Chapter 53

Talagrand's Inequality

By Sariel Har-Peled, April 26, 2022[1]

> At an archaeological site I saw fragments of precious vessels, well cleaned and groomed and oiled and spoiled. And beside it I saw a heap of discarded dust which wasn't even good for thorns and thistles to grow on.
>
> I asked: What is this gray dust which has been pushed around and sifted and tortured and then thrown away?
>
> I answered in my heart: This dust is people like us, who during their lifetime lived separated from copper and gold and marble stones and all other precious things - and they remained so in death. We are this heap of dust, our bodies, our souls, all the words in our mouths, all hopes.
>
> At an archaeological site, Yehuda Amichai

53.1. Introduction

Here, we want to prove a strong concentration inequality that is stronger than Azuma's inequality because it is independent of the underlying dimension of the process. This inequality is quite subtle, so we will need a quite elaborate way to get to it – be patient.

53.1.1. Talangrand's inequality, and the T-distance

For two numbers x, y, let $[x \neq y]$ be 1 if $x \neq y$, and 0 otherwise. For two points $\mathsf{p} = (p_1, \ldots, p_d), \mathsf{u} = (q_1, \ldots, q_d) \in \mathbb{R}^d$, let $H(\mathsf{p}, \mathsf{u})$ be the binary vector in $\{0,1\}^n$ that encodes the coordinates where they are different. Formally, we have

$$H(\mathsf{p}, \mathsf{u}) = \Big([p_1 \neq q_1], [p_2 \neq q_2], \ldots, [p_d \neq q_d] \Big). \tag{53.1}$$

For example $H\big((1,2,3),(0.1,2,-1)\big) = (1,0,1)$. Given a set $S \subseteq \mathbb{R}^d$, and a point $\mathsf{p} \in \mathbb{R}^d$, let

$$H(\mathsf{p}, S) = \{ H(\mathsf{p}, \mathsf{u}) \mid \mathsf{u} \in S \}. \tag{53.2}$$

To understand this mysterious set, consider a point $\mathsf{p} \in \mathbb{R}^d$. If $\mathsf{p} \in S$, then $(0,\ldots,0) = H(\mathsf{p},\mathsf{p}) \in H(\mathsf{p}, S)$ (which would be an uninteresting case). Otherwise, every binary point $\mathsf{x} \in H(\mathsf{p}, S)$ specifies which coordinates in p one has to change, so that one can move to a point that belongs to S.

A natural measure of the distance of p from S, is then to ask for the vector that minimizes the *Hamming distance* from the origin to $H(\mathsf{p}, S)$ – that is, the minimum number of coordinates one has to change in p to get to a point of S.

This distance measure is not informative. Think about $H(\mathsf{p}, S) = \{(0,1)\}$ and $H(\mathsf{p}, S') = \{(0,1), (1,0)\}$. In this case, the Hamming measure would rank both sets as being of equal quality (i.e., 1). But clearly, S' is closer – there are two different ways to get from p to some point of S' by changing a single coordinate.

To capture this intuition, we consider the ***convex-hull*** of these sets:

$$C(\mathsf{p}, S) = \mathcal{CH}\big(\{ H(\mathsf{p}, \mathsf{u}) \mid \mathsf{u} \in S \} \big).$$

And the corresponding ***T-distance***

$$\rho(\mathsf{p}, S) = \min_{\mathsf{u} \in C(\mathsf{p}, S)} \|\mathsf{u}\| . \qquad (53.3)$$

Observation 53.1.1. *An easy upper bound on the T-distance of p to a set S (i.e., $\rho(\mathsf{p}, S)$) is the minimum number of coordinates one has to change in p to get to a point in S. As the next example shows, however, things are more subtle – if there are many different ways to get from p to a point in S, then the T-distance is going to be significantly smaller.*

Example 53.1.2. It would be useful to understand this somewhat mysterious T-distance. To this end, consider the ball in $\mathbb{R}^d$ of radius $100d$ centered in the origin, denote by $\mathbf{b}$, and let $S = \partial\mathbf{b}$ be its boundary sphere. For a point $\mathsf{p} \in \mathrm{int}\mathbf{b}$, we have that

$$H = H(\mathsf{p}, S) = \{0, 1\}^d \setminus \{(0, 0, \ldots, 0)\} .$$

As such $C = H(\mathsf{p}, S)$ is the convex-hull of all the hypercube vertices, excluding the origin. it is easy to check that the closest point in C to the origin is the point $\mathsf{u} = (1/d, 1/d, \ldots, 1/d)$, As such, we have that $\rho(\mathsf{p}, S) = \|\mathsf{u}\| = \sqrt{1/d} = 1/\sqrt{d}$.

In particular, by monotonicity this implies that for any set T in $\mathbb{R}^d$ we have that $\rho(\mathsf{p}, T)$ is either 0 (i.e., $\mathsf{p} \in T$), or alternatively, $\rho(\mathsf{p}, T) \geq 1/\sqrt{d}$. Similarly, $\rho(\mathsf{p}, T) \leq \sqrt{d}$ as this is the maximum distance from the origin to any vertex of the hypercube $\{0, 1\}^d$.

As a concrete example, for the set $S = \partial\mathbf{b}$, and the point $\mathsf{p} = (200d, \ldots, 200d)$, we have $\rho(\mathsf{p}, S) = \sqrt{d}$.

In the following, think about the dimension d as being quite large. As such, the distance $1/\sqrt{d}$ is quite small. In particular, for a set $S \subseteq \mathbb{R}^d$, let

$$S_t = \left\{ \mathsf{p} \in \mathbb{R}^d \;\middle|\; \rho(\mathsf{p}, S) \leq t \right\} ,$$

be the expansion of S by including points that are in distance at most t from S in the T-distance.

Since we interested in probability here, consider $\mathbb{R}^d$ to be the product of d probability spaces. Formally, let Ω_i be a probability space, and consider the product probability space $\Omega = \prod_{i=1}^d \Omega_i$. As we are given a probability measure on each Ω_i, this defines a natural probability measure on Ω. That is, a point from Ω is generated by packing each of its coordinates independently from Ω_i. for $i = 1, \ldots, d$.

The ***volume*** of a set $S \subseteq \Omega$ is thus $\mathbb{P}[S]$. We are now ready to state Talagrand inequality (not that it is going to help us much).

