" to the topic of concentration bounds to the researcher in the field of theoretical computer science, without strong mathematical background.

2. BASIC TERMINOLOGY AND RESULTS

Let us review some basic definitions and facts from probability theory; for a more complete presentation, refer to any standard reference [10, 14]. For any $n \in \mathbb{N}$, let $[n]$ denote the set $\{1, 2, \dots, n\}$. For any set A, let $\mathcal{P}(A)$ denote the *power set* of A. For any Boolean expression B, we will denote by (B) the indicator function of B, which equals 1 if B and 0 if $\neg B$.

A *sample space* Ω is the set of possible outcomes of an experiment. Any element $\omega \in \Omega$ is a *basic event* and any subset $A \subseteq \Omega$ is an *event*. A *σ-field* $\mathcal{A}$ is a nonempty set of subsets of Ω such that:

1. if $A \in \mathcal{A}$ then $\overline{A} \in \mathcal{A}$,
2. the countable union of elements in $\mathcal{A}$ also is in $\mathcal{A}$, and
3. $\emptyset \in \mathcal{A}$.

A *probability measure* $\mathbf{Pr} : \mathcal{A} \to \mathbb{R}$ is a function that satisfies the following conditions:

1. $0 \leq \mathbf{Pr}[E] \leq 1$, for any $E \in \mathcal{A}$
2. $\mathbf{Pr}[\Omega] = 1$, and
3. $\mathbf{Pr}[\cup_{n\geq 1} E_n] = \sum_{n\geq 1} \mathbf{Pr}[E_n]$ for any sequence $(E_n)_{n\geq 1}$ of events in $\mathcal{A}$ such that $E_n \cap E_m = \emptyset$ for $n \neq m$.

A *probability space* is a triple $(\Omega, \mathcal{A}, \mathbf{Pr})$ where Ω is a sample space, $\mathcal{A}$ is a σ-field and $\mathbf{Pr}$ is a probability measure. Observe that $\mathcal{P}(\Omega)$ is a σ-field.

Boole's inequalities are useful to bound the probability that, given a set of events, at least one of them happens:

Proposition 1.1 (Boole's inequalities). *Let $E_1, \ldots, E_n$ be a collection of events; then*

$$\mathbf{Pr}\left[\bigcup_{i=1}^{n} E_i\right] \leq \sum_{i=1}^{n} \mathbf{Pr}\left[E_i\right] \quad \textit{and} \quad \mathbf{Pr}\left[\bigcap_{i=1}^{n} E_i\right] \geq 1 - \sum_{i=1}^{n} \mathbf{Pr}\left[\overline{E_i}\right].$$

Assume that a probability space $(\Omega, \mathcal{A}, \mathbf{Pr})$ is fixed. A collection of events $A_1, \ldots, A_n$ is *independent* if $\mathbf{Pr}\left[\cap_{i=1}^{n} A_i\right] = \prod_{i=1}^{n} \mathbf{Pr}\left[A_i\right]$. Given two events A and B such that $\mathbf{Pr}\left[B\right] \neq 0$, the *conditional probability* of event A given event B is defined as

$$\mathbf{Pr}\left[A|B\right] = \frac{\mathbf{Pr}\left[A \cap B\right]}{\mathbf{Pr}\left[B\right]}.$$

A *random variable* $X : \Omega \to \mathbb{R}$ is a real-valued function such that, for any $x \in \mathbb{R}$, the set $\{\omega \in \Omega \mid X(\omega) < x\} \in \mathcal{A}$. We shall give basic terminology only for discrete random variables, that is, when the set $X(\Omega)$ is discrete. Recall that the set $\{X = x\} = \{\omega \in \Omega \mid X(\omega) = x\}$ is an event and $\mathbf{Pr}\left[X = x\right] = \sum_{X(\omega)=x} \mathbf{Pr}\left[\omega\right]$. Given a random variable X, the function $f_X(x) = \mathbf{Pr}\left[X = x\right]$ is called the *density function* of X (some authors call it *probability mass function* to differentiate it from the continuous case).

A sequence of random variables $X_1, \ldots, X_n$ is *independent* if, for all vectors $(x_1, \ldots, x_n)$, the collection of events $(\{X_i = x_i\})_{i \in [n]}$ is independent.

Given two random variables X and Y on $(\Omega, \mathcal{A}, \mathbf{Pr})$ their *joint distribution* is defined by $f_{X,Y}(x, y) = \mathbf{Pr}\left[\{X = x\} \cap \{Y = y\}\right]$; and the *conditional distribution function* of Y given X is defined by $f_{Y|X}(x|y) = \mathbf{Pr}\left[Y = y|X = x\right]$ for any $x \in \mathbb{R}$ such that $\mathbf{Pr}\left[X = x\right] > 0$. From the definitions, we have

$$f_{Y|X}(y|x) = \frac{\mathbf{Pr}\left[\{X = x\} \cap \{Y = y\}\right]}{\mathbf{Pr}\left[X = x\right]} = \frac{f_{X,Y}(x, y)}{f_X(x)}. \tag{1.1}$$

Given a random variable X, its *expectation*, *expected* or *mean* value is defined as $\mathbf{E}\left[X\right] = \sum_{x \in X(\Omega)} x \, \mathbf{Pr}\left[X = x\right]$, its *variance* is defined as $\mathbf{Var}\left[X\right] = \mathbf{E}\left[(X - \mathbf{E}\left[X\right])^2\right]$ and its *standard deviation* is $\sigma_X = \sqrt{\mathbf{Var}\left[X\right]}$. If k is a positive integer, the k-th *moment* of X is $\mathbf{E}\left[X^k\right]$. The *median* of X is defined to be the set of all $x \in X(\Omega)$ such that $\mathbf{Pr}\left[X \leq x\right] \geq 1/2$ and $\mathbf{Pr}\left[X \geq x\right] \geq 1/2$. In our examples, the set of median values will have an unique element that will be refereed to by $\mathbf{M}[X]$.

The following proposition, whose proof is left to the reader, presents some basic properties of expectations and variances.

Proposition 1.2. *For any random variables X and Y and any $c \in \mathbb{R}$,*

$$\begin{aligned} \mathbf{E}[X+Y] &= \mathbf{E}[X] + \mathbf{E}[Y], \\ \mathbf{E}[cX] &= c\,\mathbf{E}[X], \\ \mathbf{Var}[X] &= \mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2, \\ \mathbf{Var}[cX] &= c^2\,\mathbf{Var}[X]. \end{aligned}$$

Moreover, if X and Y are independent,

$$\begin{aligned} \mathbf{E}[XY] &= \mathbf{E}[X]\,\mathbf{E}[Y], \\ \mathbf{Var}[X+Y] &= \mathbf{Var}[X] + \mathbf{Var}[Y]. \end{aligned}$$

The *conditional expectation* of a random variable Y given the random variable X is defined for every $x \in X(\Omega)$ such that $\mathbf{Pr}[X = x] > 0$ as:

$$\mathbf{E}[Y|X=x] = \sum_{y \in Y(\Omega)} y f_{Y|X}(y|x) = \sum_{y \in Y(\Omega)} y\,\mathbf{Pr}[Y=y|X=x].$$

Therefore, $\mathbf{E}[Y|X]$ as function of X is a random variable.

Some useful properties of the conditional expectation are given in the following proposition.

Proposition 1.3. *For any random variables X and Y, and any function f for which the corresponding expectations exist,*

$$\begin{aligned} \mathbf{E}[\mathbf{E}[Y|X]] &= \mathbf{E}[Y], \\ \mathbf{E}[X\,Y] &= \mathbf{E}[X\,\mathbf{E}[Y|X]], \\ \mathbf{E}[f(X,Y)|X=x] &= \mathbf{E}[f(x,Y)|X=x] \\ \mathbf{E}[Y|X=x] &= \mathbf{E}[Y], \qquad \textit{if } X \textit{ and } Y \textit{ are independent.} \end{aligned}$$

Proof. To prove the first equality, consider any basic event $\omega \in \Omega$ such that $X(\omega) = x$ and $Y(\omega) = y$. Then, $\mathbf{E}[Y|X=x] = \sum_{y \in Y(\Omega)} y f_{Y|X}(y|x)$, so that

$$\begin{aligned}\mathbf{E}\left[\mathbf{E}\left[Y|X\right]\right] &= \sum_{x\in X(\Omega)} \mathbf{E}\left[Y|X=x\right] f_X(x) \\ &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} y\, f_{Y|X}(y|x) f_X(x) \\ &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} y\, f_{X,Y}(x,y) \\ &= \sum_{y\in Y(\Omega)} y\, f_Y(y) = \mathbf{E}\left[Y\right].\end{aligned}$$

The proofs of the remaining inequalities are left as exercises. ✌

Later, we shall need the notion of conditional expectation of a multivariate random variable Y defined on a product space Ω^n given the values of the variables in a random vector $\mathbf{X} = (X_1, \ldots, X_n)$. In such a case, the definition of conditional expectation is extended, in the obvious way, to $\mathbf{E}\left[Y|X_1 = x_1, \ldots, X_i = x_i\right]$. Observe that $\mathbf{E}\left[Y|X_1 = x_1, \ldots, X_i = x_i\right]$ is a function on $\mathbb{R}^{n-i}$. To simplify notation, given a n-dimensional vector $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{X}_i$ will denote the vector formed by the first i components. The equality $\mathbf{X}_i = \mathbf{x}_i$ represents $X_1 = x_1, \ldots, X_i = x_i$, and the inequality $\mathbf{X}_i > \mathbf{x}_i$ represents $X_1 > x_1, \ldots, X_i > x_i$. The following proposition extends Proposition 1.3.

Proposition 1.4. *For any random variable Y defined on Ω^n, any random vector $\mathbf{X} = (X_1, \ldots, X_n)$ and any function $f : \mathbb{R}^n \to \mathbb{R}$ for which the corresponding expectations exist,*

$$\begin{aligned}\mathbf{E}\left[\mathbf{E}\left[Y|\mathbf{X}\right]\right] &= \mathbf{E}\left[Y\right], \\ \mathbf{E}\left[\mathbf{E}\left[Y|\mathbf{X}_i\right]|\mathbf{X}_j\right] &= \mathbf{E}\left[X|X_1, \ldots, X_j\right], \qquad 1 \le j < i \le n, \\ \mathbf{E}\left[Y f(\mathbf{X})|\mathbf{X} = \mathbf{x}\right] &= f(\mathbf{x})\mathbf{E}\left[Y|\mathbf{X} = \mathbf{x}\right]\end{aligned}$$

Let us now recall the definition and properties of some probability distributions. A random variable X has the *Bernoulli distribution* with parameter $p \in [0,1]$ if $\mathbf{Pr}\left[X = 1\right] = p$ and $\mathbf{Pr}\left[X = 0\right] = 1 - p = q$. This distribution has expectation p and variance pq. When $X_1, \ldots, X_n$ is a sequence of independent Bernoulli random variables, the random variable $X = \sum_{i=1}^n X_i$ has the *binomial distribution* with parameters n and p and it is denoted as $\mathrm{BIN}(n,p)$. The binomial distribution has expectation np and variance npq.

A continuous random variable with the *normal distribution* with mean μ and standard deviation σ has density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

and it is denoted as $N(\mu, \sigma^2)$. As a particular case, $\Phi(x)$ will denote the distribution function of $N(0,1)$, that is,

$$\Phi(x) = \mathbf{Pr}[N(0,1) \le x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}u^2)\mathrm{d}u.$$

A sequence of events $(E_i)_{i \ge 1}$ occurs *with high probability* (whp) if $\mathbf{Pr}[E_n] \to 1$ as $n \to \infty$. Some authors use the term whp when the condition $\mathbf{Pr}[E_n] \ge 1 - n^{-c}$ holds for some $c > 1$, and the term *with overwhelming probability* when $\mathbf{Pr}[E_n] \ge 1 - 2^{-cn}$ for some $c > 0$ [18].

Let us finish this review with three classical limit theorems of probability theory (see for example [10, 14]). Recall that, given a sequence of random variables $X_1, X_2, \ldots$ and a random variable X, X_n *converges in distribution* to X if, for all x, $\mathbf{Pr}[X_n < x] \to \mathbf{Pr}[X < x]$ as $n \to \infty$.

Theorem 1.1 (Law of large numbers). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with finite expectation μ. For each $n \ge 1$, set $S_n = \sum_{i=1}^{n} X_i$. Then, S_n/n converges in distribution to μ.*

Theorem 1.2 (Central limit theorem). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with finite expectation μ and finite variance σ^2. For each $n \ge 1$, set $S_n = \sum_{i=1}^{n} X_i$. Then, $(S_n - n\mu)/\sigma\sqrt{n}$ converges in distribution to $N(0,1)$.*

Theorem 1.3 (de Moivre–Laplace limit theorem). *Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed Bernoulli random variables with parameter p. For each $n \ge 1$, set $S_n = \sum_{i=1}^{n} X_i$. Then, for any fixed constants $c_1 < c_2$, as $n \to \infty$,*

$$\mathbf{Pr}[np + c_1\sqrt{npq} \le S_n \le np + c_2\sqrt{npq}] \to \Phi(c_2) - \Phi(c_1),$$

where $q = 1 - p$ and Φ denotes $N(0,1)$.

3. BASIC TAIL BOUNDS FOR A RANDOM VARIABLE

Let X be a random variable and let x be a real value, then $\mathbf{Pr}[X > x]$ and $\mathbf{Pr}[X < x]$ are called the upper and lower tails, respectively. Figure 1.1 illustrates this definition. Concentration bounds give upper bounds for the tails of the random variable $X - \mathbf{E}[X]$. We start by presenting the most basic tail bounds.

Theorem 1.4. *Let $t > 0$ and let X be a random variable, defined on a probability space Ω. Then, for any integer $k > 0$,*

$$\mathbf{Pr}[|X| > t] \le \frac{\mathbf{E}\left[|X|^k\right]}{t^k}.$$

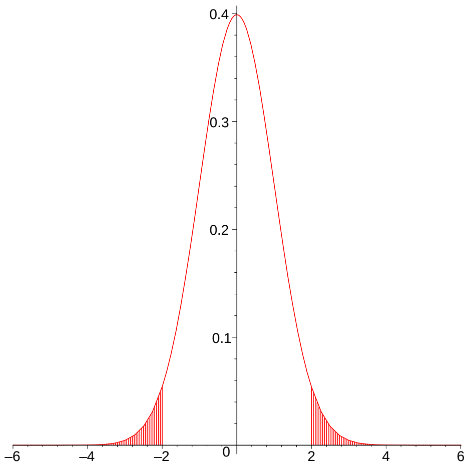

Figure 1.1 Tails of a random variable.

Proof.

$$\begin{aligned}
\mathbf{E}\left[|X|^k\right] &= \sum_{x \in X(\Omega)} |x|^k \mathbf{Pr}[X = x] \\
&= \sum_{|x| \le t} |x|^k \mathbf{Pr}[X = x] + \sum_{|x| > t} |x|^k \mathbf{Pr}[X = x] \\
&\ge \sum_{|x| > t} |x|^k \mathbf{Pr}[X = x] \\
&\ge t^k \sum_{|x| > t} \mathbf{Pr}[X = x] \\
&= t^k \mathbf{Pr}[|X| > t].
\end{aligned}$$

✌

The first classical bound on the deviation from the expected value follows from the previous theorem. It states a well known bound for the upper tail and is usually referred as the *first moment method* or *Markov inequality.*

Theorem 1.5 (Markov inequality). *Let X be a positive random variable. Then, for any $t > 0$,*

$$\mathbf{Pr}[X > t\mathbf{E}[X]] \le 1/t.$$

Proof. Take $k = 1$ and substitute t by $t\mathbf{E}[X]$ in Theorem 1.4. ✌

Markov inequality does not provide information on the other tail. Notice that this inequality cannot be used with the random variable $-X$; even if we would consider the random variable $Y = -X + \max_{\omega\in\Omega} X(\omega)$ (which is a positive random variable), Markov inequality gives no information on the lower tail. On the other hand, Markov inequality becomes vacuous when $t \leq 1$. Markov inequality is mainly an intermediate tool to get better bounds.

