Lecture 10: Talagrand Inequality and Applications
Today we shall see (without proof) a concentration inequality called the “Talagrand Inequality”

This result shall help us prove concentration of a large class of problems around its median

As an application, we shall see a concentration result for the longest increasing subsequence
Recall the definition of the Hamming distance between two elements \( x, y \in \Omega := \Omega_1 \times \cdots \times \Omega_n \)

\[
|\{i : 1 \leq i \leq n \text{ and } x_i \neq y_i\}|
\]

Intuitively, the strings get penalized “1” for every index \( i \) where \( x_i \) and \( y_i \) are different.

We can consider a weighted variant of this distance where every index \( i \) has its own associated penalty \( \alpha_i \).

Before we proceed to developing this new notion of distance, let us first normalize the Hamming distance. Consider the following redefinition. Let \( \alpha = (\alpha_1, \ldots, \alpha_n) = \left(\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}}\right) \).

We define

\[
d_H(x, y) = \sum_{1 \leq i \leq n : x_i \neq y_i} \alpha_i
\]
For the sake of completeness, we write down the inequality that we saw on Hamming distance in this new form

$$\mathbb{P} [X \in A] \cdot \mathbb{P} [d_H(X, A) \geq E] \leq \exp(-E^2/2)$$

Now, we are at a position to generalize the notion of distance to any vector \( \alpha \) with norm 1. That is, consider \( \alpha = (\alpha_1, \ldots, \alpha_n) \) such that

- \( \alpha_1, \ldots, \alpha_n \geq 0 \), and
- \( \sum_{i=1}^{n} \alpha_i^2 = 1 \).

We define the following distance between \( x, y \in \Omega \) with respect to \( \alpha \) as follows

$$d_{\alpha}(x, y) := \sum_{1 \leq i \leq n: x_i \neq y_i} \alpha_i$$

Intuitively, this captures the fact that every coordinate \( i \) could possibly be penalized differently as compared to other coordinates.
Now, for a pair \(x, y\) we consider the “most severe penalty.”

**Definition (Convex Distance)**

For \(x, y \in \Omega\), we define the convex distance between \(x\) and \(y\) as follows

\[
d_T(x, y) := \sup_{\alpha: \|\alpha\|_2 = 1} d_{\alpha}(x, y)
\]

Similar to the case of Hamming distance, we can define the distance of \(x \in \Omega\) from a set \(A \subseteq \Omega\)

\[
d_T(x, A) = \min_{y \in A} d_T(a, y)
\]

So, if \(d_T(x, A) \geq t\), then we have \(d_T(x, y) \geq t\), for all \(y \in A\).
Talagrand Inequality

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random variable over $\Omega$, such that each $X_i$ is independent of the others and $X_i \in \Omega_i$
- Let $f : \Omega \to \mathbb{R}$
- Talagrand inequality states that if any $A \subseteq \Omega$ is dense, then it is unlikely that $\mathbf{X}$ is far (w.r.t. the $d_T(\cdot, \cdot)$ distance) from $A$

Theorem (Talagrand Inequality)

For any $A \subseteq \Omega$, we have

$$\mathbb{P}[\mathbf{X} \in A] \cdot \mathbb{P}[d_T(\mathbf{X}, A) \geq E] \leq \exp(-E^2/4)$$
Let us first formulate the longest increasing subsequence problem. Suppose $X = (X_1, \ldots, X_n)$, where each $X_i$ is independent and uniformly distribution over $\Omega_i = [0, 1)$.

We are interested in $f(X)$, the length of the longest increasing subsequence in $(X_1, \ldots, X_n)$.

Let us try to understand the expected value $\mathbb{E}[f(X)]$ and its concentration that we can conclude from the previous tools that we have studied.