Theorem 53.1.3 (Talagrand's inequality). *For any set $S \subseteq \Omega$, we have*

$$\mathbb{P}[S]\, \mathbb{P}\big[\overline{S_t}\big] = \mathbb{P}[S]\big(1 - \mathbb{P}[S_t]\big) \leq \exp(-t^2/4).$$

Example 53.1.4. To see why this inequality interesting, consider $\Omega = [0, 100]^d$ with uniform distribution on each coordinate. The probability measure of a set $S \subseteq \Omega$ is $\mathbb{P}[S] = \mathrm{vol}(S)/100^d$. Let

$$S = \left\{ \mathsf{p} = (p_1, \ldots, p_d) \in [0, 100]^d \;\middle|\; \sum_i p_i \leq \frac{100d}{2} \right\} .$$

It is easy to verify that $\mathrm{vol}(S) = \mathrm{vol}\big([0, 100]^d\big)/2$. Let $t = 4\sqrt{\ln d}$, and consider the set $\overline{S_t}$. Intuitively, and not quite correctly, it the set of all points in $[0, 100]^d$, such that one needs to change more than $4 \ln d$ coordinates before one can get to a point of S. These points are t-far from being in S.

By Talagrand inequality, we have that $\mathbb{P}\Big[\overline{S_t}\Big]/2 = \mathbb{P}[S]\big(1 - \mathbb{P}[S_t]\big) \leq \exp(-t^2/4) = 1/d^4$. Namely, only a tiny fraction of the cube is more than T-distance $4\sqrt{\ln d}$ from S!

Let us try to restate this – for any set S that is half the volume of the hypercube $[0, 100]^d$, the set of points in T-distance $\leq 4\sqrt{\ln d}$ in this hypercube is small.

53.1.2. On the way to proving Talagrand's inequality

The following helper result is the core of the proof of Talagrand's inequality. The reader might want to skip reading the proof of this claim, at least at first reading.

Theorem 53.1.5. *For any set $S \subseteq \Omega = \prod_{i=1}^{d} \Omega_i$, we have*

$$\mathbb{E}\left[\exp\left(\frac{\rho^2(\mathsf{p}, S)}{4}\right)\right] = \int_{\mathsf{p}\in\Omega} \exp\left(\frac{\rho^2(\mathsf{p}, S)}{4}\right) d\mathsf{p} \leq \frac{1}{\mathbb{P}[S]}.$$

Proof: The proof is by induction on the dimension d. For $d = 1$, then $\rho(\mathsf{p}, S) = 0$ if $\mathsf{p} \in S$, and $\rho(\mathsf{p}, S) = 1$ if $\mathsf{p} \notin S$. As such, we have

$$\gamma = \mathbb{E}\left[\exp\left(\frac{\rho^2(\mathsf{p}, S)}{4}\right)\right] = e^{0^2/4}\,\mathbb{P}[S] + e^{1^2/4}(1 - \mathbb{P}[S]) = \mathbb{P}[S] + e^{1/4}(1 - \mathbb{P}[S]) = f(\mathbb{P}[S]),$$

where $f(x) = x + e^{1/4}(1 - x)$. An easy argument (see Tedium 53.1.6) shows that $f(x) \leq 1/x$, which implies that $\gamma = f(\mathbb{P}[S]) \leq 1/\mathbb{P}[S]$, as claimed.

For $d = n + 1$, let $\mathcal{O} = \prod_{i=1}^{d} \Omega_i$, and $\mathcal{N} = \Omega_{d+1}$. Clearly, $\Omega = \prod_{i=1}^{d+1} \Omega_i = \mathcal{O} \times \mathcal{N}$.

$$S_{\mathcal{O}} = \{\mathsf{p} \in \mathcal{O} \mid (\mathsf{p}, y) \in S, \text{ for some } y \in \mathcal{N}\}.$$

For $\nu \in \mathcal{N}$, let

$$S(\nu) = \{\mathsf{p} \in \mathcal{O} \mid (\mathsf{p}, \nu) \in S\} \subseteq S_{\mathcal{O}}.$$

Given a point $\mathsf{z} = (\mathsf{p}, y) \in \Omega$, we can get to a point in S, either by changing the new coordinate and then moving inside the old space $\mathcal{O}$, or alternatively, keeping the new coordinate ν fixed and moving only in the old coordinates. In particular, we have that if

$$\begin{aligned} s &\in H(\mathsf{p}, S_{\mathcal{O}}) \subseteq \{0,1\}^d &&\implies && (s, 1) \in H(\mathsf{z}, S) && \text{(see Eq. (53.2))}\\ s' &\in H(\mathsf{p}, S(\nu)) &&\implies && (s', 0) \in H(\mathsf{z}, S). \end{aligned}$$

And similarly, for the corresponding convex-hulls, we have

$$s \in C(\mathsf{p}, S_{\mathcal{O}}) \implies (s, 1) \in C(\mathsf{z}, S) \qquad \text{and} \qquad s' \in C(\mathsf{p}, S(\nu)) \implies (s', 0) \in C(\mathsf{z}, S).$$

In particular, for s, s' as above, we have (by convexity) that for any $\lambda \in [0, 1]$, the point

$$h(\lambda) = (1 - \lambda)(s, 1) + \lambda(s', 0) = \big((1 - \lambda)s + \lambda s', 1 - \lambda\big) \in C(\mathsf{z}, S) \subseteq [0, 1]^{d+1}.$$

The function $\widehat{h}(\lambda) = \|(1 - \lambda)s + \lambda s'\|^2$ is convex, see Tedium 53.1.7. We thus have

$$\rho^2(\mathsf{z}, S) = \left(\min_{\mathsf{p}\in C(\mathsf{z},S)} \|\mathsf{p}\|^2\right) \leq \|h(\lambda)\|^2 = \|(1 - \lambda)s + \lambda s'\|^2 + (1 - \lambda)^2 \leq (1 - \lambda)\,\|s\|^2 + \lambda\,\|s'\|^2 + (1 - \lambda)^2.$$

We are still at the liberty of choosing s and s'. Let s be the point realizing $\rho(\mathsf{p}, S_{\mathcal{O}})$ – this is the closest point in $C(\mathsf{p}, S)$ to the origin (i.e., $\|s\| = \rho(\mathsf{p}, S_{\mathcal{O}})$). Similarly, let s' be the point realizing $\rho(\mathsf{p}, S(\nu))$. Plugging these two points into the above inequality, we have

$$\rho^2(\mathsf{z}, S) \leq (1 - \lambda)\rho(\mathsf{p}, S_{\mathcal{O}})^2 + \lambda\rho(\mathsf{p}, S(\nu))^2 + (1 - \lambda)^2.$$

