MA-862		2024
	Talagrand's Inequality	
Lecturer: Niranjan Balachandran		Scribe: Om Swostik and Pradyumna

1 Lecture 1

Suppose $(\Omega_i, \mathbb{P}_i, \rho_i)$ are metric spaces, where ρ_i are metrics. We have, $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \rho_i, \prod_{i=1}^n \mathbb{P}_i)$ is the product (METRIC) probability space. Recall McDiarmid's inequality,

Theorem 1 (McDiarmid). *Let $f : \prod_{i=1}^n \Omega_i \to \mathbb{R}$ be lipschitz, i.e.*

$$|f(\underset{\sim}{x}) - f(\underset{\sim}{y})| \le \rho(\underset{\sim}{x}, \underset{\sim}{y}), \qquad \forall\, \underset{\sim}{x}, \underset{\sim}{y} \in \prod_{i=1}^n \Omega_i.$$

If $\underset{\sim}{x} \in \prod_{i=1}^n \Omega_i$ is picked according to $\prod_{i=1}^n \mathbb{P}_i$, and if f is bounded, then

$$\mathbb{P}(|f(x) - \mathbb{E}f(x)| > t) \le 2e^{-t^2/2n}.$$

Remark. *The above says that sufficiently smooth functions are heavily concentrated around their mean in these product spaces.*

Suppose $(\Omega_i, \mathbb{P}_i)$ are probability spaces, consider the product space $(\prod_{i=1}^n \Omega_i, \prod_{i=1}^n \mathbb{P}_i)$. A "natural" metric for this comes from the Hamming metric (counting coordinates where they differ),

$$d_H(\underset{\sim}{x}, \underset{\sim}{y}) = \#\{i : x_i \neq y_i\}.$$

This may not always be a smart choice! Suppose $\Omega_i = \{0, 1\}, \forall\, 1 \le i \le n$. Let $\Omega = \prod_i \Omega_i = \{0, 1\}^n$ and $\mathbb{A} = \{\underset{\sim}{x} \in \{0, 1\}^n : |\underset{\sim}{x}| \le (n/2)\}$, where $|\underset{\sim}{x}| = \#\{i : x_i \neq 0\}$. We have, $\mathbb{P}(\mathbb{A}) = 1/2$. Pick $\underset{\sim}{x}$ at random according to $\mathbb{P}$. Since the random variable $|\underset{\sim}{x}|$ is binomial, an application of Chernoff gives,

$$\mathbb{P}(|\underset{\sim}{x}| > n/2 + t) \le e^{-t^2/2n}.$$

The above bound doesn't take into account that there might be a LOT of points with $|\underset{\sim}{x}| = (n/2+t)$! We want a notion of distance which takes this information into account. This motivates the following,

Definition 2 (Talagrand convex distance). *Given $(\Omega_i, \mathbb{P}_i)$, let $\mathbb{A} \subseteq \Omega = \prod_{i=1}^n \Omega_i$ and $\underset{\sim}{x} \in \Omega$. Let $r \in \mathbb{R}^n$ such that $r \ge 0$ and $\|r\|_2 = 1$. We define,*

$$\rho_0(\underset{\sim}{x}, \mathbb{A}) := \max_r \min_{y \in \mathbb{A}} \langle r, h(x, y) \rangle.$$

Here $r = (r_1, r_2, \ldots, r_n)$ is the 'cost' vector.

We define the set $A_t = \{\underset{\sim}{y} \in \Omega : \rho_0(\underset{\sim}{y}, \mathbb{A}) \le t\}$, for $t > 0$. Following is the main version of Talagrand's inequality [1],

Theorem 3 (Talangrand (95)). *Let $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. We have,*

$$\mathbb{P}(\mathbb{A}) \cdot (1 - \mathbb{P}(A_t)) \le e^{-t^2/4}.$$

1.1 A Combinatorialist Version of Talagrand

For the purpose of applications, we will look at a different formulation of the Talagrand bound.

Definition 4. *A random variable $X : \Omega \to \mathbb{R}$ is f-certifiable (for a function f) if whenever $X \geq s$, then there exists,*

- $X(w_1, w_2, \ldots, w_n) \geq s$
- $I \subseteq [n]$ *with* $|I| \leq f(s)$ *s.t. for any* w' *with* $w'_i = w_i \;\forall\, i \in I$, $X(w') \geq s$.