Note that $f$ is $(1, 1, \ldots, 1)$ bounded difference function, because changing one entry in $X$ can change the longest increasing subsequence by at most 1. So, we can apply the independent bounded difference inequality to conclude the following

$$\mathbb{P}\left[f(X) \geq \mathbb{E}[f(X)] + E\right] \leq \exp\left(-\frac{2E^2}{n}\right)$$

Talagrand Inequality
Note that the radius of concentration that we obtain from the inequality is (roughly) $\sqrt{n}$.

Although, this result is non-trivial, it is useless. Because we have $E[f(X)] = \Theta(\sqrt{n})$. Students are highly encouraged to prove this result.

Our objective is to use the Talagrand inequality to prove a concentration of $f(X)$ around its median $m$ with radius of concentration $\sqrt{m}$. Note that by the Markov inequality, we have $m \leq 2E[f(X)]$, hence, $m$ and $E[f(X)]$ have the same order. Therefore, the radius of concentration is $\Theta(n^{1/4})$. Now, this result is useful.
Our objective is to get a concentration inequality of $f(X)$.

- Define $B_a = \{y : y \in \Omega \text{ and } f(y) \leq a\}$
- Suppose we prove the following claim

**Claim (A Technical Claim)**

$$P[f(X) \leq a] \cdot P[f(X) \geq a + E] \leq P[X \in B_a] \cdot P[d_T(X, B_a) \geq \frac{E}{\sqrt{a + E}}].$$

- Using this technical claim, let us get our concentration inequalities for the distribution $f(X)$
- Note that Talagrand inequality is applicable to the right-hand side of the claim. Therefore, we get

$$P[f(X) \leq a] \cdot P[f(X) \geq a + E] \leq P[X \in B_a] \cdot P\left[d_T(X, B_a) \geq \frac{E}{\sqrt{a + E}}\right] \leq \exp\left(-\frac{E^2}{4(a + E)}\right).$$

**Talagrand Inequality**
Bounding the upper tail. Set $a = m$, the median of the distribution $f(X)$. Then, we have

$$P[f(X) \leq a] = P[f(X) \leq m] \geq 1/2.$$  

Next, using the inequality, we get

$$P[f(X) \geq m + E] \leq \frac{\exp \left( - \frac{E^2}{4(m+E)} \right)}{P[f(X) \leq m]} \leq 2 \exp \left( - \frac{E^2}{4(m + E)} \right).$$
Bounding the lower tail. Set $a + E = m$, the median of the distribution $f(X)$. Then, we have
\[ \mathbb{P} [ f(X) \geq a + E ] = \mathbb{P} [ f(X) \geq m ] \geq 1/2. \] Next, using the inequality, we get
\[
\mathbb{P} [ f(X) \leq a ] = \mathbb{P} [ f(X) \leq m - E ] \\
\leq \exp \left( -\frac{E^2}{4m} \right) \\
\leq \frac{\mathbb{P} [ f(X) \geq m ]}{\mathbb{P} [ f(X) \geq m ]} \\
\leq 2 \exp \left( -\frac{E^2}{4m} \right). 
\]

Therefore, all that remains is to prove the technical claim.

Remark. We did not use any “special property” of the function $f(\cdot)$. For a particular function $f(\cdot)$, if we can prove the technical claim, then we are done!
Remark. This concentration is around the median (*not the mean*). However, by Markov inequality, we know that the median cannot be much larger than the mean.
In this part of the lecture we will prove the technical claim for the particular function $f(\cdot)$ that outputs the length of the longest subsequence of its input bitstring.

**Proof outline.**

- Recall that we need to prove

$$
P [ f(X) \leq a ] \cdot P [ f(X) \geq a + E ] \leq P [ X \in B_a ] \cdot P \left[ d_T(X, B_a) \geq \frac{E}{\sqrt{a + E}} \right].$$

- By definition, the event “$f(X) \leq a$” is equivalent to the event “$X \in B_a$.” Therefore, proving the technical claim is equivalent to proving the inequality

$$
P [ f(X) \geq a + E ] \leq P \left[ d_T(X, B_a) \geq \frac{E}{\sqrt{a + E}} \right].$$
Observe that if an event $A$ implies an event $B$, then $\mathbb{P}[A] \leq \mathbb{P}[B]$. Therefore, it suffices to prove that the event 
"$f(X) \geq a + E$" implies the event 
"$d_T(X, B_a) \geq \frac{E}{\sqrt{a+E}}$"

In the rest of the lecture, we prove this implication

Proof.