Now, fix ν, and ride the following little integral:

$$
\begin{aligned}
F(\nu) &= \int_{\mathsf{p}} \exp\left(\frac{\rho^2((\mathsf{p},\nu), S)}{4}\right) \leq \int_{\mathsf{p}} \exp\left(\frac{(1-\lambda)\rho(\mathsf{p}, S_O)^2 + \lambda\rho(\mathsf{p}, S(\nu))^2 + (1-\lambda)^2}{4}\right) \\
&\leq e^{(1-\lambda)^2/4} \int_{\mathsf{p}} \exp\left(\frac{1}{4}\rho(\mathsf{p}, S_O)^2\right)^{1-\lambda} \exp\left(\frac{1}{4}\rho(\mathsf{p}, S(\nu))^2\right)^{\lambda} \\
&\leq e^{(1-\lambda)^2/4} \left[\int_{\mathsf{p}} \exp\left(\frac{1}{4}\rho(\mathsf{p}, S_O)^2\right)\right]^{(1-\lambda)} \left[\int_{\mathsf{p}} \exp\left(\frac{1}{4}\rho(\mathsf{p}, S(\nu))^2\right)\right]^{\lambda} && \text{(by Hölder's ineq (53.4))} \\
&\leq e^{(1-\lambda)^2/4} \left(\frac{1}{\mathbb{P}[S_O]}\right)^{(1-\lambda)} \left(\frac{1}{\mathbb{P}[S(\nu)]}\right)^{\lambda} = e^{(1-\lambda)^2/4} \frac{1}{\mathbb{P}[S_O]} \left(\frac{\mathbb{P}[S(\nu)]}{\mathbb{P}[S_O]}\right)^{-\lambda} && \text{(induction)} \\
&= \frac{1}{\mathbb{P}[S_O]} \cdot e^{(1-\lambda)^2/4} r^{-\lambda}, && \text{for } r = \frac{\mathbb{P}[S(\nu)]}{\mathbb{P}[S_O]}
\end{aligned}
$$

Observe that $\mathbb{P}[S_O] \geq \mathbb{P}[S(\nu)]$, and thus $r \leq 1$. To minimize the above, consider the function $f_3(\lambda, r) = \exp\big((1-\lambda)^2/4\big) r^{-\lambda}$. Easy calculation shows that $f_3(\lambda, r)$ is minimized, for a fixed r, by choosing

$$
\lambda(r) = \begin{cases} 1 + 2\ln r & r \in [e^{-1/2}, 1] \\ 0 & r \in [0, e^{-1/2}] \end{cases},
$$

see Tedium 53.1.9 (A). Furthermore, for this choice of λ, easy calculations shows that $f_4(r) = f_r(\lambda(r), r) \leq 2 - r$, see Tedium 53.1.9 (B). As such, we have

$$
F(\nu) \leq \frac{1}{\mathbb{P}[S_O]} f_4(r) \leq \frac{1}{\mathbb{P}[S_O]} \left(2 - \frac{\mathbb{P}[S(\nu)]}{\mathbb{P}[S_O]}\right)
$$

We remind the reader that our purpose is to bound

$$
\begin{aligned}
&\int_{\mathsf{z}} \exp\left(\frac{\rho^2(\mathsf{z}, S)}{4}\right) = \int_{\nu\in\mathcal{N}} \int_{\mathsf{p}\in O} \exp\left(\frac{\rho^2((\mathsf{p},\nu), S)}{4}\right) \leq \int_{\nu\in\mathcal{N}} F(\nu) \leq \int_{\nu\in\mathcal{N}} \frac{1}{\mathbb{P}[S_O]} \left(2 - \frac{\mathbb{P}[S(\nu)]}{\mathbb{P}[S_O]}\right) \\
&= \frac{1}{\mathbb{P}[S_O]} \left(2 - \frac{\int_{\nu\in\mathcal{N}} \mathbb{P}[S(\nu)]}{\mathbb{P}[S_O]}\right) = \frac{1}{\mathbb{P}[S_O]} \left(2 - \frac{\mathbb{P}[S]}{\mathbb{P}[S_O]}\right) = \frac{1}{\mathbb{P}[S]} \cdot \frac{\mathbb{P}[S]}{\mathbb{P}[S_O]} \left(2 - \frac{\mathbb{P}[S]}{\mathbb{P}[S_O]}\right) \leq \frac{1}{\mathbb{P}[S]},
\end{aligned}
$$

since for $x = \mathbb{P}[S]/\mathbb{P}[S_O]$, we have $x(2-x) \leq 1$, for any value of x (see Tedium 53.1.10). ■

53.1.2.1. The low level details used in the above proof

Tedium 53.1.6. Let $f(x) = x + e^{1/4}(1-x)$. We claim that, for $x \in (0,1]$, $f(x) \leq 1/x$. Indeed, set $g(x) = 1/x$, and observe that $f(1) = 1 = 1/1 = g(1)$. We have that $f'(x) = e^{1/4} - 1 \approx -0.284$ and $g'(x) = -1/x^2$. In particular, $g'(x) \leq f'(x)$, for all $x \in (0,1)$. Since and $f(1) = g(1)$. it follows that $f(x) \leq g(x)$, for $x \in (0,1]$.

Tedium 53.1.7. For any $\mathsf{p}, \mathsf{u} \in \mathbb{R}^d$, the function $f(\lambda) = \|(1-\lambda)\mathsf{p} + \lambda\mathsf{u}\|^2$ is convex. Indeed, let $f_i(\lambda) = \big((1-\lambda)p_i + \lambda q_i\big)^2$, for $i = 1, \ldots, d$. Observe that $f(\lambda) = \sum_i f_i(\lambda)$, and as such it is sufficient to prove that f_i is convex. We have $f_i'(\lambda) = 2(q_i - p_i)\big((1-\lambda)p_i + \lambda q_i\big)$, and $f_i''(\lambda) = 2(q_i - p_i)^2 > 0$, which implies convexity.

Fact 53.1.8 (Hölder's inequality.). *Let $p, q \geq 1$ be two numbers such that $1/p + 1/q = 1$. Then, for any two functions f, g, we have $\|fg\|_1 \leq \|f\|_p \|g\|_q$. Explicitly, stated as integrals, Hölder's inequality is $\int |f(x)g(x)| \mathrm{d}x \leq \left(\int |f(x)|^p \mathrm{d}x\right)^{1/p} \left(\int |f(x)|^q \mathrm{d}x\right)^{1/q}$. In particular, for $\lambda \in (0,1)$, $p = 1/(1-\lambda)$ and $q = 1/\lambda$, we have that*

$$\int \left|f^{\lambda}(x) g^{1-\lambda}(x)\right| \mathrm{d}x \leq \left(\int |f(x)| \mathrm{d}x\right)^{1-\lambda} \left(\int |f(x)| \mathrm{d}x\right)^{\lambda}. \tag{53.4}$$

Tedium 53.1.9. (A) We need to find the minimum of the following function $f(\lambda) = \exp\left((1-\lambda)^2/4\right) r^{-\lambda} = \exp\left((1-\lambda)^2/4 - \lambda \ln r\right)$. We have $f'(\lambda) = f_3(\lambda)\left((1-\lambda)/2 - \ln r\right)$. Solving for $f'(\lambda) = 0$, we have $(1-\lambda)/2 - \ln r = 0 \implies 1 - \lambda = 2 \ln r 0 \implies \lambda = 1 - 2\ln r$, which works as long as $r \geq e^{-1/2}$. Otherwise, we set $\lambda = 0$.

