9 Concentration of measure

Recall that Chernoff bound allows to prove exponential tail bounds for sums of **independent** random variables. For example, if Z is a sum of n Bernoulli random variables, then

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq t\sqrt{n}) \leq 2e^{-2t^2/n}.$$

As a matter of terminology (which is convenient though we will largely not use), random variables Z that satisfy $\mathbb{P}(|Z| \geq t) \leq 2e^{-ct^2}$ for all $t \geq 0$ and constant $c > 0$ are called **sub-gaussian**. We usually are not too concerned about optimizing the constant c in the exponent of bound.

In this chapter, we develop tools for proving similar sub-gaussian tail bounds for other random variables that do not necessarily arise as a sum of independent random variables.

Here is the general principle:

> **A Lipschitz function of many *independent* random variables is concentrated.**

We will prove the following important and useful result, known by several names: **McDiarmid's inequality**, **Azuma–Hoeffding inequality**, and **bounded differences inequality**.

Theorem 9.0.1 (Bounded differences inequality)**.** Let $X_1 \in \Omega_1, \dots, X_n \in \Omega_n$ be **independent** random variables. Suppose $f\colon \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ satisfies

$$|f(x_1, \dots, x_n) - f(x'_1, \dots, x'_n)| \leq 1 \tag{9.1}$$

whenever $(x_1, \dots, x_n)$ and $(x'_1, \dots, x'_n)$ differ on exactly one coordinate. Then the random variable $Z = f(X_1, \dots, X_n)$ satisfies, for every $\lambda \geq 0$,

$$\mathbb{P}(Z - \mathbb{E}Z \geq \lambda) \leq e^{-2\lambda^2/n} \quad \text{and} \quad \mathbb{P}(Z - \mathbb{E}Z \leq -\lambda) \leq e^{-2\lambda^2/n}.$$

In particular, we can apply the above inequality to $f(x_1, \dots, x_n) = x_1 + \cdots + x_n$ to recover the Chernoff bound. The theorem tells us that the window of fluctuation of Z has length $O(\sqrt{n})$.

Example 9.0.2 (Coupon collector)**.** Let $s_1, \dots, s_n \in [n]$ chosen uniformly and independently at random. Let

$$Z = |[n] \setminus \{s_1, \dots, s_n\}|.$$

Then

$$\mathbb{E}Z = n\left(1-\frac{1}{n}\right)^n \in \left[\frac{n}{e}, \frac{n-1}{e}\right].$$

Note that changing one of the $s_1, \dots, s_n$ changes Z by at most 1, so we have

$$\mathbb{P}\left(|Z - n/e| \geq \lambda\sqrt{n} + 1\right) \leq \mathbb{P}\left(|Z - \mathbb{E}Z| \geq \lambda\sqrt{n}\right) \leq 2e^{-2\lambda^2}.$$

Definition 9.0.3 (Lipschitz functions)**.** Given two metric spaces (X, d_X) and (Y, d_Y), we say that a function $f\colon X \to Y$ is **C-Lipschitz** if

$$d_Y(f(x), f(x')) \leq C d_X(x, x') \qquad \text{for all } x, x' \in X.$$

Then (9.2) says that $f\colon \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ is 1-Lipschitz with respect to the Hamming distance on $\Omega_1 \times \cdots \times \Omega_n$.

Note that while it may be tempting to think about the cases $\Omega_i = \{0, 1\}$, it will be crucial for us to consider more general Ω_i for our applications.

Theorem 9.0.1 holds more generally allowing the bounded difference to depend on the coordinate.

Theorem 9.0.4 (Bounded differences inequality)**.** Let $X_1 \in \Omega_1, \dots, X_n \in \Omega_n$ be **independent** random variables. Suppose $f\colon \Omega_1 \times \cdots \times \Omega_n \to \mathbb{R}$ satisfies

$$|f(x_1, \dots, x_n) - f(x_1', \dots, x_n')| \leq c_i \tag{9.2}$$

whenever $(x_1, \dots, x_n)$ and $(x_1, \dots, x_n)$ differ only on the i-th coordinate. Then the random variable $Z = f(X_1, \dots, X_n)$ satisfies, for every $\lambda \geq 0$,

$$\mathbb{P}(Z - \mathbb{E}Z \geq \lambda) \leq \exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right)$$

and

$$\mathbb{P}(Z - \mathbb{E}Z \leq -\lambda) \leq \exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right).$$

We will prove these inequality using martingales.

9.1 Martingales concentration inequalities

Definition 9.1.1. A **martingale** is a random real sequence $Z_0, Z_1, \ldots$ such that for every Z_n, $\mathbb{E}|Z_n| < \infty$ and

$$\mathbb{E}[Z_{n+1}|Z_0, \ldots, Z_n] = Z_n.$$

(To be more formal, we should talk about filtrations of a probability space . . .)

Example 9.1.2 (Random walks with independent steps)**.** If $(X_i)_{i \geq 0}$ is a sequence of independent random variables with $\mathbb{E}X_i = 0$ for all i, then the partial sums $Z_n = \sum_{i \leq n} X_i$ is a Martingale.

Example 9.1.3 (Betting strategy)**.** Betting on a sequence of fair coin tosses. After round, you are allow to change your bet. Let Z_n be your balance after the n-th round. Then Z_n is always a martingale regardless of your strategy.

Originally, the term "martingale" referred to the betting strategy where one doubles the bet each time until the first win and then stop betting. Then, with probability 1, $Z_n = 1$ for all sufficiently large n. (Why does this "free money" strategy not actually work?)

The next example is especially important to us.

Example 9.1.4 (Doob martingale)**.** Given some underlying random variables $X_1, \ldots, X_n$ (not necessarily independent, though they often are independent in practice), and a function $f(X_1, \ldots, X_n)$. Let Z_i be the expected value of f after "revealing" (exposing) $X_1, \ldots, X_i$, i.e.,

$$Z_i = \mathbb{E}[f(X_1, \ldots, X_n)|X_1, \ldots, X_i].$$

So Z_i is the expected value of the random variable $Z = f(X_1, \ldots, X_n)$ after seeing the first i arguments, and letting the remaining arguments be random. Then $Z_0, \ldots, Z_n$ is a martingale (why?). It satisfies $Z_0 = \mathbb{E}Z$ (a non-random quantity) and $Z_n = Z$ (the random variable that we care about), and thereby offering a way to interpolate between the two.

Example 9.1.5 (Edge-exposure martingale)**.** We can reveal the random graph $G(n, p)$ by first fixing an order on all unordered pairs of $[n]$ and then revealing in order whether each pair is an edge. For any graph parameter $f(G)$ we can produce a martingale $X_0, X_1, \ldots, X_{\binom{n}{2}}$ where Z_i is the conditional expectation of $f(G(n, p))$ after revealing whether there are edges for first i pairs of vertices. See Figure 5 for an example.

Example 9.1.6 (Vertex-exposure martingale)**.** Similar to the previous example, except that we now first fix an order on the vertex set, and, at the i-th step, with $0 \leq i \leq n$, we

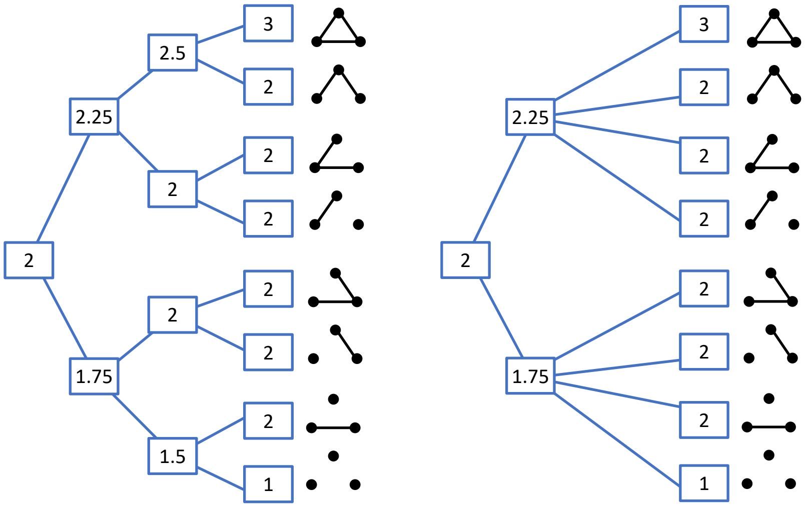


Figure 5: The edge-exposure martingale (left) and vertex-exposure martingale (right) for the chromatic number of $G(n, 1/2)$ with $n = 3$. The martingale is obtained by starting at the leftmost point, and splitting at each branch with equal probability.

reveal all edges whose endpoints are contained in the first i vertices. See Figure 5 for an example.

Sometimes it is better to use the edge-exposure martingale and sometimes it is better to use the vertex-exposure martingale. It depends on the application. There is a trade-off between the length of the martingale and the control on the bounded differences.

The main result is that a martingale with *bounded differences* must be concentrated. The following fundamental result is called Azuma's inequality or the Azuma–Hoeffding inequality.

Theorem 9.1.7 (Azuma's inequality)**.** Let $Z_0, Z_1, \ldots, Z_n$ be a martingale satisfying

$$|Z_i - Z_{i-1}| \leq 1 \quad \text{for each } i \in [n].$$

Then for every $\lambda > 0$,

$$\mathbb{P}(Z_n - Z_0 \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}.$$

Note that this is the same bound that we derived in Section 5 for $Z_n = X_1 + \cdots X_n$ where $X_i \in \{-1, 1\}$ uniform and iid.

More generally, allowing different bounds on different steps of the martingale, we have the following.