The following example is taken from [6].

Example 1.1 (Points into boxes). Let us consider a collection of n points in the unit square ($[0,1]^2$). Divide the unit square in $n/\log^2 n$ squares boxes, each of side $\log n/\sqrt{n}$. For the sake of simplicity, assume $\sqrt{n}/\log n$ is an integer. Given $\epsilon \in (0,1)$, we say that a the collection of points is *ϵ-nice* if each box contains at least $(1-\epsilon)\log^2 n$ points and at most $(1+\epsilon)\log^2 n$ points. We would like to prove that, with high probability, a collection of n points independently and uniformly distributed in the unit square is ϵ-nice.

Let X be a random variable counting the number of points of a fixed box. Then X follows a binomial distribution $\mathrm{BIN}(n,p)$ with $p = \log^2 n/n$. Therefore, $\mu = \mathbf{E}[X] = np = \log^2 n$. According to Markov inequality,

$$\mathbf{Pr}[X > (1+\epsilon)\mu] \leq \frac{1}{1+\epsilon},$$

which is independent of n. With that bound we cannot even say that a single fixed box contains less than $(1+\epsilon)\mu$ points with high probability!

Let us derive another classical bound, attributed to Chebyshev.

Theorem 1.6 (Chebyshov inequality). *Let X be a random variable with expected value μ and standard deviation σ. Then, for all $t > 0$,*

$$\mathbf{Pr}[|X-\mu| > t\sigma] \leq 1/t^2.$$

Proof. In Theorem 1.4, substitute t by $t\sigma$, X by $X - \mathbf{E}[X]$, k by 2 and use the definition of variance.

The use of Chebyshov inequality indicates that with probability greater than $1 - 1/t^2$, a random variable X will fall within t times the standard deviation around μ. For example, 75% of the times a random value falls in the interval $[\mu - 2\sigma, \mu + 2\sigma]$.

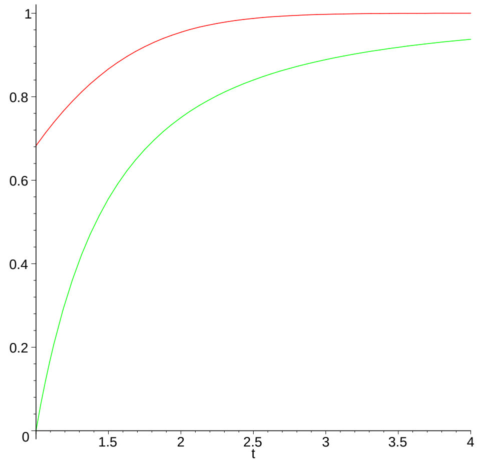

Figure 1.2 Comparing Chevyshov (thin line) and de Moivre–Laplace (thick line) bounds for Example 1.2

⇒ **Example 1.2 (Coin tosses).** Let us study the concentration of the random variable that counts the number of heads after tossing n times a fair coin. Let $X_1, \ldots, X_n$ be the random variables produced by the outcome of each toss, where $X_i =$(the i-th toss is a head). We wish to study the concentration of $S_n = \sum_{i=1}^{n} X_i$. Observe that $\mathbf{E}[S_n] = \frac{1}{2}n$ and $\mathbf{Var}[S_n] = \frac{1}{4}n$.

If we apply Chebyshov inequality, we get

$$\mathbf{Pr}[|S_n - \mathbf{E}[S_n]| > t\sigma] < 1/t^2, \qquad \forall t > 0.$$

If we apply the classical de Moivre–Laplace Theorem, then

$$\mathbf{Pr}[|S_n - \mu| \leq t\sigma] \to \Phi(t) - \Phi(-t), \qquad \forall t > 0.$$

Comparing the two bounds (see Figure 1.2), we can observe that, asymptotically, we get a better concentration bound for S_n around its expectation using the de Moivre–Laplace theorem than with Chebyshov inequality. But while the application of Chebyshov inequality is valid for all n, the de Moivre–Laplace Theorem only applies in the limit. ⇐

⇒ **Example 1.1 (cont).** Recall that X counts the number of points in a fixed box. As X is $\mathrm{BIN}(n,p)$, $\mu = \mathbf{E}[X] = \log^2 n$ and $\sigma^2 = \sigma_X^2 = npq = \log^2 n(1 - \log^2 n/n)$. Using Chebyshov inequality, we have

$$\mathbf{Pr}[|X - \mu| > \epsilon\mu] \leq \frac{\sigma^2}{\epsilon^2\mu^2} = \frac{n - \log^2 n}{\epsilon^2 n \log^2 n},$$

which tends to 0 as $n \to \infty$. Therefore, we have that, with high probability, a fixed box has at least $(1-\epsilon)\log^2 n$ points and at most $(1+\epsilon)\log^2 n$ points.

Unfortunately, we cannot yet affirm that this happens for all boxes: If we use Boole's inequality, we have that the probability that some box has less than $(1-\epsilon)\log^2 n$ points or more than $(1+\epsilon)\log^2 n$ points is less than

$$\frac{n}{\log^2 n} \cdot \frac{n - \log^2 n}{\epsilon^2 n \log^2 n} = \frac{n - \log^2 n}{\epsilon^2 \log^4 n}$$

which diverges when n goes to infinity. $\Leftarrow$

Observe that Chebyshov inequality justifies the fact that if the variance of a random variable goes to zero then, its real value is close to its expectation with high probability. Unfortunately, this happens rarely, and usually, variances are difficult to compute.

4. SUMS OF BOUNDED INDEPENDENT RANDOM VARIABLES

Our purpose in this section is to obtain concentration inequalities for the sum of bounded independent random variables. Assuming that $X_1, \ldots, X_n$ are independent identical distributed Bernoulli random variables, with probability of success p, an application of Chebyshov inequality gives

$$\mathbf{Pr}\left[\sqrt{\frac{n}{p(1-p)}}\left|\frac{1}{n}\sum_{i=1}^{n} X_i - p\right| > t\right] \leq \frac{1}{t^2}.$$

By the central limit theorem, we know that $\sum_{i=1}^{n} X_i$ approaches the normal distribution; so one may expect that

$$\mathbf{Pr}\left[\sqrt{\frac{n}{p(1-p)}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right) > t\right] \longrightarrow 1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}}\frac{e^{-t^2/2}}{t},$$

showing that the bound given by using Chebyshov inequality is far off from what one should see asymptotically, according to the central limit theorem.

The next method is an extension of Markov inequality. In this case, the bounds of the two distribution tails will be a negative exponential function to the number of trials. The bound for independent Bernoulli random variables is due to Chernoff [4], although it seems that the idea and the proof technique was already present in the work of Bernstein [28, Page 161]. Chernoff's bounds have become standard material in most books and surveys on probabilistic techniques for combinatorics and computer science (see [1, 32, 29, 31, 12] among others). In particularly, a good survey on the relations between the most used versions of Chernoff's bounds is presented in [15].

Lemma 1.1. *Let $X_1, \ldots, X_n$ be a sequence of independent random variables corresponding to Bernoulli experiments, each variable with probability p_i of success and probability $q_i = 1 - p_i$ of failure. Let X be the random variable defined by $X = \sum_{i=1}^{n} X_i$. Then, for any $c > 0$ and any $t > 0$,*

$$\mathbf{Pr}[X > t] \leq e^{-ct} \prod_{i=1}^{n} (p_i e^c + q_i) \tag{1.2}$$

and

$$\mathbf{Pr}[X < t] \leq e^{ct} \prod_{i=1}^{n} \left(p_i e^{-c} + q_i\right). \tag{1.3}$$

Proof. The key of the proof is to consider the random variable e^{cX} instead of X. Notice that for $c > 0$, e^{cX} is a positive monotone increasing function of X and therefore $\mathbf{Pr}[X > t] = \mathbf{Pr}[e^{cX} > e^{ct}]$. Using Markov inequality,

$$\mathbf{Pr}[X > t] = \mathbf{Pr}[e^{cX} > e^{ct}] \leq \frac{\mathbf{E}\left[e^{cX}\right]}{e^{ct}}$$

but

$$\mathbf{E}\left[e^{cX}\right] = \mathbf{E}\left[\exp\left(c \sum_{i=1}^{n} X_i\right)\right] = \mathbf{E}\left[\prod_{i=1}^{n} e^{cX_i}\right].$$

As $X_1, \ldots, X_n$ are independent, then $e^{cX_1}, \ldots, e^{cX_n}$ are also independent, and so we have $\mathbf{E}\left[\prod_{i=1}^{n} e^{cX_i}\right] = \prod_{i=1}^{n} \mathbf{E}\left[e^{cX_i}\right]$. This implies that

$$\mathbf{Pr}[X > t] \leq e^{-ct} \prod_{i=1}^{n} \mathbf{E}\left[e^{cX_i}\right].$$

For each $i \in [n]$, with probability p_i we have $e^{cX_i} = e^c$ and with probability q_i we have $e^{cX_i} = 1$. The statement of the lemma follows from the definition of the expected value.

To get the bound on the lower tail, we take into account the equality $\mathbf{Pr}[X < t] = \mathbf{Pr}[-X > -t]$, so we can work as before starting from $\mathbf{Pr}\left[e^{-cX} > e^{-ct}\right]$ to get the bound for the lower tail. ✌

The following result gives some other bounds for the concentration around the expected value of a sum of Bernoulli variables.

Theorem 1.7 (General Chernoff's bounds). *Let $X_1, \ldots, X_n$ be independent random variables corresponding to Bernoulli experiments, each variable with probability p_i of success and probability $q_i = 1 - p_i$ of failure. Let $X = \sum_{i=1}^n X_i$ be a random variable with expected value $\mu = \sum_{i=1}^n p_i$. Then, for any $d > 0$,*

$$\mathbf{Pr}[X > (1+d)\mu] \leq \left(\frac{e^d}{(1+d)^{1+d}}\right)^\mu . \tag{1.4}$$

Moreover, in the case $0 < d < 1$,

$$\mathbf{Pr}[X < (1-d)\mu] \leq \left(\frac{e^{-d}}{(1-d)^{1-d}}\right)^\mu . \tag{1.5}$$

Proof. Let us start with the upper tail. Observe that for every $x > 0$ we have that $\ln(1+x) < x$ and that $1 + x \leq e^x$.

Substitute $t = (1+d)\mu$ and $c = \ln(1+d)$ in Equation (1.2) in Lemma 1.1. Then, we have

$$\begin{aligned}\mathbf{Pr}[X > (1+d)\mu] &\leq e^{-\ln(1+d)(1+d)\mu} \prod_{i=1}^n \mathbf{E}\left[e^{\ln(1+d)X_i}\right] \\ &= (1+d)^{-(1+d)\mu} \prod_{i=1}^n \mathbf{E}\left[(1+d)^{X_i}\right] .\end{aligned}$$

As $\mathbf{E}\left[(1+d)^{X_i}\right] = p_i(1+d) + (1-p_i) = 1 + d\,p_i \leq e^{d\,p_i}$ we get

$$\begin{aligned}\mathbf{Pr}[X > (1+d)\mu] &\leq (1+d)^{-(1+d)\mu} \prod_{i=1}^n e^{dp_i} \\ &\leq \left(\frac{e^d}{(1+d)^{1+d}}\right)^\mu .\end{aligned}$$

For $d < 1$ the value $t = (1-d)\mu$ is positive, so, taking $c = -\ln(1-d)$ and using the same argument, we get the bound on the lower tail. ✌

The previous theorem is the basis to develop customized Chernoff's bounds, which usually are weaker but easier to apply.

Corollary 1.1. *Let $X_1, \ldots, X_n$ be independent random variables corresponding to Bernoulli experiments, each variable with probability p_i of* *success and probability $q_i = 1 - p_i$ of failure. Let $X = \sum_{i=1}^{n} X_i$ be a random variable with expectation $\mu = \sum_{i=1}^{n} p_i$. Then, for any $d \in (0,1)$,*

$$\mathbf{Pr}[X < (1-d)\mu] \le e^{-d^2\mu/2} \quad \textit{and} \quad \mathbf{Pr}[X > (1+d)\mu] \le e^{-d^2\mu/3}.$$

Therefore,

$$\mathbf{Pr}[|X - \mu| > d\mu] \le 2e^{-d^2\mu/3} \tag{1.6}$$

Proof. Using Taylor's expansion around 0 we have

$$(1-d)\ln(1-d) = -d + \tfrac{1}{2}d^2 + O(d^3),$$

which implies $(1-d)^{1-d} > e^{d^2/2-d}$. Substituting in the denominator of (1.5) we get

$$\mathbf{Pr}[X < (1-d)\mu] \le \left(\frac{e^{-d}}{e^{d^2/2-d}}\right)^{\mu} \le e^{-\frac{d^2\mu}{2}}.$$

We claim that for $0 \le d \le 1$ we have $d - (1+d)\ln(1+d) + d^2/3 \le 0$. It is true for $d = 0$ and $d = 1$. To prove it for $d \in (0,1)$, let $f(d) = d - (1+d)\ln(1+d) + d^2/3$. Then, $f'(d) = \frac{3}{2}d - \ln(1+d)$. But as f is continuous, $f(0) = 0$, $f(1) = -0.529$ and the roots of f' are $d = 0$ and $d = 1.144$, f must be decreasing in the real interval $(0,1)$, which together with the fact that $f(0) = 0$ implies $f(d) \le 0$ for $0 \le d \le 1$.

As a consequence, we have $(1+d)\ln(1+d) \ge d + d^2/3$ and therefore, taking exponentials, $(1+d)^{(1+d)} \ge e^{d+d^2/3}$. Substituting in equation (1.4) in Theorem 1.7,

$$\mathbf{Pr}[X > (1+d)\mu] \le \left(\frac{e^{d}}{e^{d+\frac{d^2}{3}}}\right)^{\mu} \le e^{-\frac{d^2\mu}{3}}.$$

Equation (1.6) is obtained using Boole's inequality.

To finish Example 1.1, we will use Corollary 1.1.

⇒ **Example 1.1 (cont).** Recall that X counts the number of points in a fixed box and that $\mu = \mathbf{E}[X] = \log^2 n$. Using Chernoff's bounds, we have

$$\mathbf{Pr}[|X - \mu| > \epsilon\mu] \le 2\exp(-\epsilon^2\mu/3) = 2n^{-\epsilon^2 \log n/3}$$

which tends to 0 as $n \to \infty$. Therefore, we have that, with high probability, a fixed box has at least $(1-\epsilon)\log^2 n$ points and at most $(1+\epsilon)\log^2 n$ points.

Using Boole's inequality, we have now that the probability that some box has less than $(1-\epsilon)\log^2 n$ points or more than $(1+\epsilon)\log^2 n$ points is less than

$$\frac{n}{\log^2 n} 2n^{-\epsilon^2 \log n/3} \leq 2n^{1-\epsilon^2 \log n/3}$$

which tends to 0. As a consequence, for any fixed ϵ, with high probability, a random configuration of points in the unit square is ϵ-nice.