(B) For $r \leq e^{-1/2}$, we have, by the above, that $f(0) = e^{1/4} \approx 1.28 \leq 1.39 \approx 2 - e^{-1/2} \leq 2 - r$. For $r > e^{-1/2}$, by the above, $\lambda = 1 - 2\ln r$, and thus

$$g(r) = f(\lambda) = \exp\left((1-\lambda)^2/4 - \lambda \ln r\right) = \exp\left((2\ln r)^2/4 + (1 - 2\ln r)\ln r\right) = \exp\left(\ln r - \ln^2 r\right) \leq 1 \leq 2 - r,$$

since $\ln r - \ln^2 r \leq \ln r \leq 0$, for $r \in (0,1]$.

Tedium 53.1.10. The function $f(x) = x(2-x) = 2x - x^2$ is a parabola with a maximum at $2x = 2 \implies x = 1 \implies \forall y \quad f(y) \leq f(1) = 1$.

53.1.3. Proving Talagrand's inequality

Proving Talagrand's inequality is now easy peasy.

Talagrand's inequality restatement (Theorem 53.1.3). For any set $S \subseteq \Omega$, we have

$$\mathbb{P}[S]\,\mathbb{P}[\overline{S_t}] = \mathbb{P}[S]\left(1 - \mathbb{P}[S_t]\right) \leq \exp(-t^2/4).$$

Proof: Consider a random point $\mathsf{p} \in \Omega$. We are interested in the probability $\mathsf{p} \notin S_t$. To this end, consider the random variable $X = \rho(\mathsf{p}, S)$. By definition, $\mathsf{p} \in \overline{S_t} \iff X \geq t$. As such, by Markov's inequality, we have

$$\mathbb{P}\left[\overline{S_t}\right] = \mathbb{P}[X \geq t] = \mathbb{P}\left[\exp\left(X^2/4\right) \geq \exp\left(t^2/4\right)\right] \leq \frac{\mathbb{E}\left[\exp\left(X^2/4\right)\right]}{\exp(t^2/4)} \leq \frac{\exp\left(-t^2/4\right)}{\mathbb{P}[S]},$$

by Theorem 53.1.5. ∎

53.2. Concentration via certification

Example 53.2.1. Consider the process of throwing m balls into n bins. The ith ball X_i is uniformly distributed in $\Omega_i = [\![n]\!]$. For $\mathbf{x} = (X_1, \ldots, X_m) \in \Omega = \prod_{i=1}^{m} \Omega_i$, let $h(\mathbf{x})$ be the number of bins that are not empty. If $h(\mathbf{x}) \geq k$, then there is a set $I = \{i_1, \ldots, i_k\}$ of k indices, such that for any two distinct $i, j \in I$, we have that $X_i \neq X_j$.

Namely, I is a "compact" proof/certificate that $h(\mathbf{x}) \geq k$. Furthermore, if for $\mathbf{y} = (Y_1, \ldots, Y_m) \in \Omega$ we have that $X_\alpha = Y_\alpha$, for all $\alpha \in I$, then $h(\mathbf{y}) \geq k$. Here, the certificate for a value k, was a set of size k.

Definition 53.2.2. Let $\Omega = \prod_{i=1}^{m} \Omega_i$. For a function $h : \Omega \to \mathbb{N}$, it is ***f-certifiable***, for a function $f : \mathbb{N} \to \mathbb{N}$, if whenever $h(\mathbf{x}) \geq k$, there exists a set $I \subseteq [\![m]\!]$, with $|I| \leq f(k)$, such that, for any $\mathbf{y} \in \Omega$, if $\mathbf{y}$ agree with $\vec{x}$ on the coordinates of I, then $h(\vec{y}) \geq k$.

Example 53.2.3. In Example 53.2.1, the function h (i.e., number of bins that are not empty) is f-certifiable, where $f(k) = k$.

Example 53.2.4. Consider the random graph $\mathsf{G}(n, p)$ over n vertices, created by picking every edge with probability p. One can interpret such a graph as a random binary vector with $\binom{n}{2}$ coordinates, where the ith coordinate is 1 $\iff$ the ith edge is in the graph (for some canonical ordering of all possible $\binom{n}{2}$ edges).

A ***triangle*** in a graph G is a triple of vertices i, j, k, such that $ij, jk, ki \in \mathsf{E}(\mathsf{G})$. For a graph G, let $h(\mathsf{G})$ be the number of distinct triangles in G. In the above interpretation as a graph as a vector $\mathbf{x} \in \{0, 1\}^{\binom{n}{2}}$, it is easy to verify that if $h(\mathsf{G}) \geq k$ then it can be certified by $3k$ coordinates. As such, the number of triangles in a graph is f-certifiable, for $f(k) = 3k$.

Note, that the certificate is only for the lower bound on the value of the function.

We need the following reinterpretation of the T-distance.

Lemma 53.2.5. *Consider a set $S \subseteq \mathbb{R}^d$ and a point $\mathsf{p} \in \mathbb{R}^d$. We have that $\rho(\mathsf{p}, S) \leq t \iff$ for all $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$, with $\|\mathbf{x}\| = 1$, there exists $\mathbf{h} \in H(\mathsf{p}, S)$, such that $\langle \mathbf{x}, \mathbf{h} \rangle \leq t$.*

Proof: The quantity $\ell = \rho(\mathsf{p}, S)$ is the distance from the origin to the convex polytope $C(\mathsf{p}, S)$. In particular, let $\mathbf{y}$ be the closest point to the origin in this polytope, and observe that $\ell = \|\mathbf{y}\| = \langle \mathbf{y}, \mathbf{y}/\|\mathbf{y}\| \rangle$. In particular, for any other vector $\mathbf{x}$, with $\|\mathbf{x}\| = 1$, we have $\langle \mathbf{y}, \mathbf{x} \rangle \leq \langle \mathbf{y}, \mathbf{y}/\|\mathbf{y}\| \rangle \leq \ell$. Since $\mathbf{y}$ is in the convex-hull of $H(\mathsf{p}, S)$, it follows that there is $\mathbf{h} \in H(\mathsf{p}, S)$ such that $\langle \mathbf{y}, \mathbf{x} \rangle \leq \langle \mathbf{h}, \mathbf{x} \rangle \leq \ell$.