Theorem 9.1.8 (Azuma's inequality)**.** Let $Z_0, Z_1, \dots, Z_n$ be a martingale satisfying

$$|Z_i - Z_{i-1}| \le c_i \quad \text{for each } i \in [n].$$

For any $\lambda > 0$,

$$\mathbb{P}(Z_n - Z_0 \ge \lambda) \le \exp\left(\frac{-\lambda^2}{2(c_1^2 + \cdots + c_n^2)}\right).$$

The above formulations of Azuma's inequality recovers the bounded differences inequality Theorems 9.0.1 and 9.0.4 up to a (usually unimportant) constant in the exponent (details shortly). To obtain the exact statement of Theorem 9.0.4, we state the following strengthening of Azuma's inequality.

Theorem 9.1.9 (Azuma's inequality)**.** Let $Z_0, Z_1, \dots, Z_n$ be a martingale such that, for each $i \in [n]$, conditioned on $(Z_0, \dots, Z_{i-1})$, the random variable Z_i lies inside an interval of length c_i (the location of the interval may depend on $Z_0, \dots, Z_{i-1}$). Then for any $\lambda > 0$,

$$\mathbb{P}(Z_n - Z_0 \ge \lambda) \le \exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right).$$

Remark 9.1.10. Applying the inequality to the martingale with terms $-Z_n$, we obtain the following lower tail bound:

$$\mathbb{P}(Z_n - Z_0 \le -\lambda) \le \exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right).$$

And we can put them together as

$$\mathbb{P}(|Z_n - Z_0| \ge \lambda) \le 2\exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right).$$

Lemma 9.1.11 (Hoeffding)**.** Let X be a real random variable contained in an interval of length ℓ. Suppose $\mathbb{E}X = 0$. Then

$$\mathbb{E}[e^X] \le e^{\ell^2/8}.$$

Proof. Suppose $X \in [a, b]$ with $a \le 0 \le b$ and $b - a = \ell$. Then since e^x is convex, using a linear upper bound on the interval $[a, b]$, we have

$$e^x \le \frac{b - x}{b - a} e^a + \frac{x - a}{b - a} e^b, \quad \forall x \in [a, b].$$

Thus

$$\mathbb{E}e^X \le \frac{b}{b-a}e^a + \frac{-a}{b-a}e^b.$$

Let $p = -a/(b-a)$. then $a = -p\ell$ and $b = (1-p)\ell$, we have

$$\log \mathbb{E}e^X \le \log\left((1-p)e^{-p\ell} + pe^{(1-p)\ell}\right) = -p\ell + \log(1-p+pe^\ell).$$

Fix $p \in [0,1]$. Let

$$\varphi(\ell) := -p\ell + \log(1-p+pe^\ell).$$

It remains to show that $\varphi(\ell) \le \ell^2/8$ for all $\ell \ge 0$, which follows from $\varphi(0) = \varphi'(0) = 0$ and $\varphi''(\ell) \le 1/4$ for all $\ell \ge 0$, as

$$\varphi''(\ell) = \left(\frac{p}{(1-p)e^{-p\ell}+p}\right)\left(1 - \frac{p}{(1-p)e^{-p\ell}+p}\right) \le \frac{1}{4},$$

since $t(1-t) \le 1/4$ for all $t \in [0,1]$. □

Proof of Theorem 9.1.9. By adding a constant to the sequence, we may assume that $Z_0 = 0$. Let

$$X_i = Z_i - Z_{i-1}$$

be the martingale difference. Let $t \ge 0$. Then the hypothesis together with Lemma 9.1.11 imply that

$$\mathbb{E}[e^{tX_i} | Z_0, \dots, Z_{i-1}] \le e^{t^2c_i^2/8}.$$

Then the moment generating function satisfies

$$\mathbb{E}[e^{tZ_n}] = \mathbb{E}\left[e^{t(X_n+Z_{n-1})}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{tX_n} \mid Z_0, \dots, Z_{n-1}\right]e^{tZ_{n-1}}\right] = e^{t^2c_n^2/8}\mathbb{E}[e^{tZ_{n-1}}].$$

Iterating, we obtain

$$\mathbb{E}[e^{tZ_n}] \le e^{t^2(c_1^2+\cdots c_n^2)/8}.$$

By Markov,

$$\mathbb{P}(Z_n \ge \lambda) \le e^{-t\lambda}\mathbb{E}[e^{tZ_n}] \le e^{-t\lambda + \frac{t^2}{8}(c_1^2+\cdots c_n^2)}.$$

Setting $t = 4\lambda/(c_1^2 + \cdots + c_n^2)$ yields the theorem. □

Let us use Azuma's inequality to prove the bounded difference inequality (Theorem 9.0.4), whose statement is copied below:

> Let $Z_0, Z_1, \dots, Z_n$ be a martingale such that, for each $i \in [n]$, conditioned on $(Z_0, \dots, Z_{i-1})$, the random variable Z_i lies inside an interval of length c_i (the

location of the interval may depend on $Z_0, \dots, Z_{i-1}$). Then for any $\lambda > 0$,

$$\mathbb{P}(Z_n - Z_0 \geq \lambda) \leq \exp\left(\frac{-2\lambda^2}{c_1^2 + \cdots + c_n^2}\right).$$

Proof of the Theorem 9.0.4. Consider the Doob martingale $Z_i = \mathbb{E}[Z|X_1, \dots, X_i]$.

By the Lipschitz condition, we see that for every $i \in [n]$ and fixed $x_1 \in \Omega_1, \dots, x_{i-1} \in \Omega_{i-1}$, we have

$$\max_{x_i \in \Omega_i} f(x_1, \dots, x_{i-1}, x_i, X_{i+1}, \dots, X_n) - \min_{x_i \in \Omega_i} f(x_1, \dots, x_{i-1}, x_i, X_{i+1}, \dots, X_n) \leq c_i$$

for every possible $X_{i+1}, \dots, X_n$, so that taking expectation of these random values shows that, conditioned on the values of $X_1, \dots, X_{i-1}$, there is an interval (possibly depending on $X_1, \dots, X_{i-1}$) of length c_i that Z_i lies in.

Since $Z_0 = \mathbb{E}Z$ and $Z_n = Z$, the desired bound follows from Azuma's inequality (Theorem 9.1.9).[3] □

9.2 Chromatic number of random graphs

9.2.1 Concentration of chromatic number

Even before Bollobás (1988) showed that $\chi(G(n, 1/2)) \sim \frac{n}{2\log_2 n}$ whp (Theorem 8.3.3), using the bounded difference inequality, it was already known that the chromatic number of a random graph must be concentrated in a $\omega(\sqrt{n})$ window around its mean. The following application shows that one can prove concentration around the mean without even knowing where is the mean!

Theorem 9.2.1 (Shamir and Spencer 1987)**.** For every $\lambda \geq 0$, $Z = \chi(G(n, p)$ satisfies

$$\mathbb{P}(|Z - \mathbb{E}Z| \geq \lambda\sqrt{n-1}) \leq 2e^{-2\lambda^2}.$$

Proof. Let $V = [n]$, and consider each vertex labeled graph as an element of $\Omega_2 \times \cdots \times \Omega_n$ where $\Omega_i = \{0, 1\}^{i-1}$ and its coordiantes correspond to edges whose larger coordinate is i (cf. the vertex-exposure martingale Example 9.1.6). If two graphs G and G' differ only in edges incident to one vertex v, then $|\chi(G) - \chi(G')| \leq 1$ since, given a proper coloring of G using $\chi(G)$ colors, one can obtain a proper coloring of G' using $\chi(G) + 1$ colors by using a new color for v. Theorem 9.0.4 implies the result. □

[3] We are cheating somewhat here, since multiple instance of $(X_1, \dots, X_i)$ can correspond to the same $(Z_0, \dots, Z_i)$. To be more correct, we should restate Theorem 9.1.9 instead of a filtration based on the Doob martingale.

Remark 9.2.2 (Non-concentration of the chromatic number)*.* Recently, a surprising breakthrough of Heckel (2019+) showed that the $\chi(G(n,1/2))$ is *not* concentrated on any interval of length $n^{1/4-\epsilon}$ for any constant $\epsilon > 0$. This was the opposite of what most experts believed in. Given the new realization, it seems reasonable to suspect that the length of the window of concentrations fluctuates between $n^{1/4+o(1)}$ to $n^{1/2+o(1)}$ depending on n.

9.2.2 Clique number, again

Previously in Section 8.3, we used Janson inequalities to prove the following exponentially small bound on the probability that $G(n,1/2)$ has small clique number. This was a crucial step in the proof of Bollobás' theorem (Theorem 8.3.3) that $\chi(G(n,1/2)) \sim n/(2\log_2 n)$ whp. Here we give a different proof using the bounded difference inequality instead of Janson inequalities. The proof below in fact was the original approach of Bollobás (1988).

Theorem 9.2.3 (Same as Theorem 8.3.2)**.** Let $k_0 = k_0(n) \sim 2\log_2 n$ be the largest positive integer so that $\binom{n}{k_0}2^{-\binom{k_0}{2}} \geq 1$. Then

$$\mathbb{P}(\omega(G(n,1/2)) < k_0 - 3) = e^{-n^{2-o(1)}}.$$

A naive approach might be to estimate the number of k-cliques in G (this is the approach taken with Janson inequalities. Here, instead, we use a very clever and non-obvious choice of a Lipschitz function of graphs.

Proof. Let $k = k_0 - 3$. Let $Y = Y(G)$ be the maximum number of edge-disjoint set of k-cliques in G. Then as a function of G, Y changes by at most 1 if we change G by one edge. (Note that the same does not hold if we change G by one vertex, e.g., when G consists of many k-cliques glued along a common vertex.)