We leave to the reader the generalization of this result when the assumption on the divisibility of n by $\log^2 n$ is removed. This can be done by slightly modifying the sides of the boxes using the floor or the ceil function to completely fit the unit square. ⇐

The next example is taken from [40].

⇒ **Example 1.3 (Mating lemma).** Let L be a sequence of n individuals, where each one can be male (M) or female (F) with probability $\frac{1}{2}$ and independently of the sex of the other members. Let us construct the subsequence S of L in the following way: an individual x of L does not belong to S when it is male and its immediate predecessor is female. We wish to prove that, with overwhelming probability, $|S| < \frac{15}{16}n$.

Let $x \in L$ be an element in an even position and let y be its predecessor in L. Then,

$$\mathbf{Pr}[x \notin S] = \mathbf{Pr}[x = M \wedge y = F] = \mathbf{Pr}[x = M]\mathbf{Pr}[y = F] = \tfrac{1}{4}.$$

Therefore, there are $\frac{1}{2}n$ candidates to be excluded from S with probability $\frac{1}{4}$. Let X be the random variable that counts the number of excluded elements. Then, $\mu = \mathbf{E}[X] \geq \frac{1}{8}n$, and using Corollary 1.1 with $d = \frac{1}{2}$, we get

$$\begin{aligned}\mathbf{Pr}\left[X > \tfrac{1}{16}n\right] &= \mathbf{Pr}\left[X > (1-\tfrac{1}{2})\tfrac{1}{8}n\right] \leq \mathbf{Pr}\left[X > (1-d)\mu\right]\\ &\leq e^{-d^2\mu/2} \leq e^{-n/64}.\end{aligned}$$

This means

$$\mathbf{Pr}[|S| > \tfrac{15}{16}n] > 1 - e^{-n/64},$$

which implies that, with overwhelming probability, $|S| < \frac{15}{16}n$ (see Page 7 for the definition of wop). ⇐

The following result gives further variations on basic Chernoff's bounds when variables are restricted to be identically distributed, that is, to the binomial distribution. Equations (1.8) and (1.9) are due to Hoeffding [17].

Corollary 1.2. *Let X be a random variable following a binomial distribution BIN(n, p) with expectation $\mu = np$. Then,*

$$\begin{aligned}
\mathbf{Pr}[|X-\mu| > t] &\leq 2e^{-\frac{t^2}{3\mu}}, & 0 < t \leq \mu, \quad (1.7)\\
\mathbf{Pr}[|X-\mu| > tn] &\leq 2e^{-2nt^2}, & t \geq 0, \quad (1.8)\\
\mathbf{Pr}[X > an] &\leq \left(\left(\frac{p}{a}\right)^a \left(\frac{1-p}{1-a}\right)^{1-a}\right)^n, & p \leq a < 1. \quad (1.9)
\end{aligned}$$

Proof. Equation (1.7) is obtained by taking $t = d\mu$ in (1.6). To prove (1.8), let $q = 1-p$ and $\delta = n(p+t)$. Observe that from (1.2) we get

$$\mathbf{Pr}[X > \delta] \leq e^{-c\delta} \prod_{i=1}^{n} (pe^c + q) = e^{-c\delta}(pe^c + q)^n.$$

As we wish to make the bound as small as possible, we want to make

$$\frac{\mathrm{d}}{\mathrm{d}c}\left(e^{c\delta}(pe^c+q)^n\right) = 0$$

but as $q = 1-p$ we obtain

$$\begin{aligned}
\delta e^{-c\delta}(pe^c+q)^n + ne^{-c\delta}pe^c(pe^c+q)^{n-1} = 0 &\implies (\delta p + np)e^c + q\delta = 0\\
&\implies e^c = \frac{-\delta q}{\delta p + np}.
\end{aligned}$$

Substituting δ by $n(p+t)$, we get

$$e^c = \frac{n(p+t)(p-1)}{n(p+t)p - np} = \frac{(p+t)(p-1)}{p(p+t-1)},$$

which holds for all values $0 < t < q$. Therefore,

$$\begin{aligned}
\mathbf{Pr}[|X-\mu| > tn] &\leq e^{-n(p+t)\ln\frac{(p+t)(1-p)}{p(p+t-1)}}\left(1 - p + pe^{\frac{(p+t)(1-p)}{p(p+t-1)}}\right)^n\\
&= e^{-n(p+t)\ln\frac{(p+t)(1-p)}{p(p+t-1)}}\left(1 - p + \frac{(p+t)(1-p)}{p+t-1)}\right)^n\\
&= \left(\frac{(p+t)(1-p)}{p(1-p-t)}\right)^{-n(p+t)}\\
&= \exp\left(-(p+t)\ln\frac{p+t}{p} - (1-p+t)\ln\frac{1-p+t}{1-p}\right).
\end{aligned}$$

Set

$$f(t) = -(p+t)\ln\frac{p+t}{p} - (1-p+t)\ln\frac{1-p+t}{1-p}.$$

Deriving we get,

$$f'(t) = \ln\frac{p+t}{p} - \ln\frac{1-p-t}{1-p} \quad \text{and} \quad f''(t) = \frac{1}{(p+t)(1-p-t)}.$$

Observe that $f(0) = f'(0) = 0$ and that $xy \leq \frac{1}{4}$ for any non-negative reals summing up to 1, using Taylor's expansion we get $f(t) \geq 2t^2$, from which Equation (1.8) follows.

To prove (1.9), let us consider a sequence $X_1, \dots, X_n$ of Bernoulli variables with success probability p. Then, $X = \sum_{i=1}^n X_i$. For any given $a \in [p, 1)$, set $d = a/p - 1$ so that $an = (1+d)\mu$. It follows from Lemma 1.1 that for any $c > 0$,

$$\mathbf{Pr}[X > an] \leq e^{-acn} \prod_{i=1}^{n} \mathbf{E}\left[e^{cX_i}\right].$$

As $\mathbf{E}\left[e^{cX_i}\right] = (pe^c + (1-p))$,

$$\mathbf{Pr}[X > an] \leq e^{-acn}(pe^c + (1-p))^n.$$

Let $c = \ln\left(\frac{a(1-p)}{p(1-a)}\right)$; taking logarithms in the previous equation, we get

$$\mathbf{Pr}[X > an] \leq \left(\frac{p(1-a)}{a(1-p)}\right)^{an}\left(\frac{a(1-p)}{1-a} + (1-p)\right)^n,$$

and (1.9) is obtained by simplifying the previous equation.

The next example is taken from [22].

⇒ Example 1.4 (MaxSat). The well known **NP**-hard problem named MAXSAT is defined as follows: Given a set of n literals and a DNF Boolean formula ϕ with m clauses, we wish to find an assignment that satisfies the maximal number of clauses. Let T^* be an optimal assignment satisfying the maximal number of clauses. For every literal x_i in ϕ, define a random assignment $T(x_i)$ where

$$\mathbf{Pr}[T(x_i) = 0] = \mathbf{Pr}[T(x_i) = 1] = \tfrac{1}{2}.$$

We wish to find the value ϵ such that we can state that, with high probability, T coincides with T^* in at least $(1-\epsilon)\frac{1}{2}n$ assignments.

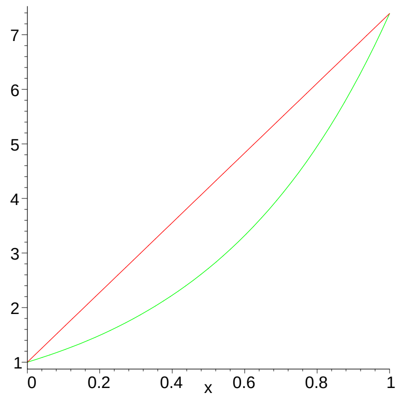

Figure 1.3 A convex function

Let Z be a random variable counting the number of literals x_i such that $T^*(x_i) = T(x_i)$. Define $Z_i = (T^*(x_i) = T(x_i))$. Then, $Z = \sum_{i=1}^n Z_i$ and $\mathbf{E}[Z] = \sum_{i=1}^n \mathbf{E}[Z_i] = \frac{1}{2}n$. By Corollary 1.1,

$$\mathbf{Pr}\left[Z < (1-\epsilon)\tfrac{1}{2}n\right] \leq e^{-\epsilon^2 n/4}.$$

Thus, in order to guarantee that the previous probability goes to zero, we need $\epsilon > \sqrt{b(n)/n}$ where $b(n) \to \infty$ and $b(n)/n \to 0$ as $n \to \infty$. $\dashv$

The inequalities of Corollary 1.2 were extended by Hoeffding to the sum of bounded independent random variables. Let us introduce first a bound on $\mathbf{E}\left[e^{tX}\right]$.

Lemma 1.2. *Let X be a random variable such that $a \leq X \leq b$ with $\mathbf{E}[X] = 0$. Then, for any $t > 0$,*

$$\mathbf{E}\left[e^{tX}\right] \leq exp\left(\frac{(b-a)^2}{8}t^2\right).$$

Proof. By the convexity of the exponential function (see Figure 1.3),

$$e^{tX} \leq \frac{x-a}{b-a}e^{tb} + \frac{x-b}{a-b}e^{ta}$$

for $a \leq x \leq b$. Using linearity of expectation and $\mathbf{E}[X] = 0$, we have

$$\mathbf{E}\left[e^{tX}\right] \leq \frac{-a}{b-a}e^{tb} + \frac{b}{b-a}e^{ta}.$$

Taking $p = -a/(b-a)$ and $1-p = b/(b-a)$, we obtain

$$\begin{aligned}
\mathbf{E}\left[e^{tX}\right] &\leq pe^{tb} + (1-p)e^{ta} \\
&\leq \left(pe^{t(b-a)} + (1-p)\right) e^{ta} \\
&\leq \left(pe^{t(b-a)} + (1-p)\right) e^{-pt(b-a)} \\
&\leq e^{-pt(b-a)+\log\left(1-p+pe^{t(b-a)}\right)}.
\end{aligned}$$

To upper bound the exponent, define $u = t(b-a)$, so that the exponent is the function

$$g(u) = -pu + \log\left(1-p+pe^u\right).$$

Therefore,

$$g'(u) = -p + \frac{p}{p+(1-p)e^{-u}} \quad \text{and} \quad g''(u) = \frac{p(1-p)e^{-u}}{(p+(1-p)e^{-u})^2} \leq \tfrac{1}{4}.$$

Using Taylor's expansions, for some $\theta \in [0, u]$, we have

$$g(u) = g(0) + ug'(0) + \tfrac{1}{2}u^2 g''(\theta) \leq \tfrac{1}{8}u^2 = \tfrac{1}{8}t^2(b-a)^2,$$

which proves the claimed result. ✌

In the case that not all the variables have the same ranges, the following result due to Hoeffding [17] might be useful.

Theorem 1.8 (Hoeffding's inequalities). *Let $X_1, \ldots, X_n$ be independent random variables such that, for all $i \in [n]$, there are values a_i and b_i such that $a_i \leq X_i \leq b_i$. Let $X = \sum_{i=1}^n X_i$ and $\mu = \mathbf{E}[X]$. Then, for any $t \geq 0$,*

$$\mathbf{Pr}[X - \mu > t] \leq exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \quad \textit{and}$$

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

Therefore,

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

Proof. By Markov inequality, and Lemma 1.2, we have that for any $\lambda > 0$,

$$\begin{aligned}
\mathbf{Pr}[X - \mu > t] &\leq e^{-\lambda t}\mathbf{E}\left[e^{\lambda(X-\mu)}\right] \\
&\leq e^{-\lambda t}\prod_{i=1}^n \mathbf{E}\left[e^{\lambda(X_i - \mathbf{E}[X_i])}\right].
\end{aligned}$$

As the random variables are independent, we get

$$\begin{aligned}\mathbf{Pr}[X-\mu>t] &\le e^{-\lambda t}\prod_{i=1}^{n} e^{\lambda^2(b_i-a_i)^2/8}\\ &= e^{-\lambda t}e^{\lambda^2\sum_{i=1}^{n}(b_i-a_i)^2/8}\end{aligned}$$

taking $\lambda = 4t/\sum_{i=1}^{n}(b_i-a_i)$, we get

$$\mathbf{Pr}[X-\mu>t] \le e^{-2t^2/\sum_{i=1}^{n}(b_i-a_i)^2}.$$

The same proof applies for the other tail, and the last inequality is obtained using Boole's inequality.

Example 1.5 (Sum of random numbers). Let $X_1,\ldots,X_n$ be independent random variables, all of them uniformly distributed in the interval $[0,1]$. Let $X=\sum_{i=1}^{n}X_i$. As for all $i\in[n]$, $\mathbf{E}[X_i]=\frac{1}{2}$, then by linearity of the expectation, $\mathbf{E}[X]=\frac{1}{2}n$. We can apply Hoeffding's inequalities to bound the probability that the sum of n random numbers deviates from its expected value taking $a_i=0$ and $b_i=1$ for all $i\in[n]$:

$$\mathbf{Pr}[|X-\mathbf{E}[X]|>t] \le 2e^{-2t^2/\sum_{i=1}^{n}(b_i-a_i)^2} \le 2e^{-2t^2/n}.$$

Observe that this probability goes to zero when $t=b(n)\sqrt{n}$ where $b(n)\to\infty$ as $n\to\infty$.

For a numeric setting, take $n=200$. Then, the probability that $90\le X\le 110$ is greater than $2/e\approx 0.735$.

The following longer example is a classical application of Chernoff's bounds taken from [26].

Example 1.6 (Maximal Independent Set). Recall that given a graph, a set of vertices is a an *independent set* if there are no edges among the elements of the set. A *maximal independent set* is an independent set that is not contained in any other independent set. It is well known that the problem of finding a maximal independent set is solvable in polynomial time (belongs to the class **P**). However, it is not know to be solvable in polylogarithmic parallel time using a polynomial number of process (in the class **NC**).

We present a simple randomized PRAM algorithm, due to Luby, to solve the problem in polylogarithmic time using a polynomial number of processors [26]. The example is self contained, and the reader does not need to know about parallel algorithms (the interested reader can find more about parallel algorithms in [19]).

```
function Luby's-Algorithm (G)
    begin
        I := ∅
        G' := G
        while V' ≠ ∅ do
            S := ∅
            for all v ∈ V' do in parallel
                with probability 1/(2d(v)) do S := S ∪ {v}
            end for
            for all v, u ∈ V' do in parallel
                if d(v) ≤ d(u) then
                    S := S − {v}
                else
                    S := S − {u}
                    I := S ∪ I
                end if
            end for
            G' := induced subgraph of G on V' − (S ∪ N(S))
        end while
        return I
    end
```

Figure 1.4 Luby's algorithm for maximal independent set

Luby's algorithm to find a maximal independent set I is given in Figure 1.4. The basic strategy is iterative, choosing at each iteration, with a small probability, a set S of vertices that are candidates to be in I.

For every vertex pair $v_1, v_2 \in S$ such that $(v_1, v_2) \in E$, discard the vertex with lower degree. Recall that for any $v \in V$, its *degree* is defined as $d(v) = |N(v)|$, where $N(v)$ denotes the set of neighbors of v in G.

Call a vertex $v \in V$ *good* when $\sum_{u \in N(v)} \frac{1}{2d(u)} \geq 1/6$. Let us first prove that in the given graph, if a vertex v is good, then $\mathbf{Pr}[v \in N(I)] \geq \frac{1}{36}$, where $N(I)$ is the set of vertices in V connected to some vertex inside of I.

We break the proof into two parts. Given v good, first we prove it for the case that $u \in N(I)$ has $d(u) \leq 2$ and after for the remaining vertices in $N(v)$ with degree greater than 2.