Suppose $X = (X_1, \ldots, X_n)$, where each $X_i$ is independent and uniformly distributed over $\Omega_i = [0, 1)$

We are interested in demonstrating a concentration bound for $f(X)$, where $f(X)$ is the longest increasing subsequence in $(X_1, \ldots, X_n)$

**Observation.** Consider any $x \in \Omega := \Omega_1 \times \cdots \times \Omega_n$. If $f(x) = k$ (i.e., the longest increased subsequence in $x$ is $k$), then there is a set $K_x = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ such that $K_x$ denotes the indices of the longest increasing subsequence in $x$
Observation. Consider any $y \in \Omega$. Note that if $y$ agrees with $x$ at all the indices in $K_x$, then we have $f(y) \geq f(x)$ (it is possible that $y$ has a longest increasing subsequence, but, definitely, it will not be shorter than the length of the longest increasing subsequence in $x$)

Observation. Let us generalize the previous observation further. Consider any $y \in \Omega$. Note that if $y$ agrees with $x$ at all indices in $K_x$ except at $\ell$ indices. Then, we have $f(y) \geq f(x) - \ell$. Formally, we can write this as follows

$$f(y) \geq f(x) - |\{i : i \in K_x \text{ and } x_i \neq y_i\}|$$
Intuitively, we incur a penalty for every $i \in K_x$ where $x$ and $y$ differ. Let us fix $\alpha_x = (\alpha_1, \ldots, \alpha_n)$ such that

$$\alpha_i = \begin{cases} 0 & i \notin K_x \\ \frac{1}{\sqrt{|K_x|}} & i \in K_x \end{cases}$$

Note that $|K_x| = f(x)$. So, we conclude that

$$f(y) \geq f(x) - \sqrt{f(x)} d_{\alpha_x}(x, y)$$

Rearranging, we get that

$$d_{\alpha_x}(x, y) \geq \frac{f(x) - f(y)}{\sqrt{f(x)}}$$
Since, $d_T(\cdot, \cdot)$ is a supremum of $d_\alpha(\cdot, \cdot)$ over all $\alpha$ with norm 1, we get that
\[
d_T(x, y) \geq \frac{f(x) - f(y)}{\sqrt{f(x)}},
\]

Define $B_a = \{y : y \in \Omega \text{ and } f(y) \leq a\}$. So, for all $y \in B_a$, we have $f(y) \leq a$. Therefore, for any $y \in B_a$, we get
\[
d_T(x, y) \geq \frac{f(x) - a}{\sqrt{f(x)}},
\]

Since, the inequality holds for all $y \in B_a$, we conclude that
\[
d_T(x, B_a) \geq \frac{f(x) - a}{\sqrt{f(x)}},
\]
Observation. If $f(x) \geq a + E$, then

$$d_T(x, B_a) \geq \frac{E}{\sqrt{a + E}}$$

This observation concludes the proof of the technical claim.
The approach of applying the Talagrand inequality to the problem of longest increasing subsequence can be generalized to several problems.

Consider the definition of $c$-configuration functions

**Definition (Configuration Functions)**

A function $f$ is a $c$-configuration function, if for every $x, y$, there exists $\alpha_{x,y}$ such that the following holds

\[ f(y) \geq f(x) - \sqrt{c \cdot f(x)} d_{\alpha_{x,y}}(x, y) \]

Note that the longest increasing subsequence defines $f(\cdot)$ that is 1-configuration function. The derivation used above can be identically used for $c$-configuration functions.