So by the bounded differences inequality, for $G \sim G(n,1/2)$,

$$\mathbb{P}(\omega(G) < k) = \mathbb{P}(Y = 0) \leq \mathbb{P}(Y - \mathbb{E}Y \leq -\mathbb{E}Y) \leq \exp\left(-\frac{2(\mathbb{E}Y)^2}{\binom{n}{2}}\right). \qquad (9.3)$$

It remains to show that $\mathbb{E}Y \geq n^{2-o(1)}$. Create an auxiliary graph $\mathcal{H}$ whose vertices are the k-cliques in G, with a pair of k-cliques adjacent if they overlap in at least 2 vertices. Then $Y = \alpha(\mathcal{H})$. We would like to lower bound the independence number of this graph based on its average degree. Here are two ways to proceed:

1. Recall the Caro–Wei inequality (Corollary 2.3.5): for every graph H with average

degree $\overline{d}$, we have

$$\alpha(H) \geq \sum_{v \in V(H)} \frac{1}{1+d_v} \geq \frac{|V(H)|}{1+\overline{d}} = \frac{|V(H)|^2}{|V(H)| + 2|E(H)|}.$$

2. Let H' be the induced subgraph obtained from H by keeping every vertex independently with probability q. We have

$$\alpha(H) \geq \alpha(H') \geq |V(H')| - |E(H')|.$$

Taking expectations of both sides, and noting that $\mathbb{E}|V(H')| = q|V(H)|$ and $\mathbb{E}|E(H')| = q^2|E(H)|$ by linearity of expectations, we have

$$\alpha(H) \geq q\mathbb{E}|V(H)| - q^2|E(H)| \qquad \text{for every } q \in [0,1].$$

Provided that $|E(H)| \geq |V(H)|/2$, we can take $q = |V(H)|/(2|E(H)|) \in [0,1]$ and obtain

$$\alpha(H) \geq \frac{|V(H)|^2}{4|E(H)|} \qquad \text{if } |E(H)| \geq \frac{1}{2}|V(H)|.$$

(This method allows us to recover Turán's theorem up to a factor of 2, whereas the Caro–Wei inequality recovers Turán's theorem exactly. For the present application, we do not care about these constant factors.)

We have, with probability $1-o(1)$, the number of k-cliques $|V(\mathcal{H})|$ satisfies

$$|V(\mathcal{H})| \sim \mu := \mathbb{E}|V(\mathcal{H})| = \binom{n}{k} 2^{-\binom{k}{2}} \geq n^{3-o(1)}$$

and the number of pairs of edge-overlapping k-cliques $|E(\mathcal{H})|$ satisfies

$$\mathbb{E}|E(\mathcal{H})| =: \frac{\Delta}{2} \sim \frac{\mu^2 k^4}{2n^2} \gg \mu$$

(details again omitted; this is the same first and second moment calculation as in Section 4.3 and Theorem 8.3.2.) Thus, with probability $1-o(1)$, we can apply either of the above lower bounds on independent sets to obtain

$$\mathbb{E}Y \gtrsim \mathbb{E}\frac{\mu^2}{|E(\mathcal{H})|} \gtrsim \frac{\mu^2}{\Delta} \sim \frac{n^2}{k^4}.$$

Thus by (9.3), we obtain

$$\mathbb{P}(\omega(G) < k) \le \exp\left(-\frac{2(\mathbb{E}Y)^2}{\binom{n}{2}}\right) \le \exp\left(-\Omega\left(\frac{n^2}{k^8}\right)\right) = \exp\left(-\Omega\left(\frac{n^2}{(\log n)^8}\right)\right). \qquad \square$$

9.2.3 Chromatic number of sparse random graphs

Let us show that $G(n,p)$ is concentrated on a constant size window if p is small enough.

Theorem 9.2.4 (Shamir and Spencer 1987)**.** Let $\alpha > 5/6$ be fixed. Then for $p < n^{-\alpha}$, $\chi(G(n,p))$ is concentrated in four values with probability $1-o(1)$, i.e., there exists $u = u(n,p)$ such that, as $n \to \infty$,

$$\mathbb{P}(u \le \chi(G(n,p)) \le u+3) = 1 - o(1).$$

Proof. Let $\epsilon = \epsilon_n > 0$ and $\epsilon \to 0$ (we'll later choose it to be arbitrarily small). Let $u = u(n,p,\epsilon)$ be the least integer so that

$$\mathbb{P}(\chi(G(n,p)) \le u) > \epsilon.$$

Now we make a clever choice of a random variable.

Let $G \sim G(n,p)$. Let $Y = Y(G)$ denote the minimum size of a subset $S \subset V(G)$ such that $G - S$ is u-colorable. Note that Y changes by at most 1 if we change the edges around one vertex of G. Thus, by applying Theorem 9.0.1 with respect to vertex-exposure (Example 9.1.6), we have

$$\begin{aligned} &\mathbb{P}(Y \le \mathbb{E}Y - \lambda\sqrt{n}) \le e^{-2\lambda^2} \\ \text{and}\qquad &\mathbb{P}(Y \ge \mathbb{E}Y + \lambda\sqrt{n}) \le e^{-2\lambda^2}. \end{aligned}$$

We choose $\lambda = \lambda(\epsilon) > 0$ so that $e^{-2\lambda^2} = \epsilon$.

First, we use the lower tail bound to show that $\mathbb{E}Y$ must be small. We have

$$e^{-2\lambda^2} = \epsilon < \mathbb{P}(\chi(G) \le u) = \mathbb{P}(Y = 0) = \mathbb{P}(Y \le \mathbb{E}Y - \mathbb{E}Y) \le \exp\left(\frac{-2(\mathbb{E}Y)^2}{n}\right)$$

so

$$\mathbb{E}Y \le \lambda\sqrt{n}.$$

Next, we apply the upper tail bound to show that Y is rarely large. We have

$$\mathbb{P}(Y \geq 2\lambda\sqrt{n}) \leq \mathbb{P}(Y \geq \mathbb{E}Y + \lambda\sqrt{n}) \leq e^{-2\lambda^2} = \epsilon.$$

Each of the following three events occur with probability at least $1-\epsilon$, for large enough n,

- By the above argument, there is some $S \subset V(G)$ with $|S| \leq 2\lambda\sqrt{n}$ and $G - S$ may be properly u-colored.
- By the next lemma, one can properly 3-color $G[S]$.
- $\chi(G) \geq u$ (by the minimality of u at the beginning of the proof).

Thus, with probability at least $1 - 3\epsilon$, all three events occur, and so we have $u \leq \chi(G) \leq u + 3$. □

Lemma 9.2.5. Fix $\alpha > 5/6$ and C. Let $p \leq n^{-\alpha}$. Then with probability $1 - o(1)$ every subset of at most $C\sqrt{n}$ vertices of $G(n,p)$ can be properly 3-colored.

Proof. Let $G \sim G(n,p)$. Assume that G is not 3-colorable. Choose minimum size $T \subset V(G)$ so that the induced subgraph $G[T]$ is not 3-colorable.

We see that $G[T]$ has minimum degree at least 3, since if $\deg_{G[T]}(x) < 3$, then $T - x$ cannot be 3-colorable either (if it were, then can extend coloring to x), contradicting the minimality of T.

Thus $G[T]$ has at least $3|T|/2$ edges. The probability that G has some induced subgraph on $t \leq C\sqrt{n}$ vertices and $\geq 3t/2$ edges is, by a union bound, (recall $\binom{n}{k} \leq (ne/k)^k$)

$$\leq \sum_{t=4}^{C\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{3t/2} \leq \sum_{t=4}^{C\sqrt{n}} \left(\frac{ne}{t}\right)^t \left(\frac{te}{3}\right)^{3t/2} n^{-3t\alpha/2}$$
$$\leq \sum_{t=4}^{C\sqrt{n}} \left(O(n^{1-3\alpha/2}\sqrt{t})\right)^t \leq \sum_{t=4}^{C\sqrt{n}} \left(O(n^{1-3\alpha/2+1/4})\right)^t$$

the sum is $o(1)$ provided that $\alpha > 5/6$. □

Remark 9.2.6. Theorem 9.2.4 was subsequently improved (by a refinement of the above techniques) by Łuczak (1991) and Alon and Krivelevich (1997), who showed two-point concentration for all $\alpha > 1/2$.

9.3 Isoperimetric inequalities: a geometric perspective

The bounded differences inequality (Theorem 9.0.1) tells that if $f\colon \{0,1\}^n \to \mathbb{R}$ is 1-Lipschitz (with respect to the Hamming distance on $\{0,1\}^n$), it must be concentrated around its mean:

$$\mathbb{P}(|f - \mathbb{E}f| \geq \lambda\sqrt{n}) \leq 2e^{-2\lambda^2}.$$

Given that the maximum possible variation in f is n, the above concentration inequality says that f is *almost constant*, which should be somewhat counterintuitive.

It turns out that similar phenomenon occurs in other spaces not just the Hamming cube. In fact, it is really a general high dimensional geometric phenomenon. In this section, we explore this concentration of phenomenon from a geometric perspective, and explain how it relates to **isoperimetric inequalities**.

Recall the classic isoperimetric theorem in $\mathbb{R}^n$ It says that among all subset of $\mathbb{R}^n$ of given volume, the ball has the smallest surface volume. (The word "isoperimetric" refers to fixing the perimeter; equivalently we fix the surface area and ask to maximize volume.)

Here is a slightly stronger formulation. Given a metric space (X, d_X) and a set $A \subset X$, we write

$$A_t := \{x \in X : d_X(x, A) := \min_{a \in A} d_X(x, a) \leq t\} \tag{9.4}$$

for set of all points within distance t from A. One can visualize by "expanding" A by distance t.