For any $v \in V$, let $L(v) = \{u \in N(v) \mid d(u) \geq d(v)\}$, then

$$\mathbf{Pr}[v \notin I | v \in S] \leq \mathbf{Pr}[\exists u \in (L(v) \cap S) | v \in S] \leq \sum_{u \in L(v)} \mathbf{Pr}[u \in S | v \in S]$$

$$= \sum_{u \in L(v)} \mathbf{Pr}[u \in S] \leq \sum_{u \in L(v)} \frac{1}{2d(v)} \leq \frac{d(v)}{2d(v)} = \tfrac{1}{2}.$$

Therefore,

$$\mathbf{Pr}[v \in I] \geq \mathbf{Pr}[v \in I \mid v \in S] \mathbf{Pr}[v \in S] \geq \frac{1}{4d(v)},$$

which when there is any $u \in N(v)$ with $d(u) \leq 2$, gives

$$\mathbf{Pr}[v \in N(v)] \geq \mathbf{Pr}[u \in I] \geq \frac{1}{4d(u)} \geq \tfrac{1}{8} > \tfrac{1}{36}.$$

For every $u \in N(v)$ with $d(u) \geq 3$, we have that $\frac{1}{2d(u)} \leq \frac{1}{6}$. But as by hypothesis v is good, then $\sum_{u \in N(v)} \frac{1}{2d(u)} \geq \frac{1}{6}$ which implies that there exists $M(v) \subset N(v)$ such that

$$\tfrac{1}{6} \leq \sum_{u \in M(v)} \frac{1}{2d(u)} \leq \tfrac{1}{3} \tag{1.10}$$

Using inclusion-exclusion we have,

$$\begin{aligned}\mathbf{Pr}[v \in N(I)] &\geq \mathbf{Pr}[\exists u \in (M(v) \cap I)] \\ &\geq \sum_{u \in N(v)} \mathbf{Pr}[u \in I] - \sum_{u \neq w \in M(v)} \mathbf{Pr}[u \in I \cap w \in I]\end{aligned}$$

but as the variables are pairwise independent, on this last term, we get

$$\mathbf{Pr}[v \in N(I)] \geq \sum_{w \in M(v)} \frac{1}{4d(u)} - \sum_{u \neq w \in M(v)} \mathbf{Pr}[u \in S] \mathbf{Pr}[w \in S],$$

using equation (1.10),

$$\mathbf{Pr}[v \in N(I)] \geq \tfrac{1}{36}.$$

Now we can prove that if X is the random variable counting the number of edges deleted at each iteration of the algorithm, then

$$\mathbf{E}[X] \geq \tfrac{1}{72}|E|.$$

Let $B \subseteq E$ be the set of good edges. For every $e \in E$ define the indicator variable $X_e = 1$ if and only if e is deleted. As $X = \sum_e X_e$, and using linearity of the expected value together with the previous lemma and the easy to prove fact that at least half of the edges in a graph are good, we get

$$\begin{aligned}\mathbf{E}[X] &= \sum_{e \in E} \mathbf{E}[X_e] \geq \sum_{e \in B} \mathbf{E}[X_e] \\ &= \sum_{e=(u,v) \in B} \mathbf{Pr}[v \in N(I)] \geq \tfrac{1}{36}|B| \geq \tfrac{1}{72}|E|.\end{aligned}$$

It remains to prove that Luby's algorithm will find a maximal independent set, within $O(\log n)$ iterations, with high probability.

Each iteration is a success if at least $\frac{|E'|}{144}$ edges are removed, where E' is the set of edges at the beginning of the iteration. Let X be the variable counting the number of edges deleted. The by Markov,

$$\mathbf{Pr}[\text{success}] = \mathbf{Pr}\left[X > \tfrac{1}{144}\right] \leq \tfrac{1}{143},$$

which is small but does not tend to 0, so we can not say that happens whp.

Consider $\log n$ iterations and use Chernoff's,

$$\mathbf{Pr}\left[\text{success} > \tfrac{1}{2}\log n \tfrac{1}{144}\right] \leq e^{-\log n/288}$$

that indeed tends to 0 as n grows. ◀

5. THE INDEPENDENT BOUNDED DIFFERENCES METHOD

In this section we present bounds on the tails of a function defined on a product space. The main result, due to McDiarmid [28], requires that the function satisfies the so-called *Lipschitz* or *independent bounded differences* condition. Whereas the canonical way to prove this result is through the use of martingales, we present here a direct proof.

Definition 1.1 (Independent bounded differences condition). *A function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the* independent bounded differences condition *or is said to be* Lipschitz, *if there exists a vector $\mathbf{c} = (c_1, \ldots, c_n)$ such that*

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq c_i$$

for all vectors $\mathbf{x}$ and $\mathbf{x}'$ that differ only at the i-th coordinate ($i \in [n]$).

═ **Example 1.2 (cont).** Consider again the example of counting the number of heads when tossing n times a fair coin. The corresponding probability space is $\Omega = \{0,1\}^n$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector where, for each $i \in [n]$, the probability that the i-th toss is a head is $\mathbf{Pr}[X_i = 1] = \frac{1}{2}$. Define $f : \Omega \to \mathbb{R}$ assigning to any $\mathbf{x} = (x_1, \ldots, x_n)$ the value $f(\mathbf{x}) = \sum_{i=1}^n x_i$, so that f is a random variable counting the number of heads. Then, f is Lipschitz, as if $\mathbf{x}$ and $\mathbf{x}'$ only differ at the i-th coordinate, this means that the number of heads in $\mathbf{x}$ and $\mathbf{x}'$ differs only by one, i.e. $|f(\mathbf{x}) - f(\mathbf{x}')| \leq 1$. ═

Lemma 1.3. *Let $X_1, \ldots, X_n$ be a collection of independent random variables. Assume $f : \mathbb{R}^n \to \mathbb{R}$ is a function that satisfies the independent bounded differences condition with $\mathbf{c} = (c_1, \ldots, c_n)$. Define the random variable Y as*

$$Y = f(X_1, \ldots, X_n) - \mathbf{E}[f(X_1, \ldots, X_n)]$$

and, for any $i \in [n]$, define the random variable Y_i as

$$Y_i = \mathbf{E}[f(X_1, \ldots, X_n)|\mathbf{X}_i] - \mathbf{E}[f(X_1, \ldots, X_n)|\mathbf{X}_{i-1}].$$

Then, $|Y_i| \leq c_i$.

Proof. Let $\mathbf{X} = (X_1, \ldots, X_n)$. According with the definition of conditional expectation, and taking into account the independence,

$$\begin{aligned}
|Y_i| &= |\mathbf{E}[f(\mathbf{X})|\mathbf{X}_i] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}]| \\
&= \Bigg| \sum_{x_{i+1},\ldots,x_n} f(\mathbf{X}_i, x_{i+1}, ..., x_n) \prod_{j=i+1}^{n} \mathbf{Pr}[X_j = x_j] \\
&\qquad - \sum_{x_i,\ldots,x_n} f(\mathbf{X}_{i-1}, x_i, ..., x_n) \prod_{j=i}^{n} \mathbf{Pr}[X_j = x_j] \Bigg| \\
&= \Bigg| \sum_{x_{i+1},\ldots,x_n} \Bigg(\sum_{y_i \neq x_i} f(\mathbf{X}_i, x_{i+1}, ..., x_n) \\
&\qquad - f(\mathbf{X}_{i-1}, x_i, ..., x_n) \Bigg) \prod_{j=i}^{n} \mathbf{Pr}[X_j = x_j] \Bigg| \\
&\leq \sum_{x_i,\ldots,x_n} c_i \prod_{j=i}^{n} \mathbf{Pr}[X_j = x_j] = c_i,
\end{aligned}$$

which proves the result. ∎

The next result is known as the *method of independent bounded differences* [28].

Theorem 1.9 (McDiarmid's bound). *Given a product space defined by* $\Omega = \prod_{i=1}^{n} \Omega_i$*, let* $X_1, \ldots, X_n$ *be independent random variables on* $\Omega_1, \ldots, \Omega_n$*. Assume* $f : \mathbb{R}^n \to \mathbb{R}$ *satisfies the independent bounded differences condition with* $\mathbf{c} = (c_1, \ldots, c_n)$*. Then, for all* $t > 0$*,*

$$\mathbf{Pr}\left[f(X_1, \ldots, X_n) - \mathbf{E}\left[f(X_1, \ldots, X_n)\right] > t\right] < exp\left(-\frac{t^2}{\sum_{i=1}^{n} c_i^2}\right).$$

Proof. Define the random variable $Y = f(X_1, \ldots, X_n) - \mathbf{E}\left[f(X_1, \ldots, X_n)\right]$ and, for any $i \in [n]$, define the random variables $Y_i = \mathbf{E}\left[f(X_1, \ldots, X_n)|\mathbf{X}_i\right]$. Then, $Y = \sum_{i=1}^{n} Y_i$ and furthermore $\mathbf{E}\left[Y_i\right] = 0$. By Lemma 1.3, we have $-c_i \leq Y_i \leq c_i$. By Lemma 1.2 we have

$$\mathbf{E}\left[e^{\lambda Y_i}\right] \leq e^{\lambda^2 (2c_i)^2/8} \leq e^{\lambda^2 c_i^2/2}.$$

Therefore, $\mathbf{Pr}[Y > t] \leq \mathbf{E}\left[e^{\lambda Y}\right]/e^{\lambda t}$. To bound this expression, observe that setting $S_i = \sum_{j=1}^{i} Y_j$ we have $Y = S_n$ and $\lambda S_i = \lambda S_{i-1} + \lambda Y_i$ for any $i \in [n]$. Thus, using the fact that, given $\mathbf{X}_i$, the random variable S_i is totally determined, we get

$$\begin{aligned}
\mathbf{Pr}[Y > t] &\leq \mathbf{E}\left[e^{\lambda S_n}\right]/e^{\lambda t} \\
&\leq \mathbf{E}\left[e^{\lambda S_{n-1}} e^{\lambda Y_n}\right]/e^{\lambda t} \\
&\leq \mathbf{E}\left[e^{\lambda S_{n-1}} \mathbf{E}\left[e^{\lambda Y_n}|X_1, \ldots, X_n\right]\right]/e^{\lambda t} \\
&\;\;\vdots \\
&\leq \prod_{i=1}^{n} \mathbf{E}\left[e^{\lambda Y_i}|\mathbf{X}_i\right]/e^{\lambda t} \\
&\leq e^{\lambda^2 \sum_{i=1}^{n} c_i^2/2}/e^{\lambda t}.
\end{aligned}$$

Taking $\lambda = 1/t$ we have the minimum in the denominator, and thus,

$$\mathbf{Pr}[Y - \mathbf{E}\left[Y\right] \geq t] < e^{-t^2/\sum_{i=1}^{n} c_i^2},$$

which proves the theorem. ∎

Observe that the inequality in the above theorem is only one-sided, but applying it to the function $-f$ we get

$$\mathbf{Pr}\left[f(X_1, \ldots, X_n) - \mathbf{E}\left[f(X_1, \ldots, X_n)\right] < -t\right] < e^{-t^2/\sum_{i=1}^{n} c_i^2},$$

and therefore,

$$\mathbf{Pr}\left[|f(X_1,\ldots,X_n)-\mathbf{E}\left[f(X_1,\ldots,X_n)\right]|<t\right]<2e^{-t^2/\sum_{i=1}^n c_i^2}.$$

In the rest of this section, we present some application examples of the above bounds.

⇒ **Example 1.2 (cont).** Recall that in the example of counting the number of heads in n tosses of a fair coin, f is Lipschitz with $\mathbf{c}=\mathbf{1}$. For any $t>0$, using McDiarmid's bound we get

$$\mathbf{Pr}\left[f(\mathbf{X})-n/2>t\right]\leq e^{-t^2/n}.$$

and using Chernoff's bounds (with $t=d\mu$ in Corollary 1.1) we get

$$\mathbf{Pr}\left[f(\mathbf{X})-n/2>t\right]\leq e^{-\frac{2}{3}t^2/n}.$$

In this case, both bounds are equivalent up to the constants. ⇐

The next example is taken from [8]. See also [1, Theorem 5.1] for a related example.

⇒ **Example 1.7 (Balls and bins).** Consider the following experiment: throw independently and uniformly m balls into n bins. We wish to count the number of empty bins.

In this case, $\Omega=[n]^m$. Let Z be a random variable counting the number of empty bins. Let $X_1,\ldots,X_m$ be a sequence of independent random variables, where each X_i takes values in $[n]$ and indicates the bin where the i-th ball falls. It is well known that

$$\mu=\mathbf{E}\left[Z\right]=n\left(1-\frac{1}{n}\right)^m.$$

Notice that Z is Lipschitz, because if the i-th ball is moved from one bin to another, keeping all other balls in the same position, the number of empty bins can increase or decrease at most by 1. So, applying McDiarmid's bound, we get that, for any $t>0$,

$$\mathbf{Pr}\left[|Z-\mu|>t\right]\leq 2e^{-t^2/m}. \tag{1.11}$$

A concentration bound tending to zero as $m\to\infty$ is obtained when t grows faster than $\sqrt{m}$. ⇐

The next example is taken from [21]

⇒ **Example 1.8 (Bin packing).** The BINPACKING problem is another classical **NP**-hard problem: Given n objects with sizes $x_1, \ldots, x_n$ (such that $0 < x_i < 1$ for all $i \in [n]$) and an unlimited collection of bins, each of size 1, what is the minimal number of bins required to pack the objects? In the randomized version of this problem, we suppose that the sizes of the objects are generated at random (uniformly and independently). Given $X_1, X_2, \ldots,$ we wish to study the random variable $Y_n = f(X_1, \ldots, X_n)$ that represents the minimal number of bins required to pack the first n items. It has been shown that Y_n grows linearly in n, that is, there exists a constant β such that

$$\frac{\mathbf{E}[Y_n]}{n} \to \beta, \qquad \text{as } n \to \infty.$$

Observe that f is Lipschitz with $\mathbf{c} = \mathbf{1}$. From Theorem 1.9, it follows that

$$\mathbf{Pr}[|Y_n - \mathbf{E}[Y_n]| > t] < 2e^{-t^2/n}.$$

As a consequence, with high probability $|Y_n - \mathbf{E}[Y_n]| > b(n)\sqrt{n}$ where $b(n)$ is any function so that $b(n) \to \infty$. ⇐

Let us consider another example the *longest increasing subsequence* problem for a random permutation [36]

⇒ **Example 1.9 (Longest increasing subsequence).** A permutation π of $[n]$ induces a sequence with a reordering of the elements $\{1, \ldots, n\}$, an *increasing subsequence* in π is a subsequence in which each element is smaller than the next. Let L_π denote the length of the longest increasing subsequence in π.

We are interested in the random variable L_π when the permutation π is random with uniform distribution. It is known that, asymptotically, $\mathbf{E}[L_\pi] = 2\sqrt{n}$ [25, 39]. Using the Hoeffding-Azuma inequality, Frieze [13] gave a concentration bound for L_π within $(2\sqrt{n})^{2/3}$ of its expected value. His proof is non trivial. However, changing a bit the setting, we could easily apply the independent bounded differences method. Consider n uniformly and independently chosen real random numbers in $[0, 1]$, $X_1, \ldots, X_n$. As they are uniform random real numbers, with probability 1, all of them would be different, therefore the order on the reals induces a permutation π of $[n]$. Let L_n be the random variable measuring the length of the longest increasing sequence on these reals. Notice that by changing only one number, the length of the longest increasing sequence may be increased or decreased by one, therefore using Theorem 1.9, for any $t \geq 0$

$$\mathbf{Pr}[|L_n - \mathbf{E}[L_n]| > t] \leq e^{-2t^2/n}.$$

Better concentration bounds will be given later. ◁

6. TAIL BOUNDS FOR DEPENDENT VARIABLES

Chernoff bounds and the independent bounded differences method require the independence of the the random variables of interest. In this section, we introduce a powerful tool, *martingale* theory, and use it to derive concentration results, for the case in which the random variables under consideration are not necessarily independent.