As for the other direction, assume that $\ell = \rho(\mathsf{p}, S) > t$, and let $\mathbf{y} \in C(\mathsf{p}, S)$ be the point realizing this distance. Arguing as above, we have that for the direction $\mathbf{y}/\|\mathbf{y}\|$, and any vertex $\mathbf{h} \in H(\mathsf{p}, S)$ we have that $\langle \mathbf{h}, \mathbf{x} \rangle \geq \langle \mathbf{h}, \mathbf{x} \rangle \geq \ell > t$. ∎

Theorem 53.2.6. *Consider a probability space $\Omega = \prod_{i=1}^{m} \Omega_i$, and let $h : \Omega \to$ be 1-Lipschitz and f-certifiable, for some function f. Consider the random variable $X = h(\mathbf{x})$, for $\mathbf{x}$ picked randomly in Ω. Then, for any positive real numbers b and t, we have*

$$\mathbb{P}\Big[X \leq b - t\sqrt{f(b)}\Big] \mathbb{P}[X \geq b] \leq \exp(-t^2/4).$$

If h is k-Lipschitz then $\mathbb{P}\Big[X \leq b - tk\sqrt{f(b)}\Big] \mathbb{P}[X \geq b] \leq \exp(-t^2/4)$.

Proof: Set $S = \left\{ \mathsf{p} \in \Omega \;\middle|\; h(\mathsf{p}) < b - t\sqrt{f(b)} \right\}$. Consider a point u, such that $h(\mathsf{u}) \geq b$. Assume for the sake of contradiction that $\mathsf{u} \in S_t$. Let $I \subseteq [\![m]\!]$ be the certificate of size $\leq f(b)$ that $h(\mathsf{u}) \geq b$. And consider the vector $\mathbf{x} = (x_1, \ldots, x_d)$, such that $x_i = 1/\sqrt{|I|}$ if $i \in I$, and $x_i = 0$ otherwise. Observe that $\|\mathbf{x}\|^2 = |I| \, (1/|I|) = 1$, and thus $\|\mathbf{x}\| = 1$. By Lemma 53.2.5, there exists $\mathbf{h} \in H(\mathsf{u}, S)$, such that $\langle \mathbf{x}, \mathbf{h} \rangle \leq t$, since by assumption $\rho(\mathsf{u}, S) \leq t$. Let $\mathsf{v} \in S$ be the point realizing $\mathbf{h}$ – that is, $H(\mathsf{p}, \mathsf{v}) = \mathbf{h}$.

Let $J \subseteq I$ be the set of indices of coordinates that are in I, such that p and v differ on this coordinate. We have by the definition of $\mathbf{x}$, that $|J| / \sqrt{|I|} \leq \langle \mathbf{x}, \mathbf{h} \rangle \leq t$, which implies that $|J| \leq t\sqrt{|I|} \leq t\sqrt{f(b)}$.

Let u' be the point that agrees with u on the coordinates of I, and agrees with v on the other coordinates. The points u' and v disagree only on coordinates in I, but such coordinates of disagreement are exactly the coordinates (in I) where u disagrees with v – which is the set J of coordinates. As such, by the 1-Lipschitz condition, we have that

$$h(\mathsf{v}) \geq h(\mathsf{u}') - |J| \geq h(\mathsf{u}) - t\sqrt{f(b)},$$

but then, by the definition of S, we have $\mathsf{v} \notin S$, which is a contradiction as $\mathsf{v} \in S$.

We conclude that $\mathsf{u} \notin S_t \implies \mathsf{u} \in \overline{S_t}$. As such, we have $\mathbb{P}[X \geq b] \leq \mathbb{P}\left[\overline{S_t}\right]$. By Talagrand inequality, we have

$$\mathbb{P}\left[X < b - t\sqrt{f(b)}\right] \mathbb{P}[X \geq b] \leq \mathbb{P}[S]\, \mathbb{P}\left[\overline{S_t}\right] \leq \exp(-t^2/4).$$

The "$<$" on the left side can be replaced by "$\leq$", as in the statement of the theorem, by using the value $t + \varepsilon$ instead of t, and taking the limit as $\varepsilon \to 0$.

The k-Lipschitz version follows by applying the above inequality to the function $h(\cdot)/k$. ■

53.3. Some examples

Definition 53.3.1. For a random variable $X \in \mathbb{R}$, let $\mathrm{med}(X)$ denote the maximum number m, such that $\mathbb{P}[X < m] \leq 1/2$ and $\mathbb{P}[X > m] \leq 1/2$. The number $\mathrm{med}(X)$ is the ***median*** of X.

53.3.1. Longest increasing subsequence

Let $\mathbf{x} = (X_1, \ldots, X_n)$ be a vector of n numbers picked randomly and uniformly from $[0, 1]$. Let $h(\mathbf{x})$ be the longest increasing subsequence in the associated sequence.

Lemma 53.3.2. *We have $h(\mathbf{x}) = \Theta(\sqrt{n})$ with high probability. Furthermore, for some constant c and any $t > 0$, we have that $\mathbb{P}\left[|h(\mathbf{x}) - \mathrm{med}(h(\mathbf{x}))| - tcn^{1/4}\right] \leq 4\exp(-t^2/4)$, Namely, the random variable $h(\mathbf{x})$ is strongly concentrated.*

Proof: Let Y_i be an indicator variable that is 1 $\iff$ $\mathbf{x}[i] \equiv x_{(i-1)\sqrt{n}+1}, \ldots, x_{i\sqrt{n}}$ contains a number in the interval $J(i) = [(i-1)/\sqrt{n}, i\sqrt{n}]$. We have $\mathbb{P}[Y_i = 1] = 1 - (1 - 1/\sqrt{n})^{\sqrt{n}} \geq 1 - 1/e \geq 1/2$, since $(1 - 1/m)^m \leq 1/e$. If Y_i happens, then we can take the number in $\mathbf{x}[i]$ that falls in $J(i)$, and add it to the generated sequence. As such, the length of the generated sequence, which is increasing, is $Y = \sum_{i=1}^{n} Y_i$. And in particular, $\mathbb{E}[Y] \geq \sqrt{n}/2$, and Chernoff's inequality implies that $\mathbb{P}\left[Y \geq (1 - \delta)\sqrt{n}/2\right] \leq \exp(-\delta^2\sqrt{n}/8)$.