Theorem 9.3.1 (Isoperimetric inequality in Euclidean space)**.** Let $A \subset \mathbb{R}^n$ be a measurable set, and let $B \subset \mathbb{R}^n$ be a ball $\operatorname{vol}(A) = \operatorname{vol}(B)$. Then, for all $t \geq 0$,

$$\operatorname{vol}(A_t) \geq \operatorname{vol}(B_t).$$

Remark 9.3.2. One can recover the classic inequality on surface volumes $\operatorname{vol}_{n-1}(\delta A) \geq \operatorname{vol}_{n-1}(\delta B)$ by noting that

$$\operatorname{vol}_{n-1}(\delta A) = \frac{d}{dt}\Big|_{t=0} \operatorname{vol}_n(A_t). \lim_{t\to 0} \frac{\operatorname{vol}(A_t) - \operatorname{vol}(A)}{t} \geq \lim_{t\to 0} \frac{\operatorname{vol}(B_t) - \operatorname{vol}(B)}{t} = \operatorname{vol}_{n-1}(\delta B).$$

We have an analogous result in the $\{0,1\}^n$ with respect to Hamming distance.In Hamming cube, **Harper's theorem** gives the exact result. Below, for $A \subset \{0,1\}^n$, we write A_t as in (9.4) for $X = \{0,1\}^n$ and d_X being the Hamming distance.

Theorem 9.3.3 (Isoperimetic inequality in the Hamming cube; Harper 1966)**.** Let $A \subset \{0,1\}^n$. Let $B \subset \{0,1\}^n$ be a Hamming ball with $|A| \geq |B|$. Then for all $t \geq 0$,

$$|A_t| \geq |B_t|.$$

Remark 9.3.4. The above statement is tight when A has the same size as a Hamming ball, i.e., when $|A| = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{k}$ for some integer k. Actually, more is true. For any value of $|A|$ and t, the size of A_t is minimized by taking A to be an initial segment of $\{0,1\}^n$ according to the *simplicial ordering*: first sort by Hamming weight, and for ties, sort by lexicographic order.

It is worth examining the sizes of the Hamming ball as a function of its radius.

Let

$$B(r) = \{x \in \{0,1\}^n : \text{weight}(x) \leq r\}$$

denote the Hamming ball of radius r. Using the central limit theorem, we find that, for every fixed $z \in \mathbb{R}$, as $n \to \infty$.

$$\frac{1}{2^n}\left|B\left(\frac{n}{2} + \frac{z\sqrt{n}}{2}\right)\right| = \frac{1}{2^n}\sum_{0 \leq i \leq \frac{n}{2} + \frac{z\sqrt{n}}{2}} \binom{n}{i} \sim \mathbb{P}_{Z \sim N(0,1)}(Z \leq t) = \frac{1}{\sqrt{2\pi}}\int_0^z e^{-x^2/2}\,dx.$$

Also, by Chernoff bound, we have

$$\frac{1}{2^n}\left|B\left(\frac{n}{2} + \frac{z\sqrt{n}}{2}\right)\right| \leq e^{-z^2/2} \qquad \text{if } z \leq 0$$

and

$$\frac{1}{2^n}\left|B\left(\frac{n}{2} + \frac{z\sqrt{n}}{2}\right)\right| \geq 1 - e^{-z^2/2} \qquad \text{if } z \geq 0.$$

Combined with the isoperimetic inequality on the cube, we obtain the following surprising consequence. Suppose we start with just half of the cube, and then expand it by a bit (recall that the diameter of the cube is n, and we will be expanding it by $o(n)$), then resulting expansion occupies nearly all of the cube.

Theorem 9.3.5. Let $t > 0$. For every $A \subset \{0,1\}^n$ with $|A| \geq 2^{n-1}$, we have

$$|A_t| > (1 - e^{-2t^2/n})2^n.$$

Proof. Let $B = \{x \in \{0,1\}^n : \text{weight}(x) < n/2\}$, so that $|B| \leq 2^{n-1} \leq |A|$. Then by

Harper's theorem (Theorem 9.3.3),

$$|A_t| \geq |B_t| = |\{x \in \{0,1\}^n : \text{weight}(x) < n/2 + t\}| > (1 - e^{-2t^2/n})2^n$$

by the Chernoff bound. □

In fact, using the above, we can deduce that even if we start with a small fraction (e.g., 1%) of the cube, and expand it slightly, then we would cover most of the cube.

Theorem 9.3.6. Let $\epsilon > 0$. If $A \subset \{0,1\}^n$ with $|A| \geq \epsilon 2^n$, then

$$\left| A_{\sqrt{2\log(1/\epsilon)n}} \right| \geq (1-\epsilon)2^n.$$

First proof via isoperimetric inequality. Let $t = \sqrt{\log(1/\epsilon)n/2}$ so that $e^{-2t^2/n} = \epsilon$. Applying Theorem 9.3.5 to $A' = \{0,1\}^n \setminus A_t$, we see that $|A'| < 2^{n-1}$ (or else $|A'_t| > (1-\epsilon)2^n$, so A'_t would intersect A, which is impossible since the distance between A and A' is greater than t). Thus $|A_t| \geq 2^{n-1}$, and then applying Theorem 9.3.5 yields $|A_{2t}| \geq (1-\epsilon)2^n$. □

Let us give another proof of Theorem 9.3.6 without using Harper's exact isoperimetric theorem in the Hamming cube, and instead use the bounded differences inequality that we proved earlier.

Second proof via the bounded differences inequality. Pick random $x \in \{0,1\}^n$ and let $X = \text{dist}(x, A)$. Note that X changes by at most 1 if a single coordinate of x is changed. Applying the bounded differences inequality, Theorem 9.0.1, we have the lower tail

$$\mathbb{P}(X - \mathbb{E}X \leq -t) \leq e^{-2t^2/n}.$$

We have $X = 0$ if and only if $x \in A$, so

$$\epsilon \leq \mathbb{P}(x \in A) = \mathbb{P}(X - \mathbb{E}X \leq -\mathbb{E}X) \leq e^{-2(\mathbb{E}X)^2/n}.$$

Thus

$$\mathbb{E}X \leq \sqrt{\frac{\log(1/\epsilon)n}{2}}.$$

Now we apply the upper tail

$$\mathbb{P}(X - \mathbb{E}X \geq t) \leq e^{-2t^2/n}$$

with

$$t = \sqrt{2(\log(1/\epsilon)n} \geq 2\mathbb{E}X$$

to yield

$$\mathbb{P}(x \notin A_t) = \mathbb{P}(X > t) < \mathbb{P}\left(X \geq \mathbb{E}X + \sqrt{\frac{\log(1/\epsilon)n}{2}}\right) \leq \epsilon. \qquad \square$$

The above expansion/isoperimetry properties turn out to be actually equivalent to the concentration of Lipschitz function phenomenon we discussed earlier, as we show next. Milman recognized the importance of this **concentration of measure phenomenon**, which he heavily promoted in the 1970's. The subject was have been since then extensively developed. It plays a central role in probability theory, the analysis of Banach spaces, and it also has been influential in theoretical computer science.

Theorem 9.3.7 (Equivalence between notions of concentration of measure)**.** Let $t, \epsilon \geq 0$. In a probability space $(\Omega, \mathbb{P})$ equipped with a metric. The following are equivalent:

1. (Expansion/approximate isoperimetry) If $A \subset \Omega$ with $\mathbb{P}(A) \geq 1/2$, then
$$\mathbb{P}(A_t) \geq 1 - \epsilon.$$
2. (Concentration of Lipschitz functions) If $f\colon \Omega \to \mathbb{R}$ is 1-Lipschitz and $m \in \mathbb{R}$ satisfies $\mathbb{P}(f \geq m) \geq 1/2$ and $\mathbb{P}(f \leq m) \geq 1/2$ (i.e., m is a **median** of f), then
$$\mathbb{P}(f > m + t) \leq \epsilon.$$

Remark 9.3.8. There always exists a median, but it might not be unique. For example, for the uniform distribution on $\{0, 1\}$, any real number in the interval $[0, 1]$ is a valid median.

Proof. (a) $\implies$ (b): Let $A = \{x \in \Omega : f(x) \leq m\}$. So $\mathbb{P}(A) \geq 1/2$. Since f is 1-Lipschitz, we have $f(x) \leq m + t$ for all $x \in A_t$. Thus by (a)

$$\mathbb{P}(f > m + \epsilon) \leq \mathbb{P}(\overline{A_t}) \leq \epsilon.$$

(b) $\implies$ (a): Let $f(x) = \text{distance}(x, A)$ and $m = 0$. Since $\mathbb{P}(f \leq 0) = \mathbb{P}(A) \geq 1/2$ and $\mathbb{P}(f \geq 0) = 1$, m is a median. Also f is 1-Lipschitz. So by (b),

$$\mathbb{P}(\overline{A_t}) = \mathbb{P}(f > m + t) \leq \epsilon. \qquad \square$$

Informally, we say that a space (or rather, a sequence of spaces), has concentration of measure if ϵ decays rapidly as a function of t in the above theorem (the notion of "Lévy family" makes this precise). Earlier we saw that the Hamming cube exhibits has concentration of measure. Other notable spaces with concentration of measure include the sphere, Gauss space, orthogonal and unitary groups, postively-curved manifolds, and the symmetric group.

Mean versus median. For a sub-gaussian random variable, very tight concentration (e.g., sub-gaussian), one can deduce that the mean and the median must be very close to each other.

Indeed, suppose there exist constants $C, \sigma > 0$ such that $\mathbb{P}(A_t) \le Ce^{-(t/\sigma)^2}$ for all A with $\mathbb{P}(A) \ge 1/2$ and $t > 0$. Then for all 1-Lipschitz function f on Ω and m a median of f, one has

$$|\mathbb{E}f - m| \le \mathbb{E}|f - m| = \int_0^\infty \mathbb{P}(|f-m| \ge t)\, dt \le \int_0^\infty 2Ce^{-(t/\sigma)^2}\, dt = C\sqrt{\pi}\sigma$$

It follows that, for all $t \ge 0$,

$$\mathbb{P}(f \ge \mathbb{E}f + (t + C\sqrt{\pi})\sigma) \ge \mathbb{P}(f \ge m + t\sigma) \le Ce^{-(t/\sigma)^2}$$

and

$$\mathbb{P}(f \le \mathbb{E}f - (t + C\sqrt{\pi})\sigma) \ge \mathbb{P}(f \le m - t\sigma) \le Ce^{-(t/\sigma)^2}.$$

Similarly, if we know that $\mathbb{P}(|f - \mathbb{E}f| \ge t) \le Ce^{-(t/\sigma)}$ for all $t > 0$, then $\mathbb{P}(|f - \mathbb{E}f| \ge t) < 1/2$ for all $t > \sqrt{\log(2C)}\sigma$, from which we deduce that every median m satisfies $|\mathbb{E}f - m| \le \sqrt{\log(2C)}\sigma$.