Remark. *The notion of f-certifiability becomes weak if the function f attains large values.*

Following is the widely used combinatorialist version of Talagrand,

Theorem 5 (Talagrand)**.** *Let $\Omega = \prod_{i=1}^n \Omega_i$ and $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. If X is lipschitz and r-certifiable (i.e. $f(s) = rs$), then*

$$\mathbb{P}(|X - \mathbb{E}X| > t + 60\sqrt{r\mathbb{E}X}) \leq e^{-t^2/8r\mathbb{E}X}.$$

If the expectation $\mathbb{E}X$ is linear in n, the above bound is similar to McDiarmid. If it isn't linear in n, the above bound is better! As some summary of our discussion so far, we have the following remark,

Remark. *The shortcoming of Hamming is that it cannot tell if there are lots of points at same distance away from the set. Talagrand takes this account and averages it out, exploiting more information and hence giving a better bound.*

2 Lecture 2

Given $(\Omega_i, \mathbb{P}_i)$, let $\Omega = \prod_{i=1}^n \Omega_i$ and $\mathbb{P} = \prod_{i=1}^n \mathbb{P}_i$. Fix $\mathbb{A} \subseteq \Omega$, $x \in \Omega$ and vector $r = (r_1, \ldots, r_n) \geq 0$. Recall the hamming difference vector $h(x, y) = (h_1, h_2, \ldots, h_n)$ such that $h_i = 1$ if $x_i \neq y_i$ and 0 otherwise. For the rest of this lecture, we denote $\|x\| = \|x\|_2$ in $\mathbb{R}^n$.

Definition 6. *Define the set,*

$$\mathcal{U}'_A(x) = \{h(x, y) \in \{0, 1\}^n \mid y \in A\}.$$

Recall Talagrand's notion of distance,

Definition 7 (Talagrand convex distance)**.**

$$\begin{aligned}\rho_r(x, A) &\coloneqq \min\{\langle r, h(x, y)\rangle : y \in A\}\\ \rho_0(x, A) &\coloneqq \max_r\{\rho_r(x, A) : r \geq 0,\ \|r\| = 1\}\end{aligned}$$

The following theorem gives an equivalent characterization for the Talagrand distance $\rho_0(x, A)$.

Theorem 8.

$$\begin{aligned}\rho_0(x, A) &= \min\{\|z\| \ : \ z \in CH(\mathcal{U}'_A(x))\}\\ &= \min\{\|z\| \ : \ z \in CH(\mathcal{U}_A(x))\}\end{aligned}$$

where $\mathcal{U}_A(x)$ denotes the UPSET[1] *generated by $\mathcal{U}'_A(x)$ and CH denotes the convex hull of a set.*

[1] $\mathcal{U}_A(x) = \{(z_1, \ldots, z_n) : \exists (w_1, \ldots, w_n) \in \mathcal{U}'_A(x) \text{ s.t. } z_i \geq w_i \ \ \forall i\}$

Proof. First we have the following claim,

Claim 9.

$$\rho_r(x, A) = \min\{\langle r, z\rangle \ : \ z \in CH(\mathcal{U}'_A(x))\} \tag{1}$$

$$= \min\{\langle r, z\rangle \ : \ z \in CH(\mathcal{U}_A(x))\}. \tag{2}$$

Proof. Suppose minimum of RHS in Equation (1) is attained at z. Then $z = \sum_i \lambda_i h(x, y_i)$, for $y_i \in A$, $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. So $\langle r, z\rangle = \sum_i \lambda_i \langle r, h(x, y_i)\rangle$, which gives that $\langle r, z\rangle \geq \langle r, h(x, y_i)\rangle$ for some i. This gives Equation (1). Similarly, suppose minimum of RHS in Equation (2) is attained at $z_0 = \sum_i \lambda_i z_i$, where $z_i \in \mathcal{U}_A(x)$. We have,

$$\langle r, z_0\rangle = \sum_i \lambda_i \langle r, z_i\rangle \geq \sum_i \lambda_i \langle r, h(x, y_i)\rangle$$

where $\langle r, z_i\rangle \in \mathcal{U}_A(x)$ and $\langle r, h(x, y_i)\rangle \in \mathcal{U}'_A(x)$. We are done as before, proving the claim. □