The upper bound is more interesting. The probability that a specific subsequence of t indices $i_1 < i_2 < \ldots < i_t$ form an increasing subsequence $X_{i_1} < X_{i_2} < \cdots < X_{i_t}$ is $1/t!$. As such, the expected number of such increasing sequences of length $\geq \ell$ is bounded by

$$\alpha = \sum_{t=\ell}^{n} \binom{n}{t} \frac{1}{t!} \leq \sum_{t=\ell}^{n} \left(\frac{ne}{t}\right)^t \frac{1}{t!} \leq \sum_{t=\ell}^{n} \left(\frac{ne}{t}\right)^t \frac{1}{(t/e)^t} = \sum_{t=\ell}^{n} \frac{n^t e^{2t}}{t^{2t}},$$

using Lemma 53.6.1. In particular, for $\ell = 4e\sqrt{n}$, we have

$$\alpha \leq \sum_{t=\ell}^{n} \frac{n^t e^{2t}}{\left(4e\sqrt{n}\right)^{2t}}, = \sum_{t=\ell}^{n} \frac{1}{4^{2t}} \leq 2/4^{8e\sqrt{n}} \ll 1.$$

By Markov's inequality, this implies that $\mathbb{P}\left[h(\mathbf{x}) \geq 4e\sqrt{n}\right] \leq 2/4^{8e\sqrt{n}}$.

The above readily implies that $\nu = \text{med}(h(\mathbf{x})) = \Theta(\sqrt{n})$. Furthermore, $h(\mathbf{x})$ is $f(x) = x$ certifiable, and it is 1-Lipschitz. Theorem 53.2.6 now implies that

$$\mathbb{P}\big[h(\mathbf{x}) \le \nu - t\sqrt{\nu}\big]/2 \le \mathbb{P}\big[h(\mathbf{x}) \le \nu - t\sqrt{\nu}\big]\mathbb{P}[X \ge \nu] \le \exp(-t^2/4).$$

As $\nu = O(n^{1/4})$, we get the following (this requires some further tedious calculations which we omit).

$$\mathbb{P}\Big[\big|h(\mathbf{x}) - \nu\big| - tcn^{1/4}\Big] \le 4\exp(-t^2/4),$$

where c is some constant. ■

53.3.2. Largest convex subset

A set of points P is in ***convex position*** if they are all vertices of the convex-hull of P.

Lemma 53.3.3. *Let P be a set of n points picked randomly and uniformly in the unit square $[0,1]^2$. Let Y be the size of the largest subset of point of P that are in convex position. Then, we have that $\mathbb{E}[Y] = \Omega(n^{1/3})$.*

Proof: Let $\mathsf{p} = (1/2, 1/2)$, and consider the regular N-gon Q, for $N = n^{1/3}$, that its vertices lie on the circle centered at p, and is of radius $r = 1/2$. Consider the triangle $\triangle_i$ formed by connecting three consecutive vertices $\mathsf{p}_{2i-1}, \mathsf{p}_{2i}, \mathsf{p}_{2i+1}$ of Q. We have that $\alpha = 2\pi/N$, and we pick n large enough, so that $\alpha \le 1/4$. We remind the reader that $1 - x^2/4 \ge \cos x \ge 1 - x^2/2$, for $x \in (0, 1/4)$. As such, we have that $\alpha^2/4 \le 1 - \cos\alpha \le \alpha^2/2$. In particular, this implies that the height of $\triangle$ is $h = r(1 - \cos(\alpha))$, and we have $\alpha^2/8 = r\alpha^2/4 \le h \le r\alpha^2/2$.

Let $\ell = \|\mathsf{p}_{2i-1} - \mathsf{p}_{2i+1}\| = 2r\sin\alpha$, since $x/2 \le \sin(x) \le x$, we have that $\alpha/2 \le \ell \le \alpha$. As such, we have that

$$\text{area}(\triangle_i) = h\ell/2 \ge (\alpha^2/8)(\alpha/2) = \alpha^3/16 = (2\pi/N)^3/16 = 8/n.$$

In particular, the probability that $\triangle_i$ does not contain a point of P is at most $(1 - \text{area}(\triangle_i))^n \le (1 - 8/n)^n \le \exp(-8)$. We conclude that, in expectation, at least $(1 - \exp(-8))N/2$ triangles contains points of P. Selecting a point of P from each such triangle results in a convex subset, which implies the claim. ■

It is not hard to show that $Y = \Omega(n^{1/3})$, with high probability, see Exercise 53.5.1. This readily implies that $\text{med}(Y) = \Omega(n^{1/3})$. It is significantly harder, but known, that $\mathbb{E}[Y] = O(n^{1/3})$, see [Val95]. We provide a weaker but easier upper bound next.

Lemma 53.3.4. *Let P be a set of n points picked randomly and uniformly in the unit square $[0,1]^2$. Let Y be the size of the largest subset of point of P that are in convex position. Then, $\mathbb{E}[Y] = O(n^{1/3}\log n/\log\log n)$, with high probability.*

Proof: Let V be a set of directions of size $O(n^c)$, where c is some constant, such that for any unit vector u, there is a vector in $\mathbf{v} \in V$, such that the angle between u and v is at most $1/n^c$. For a vector $\mathbf{v} \in V$, consider the grid $G(\mathbf{v})$ with directions $\mathbf{v}$, and orthogonal direction $\mathbf{v}^\perp$. Every cell of this grid is a rectangle with sidelength $1/n^{1/3}$ in the direction of $\mathbf{v}$, and $1/n^{2/3}$ in the orthogonal direction. In addition the origin is a vertex of $G(\mathbf{v})$. This grid is uniquely defined, and every cell in this grid has sidelength 1. The of number of grid cells of this grid intersecting the unit square is $O(n)$, as can be easily verified.

Let $\mathcal{F}$ be the set of all rectangles in all these grids that intersect the unit square. Clearly, the number of such cells is $O(|V| n) = O(n^{c+1})$. Each rectangle in $\mathcal{F}$ has area $1/n$, and as such by expectation it contains ≤ 1 point of P (the inequality is there because the rectangle might be partially outside the unit square). A standard application of Chernoff's inequality implies that the probability that a rectangle of $\mathcal{F}$ contains more than $10c \log n/\log\log n$ points of P is $\leq 1/n^{2c}$. As such, with high probability no rectangle in $\mathcal{F}$ contains more than $O(\log n/\log\log n)$ points of P.