The study of martingales is an old topic in probability theory and provides one of the most useful tools to establish limit theorems, because the convergence of martingales is guaranteed. We will restrict ourselves to finite martingale sequences, as we are not interested in convergence results. We shall introduce only the basic definitions and results needed for our purposes. For a more complete introduction to martingales see [14].

The name martingale comes from the old game of martingale: A gambler has a large amount of money, he puts 1 euro in the first bet. While he loses the bet, he doubles the amount in the next bet. The play continues until the gambler wins a bet. Observe that if the gambler arrives to the n-th play, then he bets 2^n euros on the $n+1$-th bet. So, if he wins at the n-th round, there will be a profit of $2^n - \sum_{i=0}^{n-1} 2^i$, that is just 1 euro!

If X_i denotes the amount that the gambler could win or lose at the i-th hand, as at any play round i, the gambler wins X_i with probability 1/2 and loses X_i with the same probability, then $\mathbf{E}[X_n|X_1,\ldots,X_{n-1}] = X_{n-1}$, for any n. The book [14] contains an amusing history about G. Casanova and the game of martingales in 18th century Venice.

Let us start by introducing the concept of martingale.

Definition 1.2. *A sequence of random variables* $Y_0,\ldots,Y_n$ *is a* martingale with respect to the sequence $X_1,\ldots,X_n$, *if for all* $i \in [n]$,

$$\mathbf{E}[Y_i|X_1,\ldots,X_{i-1}] = Y_{i-1}.$$

The canonical example of martingale is the sum of bounded independent random variables with expected value 0.

▷ **Example 1.10 (Independent sums).** Let $X_1,\ldots,X_n$ be a sequence of bounded independent variables with expected value 0. For instance $X_i = 1$ if the toss of a fair coin results in a head and $X_i = -1$ otherwise. For any i, $0 \le i \le n$, define $S_i = \sum_{k=1}^{i} X_k$. Notice that in this example the sum is the function that defines S_n. To prove that

$S_0, \ldots, S_n$ is a martingale with respect to $X_1, \ldots, X_n$, we have to show that for any i, $\mathbf{E}[S_i|X_1, \ldots, X_{i-1}] = S_{i-1}$.

$$\begin{aligned}\mathbf{E}[S_i|X_1, \ldots, X_{i-1}] &= \mathbf{E}[S_{i-1} + X_i|X_1, \ldots, X_{i-1}] \\ &= \mathbf{E}[S_{i-1}|X_1, \ldots, X_{i-1}] + \mathbf{E}[X_i|X_1, \ldots, X_{i-1}].\end{aligned}$$

As X_i is independent from $X_1, \ldots, X_{i-1}$, we get

$$\mathbf{E}[S_i|X_1, \ldots, X_{i-1}] = \mathbf{E}[S_{i-1}|X_1, \ldots, X_{i-1}] + \mathbf{E}[X_i] = S_{i-1} + 0.$$

Often, it will be the case that we have a martingale with respect to itself.

Definition 1.3. *A sequence of random variables* $Y_0, Y_1, \ldots, Y_n$ *is a* martingale *if for all* $i \in [n]$,

$$\mathbf{E}[Y_i|Y_1, \ldots, Y_{i-1}] = Y_{i-1}.$$

Example 1.10 (cont). It is easy to show that the process $S_0, \ldots, S_n$ defined as the prefix sums of a sequence of bounded independent variables with mean 0, is a martingale with respect to itself. In fact the events $\mathbf{X}_i = \mathbf{x}_i$ and $\mathbf{S}_i = \mathbf{s}_i$ are in a one to one correspondence.

Let us end our presentation of martingales with two properties:

Proposition 1.5. *Let* $Y_0, \ldots, Y_n$ *be a martingale with respect to the sequence* $X_1, \ldots, X_n$. *Then, for all* $i \in [n]$

1. $\mathbf{E}[Y_{i+j}|X_1, \ldots, X_i] = Y_i$, *for any* j *such that* $i + j \in [n]$, *and*
2. $\mathbf{E}[Y_i] = \mathbf{E}[Y_0]$.

Proof. By Proposition 1.4

$$\mathbf{E}[Y_{i+j}|X_1, \ldots, X_{i+j}] = \mathbf{E}[\mathbf{E}[Y_{i+j}|X_1, \ldots, X_{i+j}]|X_1, \ldots, X_i]$$

by definition of martingale we get

$$\mathbf{E}[Y_{i+j}|X_1, \ldots, X_{i+j}] = \mathbf{E}[Y_{i+j-1}|X_1, \ldots, X_i].$$

An iteration of the above argument gives the first claim.

Recall the $\mathbf{E}[Y_i] = \mathbf{E}[\mathbf{E}[Y_i|X_1, \ldots, X_j]]$ for any $j < i$, therefore, using the first part, we get $\mathbf{E}[Y_i] = \mathbf{E}[\mathbf{E}[Y_i|X_1]] = \mathbf{E}[\mathbf{E}[Y_1|X_1]] = \mathbf{E}[Y_0]$.

Another useful form of a martingale is presenting the process as a sequence of differences.

Definition 1.4. *A process* $Y_1, \ldots, Y_n$ *is a* martingale differences sequence *with respect to* $X_1, \ldots, X_n$ *if for every* $i \in [n]$*, the random variable* Y_i *verifies* $\mathbf{E}[Y_i | X_1, \ldots, X_{i-1}] = 0$.

The name of differences comes from the fact that any martingale sequence $Y_0, \ldots, Y_n$ with respect to a sequence $X_1, \ldots, X_n$ provides a natural way to define a sequence of differences, $Z_1, Z_2, \ldots, Z_n$, by setting $Z_i = Y_i - Y_{i-1}$ for any $i \in [n]$. Indeed, the sequence $Z_1, \ldots, Z_n$ is a martingale differences:

$$\mathbf{E}[Z_i | X_1, \ldots, X_{i-1}] = \mathbf{E}[Y_i - Y_{i-1} | X_1, \ldots, X_{i-1}]$$

by linearity of expectation, the definition of martingale and Proposition 1.4

$$\begin{aligned} \mathbf{E}[Z_i | X_1, \ldots, X_{i-1}] &= \mathbf{E}[Y_i | X_1, \ldots, X_{i-1}] - \mathbf{E}[Y_{i-1} | X_1, \ldots, X_{i-1}] \\ &= Y_{i-1} - Y_{i-1} = 0. \end{aligned}$$

We will refer to such construction as the *martingale differences sequence*.

Conversely, from any given random variable X_0 and a martingale differences sequence $Y_1, \ldots, Y_n$, we obtain a martingale by setting $X_i = X_0 + \sum_{k=1}^{i} Y_k$. Furthermore, observe that any sequence of independent random variables with mean expected value zero is a martingale differences sequence, because $\mathbf{E}[X_{i+1} | X_i, \ldots, X_1] = \mathbf{E}[X_{i+1}] = 0$.

6.1 EXPOSURE MARTINGALE

Our usual setting will be a random structure (a graph, a tree, some balls and bins experiment, etc.) on which we have defined a measure. Often the structure can be described by an incremental process, in such a way that at each step some portion of the structure is revealed. The conditional expectation given the so far exposed parts allows the definition of a random process, that is usually referred to as the *Doob's process*. The formal setting to introduce this process is when studying a random variable on a product space.

Definition 1.5 (Doob's process). *Let* $X_1, \ldots, X_n$ *be a sequence of random variables, where, for any* $i \in [n]$*,* X_i *is defined on the probability space* $(\Omega_i, \mathcal{A}_i, \mathbf{Pr}_i)$*. Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a function. The process* $Y_0, \ldots, Y_n$ *defined as*

$$Y_i = \mathbf{E}[f(\mathbf{X}) | X_1, \ldots, X_i], \qquad \text{for } 0 \leq i \leq n \tag{1.12}$$

is called the Doob's process.

Observe that for any i, $0 \leq i \leq n$, Y_i is a random variable on the product space $\Omega = \prod_{i=1}^{n} \Omega_i$.

Using Proposition 1.4 we have that

$$\mathbf{E}\left[Y_i|\mathbf{X}_{i-1}\right] = \mathbf{E}\left[\mathbf{E}\left[f(\mathbf{X})|\mathbf{X}_i\right]|\mathbf{X}_{i-1}\right] = \mathbf{E}\left[f(\mathbf{X})|\mathbf{X}_{i-1}\right] = Y_{i-1}.$$

Therefore the Doob's process $Y_0, \ldots, Y_n$ is indeed a martingale with respect to $X_1, \ldots, X_n$. We will refer to such construction as a *Doob's martingale.* Each Y_i measures the expected value of $f(\mathbf{X})$ after discovering the first i portions of the structure.

Thus, Doob's construction is useful to define a martingale from any function of a combinatorial object, provided that the random construction of the object can be performed incrementally.

⇒ **Example 1.11 (Three tosses).** Consider the experiment of tossing three times a fair coin. Assume that we are interested in knowing the number of heads after the three tosses.

To analyze this random variable, we will define a Doob's martingale. To do so we first need to describe the sample space as a product space. In this particular case, the task is easy, as we can describe the outcome of the experiment using for each $i \in [3]$ the indicator variables X_i of the event: a head in the i-th toss. For instance, the event $\mathbf{X} = (0, 1, 0)$ indicates tail, head, tail.

Let f be the sum function; then $f(X_1, X_2, X_3)$ counts the number of heads. The Doob's process Y_0, Y_1, Y_2, Y_3 corresponds to follow the experiment stopping after each toss, to compute the expected number of heads, given that we already known the outcome of this and the previous tosses. The Doob's martingale sequence is the following:

$$\begin{aligned}
Y_0(\omega) &= \mathbf{E}\left[f(\mathbf{X})\right] = 3/2,\\
Y_1(\omega) &= \mathbf{E}\left[f(\mathbf{X})|X_1\right](\omega) = \begin{cases} 2 & X_1(\omega) = 1 \\ 1 & X_1(\omega) = 0 \end{cases}\\
Y_2(\omega) &= \mathbf{E}\left[f(\mathbf{X})|X_1, X_2\right](\omega) = \begin{cases} 5/2 & X_1(\omega) = 1, X_2(\omega) = 1 \\ 3/2 & X_1(\omega) = 1, X_2(\omega) = 0 \\ 3/2 & X_1(\omega) = 0, X_2(\omega) = 1 \\ 1/2 & X_1(\omega) = 0, X_2(\omega) = 0 \end{cases}\\
Y_3(\omega) &= \mathbf{E}\left[Y_3|X_1, X_2, X_3\right](\omega) = f(X_1(\omega), X_2(\omega), X_3(\omega)).
\end{aligned}$$

Notice that the last random variable in the sequence corresponds to the function in whose analysis we are interested, and the first one (Y_0) to its expected value. ⇐

⇒ **Example 1.7 (cont).** The random variable X that counts the number of empty bins, after throwing independently and uniformly at random m balls into n bins, can be described incrementally by the process of throwing the balls one after the other. Throwing a ball can be represented by a random variable X_i uniformly distributed on $[n]$ where $X_i = b_i$ denotes the event: "the i-th ball falls into bin b_i", for any $i \in [n]$. From this representation we can consider the Doob's martingale $Y_0, \ldots, Y_n$ where

$$Y_i = \mathbf{E}\left[X | X_1, \ldots X_i\right],$$

that will allow us to study concentration of the random variable X, by analyzing the incremental process. ⇐

Random graphs provide a natural setting to define exposure martingales. In fact there are two usual ways to grow up a graph, by discovering edges or vertices.

⇒ **Example 1.12 (Edge exposure).** Let $G = (V, E)$ be a random graph in the $G_{n,p}$ model, with vertex set $V = [n]$. Recall that in the $G_{n,p}$ a graph with n vertices is obtained by the independent selection of edges. Each edge appears with probability p.

Taking $m = \binom{n}{2}$, and labeling the edges with numbers in $[m]$, any graph G with n vertices can be represented by an m-tuple $(a_1, a_2, \ldots, a_m)$ where $a_j = 1$ if and only if the edge j appears in the graph. The sequence of variables $X_1, \ldots, X_m$, where for each $j \in [m]$, X_j is the indicator variable of the event: "the edge j appears in the graph", is the basis to define the associated Doob's martingale for any graph theoretical function f. In this case each random variable X_j, is a Bernoulli variable with probability of success p.

The *Doob's edge exposure martingale* for a given graph function f, is defined as usual by $Y_j = E[f(G)|\mathbf{X}_j]$, $0 \leq j \leq m$. Observe that the random variables Y_j's measure the expected value of $f(G)$, when the existence of the first i edges of the graph is known. ⇐

⇒ **Example 1.13 (Vertex exposure).** Let $G = (V, E)$ be a random graph in the $G_{n,p}$ model. We now want to consider a way of generating the graph by exposing its vertices. Starting from 1, add a new vertex i and the edges that connect i with the previous $i - 1$ vertices. When adding vertex i, the set of possible end-points of its edges is $[i - 1]$, this can be represented by a string of $i - 1$ bits. For each $i \in [n]$ define the random variable X_i uniformly distributed on the set $B_i = \{0, 1\}^{i-1}$ of numbers represented with $i-1$ bits. So, any graph G with n vertices will be specified by a sequence of bit strings, one for each addition of a vertex.

Given any graph theoretical function f, we can consider the *Doob's vertex exposure martingale* by $Y_i = E[f(G)|\mathbf{X}_i]$. In this case the random variable Y_i, gives the expected value of $f(G)$, when the first i vertices and their corresponding internal edges are known. It is not difficult to see, that with an appropriate ordering of the set of possible edges, the vertex exposure martingale is a subsequence of the edge exposure martingale. $\Leftarrow$

The following construction was used in [7] to show that the *monotone circuit value problem* is in Average **NC**.

$\Rightarrow$ **Example 1.14 (Circuit's representation depth).** Consider the following process to generate an unlabeled binary random circuit. Gate 1 holds a 0 and gate 2 holds a 1, these are the input gates. At step i ($i > 2$), select two nodes from $[i-1]$ as inputs to gate i. The depth of a gate g is the length of the longest directed path from g to the inputs. We are interested in analyzing the depth of the circuit, that is the maximum of the depths of its gates. This measure corresponds to the time needed to evaluate the circuit in presence of parallelism.

The incremental process suggests the use of a vertex exposure filter and the definition of the corresponding martingale. $\Leftarrow$

6.2 THE BOUNDED DIFFERENCES CONDITION

For the case of independent random variables, we asked for the Lipschitz condition. In the case of dependent random variables, that form a martingale, in order to get concentration results, we need some sort of guarantee of bounded increments. We will only need that the process does not make big jumps in one step.

Definition 1.6 (Bounded differences condition). *A martingale sequence $Y_0, \ldots, Y_n$ is said to satisfy the* bounded differences condition *if for any $i \in [n]$ there exists a vector $\mathbf{c} = (c_1, \ldots, c_n)$ such that*

$$|Y_i - Y_{i-1}| \leq c_i.$$

The next theorem says that when the bounded differences condition holds, the process itself does not go far away from its starting point. The theorem is known as Hoeffding-Azuma's inequality and has appeared in different forms in papers by Hoeffding [17], Azuma [2], Steiger [35], Freedman [11] and others.