We have the following claim which completes the proof of the theorem,

Claim 10.

$$\rho_0(x, A) = \max_r\{\rho_r(x, A) \ : \ r \geq 0, \|r\| = 1\} \tag{3}$$

$$= \min\{\|z\| \ : \ z \in CH(\mathcal{U}_A(x))\}. \tag{4}$$

Proof. Equation (3) follows from the definition of $\rho_0(x, A)$. For $r \geq 0$, $\|r\| = 1$ and $z \in \mathrm{CH}(\mathcal{U}_A(x))$, we have

$$\langle r, z\rangle \leq \|z\|$$

from Cauchy-Schwarz. This gives $\rho_0(x, A) \leq \min\{\|z\| \ : \ z \in \mathrm{CH}(\mathcal{U}_A(x))\}$. Suppose the minimum in Equation (4) is attained at $z_0 \in \mathrm{CH}(\mathcal{U}_A(x))$. Then for any $z \in \mathrm{CH}(\mathcal{U}_A(x))$ and $\forall\, 0 \leq \lambda \leq 1$, we have

$$\lambda z + (1 - \lambda) z_0 \in \mathrm{CH}(\mathcal{U}_A(x)).$$

Since $\|z_0\|$ is a minimum,

$$\|z_0\|^2 \leq \|\lambda z + (1 - \lambda) z_0\|^2.$$

Following this, set $p(\lambda) = \|\lambda z + (1-\lambda) z_0\|^2$ which is a quadratic in λ. Differentiating and evaluating at the minimum of $p(\lambda)$ gives,

$$\|z_0\|^2 \leq \langle z, z_0\rangle.$$

Since $\rho_0(x, A) \leq z_0$, $z_0 = 0$ gives $\rho_0(x, A) = 0$ and the claim is proved. Therefore, WLOG let $z_0 \neq 0$. Set $r = z_0/\|z_0\|$. Since $\rho_r(x, A) = \min\{\langle r, z\rangle : z \in \mathrm{CH}(\mathcal{U}_A(x))\}$, we have

$$\frac{\langle z_0, z\rangle}{\|z_0\|} \geq \|z_0\|$$

$$\Rightarrow \rho_r(x, A) \geq \|z_0\|$$

$$\Rightarrow \rho_0(x, A) \geq \|z_0\|.$$

This completes the proof of the claim. □

From the two claims above, proof of Theorem 8 follows. □

2.1 Talagrand's inequality

For any $A \subseteq \Omega$ and $t \geq 0$, let $A_t = \{w \in \Omega : \rho_0(w, A) \leq t\}$. Recall Talagrand's inequality,

Theorem 11 (Talagrand)**.** *For A, A_t as above, we have*

$$\mathbb{P}(A).\mathbb{P}(\overline{A_t}) \leq e^{-t^2/4}.$$

Let $w \in \Omega$ and $f : \mathbb{R} \to \mathbb{R}^+$ be some function. We have the following definition,

Definition 12 (f-certifiability)**.** *A random variable X $(= X(w))$ is said to be f-certifiable if $X(w) \geq s \Rightarrow \exists I \subseteq [n]$ with $|I| \leq f(s)$ s.t. for ANY w' that agrees with w on I, $X(w') \geq s$ as well.*

The following is another version of the inequality,

Theorem 13. *If X is 1-lipschitz and f-certifiable, then for any b, we have*

$$\mathbb{P}\big(X < b - t\sqrt{f(b)}\big).\mathbb{P}\big(X \geq b\big) \leq e^{-t^2/4}.$$