Consider any convex body $C \subseteq [0,1]^2$. The key observation is that ∂C can be covered by $O(n^{1/3})$ rectangles of $\mathcal{F}$. Indeed, the perimeter of C is at most 4. As such, place $O(n^{1/3})$ points along ∂C that are at distance at most $1/(10n^{1/3})$ from each other. Similarly, place additional $O(n^{1/3})$ points on ∂C, such that the angle of the tangents between two consecutive points is at most $1/n^{1/3}$ (in radians) [a vertex of C might be picked repeatedly]. Let Q be the resulting set of points. Consider two consecutive points $\mathsf{p}, \mathsf{u} \in \mathsf{Q}$ along ∂C, and observe that the distance between them is at most $1/(10n^{1/3})$, and the angle between their two tangents is at most $\alpha = 1/n^{1/3}$. consider the triangle $\triangle$ formed by the two tangents to ∂C at p, u, and the segment p, u. This triangle has height bounded by $\|\mathsf{p} - \mathsf{u}\| \sin \alpha \leq 1/(10n^{2/3})$. It is now straightforward, if somewhat tedious to argue that one of the rectangles of $\mathcal{F}$ must contain $\triangle$.

Now we are almost done – if the maximum cardinality convex subset $\mathsf{Q} \subseteq \mathsf{P}$ was larger than $c' n^{1/3} \log n/\log\log n$, for some constant c', then let C be the convex-hull of this large subset. The above would imply that one of the rectangles of $\mathcal{F}$ must contain at least $\Omega(c' \log n/\log\log n)$ points of P, but this does not happen with high probability, for c' sufficiently large. Thus implying the claim. ■

In particular, the above implies that $\mathrm{med}(Y) = O(n^{1/3} \log n)$.

Theorem 53.3.5. *Let P be a set of n points picked randomly and uniformly in the unit square $[0,1]^2$. Let Y be the size of the largest subset of point of P that are in convex position. Then, for any $t > 0$, we have*

$$\mathbb{P}\Big[|Y - \mathrm{med}(Y)| \geq tcn^{1/6} \log^{1/2} n\Big] \leq 2\exp(-t^2/4),$$

for some constant c.

Proof: Observe that Y is 1-Lipschitz (i.e., changing the location of one point in P can decrease or increase the value of Y by at most 1. In addition Y is 1-certifiable, since we only need to list the points that form the convex subset. As such, Theorem 53.2.6 applies. Setting $b = \mathrm{med}(Y)$, we have by the above that $\mathrm{med}(Y) = \Omega(n^{1/3})$ and $\mathrm{med}(Y) = O(n^{1/3} \log n)$. As such, we have

$$\mathbb{P}\Big[Y \leq \mathrm{med}(Y) - t\sqrt{cn^{1/3} \log n}\Big]\mathbb{P}[X \geq \mathrm{med}(Y)] \leq \exp(-t^2/4).$$

Similarly, setting $b = \mathrm{med}(Y) + t\sqrt{cn^{1/3} \log n} \leq 2\mathrm{med}(Y)$, we have

$$\mathbb{P}[Y \leq \mathrm{med}(Y)]\mathbb{P}\Big[X \geq \mathrm{med}(Y) + t\sqrt{cn^{1/3} \log n}\Big] \leq \exp(-t^2/4).$$

Putting the two inequalities together, we get

$$\mathbb{P}\Big[|Y - \mathrm{med}(Y)| \geq t\sqrt{cn^{1/3} \log n}\Big] \leq 2\exp(-t^2/4).$$ ■

53.3.3. Balls into bins revisited

Given n balls, one throw them into b bins, where $b \geq n$. A ball that falls into a bin with i or more balls is ***i-heavy***. Let $h_{\geq i}$ be the number of i-heavy balls. It turns out that a strong concentration on $h_{\geq i}$ follows readily from Talagrand's inequality.

Lemma 53.3.6. *Consider throwing n balls into b bins, where $b \geq 3n$. Then, $e^{-2}F_i \leq \mathbb{E}[h_{\geq i}] \leq 6e^{i-1}F_i$, where $h_{\geq i}$ is the number of i-heavy balls, and $F_i = n(n/ib)^{i-1}$. Let β_i denote the expected number of pairs of i-heavy balls that are colliding. We have that $\beta_i = O\bigl(ni(en/ib)^{i-1}\bigr)$.*

Proof: Let $p = 1/b$. A specific ball falls into a bin with exactly i balls, if there are $i-1$ balls, of the remaining $n-1$ balls that falls into the same bin. As such, the probability for that is $\gamma_i = p^{i-1}(1-p)^{n-i}\binom{n-1}{i-1}$. As such, a specific ball is i-heavy with probability

$$\alpha = \sum_{j=i}^{n} \gamma_j = \sum_{j=i-1}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-j-1} \leq \sum_{j=i-1}^{n-1} \left(\frac{e(n-1)}{jb}\right)^j \leq 2\left(\frac{en}{b(i-1)}\right)^{i-1} \leq 6\left(\frac{en}{ib}\right)^{i-1},$$

as $(n/i)^i \leq \binom{n}{i} \leq (en/i)^i$. Since $(1-p)^{n-j-1} \geq (1-1/b)^{b-1} \geq 1/e$, we have

$$\alpha \geq \frac{1}{e}\sum_{j=i-1}^{n-1} \left(\frac{n-1}{jb}\right)^j \geq \frac{1}{e}\left(\frac{n-1}{n} \cdot \frac{n}{(i-1)b}\right)^{i-1} \geq \frac{1}{e^2}\left(\frac{n}{ib}\right)^{i-1}.$$

As such, we have $\mathbb{E}[h_{\geq i}] = n\alpha = \Theta(n(n/b)^{i-1})$.

If a ball is in a bin with exactly j balls, for $j \geq i$, then it collides directly with $j-1$ other i-heavy balls. Thus, the expected number of collisions that a specific ball has with i-heavy balls is in expectation $\sum_{j=i}^{n}(j-1)\gamma_j = \sum_{j=i-1}^{n-1} j\gamma_{j+1}$. Summing over all balls, and dividing by two, as every i-heavy collision is counted twice, we have that the expected overall number of such collisions is

$$\beta_i = \frac{n}{2}\sum_{j=i-1}^{n-1} j\gamma_{j+1} = \frac{n}{2}\sum_{j=i-1}^{n} j\binom{n-1}{j}p^j(1-p)^{n-j-1} = O\left(ni\left(\frac{en}{ib}\right)^{i-1}\right). \qquad \blacksquare$$

Lemma 53.3.7. *Consider throwing n balls into b bins, where $b \geq 3n$. Let i be a small constant integer, $h_{\geq i}$ be the number of i-heavy balls, and let $\nu_i = \mathrm{med}(h_{\geq i})$. Assume that $\nu_i \geq 16i^2c\log n$, where c is some arbitrary constant. Then, for some constant c', we have that $|\nu_i - \mathbb{E}[h_{\geq i}]| \leq c'i\sqrt{\nu_i}$, and*

$$\mathbb{P}\Bigl[|h_{\geq i} - \nu_i| \geq 6i\sqrt{c\nu_i \ln n}\Bigr] \leq \frac{1}{n^c} \qquad \textit{and} \qquad \mathbb{P}\Bigl[\bigl|h_{\geq i} - \mathbb{E}[h_{\geq i}]\bigr| \geq c'i\sqrt{\nu_i} + 6i\sqrt{c\nu_i \ln n}\Bigr] \leq \frac{1}{n^c}.$$

Proof: Observe that $h_{\geq i}$ is 1-certifiable – indeed, the certificate is the list of indices of all the balls that are contained in bins with i or more balls. The variable $h_{\geq i}$ is also i-Lipschitz. Changing the location of a single ball, can make one bin that contains i balls, into a bin that contains only $i-1$ balls, thus decreasing $h_{\geq i}$ by i.