Theorem 1.10 (Hoeffding-Azuma's inequality.). *Let* $Y_0, \dots, Y_n$ *be a martingale with respect to* $X_1, \dots, X_n$ *that satisfies the bounded differences condition with* $\mathbf{c} = (c_1, \dots, c_n)$*. Then, for any* $t > 0$*,*

$$\mathbf{Pr}\left[|Y_n - Y_0| > t\right] \le 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n c_i^2}\right).$$

Proof. Let us prove the upper bound $\mathbf{Pr}[Y_n > Y_0 + t] \le 2\exp(-\frac{t^2}{2c})$. Without loss of generality we can assume $Y_0 = 0$. Define the martingale differences $D_i = Y_i - Y_{i-1}$, so $Y_n = Y_{n-1} + D_n$. Using Markov inequality we get that, for all $\lambda > 0$, $\mathbf{Pr}[Y_n > t] \le \mathbf{E}\left[e^{\lambda Y_n}\right]/e^{\lambda t}$. We wish to find the value of λ that minimizes this ratio.

$$\mathbf{E}\left[e^{\lambda Y_n}\right] = \mathbf{E}\left[e^{\lambda(Y_{n-1}+D_n)}\right] = \mathbf{E}\left[\mathbf{E}\left[e^{\lambda(Y_{n-1}+D_n)}|X_0, \dots, X_{n-1}\right]\right]$$

but as Y_{n-1} is a function of $X_0, \dots, X_{n-1}$ using Proposition 1.4 we get

$$\mathbf{E}\left[e^{\lambda Y_n}\right] = \mathbf{E}\left[e^{\lambda Y_{n-1}}\mathbf{E}\left[e^{\lambda D_n}|X_0, \dots X_{n-1}\right]\right]. \tag{1.13}$$

Now, $\mathbf{E}\left[\mathbf{E}\left[D_n|X_0, \dots X_{n-1}\right]\right] = 0$ and $|\mathbf{E}\left[D_n|X_0, \dots X_{n-1}\right]| \le c_n$. Applying Lemma 1.2 to $\mathbf{E}\left[D_n|X_0, \dots X_{n-1}\right]$ we get,

$$\mathbf{E}\left[e^{\lambda D_n}|X_0, \dots X_{n-1}\right] \le e^{\lambda^2 c_n^2/2}.$$

Substituting this expression and iterating we get

$$\mathbf{E}\left[e^{\lambda Y_n}\right] \le ne^{\lambda^2 c/2},$$

where $c = \sum_{i=1}^n c_i^2$. But $\frac{e^{\lambda^2 c/2}}{e^{\lambda t}}$ attains the minimum at $\lambda = t/c$ and the proof of the upper tail follows.

In order to get the proof of the lower tail bound, just consider the symmetrical martingale $-Y_0, \dots, -Y_n$. ✌

The main point of the Hoeffding-Azuma's inequality with respect to the bounds in the previous section are the following: First, we don't need to make any assumption on the independence of the random variables. Second, we stress that the c_i in the bounded differences condition may depend on i.

A theorem that is often refereed as Azuma's inequality is the particular case of Theorem 1.10, when the martingale is defined with respect to itself. See for example [29, Theorem 3.10] or [32, Theoremm 4.16]. Observe that in the case of the Doob's Martingale, we have $Y_0 = \mathbf{E}[f(\mathbf{X})]$, so Theorem 1.10, gives a concentration around the mean result.

An useful corollary to Hoeffding-Azuma's inequality arises when there is a value c for which $c_i = c$ for any $i \in [n]$ [21]:

Corollary 1.3. *Let $Y_0, \ldots, Y_n$ be a martingale satisfying the bounded differences condition such that for a vector $\mathbf{c} = (c, \ldots, c)$. Then for any $t > 0$*

$$\mathbf{Pr}\left[|Y_n - Y_0| \geq ct\sqrt{n}\right] \leq 2\exp\left(-\tfrac{1}{2}t^2\right).$$

⇒ **Example 1.2 (cont).** We can use the previous inequality to study the concentration of the random variable that counts the number of heads after tossing a fair coin n times. Let $Y_0, \ldots, Y_n$, be the corresponding Doob's martingale with respect to $X_1, \ldots, X_n$, the indicator variables that record the outcome of each toss. Changing the value of X_i can affect the difference $Y_i - Y_{i-1}$ by at most by 1. Therefore $|Y_i - Y_{i-1}| \leq 1$ and we can apply the previous corollary with $c = 1$ and $Y_0 = n/2$, to get

$$\mathbf{Pr}[|Y_n - \mu| \geq t\sqrt{n}] \leq 2e^{-\frac{t^2}{2}}. \tag{1.14}$$

To compare with Chernoff and the bound obtained from the independent bounded condition, take $t = \frac{d\sqrt{n}}{2}$ so, we get,

$$\mathbf{Pr}\left[|S_n - \mathbf{E}[S_n]| \geq dn/2\right] \leq \begin{cases} 2\exp(-\frac{1}{8}nd^2) & \text{Hoeffding-Azuma's inequality} \\ 2\exp(-\frac{2}{3}nd^2) & \text{Chernoff inequality} \\ 2\exp(-\frac{1}{4}nd^2) & \text{independent bounded differences} \end{cases}$$

The three bounds give the same behavior, but with different constants in the exponent. The better bound in this case is obtained with Chernoff inequality. ⇐

In general, for binomial distribution, the appropriate use of Chernoff bounds will yield better concentration bounds than Hoeffding-Azuma's inequality. On the other hand, Hoeffding-Azuma's does not require independence.

As in most situations we will have a set of random variables and a function of them, we state the bounded differences condition in such terms.

Definition 1.7. *A sequence of random variable* $X_1, \ldots, X_n$ *and a function* $f : \mathbb{R}^n \to \mathbb{R}$ *satisfy the* bounded differences condition *if, for any* $i \in [n]$, *there is a vector* $\mathbf{c} = (c_1, \ldots, c_n)$ *such that*

$$|\mathbf{E}[f(\mathbf{X})|\mathbf{X}_i] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}]| < c_i.$$

This leads the Hoeffding-Azuma's inequality given in Theorem 1.10, for the case of the Doob's martingale.

Corollary 1.4. *Let* $X_1, \ldots, X_n$ *be a sequence of random variables and let* $f : \mathbb{R}^n \to \mathbb{R}$ *be a function that satisfies the bounded differences condition for* $\mathbf{c} = (c_1, \ldots, c_n)$. *Then, for any* $t > 0$,

$$\mathbf{Pr}\left[|f(\mathbf{X}) - \mathbf{E}[f(\mathbf{X})]| > t| \leq 2exp\left(-\frac{t^2}{2c}\right)\right]$$

where $c = \sum_{i=1}^{n} c_i^2$.

⇒ **Example 1.7 (cont).** The Doob's martingale for the number of empty bins, after throwing m balls into n bins, verifies the bounded differences condition with all $c_i = 1$, because by throwing an additional ball we may only fill an empty bin. It is well known that $\mathbf{E}[X] = ne^{-m/n}$, therefore,

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

Letting t growing to ∞ a bit faster than $\sqrt{m}$ we get a sharp concentration result. ⇐

Doob's construction for functions that satisfy the bounded differences condition give directly concentration results. However, it does not give any information about the value of the expected value. The following example is from [33].

⇒ **Example 1.15 (Chromatic number).** Given a graph G its chromatic number $\chi(G)$ is the minimum number of colors that are needed to color the vertices of G, in such a way that no two adjacent vertices have the same color.

Let us consider the $G_{n,p}$ model, according to Example 1.13, for any graph G, we can define a Doob's vertex exposure martingale for the function $\chi(G)$. Observe that a single additional vertex can increase only by one the number of colors needed to color the graph exposed so far. Therefore, the Doob's vertex exposure martingale for chromatic number satisfies the bounded differences condition, with $c_i = 1$ for any $i \in [n]$.

Therefore taking $t = d\sqrt{n}$, we can apply Corollary 1.4 to say:

$$\mathbf{Pr}\left[|\chi(G) - \mathbf{E}\left[\chi(G)\right]| > d\sqrt{n}\right] \leq 2e^{-d^2/2}$$

Letting d go arbitrary slowly to ∞ we get a sharp concentration result. However there is no no hint on the value of $\mathbf{E}\left[\chi(G)\right]$. ⇐

Sometimes, it is necessary to give a clever definition of function to guarantee the bounded differences condition. The following example is from [1]

⇒ **Example 1.16 (Clique number).** Given a graph $G = (V, E)$ its clique number $w(G)$ is the size of the maximum clique in G. Notice that by the addition of a single edge or vertex, we can join many vertices, increasing in more than a constant the size of the biggest clique. Therefore, the vertex or edge exposure martinagale for the clique number will not show concentration.

Let $f(G)$ be the maximal size of a family of edge disjoint cliques of size k in G. Let us consider the $G \in G_{n,p}$ model, according to Example 1.12 define the Doob's edge exposure martingale for f. Adding a single edge can only add at most one clique to a family of edge disjoint cliques. Therefore Corollary 1.4 gives that the clique number is concentrated around its expected value.

When $G \in G_{n,1/2}$ it can be shown, that

$$\mathbf{E}\left[f(G)\right] \geq \frac{n^2}{2k^4}(1 + o(1)),$$

and combining this fact with the concentration result we get,

$$\mathbf{Pr}\left[w(G) < k\right] < \exp\left(-(d + o(1))\frac{n^2}{\ln^8 n}\right)$$

for a positive constant d. This is the basic step used by Bollobás [3] to show that $\chi(G)$ is almost surely $n/2\log n$ where G is a random graph with n nodes and $p = 1/2$. ⇐

⇒ **Example 1.14 (cont).** In the greedy process of constructing a representation of a circuit, adding a new gate may only increase the depth of the circuit by one, in the worst case. Therefore we can obtain tight concentration results. Again, Hoeffding-Azuma inequality does not give any information on the expected value. It is known that the expected depth of the representation of a circuit is a polylogarithmic function of the number of gates [7, 38]. ⇐

It is possible to consider an intermediate notion of bounded condition, that falls between the independent bounded differences condition, and the bounded differences condition.

Definition 1.8. *A sequence of random variables $X_1, \ldots, X_n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$ satisfy the the* average bounded differences condition *if, for any $i \in [n]$, there exists a vector $\mathbf{c} = (c_1, \ldots, c_n)$ such that for any $a_i, b_i \in X_i(\Omega_i)$*

$$|\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = a_i] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = b_i]| < c_i.$$

The average bounded differences condition implies the bounded differences condition. Therefore, as a corollary of Theorem 1.10 we get,

Corollary 1.5. *Let $X_1, \ldots, X_n$ be a sequence of random variables and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that satisfy the average bounded differences condition for $\mathbf{c} = (c_1, \ldots, c_n)$, then for any $t > 0$,*

$$\mathbf{Pr}[|f(\mathbf{X}) - \mathbf{E}[f(\mathbf{X})]| > t] \leq 2exp\left(-\frac{t^2}{2c}\right),$$

where $c = \sum_{i=1}^n c_i^2$

Proof. Let us show that the average bounded differences condition implies the bounded differences condition, with the same constants. Assume that for any $i \in [n]$ suitable c_i and any $a_i, b_i \in X_i(\Omega_i)$ we have

$$|\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = a_i] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = b_i]| \leq c_i$$

Then we can express

$$\begin{aligned}
\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}] &= \sum_{a \in X_i(\Omega_i)} \mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = a]\, \mathbf{Pr}[X_i = a|X_{i-1}] \\
\mathbf{E}[f(\mathbf{X})|\mathbf{X}_i] &= \sum_{a \in X_i(\Omega_i)} \mathbf{E}[f(\mathbf{X})|\mathbf{X}_i]\, \mathbf{Pr}[X_i = a|X_{i-1}]
\end{aligned}$$

Therefore

$$\begin{aligned}
&|\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_i]| \\
&\quad = \left|\sum_a (\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = a] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_i])\mathbf{Pr}[X_i = a|X_{i-1}]\right| \\
&\quad \leq \sum_a |\mathbf{E}[f(\mathbf{X})|\mathbf{X}_{i-1}, X_i = a] - \mathbf{E}[f(\mathbf{X})|\mathbf{X}_i]|\, \mathbf{Pr}[X_i = a|X_{i-1}] \\
&\quad \leq \sum_a c_i \mathbf{Pr}[X_i = a|X_{i-1}] = c_i,
\end{aligned}$$

which proves the statement.

The previous corollary is refereed to as *the method of bounded average differences*. Notice that the real improvement is not in the application of Hoeffding-Azuma's inequality but in the fact the value c in the average bounded differences method is usually better that the corresponding value in the bounded differences method.

⇒ **Example 1.7 (cont).** To apply the average bounded differences condition, to the function that counts the number of empty bins, we will follow [8]. In order to use the method of bounded average differences, we need to compute, for any $i \in [m]$,

$$c_i = \left|\mathbf{E}\left[Z|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b_i\right] - \mathbf{E}\left[Z|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b'_i\right]\right|,$$

for fixed $b_1, \ldots, b_{i-1}, b_i, b'_i \in [n]$ and $b_i \neq b'_i$. The variable Z can be decomposed as the sum of indicator random variables, that is for $i \in [n]$ set Z_i to be 1 when the bin i is empty and 0 when it is not.

Let $S = \{b_1, \ldots, b_{i-1}\}$. As $Z = \sum_{j\in[n]} Z_j$, by linearity of the expectation, the computation of c_i boils down to compute for each $j \in [n]$,

$$c_{ij} = \left|\mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b_i\right] - \mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b'_i\right]\right|.$$

so $c_i = \sum_j c_{ij}$. Let us consider some cases:

Case 1: For $j \in S$ then

$$\mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b_i\right] = \mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b'_i\right],$$

so $c_{ij} = 0$.

Case 2: For $j \notin S \cup \{b'_i, b_i\}$

$$\mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b_i\right] \mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b'_i\right] = (1 - 1/n)^{m-i}.$$

Therefore, again $c_{ij} = 0$.

Case 3: For $j = b_i \notin S$, and $b'_i \notin S$ then

$$\mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b_i\right] = 0$$

and

$$\mathbf{E}\left[Z_j|\mathbf{X}_{i-1} = \mathbf{b}_{i-1}, X_i = b'_i\right] = (1 - 1/n)^{m-i},$$

and in this case $c_{ij} = (1 - 1/n)^{m-i}$.

Putting all together, we get $c_i = \sum_{j,i} c_{ij} \leq (1 - 1/n)^{m-i}$, so that

$$c = \sum_{i\in[m]} c_i^2 \leq \frac{1 - (1/n)^{2m}}{1 - (1 - 1/n)^2} = \frac{n^2 - (n(1/n)^m)^2}{2n - 1} = \frac{n^2 - \mu^2}{2n - 1}.$$

Thus taking $\mu = \mathbf{E}[Z] = ne^{-m/n}$, from Corollary 1.4 we get,

$$\mathbf{Pr}[|Z - \mu| \geq t] = \mathbf{Pr}[|Z - \mu| \geq t] \leq 2\exp\left(-\frac{t^2(2n-1)}{2(n^2 - \mu^2)}\right).$$

Recall that using the method of bounded differences we got

$$\mathbf{Pr}[|Z - \mu| > t] \leq 2\exp\left(\frac{-t^2}{2m}\right).$$

As indicated in [8], if we take $t = d\mu$ we get the same bounds that in [20]. $\Leftarrow$

7. ISOPERIMETRIC INEQUALITIES

Isoperimetric inequalities have become an important topic in graph theory. Moreover, some of the techniques used to show isoperimetric inequalities provide a nice example for the application of the independent bounded differences method. We start this section by presenting a brief introduction to isoperimetry. After we move to present Talagrand's isoperimetric inequality and show some of its applicability. For the interested reader, we recommend [23, 24] and [28, Sec. 7].