There can indeed exist an order σ difference between the mean and the median in the setup above. For example, treating the cube as $\{-1,1\}^n$, and taking

$$f(x_1, \dots, x_n) = \max\{x_1 + \cdots + x_n, 0\},$$

we see that by the central limit theorem

$$\lim_{n\to\infty} \frac{|\mathbb{E}f - \operatorname{median}(f)|}{\sqrt{n}} = \mathbb{E}_{Z \sim N(0,1)}[\max\{Z, 0\}] = \frac{1}{\sqrt{2\pi}}.$$

9.3.1 The sphere and Gauss space

We discuss analogs of the concentration of measure phenomenon in high dimensional geometry. This is rich and beautiful subject. An excellent introductory to this topic is the survey *An Elementary Introduction to Modern Convex Geometry* by Ball (1997).

Recall the isoperimetric inequality in $\mathbb{R}^n$ says:

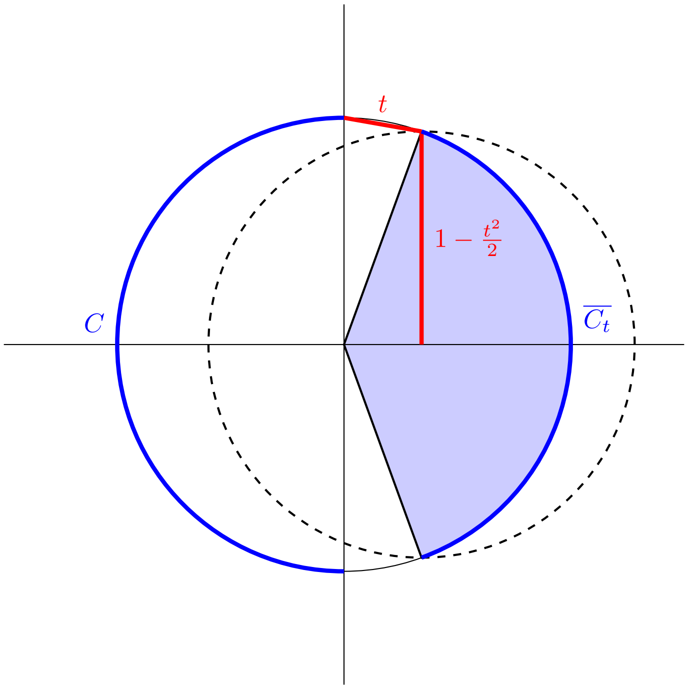
> If $A \subset \mathbb{R}^n$ has the same measure as ball B, then $\operatorname{vol}(A_t) \ge \operatorname{vol}(B_t)$ for all $t \ge 0$.

Analogous exact isoperimetric inequalities are known in several other spaces. We already saw it for the boolean cube (Theorem 9.3.3). The case of sphere and gaussian space are particularly noteworthy. The following theorem is due to Lévy.

Theorem 9.3.9 (Spherical isopeimetric inequality)**.** Inside S^{n-1} (equipped with the natural measure and distance), let A be a subset and B a spherical cap with $\operatorname{vol}_{n-1}(A) = \operatorname{vol}_{n-1}(B)$. Then for all $t \geq 0$,

$$\operatorname{vol}_{n-1}(A_t) = \operatorname{vol}_{n-1}(B_t).$$

Suppose C is a hemisphere in $S^{n-1} \subset \mathbb{R}^n$. Let us estimate $\operatorname{vol}_{n-1}(C)$. As in the diagram below, in the planar cross-section, the chord of length t subtends an angle of $\theta = 2\arcsin(t/2)$, so the vertical bolded segment has length $\cos\theta = 1 - 2\sin^2(\theta/2) = 1 - t^2/2$.



By considering the fraction of the ball subtended by $\overline{C_t}$ (i.e., the shaded wedge-sector above), which is contained in the smaller dashed ball or radius $1 - t^2/2$, we see that

$$1 - \frac{\operatorname{vol}_{n-1}(C_t)}{\operatorname{vol}_{n-1}(S^{n-1})} = \frac{\operatorname{vol}_{n-1}(\overline{C_t})}{\operatorname{vol}_{n-1}(S^{n-1})} \leq \left(1 - \frac{t^2}{2}\right)^n \leq e^{-nt^2/2}.$$

Corollary 9.3.10 (Concentration of measure on a sphere)**.** There exists some constant $c > 0$ so that

- If $A \subseteq S^{n-1}$ has $\operatorname{vol}_{n-1}(A)/\operatorname{vol}_{n-1}(S^{n-1}) \geq 1/2$, then
$$\frac{\operatorname{vol}_{n-1}(A_t)}{\operatorname{vol}_{n-1}(S^{n-1})} \geq 1 - e^{-t^2 n/2}.$$
- If $f \colon S^{n-1} \to \mathbb{R}$ is 1-Lipschitz, then there is some real m (e.g., a median) so that
$$\mathbb{P}(|f - m| > t) \leq 2e^{-nt^2/2}.$$

Second statement may be interpreted as "every Lipschitz function on a high dimensional sphere is nearby constant almost everywhere"

Another related setting is the **Gauss space**, which is $\mathbb{R}^n$ equipped with the the probability measure induced by the Gaussian random vector whose coordinates are n iid standard normals, i.e., the normal random vector in $\mathbb{R}^n$ with covariance matrix I_n. Its probability density function $(2\pi)^{-n}e^{-|x|^2/2}$ for $x \in \mathbb{R}^n$. Let λ denote the Gaussian measure on $\mathbb{R}^n$. The metric on $\mathbb{R}^n$ is the usual Euclidean metric.

What would an isoperimetric inequality in Gauss space look like?

A naive guess, inspired by $\mathbb{R}^n$, may be that disks minimize perimeter. But this is actually not the case. It turns out that the Hamming cube is a better model for the Gauss space. Indeed, consider $\{-1, 1\}^{mn}$, where both m and n are large. Let us group the coordinates of $\{-1, 1\}^{mn}$ into block of length m. The sum of entries in each block (after normalizing by $\sqrt{m}$) approximates normal random variable by the central limit theorem.

In the Hamming cube, Harper's theorem tells us Hamming balls are isoperimetric optimizers. Since a Hamming ball in $\{-1, 1\}^{mn}$ is given by all points whose sum of coordinates is below a certain threshold, we should look at the analogous subset in the Gauss space, which would then consist of all points whose sum of coordinates is below a certain threshold.

Note that the Gaussian measure is radially symmetric. So the above heuristic (which can be made rigorous) suggests that for the Gaussian isoperimetric inequality, we should look for **half-spaces**, i.e., points on one side of some hyperplane. This is indeed the case, as first shownindependently by Borell (1975) and Sudakov and Tsirel'son (1974).

Theorem 9.3.11 (Gaussian isoperimetric inequality)**.** If $A, H \subset \mathbb{R}^n$, H a half-space, and $\lambda(A) = \lambda(H)$, then $\lambda(A_t) \geq \lambda(H_t)$ for all $t \geq 0$, where λ is the Gauss measure.

Consequently, if $\mathbb{P}(A) \geq 1/2$, then $\mathbb{P}(\overline{A_t}) \leq \mathbb{P}(Z_1 > t) \leq e^{-t^2/2}$. And, if $f\colon \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz, and z is a vector of iid standard normals, then $X = f(z)$ satisfies

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 2e^{-t^2/2}$$

The sphere as approximately a sum of independent Gaussians. The gauss space is a nice space to work with because a standard normal vector simultaneously possesses two useful properties (and it is essentially the only such random vector to have both properties):

(a) Rotational invariance

(b) Independence of coordinates

Furthermore, the length of a random gaussian vector is given by $\sqrt{Z_1^2 + \cdots + Z_n^2}$ for iid $Z_1, \ldots, Z_n \in N(0,1)$, which is concentrated around $\sqrt{n}$ (e.g., by a straight forward adaptation of Chernoff bound. In fact, since $\sqrt{n + O(\sqrt{n})} = \sqrt{n} + O(1)$, the length of gaussian vector has a $O(1)$-length window of typical fluctation). So most of the distribution in the gauss space lies lie to a sphere of radius $\sqrt{n}$. Due to rotational invariance, we see that a gaussian distribution approximates the uniform distribution on sphere of radius $\sqrt{n}$ in high dimensions. Random gaussian vectors give us a convenient method to analyze the concentration of measure phenomenon on the sphere. (It should now be satisfying to see how half-spaces in the gauss space intersect the sphere in a spherical cap, and both objects are isoperimetric optimzers in their respective spaces).

9.3.2 Johnson–Lindenstrauss Lemma

The next theorem is a powerful in many areas. For example, it is widely used in computer science as a means of dimension reduction.

Theorem 9.3.12 (Johnson and Lindenstrauss 1982)**.** Let $s_1, \ldots, s_N \in \mathbb{R}^n$. Then there exists $s_1', \ldots, s_N' \in \mathbb{R}^m$ where $m = O(\epsilon^{-2} \log N)$ and such that, for every $i \neq j$,

$$(1-\epsilon)|s_i - s_j| \leq |s_i' - s_j'| \leq (1+\epsilon)|s_i - s_j|.$$

Remark 9.3.13. Here m is optimal up to a constant factor (Larsen and Nelson 2017).