We require that $ti\sqrt{\nu_i} \leq \nu_i/2 \implies t \leq \sqrt{\nu_i}/(2i)$. Theorem 53.2.6 implies that

$$\mathbb{P}\Bigl[h_{\geq i} \leq \nu_i - ti\sqrt{\nu_i}\Bigr] \leq 2\exp(-t^2/4). \tag{53.5}$$

Setting $b = \nu_i + 2ti\sqrt{\nu_i}$, we have that

$$b - ti\sqrt{b} \geq b - ti\sqrt{\nu_i + 2ti\sqrt{\nu_i}} \geq b - ti\sqrt{2\nu_i} = \nu_i + 2ti\sqrt{\nu_i} - ti\sqrt{2\nu_i} \geq \nu_i.$$

This implies that $\mathbb{P}[h_{\geq i} \geq b]/2 \leq \mathbb{P}[h_{\geq i} \leq \nu_i]\,\mathbb{P}[X \geq b] \leq \mathbb{P}\Big[h_{\geq i} \leq b - tk\sqrt{b}\Big]\mathbb{P}[h_{\geq i} \geq b] \leq \exp(-t^2/4)$. We conclude that

$$\mathbb{P}\Big[h_{\geq i} \geq \nu_i + 2ti\sqrt{\nu_i}\Big] \leq 2\exp(-t^2/4). \tag{53.6}$$

Combining the above, we get that

$$\mathbb{P}\Big[|h_{\geq i} - \nu_i| \geq 2ti\sqrt{\nu_i}\Big] \leq 4\exp(-t^2/4)$$

We require that $4\exp(-t^2/4) \leq 1/\mathsf{n}^c$, which holds for $t = 3\sqrt{c\ln n}$. We get the inequality $\mathbb{P}\Big[|h_{\geq i} - \nu_i| \geq 6i\sqrt{c\nu_i \ln n}\Big] \leq 1/\mathsf{n}^c$, as claimed.

This in turn translates into the requirement that $3\sqrt{c\ln n} \leq \sqrt{\nu_i}/(2i)$. $\implies$ $6i\sqrt{c\ln n} \leq \sqrt{\nu_i}$. $\implies$ $36i^2 c\ln n \leq \nu_i$.

Next, we estimate the expectation. We have that

$$\mathbb{E}[h_{\geq i}] \geq \nu_i - \sum_{t=1}^{\infty} ti\sqrt{\nu_i}\,\mathbb{P}\Big[h_{\geq i} \leq \nu_i - (t-1)i\sqrt{\nu_i}\Big] \geq \nu_i - i\sqrt{\nu_i}\sum_{t=1}^{\infty} t2\exp(-(t-1)^2/4) \geq \nu_i - 10i\sqrt{\nu_i},$$

by Eq. (53.5). Similarly, by Eq. (53.6), we have

$$\mathbb{E}[h_{\geq i}] \leq \nu_i + \sum_{t=1}^{\infty} 2ti\sqrt{\nu_i}\,\mathbb{P}\Big[h_{\geq i} \geq \nu_i + (t-1)i\sqrt{\nu_i}\Big] \leq \nu_i + 4i\sqrt{\nu_i}\sum_{t=1}^{\infty} t\exp(-(t-1)^2/4) \leq \nu_i + 20i\sqrt{\nu_i},$$

As such, we have that $|\mathbb{E}[h_{\geq i}] - \nu_i| \leq 30i\sqrt{\nu_i}$, namely $c' \leq 30$. Combining the above inequalities implies the statement of the lemma. ∎

Example 53.3.8. Consider throwing n into $\mathsf{b} = \mathsf{n}^{4/3}$ bins. Lemma 53.3.6 implies that $e^{-2}F_i \leq \mathbb{E}[h_{\geq i}] \leq 6e^{i-1}F_i$, where $F_i = \mathsf{n}/(i\mathsf{b}^{1/3})^{i-1}$. As such $\mathbb{E}[h_{\geq 2}] = \Theta(\mathsf{n}^{2/3})$, $\mathbb{E}[h_{\geq 3}] = \Theta(\mathsf{n}^{1/3})$, and $\mathbb{E}[h_{\geq 4}] = \Theta(1)$.

Applying Lemma 53.3.7, we get that the number of balls that collides (i.e., $h_{\geq 2}$), is strongly concentrated around some value $\nu_2 = \Theta(\mathsf{n}^{2/3})$, with the interval where it lies being of length $O\big(\mathsf{n}^{1/3}\sqrt{\log n}\big)$.

53.4. Bibliographical notes

Our presentation follows closely Alon and Spencer [AS00]. Section 53.3.3 is from Har-Peled and Jones [HJ18].

53.5. Problems

Exercise 53.5.1. Elaborating on the argument of Lemma 53.3.3, prove that, with high probability, a random set of points picked uniformly in the unit square contains a convex subset of size $\Omega(n^{2/3})$.

53.6. From previous lectures

Lemma 53.6.1. *For any positive integer n, we have:*

(i) $(1 + 1/n)^n \leq e$.

(ii) $(1 - 1/n)^{n-1} \geq e^{-1}$.

(iii) $n! \geq (n/e)^n$.

(iv) For any $k \leq n$, we have: $\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$.

References

[AS00] N. Alon and J. H. Spencer. *The probabilistic method*. 2nd. Wiley InterScience, 2000.

[HJ18] S. Har-Peled and M. Jones. *On separating points by lines*. *Proc. 29th ACM-SIAM Sympos. Discrete Algs.* (SODA), 918–932, 2018.

[Val95] P. Valtr. *Probability that n random points are in convex position*. *Discrete Comput. Geom.*, 13(3): 637–643, 1995.