In graph theoretical terms, given a graph, an isoperimetric inequality of a subgraph is an inequality relating the size of a subgraph with the size of its boundary. To formalize and generalize the above intuition, let (Ω, d) be a metric space with distance d, together with a measure ν. Recall that for any subset $A \subseteq \Omega$ the distance of a point $\mathbf{x}$ to A is defined as

$$d(\mathbf{x}, A) = \inf\{d(\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in A)\},$$

and the distance between two subsets A and B of Ω is

$$d(A, B) = \inf\{d(\mathbf{x}, B) \mid \mathbf{x} \in A\}.$$

For a given $A \subseteq \Omega$ and any real $t > 0$, define the *t-neighborhood* A_t of A as

$$A_t = \{\mathbf{x} \in \Omega \mid d(\mathbf{x}, A) \leq t\}.$$

Such a subset A_t is also known in the literature as *t-fattening* [28] or *t-boundary* [23]. Notice that $A \subseteq A_t$. An *isoperimetric inequality* is a lower bound of the size $\nu(A_t)$ depending on $\nu(A)$ and t. This relationship is usually presented as an upper bound on the expression $\nu(A)(\nu(\Omega) - \nu(A_t))$.

Considering again graphs, let $G = (V, E)$ be a finite graph; for any $u, v \in V$, let $d(u, v)$ be the shortest distance between u and v in G, and for any $A \subseteq V$, let the measure $\nu(A)$ be the number of vertices in A. For any A, the set A_t consists of the vertices that can be joined to a vertex in A by a path of length at most t. An isoperimetric inequality would be an inequality of the form $|A_t| \geq f(|A|, t)$, where we would like to have the value of the function $f(|A|, t)$ as close to the actual value of $|A_t|$ as possible. Recall that the *boundary* ∂ of a subgraph are the set of vertices connected by an edge to some vertex in A. Therefore, in this particular example, $\partial(A) = A_1 \setminus A$.

Let us begin by considering the case of isoperimetric inequalities in a product space, where the distance is the Hamming distance and the probability of a set gives its measure. Formally, let $\Omega_1, \ldots, \Omega_n$ be probability spaces, and let $\Omega = \prod_{i=1}^{n} \Omega_i$ be the corresponding product space. The Hamming distance between any two points $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ in Ω, denoted by $d_H(\mathbf{x}, \mathbf{y})$, is the number of indices i such that $x_i \neq y_i$. Given $\mathbf{x}$ and $\mathbf{y}$ as before, for each $i \in [n]$, define the function $\delta_i(\mathbf{x}, \mathbf{y}) = 1$ if $x_i \neq y_i$ and $\delta_i(\mathbf{x}, \mathbf{y}) = 0$ otherwise. Thus we can express the Hamming distance as

$$d_H(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \delta_i(\mathbf{x}, \mathbf{y}).$$

The t-neighborhood of a set A under the Hamming distance will be denoted by A_t^H. For any $A \in \Omega$ consider the measure $\nu(A) = \mathbf{Pr}[\mathbf{X} \in A]$. In this case, the independent bounded differences inequality can be used to show that when the probability that a point belongs to a set A is not to small, then with probability tending to 1, any random point is close to A. The following isoperimetric inequality is from McDiamird [29].

Theorem 1.11 (McDiamird). *Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sequence of random variables, each X_i defined on Ω_i. Then for any $t > 0$,*

$$\mathbf{Pr}[\mathbf{X} \in A] \left(1 - \mathbf{Pr}\left[\mathbf{X} \in A_t^H\right]\right) \leq e^{-t^2/4n}.$$

Proof. Notice that the function $d_H(\mathbf{X}, A)$ is a random variable defined on Ω. As usual let $\mu = \mathbf{E}[d_H(\mathbf{X}, A)]$. By Theorem 1.9, for $t > 0$, we have

$$\mathbf{Pr}[d_H(\mathbf{X}, A) - \mu < -t] \leq e^{-2t^2/n}.$$

Notice that $d_H(\mathbf{X}, A) = 0$ only holds for those points that belong to A. Taking $t = \mu$ we have

$$\mathbf{Pr}[\mathbf{X} \in A] = \mathbf{Pr}[d_H(\mathbf{X}, A) - \mu \leq -\mu] \leq e^{-2\mu^2/n},$$

making

$$t_0 = \left(\tfrac{1}{2} n \ln\left(\frac{1}{\mathbf{Pr}[\mathbf{X} \in A]}\right)\right)^2,$$

we have $\mu \leq t_0$. Therefore, for $t > t_0$ we have

$$\mathbf{Pr}[d_H(\mathbf{X}, A) > t] \leq e^{-2(t-t_0)^2/n}.$$

As for $t \geq 2a$, $(t-a)^2 \geq t^2/4$ for $t \geq 2a$, we have that for $t \geq 2t_0$,

$$\mathbf{Pr}[d_H(\mathbf{X}, A) > t] \leq e^{-2(t-t_0)^2/n}.$$

When $0 \leq t \leq 2t_0$, $\mathbf{Pr}[d_H(\mathbf{X}, A) > t] \geq \mathbf{Pr}[\mathbf{X} \in A]$. Thus

$$\min\{\mathbf{Pr}[d_H(\mathbf{X}, A) > t], \mathbf{Pr}[\mathbf{X} \in A]\} \leq e^{-t^2/2n},$$

which is the isoperimetric inequality in the statement of the theorem.

In plain words Theorem 1.11 says that for any subset $A \subseteq \Omega$ such that $\mathbf{Pr}[\mathbf{X} \in A]$ is not too small, say at least 1/2, with high probability, a random point is close to A under the Hamming distance, that is

$$\mathbf{Pr}\left[\mathbf{X} \in A_t^H\right] \geq 1 - 2e^{-t^2/4n}. \tag{1.15}$$

Example 1.17 (Isoperimetry on the Hypercube). Let us consider the particular example of the n-dimensional discrete hypercube Q_n with vertex set $\{0,1\}^n$ and edges between pairs of vertices whose Hamming distance is 1. Consider the probability space $\Omega = \prod_{i=1}^n \Omega_i$ where $\Omega_i = \{0,1\}$ for all $i \in [n]$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vertex of Q_n such that $\mathbf{Pr}[X_i = 0] = \mathbf{Pr}[X_i = 1] = \frac{1}{2}$ for all $i \in [n]$.

Then, for any set A of Ω, $\mathbf{Pr}[X \in A] = |A|/2^n$. Thus, if we choose $|A| = 2^{n-1}$ so $\mathbf{Pr}[X \in A] \geq 1/2$, by equation (1.15), we get the following isoperimetric inequality for the hypercube: For any $t \geq 0$,

$$|A_t| \geq 2^n(1 - 2e^{-t^2/4n}). \tag{1.16}$$

Historically, L. Harper [16], was the first to prove a tight isoperimetric inequality and he did it for the n-hypercube Q_n. His technique was based in the following: Given a hypercube Q_n and a real $t \geq 0$, define the *Hamming ball* of radius t centered at a vertex $\mathbf{x}$ as the set of vertices in Q_n at distance less or equal than t. Harper showed that for all two subsets $A_1, A_2 \subset Q_n$, there exist two Hamming balls B_1 and B_2 centered at vertex $\mathbf{0}$ and $\mathbf{1}$ respectively, such that $|B_1| = |A_1|$, $|B_2| = |A_2|$ and $d_H(B_1, B_2) \geq d_H(A_1, A_2)$. His final result is the following,

Theorem 1.12 (Harper). *For any given* $0 \le r \le n$*, let* $A \subseteq Q_n$ *be such that* $|A| \ge \sum_{k=0}^{r} \binom{n}{k}$*. Then, for any* $t > 0$*, we have* $|A_t^H| \ge \sum_{k=0}^{r+t} \binom{n}{k}$

Harper's result is deterministic, with $\nu(A) = |A|$. If we consider the probabilistic setting in which $\mathbf{Pr}[\mathbf{X} \in A] = 1/2$, a straightforward application of Chernoff's bounds (in the form of equation (1.8)), gives the following isoperimetric inequality for Q_n and any $t > 0$,

$$|A_t^H| \ge 2^n \left(1 - 2e^{-2t^2/n}\right). \tag{1.17}$$

Notice that this last equation is an improvement on (1.16). ◀

Let us consider now another definition of distance, for which we are going to derive an isoperimetric inequality due to Talagrand [36, 37]. We shall survey the main features of this inequality and present some of its applications. In order to keep the exposition at a basic level, we will skip the proof of the inequality. We refer the interested reader to [34, Chap. 6], [41, Chap. 6] or [29, Sec. 4] for details of the proofs. Most of our presentation follows the line in the last reference.

Recall that the L_2 norm of a vector $\mathbf{a} = (a_1, \ldots, a_n) \ge \mathbf{0}$, is defined as

$$\|\mathbf{a}\| = \left(\sum_i^n a_i^2\right)^{1/2}.$$

As usual, let $\Omega = \prod_{i=1}^n \Omega_i$ be a probability product space, let $\mathbf{X}$ be a sequence of random variables $(X_1, \ldots, X_n)$ in the corresponding spaces Ω_i, and let $\mathbf{a} = (a_1, \ldots, a_n)$ be a point in $[0, \infty)^n$, such that not all its components are 0. For any two points in Ω, $\mathbf{x} = (x_1, \ldots, x_n)$ and any $\mathbf{y} = (y_1, \ldots, y_n)$, define the distance $d_{\mathbf{a}}$ as:

$$d_{\mathbf{a}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n a_i\, \delta_i(\mathbf{x}, \mathbf{y}).$$

Notice, that if we consider the specific vector $\mathbf{a} = \mathbf{1} = (1, 1, \ldots, 1)$ then,

$$d_{\mathbf{a}}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \delta_i(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{x}, \mathbf{y}).$$

Given $A \subseteq \Omega$ define $d_{\mathbf{a}}(\mathbf{x}, A)$ as usual. The *Talagrand's convex distance* $d_T(\mathbf{x}, A)$ from a point $\mathbf{x}$ to a subset A, is the supremum over all possible

n-dimensional vectors $\mathbf{a}$ with $||\mathbf{a}|| = 1$ of $d_{\mathbf{a}}(\mathbf{x}, A)$, that is,

$$d_T(\mathbf{x}, A) = \sup_{||\mathbf{a}||=1} \left\{ \inf_{\mathbf{y}\in A} \left\{ \sum_{i=1}^{n} a_i\, \delta_i(\mathbf{x}, \mathbf{y}) \right\} \right\}.$$

In other words, the Talagrand's distance is the best convex distance over *all* choices of non-negative unit vectors.

Given $A \subseteq \Omega$ and $t > 0$, the *Talagrand's t-neighborhood of A*, A_t^T, is defined in the usual way:

$$A_t^T = \{\mathbf{x} \in \Omega : d_T(\mathbf{x}, A) \leq t\}.$$

With these definitions, we are ready to state Talagrand's inequality [36]:

Theorem 1.13 (Talagrand's isoperimetric inequality). *Let Ω be a product space, let $\mathbf{X} = (X_1, \dots, X_n)$ be a sequence of random variables, with each X_i defined on Ω_i. Then, for any $A \subseteq \Omega$ and any $t \geq 0$,*

$$\mathbf{Pr}[\mathbf{X} \in A]\left(1 - \mathbf{Pr}\left[\mathbf{X} \in A_t^T\right]\right) \leq e^{-t^2/4}.$$

In plain words, Talagrand's inequality states that under the Talagrand's convex distance, if $A \subseteq \Omega$ is large enough, then A_t^T covers a large portion of Ω.

Let us give some intuition for the inequality in the previous Theorem with respect to the one obtained in Theorem 1.11. If we consider the vector $\mathbf{a}' = (1/\sqrt{n}, \dots, 1/\sqrt{n})$, then $||\mathbf{a}'|| = 1$, and for any $\mathbf{x} \in \Omega$

$$\begin{aligned} d_T(\mathbf{x}, A) &= \sup_{||\mathbf{a}'||=1} \inf_{\mathbf{y}\in A} \left\{ \sum_{i=1}^{n} a_i \delta_i(\mathbf{x}, \mathbf{y}) \right\} \\ &\geq \inf_{\mathbf{y}\in A} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \delta_i(\mathbf{x}, \mathbf{y}) \\ &= \inf_{\mathbf{y}\in A} d_H(\mathbf{y}, A) = \frac{1}{\sqrt{n}} d_H(\mathbf{x}, A), \end{aligned}$$

therefore, for any $t > 0$, we get

$$A_{t/\sqrt{n}}^T \subseteq A_t^H. \tag{1.18}$$

= **Example 1.17 (cont).** Let us return to the hypercube. Consider a random vertex $\mathbf{X} \in \Omega = Q_n$. Given a set $A \subseteq \Omega$ with $\mathbf{Pr}[\mathbf{X} \in A] \geq 1/2$, and using Talagrand's inequality together with (1.18), we have

$$\mathbf{Pr}\left[\mathbf{X} \in A_t^H\right] \geq \mathbf{Pr}\left[\mathbf{X} \in A_{t/\sqrt{n}}^T\right] \geq 1 - 2e^{t^2/4n}.$$

Therefore, we obtain the isoperimetric inequality of equation (1.17). ⫤

Once again, we stress than the real strength of Talagrand's inequality over the inequality in Theorem 1.11 lies in the fact that Talagrand's refers simultaneously to all possible unit vectors $\mathbf{a}$.

Talagrand's isoperimetric inequality has proved to be a useful tool to simplify concentration bounds on several kind of problems. However, its application is far from obvious. The basic setting to use Talagrand's is the following: given a product space Ω, and given function $f : \mathbb{R}^n \to \mathbb{R}$, and a collection of random variables $\mathbf{X} = (X_1, \ldots, X_n)$, we wish to study the concentration of $f(\mathbf{X})$ around the median of $f(\mathbf{X})$. Notice that the set of points with measure above the median has probability at least 1/2.

To apply the method, we need first to find a bound on the Talagrand's distance between any two $\mathbf{x}, \mathbf{y}$ points in Ω. Then, considering the set A of points such that $f(\mathbf{x}) \leq \mathbf{M}[f(\mathbf{X})]$ and choosing a $t > 0$, we can apply Theorem 1.13, to show that the number of $\mathbf{x} \notin A_t^T$ is very small, so most of the points in Ω are in A_t^T, which implies that whp $f(\mathbf{x})$ and $\mathbf{M}[f(\mathbf{X})]$ are within distance t.

Let us return to the example of the longest increasing subsequence in a random permutation.

⫢ **Example 1.9 (cont).** We have seen, that the problem is equivalent to the problem of generate independently and uniformly n random reals in $[0, 1]$ and compute the length of the longest increasing subsequence. Let $\Omega = [0, 1]^n$, and let $\mathbf{X} \in \Omega$, be a random real sequence $\mathbf{X} = (X_1, \ldots, X_n)$. We wish to bound the concentration of the random variable $L(\mathbf{X})$ measuring the longest increasing subsequence.