The theorem is proved by obtaining the new points $s_j' \in \mathbb{R}^m$ by taking a projection onto a uniform random m-dimensional subspace (and the scaling by $\sqrt{n/m}$). We would like to know that these projects roughly preserve the length of vectors. Once we have the following lemma, set $s_i' = \sqrt{m/n} P s_i$, and we can apply the lemma to $z = s_i - s_j$ for every pair (i,j) and apply the union bound to use that, with probability at least $1-CN^2e^{-c\epsilon^2 m}$, one has $(1-\epsilon)|s_i - s_j| \le |s'_i - s'_j| \le (1+\epsilon)|s_i - s_j|$ for all (i,j).

Lemma 9.3.14 (Random projection)**.** Let P be a projection from $\mathbb{R}^n$ onto a random m-dimensional subspace. Let $z \in \mathbb{R}^n$ (fixed) and $y = Pz$. Then

$$\mathbb{E}[|y|^2] = \frac{m}{n}|z|^2$$

and, with probability $\ge 1 - 2e^{-c\epsilon^2 m}$ for some constant $c > 0$,

$$(1-\epsilon)\sqrt{\frac{m}{n}}\,|z| \le |y| \le (1+\epsilon)\sqrt{\frac{m}{n}}\,|z|\,.$$

Proof. By rescaling we may assume that $|z| = 1$.

The distribution of $Y = |y|$ does not change if we instead fix P to be the orthogonal projection onto the subspace spanned by the first m coordinate vectors, and z vary uniformly over the unit sphere.

Writing $z = (z_1, \dots, z_n)$, by symmetry we have $\mathbb{E}[z_1^2] = \cdots = \mathbb{E}[z_n^2]$. Since $z_1^2 + \cdots + z_n^2 = 1$, we have $\mathbb{E}[z_i^2] = 1/n$ for each i. Thus

$$\mathbb{E}[Y^2] = \mathbb{E}[z_1^2 + \cdots + z_m^2] = \frac{m}{n}.$$

Since the map $z \mapsto |y|$ is 1-Lipschitz, by Lévy concentration (Corollary 9.3.10),

$$\mathbb{P}\left(|Y - \mathbb{E}Y| \ge t\right) \le 2e^{-nt^2/2}, \quad \text{for all } t \ge 0.$$

In particular, we have that

$$\mathbb{E}[Y^2] - (\mathbb{E}Y)^2 = \operatorname{Var} Y = \int_0^\infty \mathbb{P}\left(|Y - \mathbb{E}Y|^2 \ge t\right)\, dt \le \int_0^\infty 2e^{-nt/2}\, dt = \frac{4}{n}.$$

So

$$\sqrt{\frac{m-4}{n}} \le \mathbb{E}Y \le \sqrt{\frac{m}{n}}.$$

This implies that, for some constants $c > 0$,

$$\mathbb{P}\left(\left|Y - \sqrt{\frac{m}{n}}\right| \ge t\right) \le 2e^{-cnt^2}, \quad \text{for all } t \ge 0.$$

Setting $t = \epsilon\sqrt{m/n}$ yields the result. □

A cute application of Johnson–Lindenstrauss (this was a starred homework exercise where you were asked to prove it using the Chernoff bound).

Corollary 9.3.15. There is a constant $c > 0$ so that for every positive integer m, there is a set of $e^{c\epsilon^2 m}$ points in $\mathbb{R}^m$ whose pairwise distances are in $[1-\epsilon, 1+\epsilon]$.

Proof. Applying Theorem 9.3.12 to the the N coordinate vectors in $\mathbb{R}^N$ yields a set of N points in $\mathbb{R}^m$ for $m = O(\epsilon^{-2}\log N)$ with pairwise distances in $[1-\epsilon, 1+\epsilon]$. □

9.4 Talagrand inequality

9.4.1 Convex Lipschitz functions of independent random variables

Problem 9.4.1. Let V be a *fixed* d-dimensional subspace. Let $x \sim \mathrm{Unif}\{-1,1\}^n$. How well is $\operatorname{dist}(x, V)$ concentrated?

Let $P = (p_{ij}) \in \mathbb{R}^{n\times n}$ be the matrix giving the orthogonal projection onto $V^\perp$. We have $\operatorname{tr} P = \dim V^\perp = n - d$. Then

$$\operatorname{dist}(x,V)^2 = |x \cdot Px| = \sum_{i,j} x_i x_j p_{ij}.$$

So

$$\mathbb{E}[\operatorname{dist}(x,V)^2] = \sum_i p_{ii} = \operatorname{tr} P = n - d.$$

How well is $\operatorname{dist}(x,V)$ concentrated around $\sqrt{n-d}$?

We say that a random variable X is **K-subgaussian** if

$$\mathbb{P}(|X - \mathbb{E}X| \ge t) \le 2e^{-t^2/K^2}.$$

Note that a K-subgaussian random variable typically has $O(K)$-fluctuation around its mean.

Let us start with some examples.

If V is some coordinate subspace, then $\operatorname{dist}(x,V)$ is a constant not depending on x.

If $V = (1,1,\dots,1)^\perp$, then $\operatorname{dist}(x,V) = |x_1 + \cdots + x_n|/\sqrt{n}$ which converge $|Z|$ for $Z \sim N(0,1)$. In particular, it is $O(1)$-subgaussian.

More generally, if for a hyperplane $V = \alpha^\perp$ for some unit vector $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{R}^n$, one has $\operatorname{dist}(x, V) = |\alpha \cdot x|$. Note that flipping x_i changes $|\alpha \cdot x|$ by at most $2|\alpha_i|$. So So the bounded differences inequality Theorem 9.0.4, for every $t \geq 0$,

$$\mathbb{P}(|\operatorname{dist}(x, V) - \mathbb{E} \operatorname{dist}(x, V)| \geq t) \leq 2 \exp\left(\frac{-2t^2}{4(\alpha_1^2 + \cdots + \alpha_n^2)}\right) \leq 2e^{-t^2/2}.$$

So again $\operatorname{dist}(x, V)$ is $O(1)$-subgaussian.

What about higher codimensional subspaces V? Then

$$\operatorname{dist}(x, V) = \sup_{\substack{\alpha \in V^\perp \\ |\alpha|=1}} |\alpha \cdot x| .$$

It is not clear how to apply the bounded difference inequality to all such α in the above supremum simultaneously.

On the other hand, if we were to ignore the α's and simply apply the bounded difference inequality to the function $x \in \{-1, 1\}^n \mapsto \operatorname{dist}(x, V)$, then, since this function is 2-Lipschitz (with respect to Hamming distance), we obtain

$$\mathbb{P}\left(|\operatorname{dist}(x, V) - \mathbb{E} \operatorname{dist}(x, V)| \geq t\right) \leq 2e^{-nt^2/2},$$

showing that $\operatorname{dist}(x, V)$ is $O(\sqrt{n})$-subgaussian—but this is a pretty bad result, as $|\operatorname{dist}(x, V)| \leq \sqrt{n}$ (half the length of the longest diagonal of the cube).

Perhaps the reason why the above bound is so poor is that the bounded difference inequality is measuring distance in $\{-1, 1\}^n$ using the Hamming distance (ℓ_1) whereas we really care about the Euclidean distance (ℓ_2).

Instead of sampling $x \in \{-1, 1\}^n$, if we had taking x to be a uniformly random point on the radius $\sqrt{n}$ sphere in $\mathbb{R}^n$ (which contains $\{-1, 1\}^n$), then Lévy concentration would imply that

$$\mathbb{P}_{x \sim \mathrm{Uniform}(\sqrt{n} S^{n-1})}(|\operatorname{dist}(x, V) - \mathbb{E} \operatorname{dist}(x, V)| \geq t) \leq 2e^{-t^2/2}.$$

So $\operatorname{dist}(x, V)$ is $O(1)$-subgaussian if x is chosen from the radius $\sqrt{n}$ sphere. Perhaps a similar bound holds when x is chosen from $\{-1, 1\}^n$?

Talagrand (1995) developed a powerful inequality that allows us to answer the above question. The most general form of Talagrand's inequality can be somewhat hard to grasp at first, though it has important combinatorial consequences. We begin with more concrete geometric special cases.

Theorem 9.4.2. Let V be a fixed d-dimensional subspace in $\mathbb{R}^n$. For uniformly random $x \in \{-1,1\}^n$, one has

$$\mathbb{P}(|\operatorname{dist}(x,V) - \sqrt{n-d}| \geq t) \leq 2e^{-ct^2}$$

where $c > 0$ is some constant.

Previously, the bounded differences inequality tells us that a Lipschitz function on $\{-1,1\}^n$ is $O(\sqrt{n})$-subgaussian.

Talagrand inequality tells us that a **convex** Lipschitz function in $\mathbb{R}^n$ is $O(1)$-subgaussian when restricted to the boolean cube. We give the precise statement below. We omit the proof of Talagrand's inequality (see Alon–Spencer textbook or Tao's blog post) and instead focus on explaining the theorem and how to apply it.

Below $\operatorname{dist}(\cdot,\cdot)$ means Euclidean distance. And $A_t = \{x : \operatorname{dist}(x,A) \leq t\}$.

Theorem 9.4.3 (Talagrand)**.** Let $A \subset \mathbb{R}^n$ be convex, and let $x \sim \operatorname{Unif}\{0,1\}^n$. Then for any $t > 0$,

$$\mathbb{P}(x \in A)\mathbb{P}(\operatorname{dist}(x,A) \geq t) \leq e^{-ct^2}$$

where $c > 0$ is some absolute constant.

Remark 9.4.4. (1) Note that A is a convex body in $\mathbb{R}^n$ and not simply a set of points in A. It may be useful to think of A as the convex hull of a set of points in $\{-1,1\}^n$. Then distance to A is not the distance to these vertices of the boolean cube, but rather distance to the convex body A.

(2) The bounded differences inequality gives us an upper bound of the form $e^{-ct^2/n}$, which is much better than Talagrand's bound.