Proof. Set $A = \{w : X(w) < b - t\sqrt{f(b)}\}$. We shall show that $\{w : X(w) \geq b\} \subseteq \overline{A_t}$. Then the conclusion will follow from Theorem 11. Suppose $X(w) \geq b$, we need to show that $w \notin A_t$. Suppose not *i.e.* $w \in A_t$, or equivalently $\rho_0(w, A) \leq t$. Since X is f-certifiable, $\exists I \subseteq [n]$ with $|I| = f(b)$ s.t. any w' agreeing with w on I must also have $X(w') \geq b$. Set

$$r = \frac{\mathbb{1}_{i \in I}}{\sqrt{|I|}}_{i=1\ldots n}.$$

By our assumption that $w \in A_t$, there exists $y \in A$ such that $\langle r, h(w, y)\rangle \leq t$. Then the number of coordinates (in I) on which y and w disagree is no more than $t\sqrt{|I|} \leq t\sqrt{f(b)}$. Now pick $z \in \Omega$ such that $z_i = y_i$ for all $i \notin I$ and $z_i = w_i$ for $i \in I$. Since z disagrees with y on no more than $t\sqrt{f(b)}$ coordinates and X is 1-lipschitz, we have $|X(z) - X(y)| \leq t\sqrt{f(b)}$. But since $y \in A$, we have $X(y) < b - t\sqrt{f(b)}$, so by the closeness of $X(y)$ and $X(z)$ we have $|X(z)| < b$. But since z agrees with w on the coordinates of I, f-certifiability guarantees that $X(z) \geq b$, and we have a contradiction. $\square$

Remark. *In particular, if $b = Med[X]$ in Theorem 13,*

$$\mathbb{P}\big(X < Med[X] - t\sqrt{f(Med[X])}\big) \leq O(e^{-t^2/4}) \text{ and}$$
$$\mathbb{P}\big(X > Med[X] + t\sqrt{f(Med[X])}\big) \leq O(e^{-t^2/4})$$

which essentially gives the concentration of the random variable X around its median. Note that $\mathbb{P}(X \leq Med[X]) = \mathbb{P}(X \geq Med[X]) = 1/2$.

We have the following Corollary to Theorem 13,

Corollary 14. *If X is lipschitz and r-certifiable (i.e. $f(s) = rs$), then*

$$\mathbb{P}\big(|X - \mathbb{E}X| \geq t + 60\sqrt{r\mathbb{E}X}\big) \leq e^{-t^2/8r\mathbb{E}X}.$$

FACT: If X is r-certifiable and lipschitz, then $|\mathbb{E}X - \text{Med}[X]| \leq O(\sqrt{\mathbb{E}X})$.

2.2 An Application: Longest Increasing Subsequence

Suppose $\pi \in S_n$ is chosen at random, let $X(\pi)$ = length of a longest monotone subsequence in π. The following theorem gives a lower bound on the longest monotone subsequence in any sequence.

Theorem 15 (Erdós-Szekeres). *Any real sequence of length (n^2+1) has a monotone subsequence of length $\geq (n+1)$.*

From Theorem 15, we have $X(\pi) \geq \sqrt{n-1}+1$. Also, we will show that $X(\pi) \leq 3\sqrt{n}$ holds WHP. Let $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{U}[0,1]$, this gives an uniform $\pi \in S_n$. Note that,

$$\mathbb{P}(X \geq k) \leq \binom{n}{k} \cdot \frac{1}{k!} \leq \frac{(en)^k}{k^k} \cdot \frac{2^k}{k^k} = \left(\frac{2en}{k^2}\right)^k.$$

Setting $k = 3\sqrt{n}$ gives an upper bound of $O((2e/9)^{\sqrt{n}})$, which tends to zero exponentially quickly with increasing n. Continuing, we also have $\mathbb{E}X$ is $O(\sqrt{n})$ since $\sqrt{n} \leq X \leq 3\sqrt{n}$ holds WHP. The random variable X is 1-lipschitz as changing the position of any one coordinate in the permutation π makes the length of the longest monotone subsequence go up or down by atmost 1. We also have that X is 1-certifiable, which follows from definition. Now we apply Talagrand bound on the random variable X. This will show that X lies in an interval of length $O(n^{1/4})$ around $\mathbb{E}X$, WHP. Set $t = Cn^{1/4}\sqrt{\log(n)}$. We have,

$$\mathbb{P}(|X - \mathbb{E}X| \geq t + 60\sqrt{r\mathbb{E}X}) \leq e^{-t^2/8r\mathbb{E}X}.$$

where $r = 1$. Since $\mathbb{E}X$ is $O(\sqrt{n})$, we have the result.

Remark. *Notice that McDiarmid is weak! Applying McDiarmid gives,*

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

To ensure concentration, we are forced to choose $t >> \sqrt{n}$.

References

[1] Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81:73–205, 1995.