Let $\mathbf{x} = (x_1, \ldots, x_n)$ a sequence of reals and let $K_\mathbf{x} \subseteq [n]$ be the set of subindex corresponding to the longest increasing subsequence, so $L(\mathbf{x}) = |K_\mathbf{x}|$. For any other sequence $\mathbf{y} \in [0, 1]^n$, we have

$$L(\mathbf{y}) \geq L(\mathbf{x}) - |\{i \in K_\mathbf{x} \mid x_i \neq y_i\}|.$$

To bound the Talagrand distance between any two points $\mathbf{x}$ and $\mathbf{y}$, let us consider the vector $\mathbf{a}' = (a'_1, \ldots, a'_n)$ where for every $i \in [n]$, $a'_i = 1/\sqrt{L(\mathbf{x})}$ if $i \in K_\mathbf{x}$ and $a'_i = 0$ otherwise. Then $||\mathbf{a}'|| = 1$. Moreover, by definition of $\mathbf{a}'$, $|\{i \in K_\mathbf{x} \mid x_i \neq y_i\}| = \sqrt{L(\mathbf{x})} d_{\mathbf{a}'}(\mathbf{x}, \mathbf{y})$ and we get

$$L(\mathbf{y}) \geq L(\mathbf{x}) - \sqrt{L(\mathbf{x})} d_{\mathbf{a}'}(\mathbf{x}, \mathbf{y}). \tag{1.19}$$

For instance, take

$$\mathbf{x} = (0.7, 0.8, 0.3, 0.1, 0.5, 0.6, 0.2, 0.7, 0.4) \quad \text{and}$$
$$\mathbf{y} = (0.3, 0.7, 0.4, 0.9, 0.5, 0.2, 0.6, 0.8, 0.1).$$

We have $K_{\mathbf{x}} = \{3, 5, 6, 8\}$, $L(\mathbf{x}) = 4$ and $\mathbf{a}' = (0, 0, \frac{1}{2}, 0, \frac{1}{2}, \frac{1}{2}, 0, \frac{1}{2}, 0)$. Furthermore, $|\{i \in K_{\mathbf{X}} \mid X_i \neq Y_i\}| = 3$ and $5 = L(\mathbf{y}) \geq 4 - 2$.

If we choose $A = \{\mathbf{x} \mid L(\mathbf{x}) \leq m\}$, where $m = \mathbf{M}[L(\mathbf{X})]$, by definition of median and for any random variable $\mathbf{X}$ on Ω,

$$\mathbf{Pr}[\mathbf{X} \in A] = \mathbf{Pr}[L(\mathbf{X}) \leq m] \geq \frac{1}{2}$$

and

$$\mathbf{Pr}[\mathbf{X} \notin A] = \mathbf{Pr}[L(\mathbf{X}) \geq m] < \frac{1}{2}.$$

For any $t \in [n]$ define the usual way A_t^T. Let $\mathbf{y} \in A$, that means $L(\mathbf{y}) \leq m$, and let $\mathbf{x} \notin A_t^T$. From Equation (1.19), taking the infimum over all possible $\mathbf{a} \in A$ and replacing $L(\mathbf{y})$ by m, we get $L(\mathbf{x}) \leq m + \sqrt{L(\mathbf{x})}\, d_{\mathbf{a}}(\mathbf{x}, A) \leq m + \sqrt{L(\mathbf{x})}\, d_T(\mathbf{x}, A)$. So,

$$d_T(\mathbf{x}, A) \geq \frac{L(\mathbf{x}) - m}{\sqrt{L(\mathbf{x})}} \geq \frac{t}{\sqrt{t+m}},$$

where this last inequality follows from the fact that the function $\frac{L(\mathbf{x})-m}{\sqrt{L(\mathbf{x})}}$ is increasing for values of $L(\mathbf{x}) > m$. Thus,

$$\begin{aligned}\mathbf{Pr}\left[\mathbf{X} \notin A_t^T\right] &= \mathbf{Pr}[L(\mathbf{X}) \geq m + t] \\ &\leq \mathbf{Pr}\left[d_T(\mathbf{X}, A) \geq \frac{t}{\sqrt{t+m}}\right] \\ &\leq e^{-\frac{t^2}{4(m+t)}}. \end{aligned} \tag{1.20}$$

Using Talagrand's inequality, for any $t \in [m]$,

$$\begin{aligned}\mathbf{Pr}[L(\mathbf{X}) \leq m]\,\mathbf{Pr}[L(\mathbf{X}) \geq m + t] &= \mathbf{Pr}[\mathbf{X} \in A]\,\mathbf{Pr}\left[\mathbf{X} \notin A_t^T\right] \\ &\leq \mathbf{Pr}[\mathbf{X} \in A]\,\mathbf{Pr}\left[d_T(\mathbf{x}, A) \geq \frac{t}{\sqrt{t+m}}\right] \\ &\leq 2e^{-\frac{t^2}{4(m+t)}}\end{aligned}$$

and, as by definition of the set A, we have $\mathbf{Pr}[L(\mathbf{x}) \leq m] \geq 1/2$, then for any $t \geq 0$

$$\mathbf{Pr}[L(\mathbf{X}) \geq m + t] \leq 2e^{-\frac{t^2}{4(m+t)}}, \tag{1.21}$$

and repeating the argument for $m-t$ in the place of m, we get that for any $t \in [m-1]$,

$$\mathbf{Pr}[L(\mathbf{X}) \geq m-t] \leq 2e^{-\frac{t^2}{4m}}, \tag{1.22}$$

which gives a good concentration result. $\Leftarrow$

We have used the median, to ensure that the set A is "large enough". However, it turns out, that under the condition of Talagrand's inequality, the same inequalities imply that the expectation μ and the median m are very close: $|\mu - m| \leq O(\sqrt{m})$. The following theorem is from [29] and generalizes this relationship. The proof uses only classical probability theory and it is left to the reader.

Theorem 1.14 (McDiarmid). *Let the random variable $f(\mathbf{X})$ have mean μ and median m, and let $a, b > 0$ be constants. Then for any $t > 0$,*

$$\mathbf{Pr}[f(\mathbf{X}) - m > t] \leq ae^{-t^2/b} \Longrightarrow |\mu - m| \leq O(a\sqrt{b})$$
$$\mathbf{Pr}[f(\mathbf{X}) - m \geq t] \leq ae^{-t^2/b(m+t)} \Longrightarrow |\mu - m| \leq O(\sqrt{m}).$$

There have been several studies on families of problems, where similar concentration theorems based on Talagrand's isoperimetric inequality, can be used. For instance, the above proof is taken from a more general theorem for functions f satisfying certain conditions, very similar to equation 1.19. See [29, Theorem 4.3]. Furthermore, for problems that could be defined in terms of hereditary properties of index sets, there exists a framework to apply a concentration theorem based in Talagrand's inequalities. Examples of problems in this class are increasing weighted sequences, balls and bins, counting extensions and representation of a graph, and graph edge coloring. See [9] and [34, Chapter 6]. Geometrical problems are another class of problems that can be treated with a unified approach using this method. Among them, the Euclidean traveling salesman problem, the Steiner tree problem and minimum spanning tree problem have received much attention. Again, see [9] and [34, Chap. 6]. A particularly handy but weaker form of Talagrand's inequality, was given in [31].

Theorem 1.15 (Molloy). *Let $f(\mathbf{X})$ be a random variable determined by n independent experiments, satisfying the following two conditions:*

1. *Changing the outcome of any one experiment can affect f by at most a small constant c, and*
2. *for any s, if $f(\mathbf{X}) \geq s$, by changing the outcomes of all the experiments, except s of them, the inequality $f(\mathbf{X}') \geq s$, still holds.*

Then, for any $0 < t \le m$,

$$\mathbf{Pr}\left[|f(\mathbf{X}) - m| > t\right] \le e^{-\frac{t^2}{16c^2m}},$$

where m *is the median of* f.

Notice, that to prove the second condition, it is sufficient to show that the outcomes of the selected s trials provide a lower bound for f. Let us apply Molloy's version of Talagrand's, to the the longest increasing subsequence example.

═ **Example 1.9 (cont).** Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sequence of reals, let $s = L(\mathbf{X})$ and $m = \mathbf{M}\left[L(\mathbf{X})\right]$. Notice that changing the value of a single X_i, it may affect $L(\mathbf{X})$ by at most one. Moreover, the $\{X_i\}$ such that $i \in K_{\mathbf{X}}$ are by themselves alone witnesses to the fact that $s = L(\mathbf{X})$. Therefore, we can apply Therefore, Theorem 1.15 to get

$$\mathbf{Pr}\left[|L(\mathbf{X}) - m| > t\right] \le 2e^{-t^2/16m}.$$

That shows concentration around the mean value m. ═

Coming back to the balls and bins, experiment, Molloy's version of Talagrand's inequality can be applied to show concentration of the number of non-empty bins.

═ **Example 1.7 (cont).** Let us consider now a symmetrical version of the balls and bins experiment. Let Y be the random variable that counts the number of non empty bins. Notice that changing the outcome of one throw only can change the function by at most 1. If we have that the number of non empty bins in a configuration is at least s, there must exits s balls that fall in s different bins, by fixing these positions, changing the outcomes of the other balls will maintain s as lower bound. From Theorem 1.15 we get

$$\mathbf{Pr}\left[|Y - \mathbf{M}\left[Y\right]| > t\right] \le 2\exp\left(-\frac{-t^2}{16\mathbf{M}\left[Y\right]}\right).$$

That shows concentration around the median. Using Theorem 1.14 and taking into account that $Z + Y = n$, the above result may be used to show concentration around the mean for the number of empty bins. ═

A final remark is that in the statement of Talagrand's isoperimetric inequality, the variables $X_1, \ldots, X_n$ must be independent. Recent work by Marton extends Talagrand's result to enable some sort of limited dependence [27].

References

[1] N. Alon, J. H. Spencer, and P. Erdős. *The probabilistic method.* Wiley-Interscience, New York, 1992.

[2] K. Azuma. Weighted sums of certain dependent random variables. *Tokuku Math. Journal*, 19:357–367, 1967.

[3] B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8:49–55, 1988.

[4] H. Chernoff. A measure of asymptotic efficiency for tests of a hypotheses based on the sum of observations. *Annals of Mathematics Statistics*, 23:493–507, 1952.

[5] T. H. Cormen, Ch. Leiserson, and R. Rivest. *Introduction to Algorithms.* The MIT Press, Cambridge, Mass., 1989.

[6] J. Díaz, M. D. Penrose, J. Petit, and M. J. Serna. Linear orderings of random geometric graphs. In P. Widmayer, G. Neyer, and S. Eidenbenz, editors, *Graph Theoretic Concepts in Computer Science*, volume 1665 of *Lecture Notes in Computer Science*, pages 291–302. Springer-Verlag, 1999.

[7] J. Díaz, M. J. Serna, P. Spirakis, J. Torán, and T. Tsukiji. On the expected depth of boolean circuits. Technical Report LSI 94-7-R, Universitat Politècnica de Catalunya, 1994.

[8] D. Dubhashi. Simple proofs of occupancy tail bounds. *Random Structures and Algorithms*, 11:119–123, 1997.

[9] D. Dubhashi. Talagrand's inequality in hereditary settings. Technical report, Dept. CS, Indian Institute of Technology, 1998.

[10] W. Feller. *An Introduction to Probability Theory and its Applications. Vol I.* J. Wiley, New York, 1957.

[11] D. A. Freedman. On tail probabilities for martingales. *Annals Probability*, 3:100–118, 1975.

[12] A Frieze and B. Reed. Probabilistic analysis of algorithms. In H. Habib, C. McDiarmid, J. Ramirez, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 36–92. Springer-Verlag, 1998.

[13] Alan Frieze. On the length of the longest monotone subsequence in a random permutation. *Annals Applied Probability*, 1:301–305, 1991.

[14] G. Grimmett and D. R. Stirzaker. *Probability and Random Processes. (2nd. edition).* Oxford University Press, 1992.

[15] T. Hagerup and C. Rub. A guided tour of Chernoff bounds. *Information Processing Letters*, 33:305–308, 1990.

[16] L. H. Harper. Optimal numberings and isoperimetric problems on graphs. *Journal of Combinatorial Theory*, 1(3):385–393, 1966.

[17] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal American Statistics Association*, 58:13–30, 1963.

[18] M. Jerrum and G. Sorkin. Simulated Annealing for Graph Bisection. In *34th Symposium on Foundations of Computer Science*, pages 94–103, 1993.

[19] J. JáJá. *An introduction to parallel algorithms.* Addison-Wesley, Reading, Mass., 1992.

[20] A. Kamath, R. Motwani, K. Palem, and P. Spirakis. Tail bounds for occupancy problems and the satisfiability conjecture. *Random Structures and Algorithms*, 7:59–80, 1995.

[21] R. M. Karp. Class notes for CS-292F. 1989.

[22] E. Koutsupias and X. Papadimitriu. On the greedy algorithm for satisfiability. *Information Processing Letters*, 43:53–55, 1992.

[23] I. Leader. Discrete isoperimetric inequalities. In B. Bollobas, editor, *Probabilistic Combinatorics and its Applications*, volume 44 of *Proceedings of Symposia in Applied Mathematics*, pages 57–80. American Mathematical Society, Rhode Island, 1991.

[24] M. Ledoux. Isoperimetry and gaussian analysis. In P. Groeneboom R. Dobrushin and M. Ledoux, editors, *Lectures in Probability Theory. Ecole d'Ete de Probabilites de Saint-Flour XXIV*, volume 1648 of *Lectures Notes in Mathematics*, pages 165–294. Springer-Verlag, Berlin, 1994.

[25] B. F. Logan and L. A. Shepp. A variational problem for Young tableaux. *Advances in Mathematics*, 26:206–222, 1977.

[26] M. Luby. A simple parallel algorithm for the maximal independent set. *SIAM Journal on Computing*, 15:1036–1056, 1986.

[27] K. Marton. On a measure concentration inequality of Talagrand for dependent random variables. Technical report, Math. Institute of the Hungarian Academy of Sciences, 1996.

[28] C. McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics, 1989*, pages 669–188. London Mathematical Society Lectures Notes Series, 1989.

[29] C. McDiarmid. Concentration. In H. Habib, C. McDiarmid, J. Ramirez, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248. Springer-Verlag, 1998.

[30] C. McDiarmid and R. B. Hayward. Large deviations for quicksort. *Journal of Algorithms*, 21:476–507, 1996.

[31] M. Molloy. The probabilistic method. In H. Habib, C. McDiarmid, J. Ramirez, and B. Reed, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 1–35. Springer-Verlag, 1998.

[32] R. Motwani and P. Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995.

[33] E. Shamir and J. Spencer. Sharp concentration of the chromatic number in random graphs. *Combinatorica*, 7:121–130, 1987.

[34] J. M. Steele. Probabilistic algorithm for the directed traveling salesman problem. *Mathematics of Operation Research*, 11:343–350, 1986.

[35] W. Steiger. Some Kolmogoroff-type inequalities for bounded random variables. *Biometrica*, 54:641–648, 1967.

[36] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S.*, 81:73–1205, 1995.

[37] M. Talagrand. A new look at independence. *Annals of Probability*, 24:1–34, 1996.

[38] T. Tsukiji and F. Xhafa. On the depth of randomly generated circuits. In J. Díaz and M. Serna, editors, *4th European Symposium on Algorithms*, volume 1136 of *Lecture Notes in Computer Science*, pages 208–219. Springer-Verlag, Berlin, 1996.

[39] A. M. Vershik and C. V. Kerov. Asymptotics of the Plancherel measure of the symmetric group and a limiting form from Young Tableau. *Dokl. Akad. Nauk. U.S.S.R.*, 233:1024–1027, 1977.

[40] U. Vishkin. Randomized speed-ups in parallel computation. In *15th. ACM Symposium on Theory of Computing*, pages 230–239. ACM Society, 1984.

[41] J. E. Yukich. *Probability Theory of classical Euclidian Optimization Problems.* Springer–Verlag, Heidelberg, 1998.