Example 9.4.5 (Talagrand's inequality fails for nonconvex sets)**.** Let

$$A = \left\{x \in \{0,1\}^n : \operatorname{wt}(x) \leq \frac{n}{2} - \sqrt{n}\right\}$$

(here A is a discrete set of points and not their convex hull). Then for every $y \in \{0,1\}^n$ with $\operatorname{wt}(y) \geq n/2$, one has $\operatorname{dist}(y,A) \geq n^{1/4}$. Using the central limit theorem, we have, for some constant $c > 0$ and sufficiently large n, for $x \sim \operatorname{Uniform}(\{-1,1\}^n)$, $\mathbb{P}(x \in A) \geq c$ and $\mathbb{P}(\operatorname{wt}(x) \geq n/2) \geq 1/2$, so the above inequality is false for $t = n^{1/4}$.

By an argument similar to our proof of Theorem 9.3.7 (the equivalence of notions of concentration of measure), one can deduce the following consequence.

Corollary 9.4.6. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and 1-Lipschitz (with respect to Euclidean distance on $\mathbb{R}^n$). Then for any $r \in \mathbb{R}$ and $t > 0$, for $x \sim \operatorname{Unif}\{0,1\}^n$

$$\mathbb{P}(f(x) \le r)\mathbb{P}(f(x) \ge r + t) \le e^{-ct^2}.$$

where $c > 0$ is some absolute constant.

Remark 9.4.7. The proof below shows that the assumption that f is convex can be weakened to f being **quasiconvex**, i.e., $\{f \le a\}$ is convex for every $a \in \mathbb{R}$.

The versions of Talagrand inequality, Theorem 9.4.3 and Corollary 9.4.6, are equivalent:

- Theorem 9.4.3 implies Corollary 9.4.6: take $A = \{x : f(x) \le r\}$. We have $f(x) \le r + t$ whenever $\operatorname{dist}(a, A) \le t$ since f is 1-Lipschitz. So $\mathbb{P}(f(x) \le r) = \mathbb{P}(x \in A)$ and $\mathbb{P}(f(x) \ge r + t) \le \mathbb{P}(\operatorname{dist}(x, A) \ge t)$.
- Corollary 9.4.6 implies Theorem 9.4.3: take $f(x) = \operatorname{dist}(x, A)$ which is convex since A is convex.

Let us write $\mathbb{M}X$ to be a **median** for the random variable X, i.e., a non-random real so that $\mathbb{P}(X \ge \mathbb{M}X) \ge 1/2$ and $\mathbb{P}(X \le \mathbb{M}X) \ge 1/2$.

Corollary 9.4.8. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be convex and 1-Lipschitz (with respect to Euclidean distance on $\mathbb{R}^n$). Let $x \sim \operatorname{Unif}(\{0,1\}^n)$. Then

$$\mathbb{P}(|f(x) - \mathbb{M}f(x)| \ge t) \le 2e^{-ct^2}$$

where $c > 0$ is an absolute constant.

Proof. Setting $r = \mathbb{M}f(x)$ in Corollary 9.4.6 yields

$$\mathbb{P}(f(x) \ge \mathbb{M}f(x) + t) \le 2e^{-ct^2},$$

and setting $r = \mathbb{M}f(x)$ in Corollary 9.4.6 yields

$$\mathbb{P}(f(x) \ge \mathbb{M}f(x) - t) \le 2e^{-ct^2}. \qquad \square$$

Putting the two inequalities together, and changing the constant c, yields the corollary.

As an immediate corollary, we deduce Theorem 9.4.2 regarding the distance from a random point $x \in \{-1,1\}^n$ to a d-dimensional subspace. The above corollary shows that $\operatorname{dist}(x, V)$ (which is a convex 1-Lipschitz function of $x \in \mathbb{R}^n$) is $O(1)$-subgaussian, which immediately implies the result (see Lemma 9.3.14 for an example of how to argue the omitted step where we replaced $\mathbb{M}X$ by $\mathbb{E}X$ and then by $(\mathbb{E}X^2)^{1/2}$).

Example 9.4.9 (Operator norm of a random matrix)**.** Let A be a random matrix whose entries are uniform iid from $\{-1,1\}$. Viewing $A \mapsto \|A\|_{\mathrm{op}}$ as a function $\mathbb{R}^{n^2} \to \mathbb{R}$, we see that it is convex (since the operator norm is a norm) and 1-Lipschitz (using that $\|\cdot\|_{\mathrm{op}} \le \|\cdot\|_{\mathrm{HS}}$, where the latter is the Hilbert–Schmidt norm, also known as the Frobenius norm, i.e., the ℓ_2-norm of the matrix entries). It follows by Talagrand's inequality (Corollary 9.4.8) that f is $O(1)$-subgaussian.

9.4.2 Convex distance

Talagrand's inequality if much more general than what we saw earlier and can be applied to a wide variety of combinatorial applications. We need a define a more subtle notion of distance.

We consider $\Omega = \Omega_1 \times \cdots \times \Omega_n$ with product probability measure (i.e., independent random variables).

Weighted hamming distance: given $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n_{\ge 0}$, $x, y \in \Omega$, we set

$$d_\alpha(x,y) = \sum_{i=1}^n \alpha_i 1_{x_i \ne y_i}.$$

and for $A \subset \Omega$,

$$d_\alpha(x,A) = \inf_{y \in A} d_\alpha(x,y)$$

Talagrand's **convex distance** between $x \in \Omega$ and $A \subset \Omega$ is defined by

$$d_T(x,A) = \sup_{\substack{\alpha \in \mathbb{R}^n_{\ge 0} \\ |\alpha|=1}} d_\alpha(x,A)$$

(here $|\alpha|^2 = \alpha_1^2 + \cdots + \alpha_n^2$).

Example 9.4.10. If $A \subset \{0,1\}^n$ and $x \in \{0,1\}^n$, then $d_T(x,A)$ is the Euclidean distance from x to the convex hull of A.

To see why this is called a convex distance, note that to compute $d_T(x,A)$, we can convert Ω to $\{0,1\}^n$ based on their agreement with x, i.e., let $\phi_x(y) \in \{0,1\}^n$ be the vector whose i-th coordinate is 1 iff $x_i \ne y_i$. Then, $d_\alpha(x,A)$ in Ω equals to $d_\alpha(\vec{0}, \phi_x(A)) = \phi_x(A) \cdot \alpha$ in $\{0,1\}^n$. Taking the supremum over α, we see, using the Example 9.4.10,

$$d_T(x,A) = \operatorname{dist}(\vec{0}, \operatorname{ConvexHull} \phi_x(A)).$$

The general form of Talagrand's inequality says the following. Note that it reduces to the earlier special case Theorem 9.4.3 if $\Omega = \{0,1\}^n$.

Theorem 9.4.11 (General form of Talagrand's inequality)**.** Let $A \subseteq \Omega = \Omega_1 \times \cdots \times \Omega_n$, with Ω equipped with a product probability measure. Let $t \geq 0$. We have

$$\mathbb{P}(A)\mathbb{P}(x \in \Omega : d_T(x, A) \geq t) \leq e^{-t^2/4}.$$

Let us see how Talagrand's inequality recovers a more general form of our geometric inequalities from earlier, extending from independent boolean random variables to independent bounded random variables.

Lemma 9.4.12 (Convex distance upper bounds Euclidean distance)**.** Let $A \subset [0,1]^n$ and $x \in [0,1]^n$. Then $\operatorname{dist}(x, \operatorname{ConvexHull} A) \leq d_T(x, A)$.

Proof. For any $\alpha \in \mathbb{R}^n$, and any $y \in [0,1]^n$, we have

$$|(x-y)\cdot\alpha| \leq \sum_{i=1}^n |\alpha_i|\,|x_i - y_i| \leq \sum_{i=1}^n |\alpha_i|\, 1_{x_i \neq y_i}.$$

First taking the infimum over all $y \in A$, and then taking the supremum over unit vectors α, the LHS becomes $\operatorname{dist}(x, \operatorname{ConvexHull} A)$ and the RHS becomes $d_T(x, A)$. □

Corollary 9.4.13 (Convex functions of independent bounded random variables)**.** Let $x = (x_1, \dots, x_n) \in [0,1]$ be independent random variables (not necessarily identical). Let $t \geq 0$. Let $A \subset [0,1]^n$ be a convex set. Then

$$\mathbb{P}(x \in A)\mathbb{P}(\operatorname{dist}(x, A) \geq t) \leq e^{-t^2/4}$$

where dist is Euclidean distance. Also, if $f\colon [0,1]^n \to \mathbb{R}$ is a convex 1-Lipschitz function, then

$$\mathbb{P}(|f - \mathbb{M}f| \geq t) \leq 4e^{-t^2/4}.$$

9.4.3 How to apply Talagrand's inequality

Theorem 9.4.14. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ equipped with the product measure. Let $f\colon \Omega \to \mathbb{R}$ be a function. Suppose for every $x \in \Omega$, there is some $\alpha(x) \in \mathbb{R}^n_{\geq 0}$ such that for every $y \in \Omega$,

$$f(x) \leq f(y) + d_{\alpha(x)}(x, y).$$

Then, for every $t \geq 0$,

$$\mathbb{P}(|f - \mathbb{M}f| \geq t) \leq 4 \exp\left(\frac{-t^2}{4 \sup_{x \in \Omega} |\alpha(x)|^2}\right).$$

Remark 9.4.15. Note that we can use a different weight $\alpha(x)$ for each x. This will be important for applications. Intuitively, it says that the smallness (or, equivalently the largeness) of $f(x)$ can be "certified" using $\alpha(x)$.

Remark 9.4.16. By considering $-f$ instead of f, we can change the hypothesis on f to

$$f(x) \geq f(y) - d_{\alpha(x)}(x, y).$$

Note that x and y play asymmetric roles.

Remark 9.4.17 (Talagrand recovers bounded differences). By choosing a fixed $\alpha \in \mathbb{R}^n_{\geq 0}$ (not varying with x), we see that Theorem 9.4.14 recovers the bounded differences inequality Theorem 9.0.4 up to an unimportant constant factor in the exponent of the bound. The power of Talagrand's inequality is that we are allowed to vary $\alpha(x)$.

Proof. Let $r \in \mathbb{R}$. Let $A = \{y \in \Omega : f(y) \leq r - t\}$. For any $x \in \Omega$, by hypothesis, there is some $\alpha(x) \in \mathbb{R}^n_{\geq 0}$ such that, for all $y \in A$,

$$f(x) \leq f(y) + d_{\alpha(x)}(x, y) \leq r - t + d_{\alpha(x)}(x, y).$$

Taking infimum over $y \in A$, we find

$$f(x) \leq r - t + d_{\alpha(x)}(x, A) \leq r - t + |\alpha(x)|\, d_T(x, A).$$

Thus, if $f(x) \geq r$, then

$$d_T(x, A) \geq \frac{t}{|\alpha(x)|} \geq \frac{t}{\sup_x |\alpha(x)|} =: s$$

And hence by Talagrand's inequality Theorem 9.4.11,

$$\mathbb{P}(f \leq r - t)\mathbb{P}(f \geq r) \leq \mathbb{P}(A)\mathbb{P}(x \in \Omega : d_T(x, A) \geq s) \leq e^{-s^2/4}.$$

Taking $r = \mathbb{M}f + t$ yields

$$\mathbb{P}(f \geq \mathbb{M}f + t) \leq 2e^{-s^2/4}$$

and taking $r = \mathbb{M}f$ yields

$$\mathbb{P}(f \leq \mathbb{M} - t) \leq 2e^{-s^2/4}.$$

Putting them together yields the final result. □

9.4.4 Largest eigenvalue of a random matrix

Theorem 9.4.18. Let $A = (a_{ij})$ be an $n \times n$ symmetric random matrix with independent entries in $[-1, 1]$. Let $\lambda_1(X)$ denote the largest eigenvalue of A. Then

$$\mathbb{P}(|\lambda_1(A) - \mathbb{M}\lambda_1(A)| \geq t) \leq 4e^{-t^2/32}.$$

Proof. We shall verify the hypotheses of Theorem 9.4.14. We would like to come up with a good choice of a weight vector $\alpha(A)$ for each matrix A so that for any other symmetric matrix B with $[-1, 1]$ entries,

$$\lambda_1(A) \leq \lambda_1(B) + \sum_{i \leq j} \alpha_{i,j} 1_{a_{ij} \neq b_{ij}}. \tag{9.5}$$

(note that in a random symmetric matrix we only have $n(n + 1)/2$ independent random entries: the entries below the diagonal are obtained by reflecting the upper diagonal entries). Let $v = v(A)$ be the unit eigenvector of A corresponding to the eigenvalue $\lambda_1(A)$. Then, by the Courant–Fischer characterization of eigenvalues,

$$v^\intercal A v = \lambda_1(A) \qquad \text{and} \qquad v^\intercal B v \leq \lambda_1(B).$$

We have

$$\lambda_1(A) = v^\intercal A v = v^\intercal B v + v^\intercal (A - B) v \leq \lambda_1(B) + \sum_{i,j} 2\,|v_i|\,|v_j|\,1_{a_{ij} \neq b_{ij}}$$

(since $|a_{ij} - b_{ij}| \leq 2$). Thus (9.5) holds for the vector $\alpha(A) = (\alpha_{ij})_{i \leq j}$ defined by

$$\alpha_{ij} = \begin{cases} 4\,|v_i|\,|v_j| & \text{if } i < j \\ 2\,|v_i|^2 & \text{if } i = j. \end{cases}$$

We have

$$\sum_{i \le j} \alpha_{ij}^2 \le 8 \sum_{i,j} |v_i|^2 |v_j|^2 = 8 \left(\sum_i |v_i|^2 \right)^2 = 8.$$

So Theorem 9.4.14 yields the result. □

Remark 9.4.19. The above method can be adapted to prove concentration of the k-th largest eigenvalue, which is not a convex function of A, so the previous method in Example 9.4.9 does not apply.

Remark 9.4.20. If A has mean zero entries, then a moments computation shows that $\mathbb{E}\lambda_1(A) = O(\sqrt{n})$ (the constant can be computed as well). A much more advanced fact is that, say for uniform $\{-1, 1\}$ entries, the true scale of fluctuation is $n^{-1/6}$, and when normalized, the distribution converges to something called a Tracy–Widom distribution.

9.4.5 Certifiable functions and longest increasing subsequence

An **increasing subsequence** of a permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ is defined to be some $(\sigma_{i_1}, \dots, \sigma_{i_\ell})$ for some $i_1 < \cdots < i_\ell$.

Question 9.4.21. How well is the length X of the longest increasing subsequence (LIS) of uniform random permutation concentrated?

While the entries of σ are not independent, we can generate a uniform random permutation by taking iid uniform $x_1, \dots, x_n \sim \mathrm{Unif}[0, 1]$ and let σ record the ordering of the x_i's. This trick converts the problem into one about independent random variables.

The probability that there exists an increasing subsequence of length k is, by union bound, at most

$$\mathbb{P}(X \ge k) \le \frac{1}{k!}\binom{n}{k} \le \left(\frac{e}{k}\right)^k \left(\frac{ne}{k}\right)^k \le \left(\frac{e^2 n}{k^2}\right)^k .$$

It follows that $\mathbb{M}X = O(\sqrt{n})$.

Changing one of the x_i's changes LIS by at most 1, so the bounded differences inequality tells us that X is $O(\sqrt{n})$-subgaussian. Can we do better?

The assertion that a permutation has an increasing permutation of length s can be checked by verifying s coordinates of the permutation. Talagrand's inequality tells us that in such situations the typical fluctuation should be on the order $O(\sqrt{\mathbb{M}X})$, or $O(n^{1/4})$ in this case.

Definition 9.4.22. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$. Let $A \subseteq \Omega$. We say that A is ***s*-certifiable** for every $x \in A$, there exists a set $I(x) \subseteq [n]$ with $|I| \le s$ such that for every $y \in \Omega$ with $x_i = y_i$ for all $i \in I(x)$, one has $y \in A$.

Theorem 9.4.23. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ be equipped with a product measure. Let $f \colon \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose that $\{f \ge r\}$ is s-certifiable. Then, for every $t \ge 0$,

$$\mathbb{P}(f \le r - t)\mathbb{P}(f \ge r) \le e^{-t^2/(4s)}.$$

Proof. Let $A, B \subset \Omega$ be given by $A = \{x : f(x) \le r - t\}$ and $B = \{y : f(y) \ge r\}$. To apply Talagrand's inequality, Theorem 9.4.11, it suffices to show that for every $y \in B$, one has $d_T(y, A) \ge t/\sqrt{s}$, i.e., there is some $\alpha(y) \in \mathbb{R}^n_{\ge 0}$ so that

$$d_\alpha(x, y) \ge t|\alpha(y)|/\sqrt{s} \qquad \forall x \in A.$$

Indeed, let $y \in B$, and let $I(y)$ be a set of s coordinates that certify $f(y) \ge r$. Let $\alpha(y)$ be the indicator vector for $I(y)$. Note that

$$d_\alpha(x, y) = |\{i \in I(y) : x_i \ne y_i\}|\,.$$

Every $x \in A$ disagrees with y on at least t coordinates of $I(y)$, or else one can change x by fewer than t coordinates to get x' that agrees with y on I, so that $f(x') \ge r$, which contradicts f being 1-Lipschitz as $f(x) \le r - t$. It follows that

$$d_\alpha(x, y) \ge t = t|\alpha(y)|/\sqrt{s}. \qquad \square$$

Corollary 9.4.24. Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ be equipped with a product measure. Let $f \colon \Omega \to \mathbb{R}$ be 1-Lipschitz with respect to Hamming distance on Ω. Suppose $\{f \ge r\}$ is r-certifiable for every r. Then for every $t \ge 0$,

$$\mathbb{P}(f \le \mathbb{M}f - t) \le 2\exp\left(\frac{-t^2}{4\mathbb{M}f}\right).$$

and

$$\mathbb{P}(f \ge \mathbb{M}f + t) \le 2\exp\left(\frac{-t^2}{4(\mathbb{M}f + t)}\right)$$

Proof. Applying the previous theorem, we have, for every $r \in \mathbb{R}$ and every $t \geq 0$,

$$\mathbb{P}(f \leq r - t)\mathbb{P}(X \geq r) \leq \exp\left(\frac{-t^2}{4r}\right).$$

Setting $r = \mathbb{M}f$, we obtain the lower tail.

$$\mathbb{P}(f \leq \mathbb{M}f - t) \leq 2\exp\left(\frac{-t^2}{4m}\right).$$

Setting $r = m + t$, we obtain the upper tail

$$\mathbb{P}(X \geq \mathbb{M}f + t) \leq 2\exp\left(\frac{-t^2}{4(\mathbb{M}f + t)}\right). \qquad \square$$

Corollary 9.4.25. Let X be the length of the longest increasing subsequence of a random permutation of $[n]$. Then for every $\epsilon > 0$ there exists $C > 0$ so that

$$\mathbb{P}(|X - \mathbb{M}X| \leq Cn^{1/4}) \geq 1 - \epsilon.$$

Remark 9.4.26. The distribution of the length X of longest increasing subsequence of a uniform random permutation is now well understood through some deep results.

Vershik and Kerov (1977) showed that $\mathbb{E}X \sim 2\sqrt{n}$.

Baik, Deift, and Johansson (1999) showed that the correcting scaling is $n^{1/6}$, and, after under this normalization, $n^{-1/6}(X - 2\sqrt{n})$ converges to the Tracy–Widom distribution, the same distribution for the top eigenvalue of a random matrix.