20.1 Recap

Previously, we proved a general bounded differences inequality: (recall the notation that $X^{(i)}$ represents all $X$ values except for $X_i$)

**Theorem 20.1.** Let $X_1, X_2, \ldots, X_n$ be independent variables, and let $Z = f(X_1, \ldots, X_n)$ obey the following property:

$$\sup_{X_1, \ldots, X_n, X'_1, \ldots, X'_n} \sum_{i=1}^{n} \left( f(X) - f(X^{(i)}, X'_i) \right)^2 \leq K^2.$$

Then

$$Z - \mathbb{E}Z \in \mathcal{G} \left( \frac{K^2}{8} \right).$$

(Recall that this is more strong than the usual Azuma-Hoeffding Bounded differences since we are taking the supremum outside that sum, rather than the supremum of each term of the sum individually).

An example application of this can be found in the Directed First Passage Percolation (DFPP)

**Example 20.2.** Directed FPP on $[n]^d$: Consider the directed grid graph on $[n]^d$. Let $X_e$ be independent random variables for each $e \in E$, assume they are bounded in $[0,1]$ (one may assume other bounds be scaling).

$$f(X) = \max_{P: 0 \cdot 1 \to n} \sum_{e \in P} X_e$$

Then 20.1 holds with $K^2 = dn$, (by taking $X_i = \{X_e\}_{e:i}$.)

This bound actually gives tail bound

$$\mathbb{P}(\bar{Z} \geq t) \leq e^{-4t^2/dn}$$

which is not the right bound. For $d = 2$, the tail is conjectured to be

$$\mathbb{P}(\bar{Z} \geq t) \leq e^{-c(t/n^{1/3})^{3/2}}$$

and respectively $e^{-ct^{O(d)}}$ for $d \geq 3$.

The Curie-Weiss model gives an exercise in this bound
Exercise 20.1. The Curie-Weiss model with $\beta = 1$: A fully connected Ising model, where $\sigma = (\sigma_1, \ldots, \sigma_n)$.

$$P(\sigma) \propto \exp\left(\frac{1}{n} \sum_{i<j} \sigma_i \sigma_j \right).$$

Show that

$$P(m(\sigma) \geq t) \leq e^{-c(n)t^4}$$

where $m(\sigma) = \frac{1}{n} \sum_i \sigma_i$.

20.1.1 Convex Lipschitz Function Bound

We also have

**Theorem 20.3.** Let $X_1, X_2, \ldots, X_n$ be independent variables $\in [0, 1]$, and let $f(X_1, \ldots, X_n)$ be coordinate-wise convex and $L$-Lipschitz. Then $f(X) \in G(L^2)$.

20.1.2 Self Bounding Functions

A final application: the longest increasing subsequence distribution. Among equivalent definitions, this is the distribution of the length of the longest increasing subsequence $Z = f(X)$ in a sequence $X_1, \ldots, X_n$ of $n$ i.i.d. $\mathcal{U}([0,1])$ samples. It is known that $\mathbb{E}Z \approx 2\sqrt{n}$.

The longest increasing subsequence function $f$ above is “self-bounding”

**Definition 20.4.** A function $f$ is “self-bounding” if there exist $f_i(x^{(i)})$ such that

$$0 \leq f - f_i \leq 1$$

and

$$\sum_{i=1}^{n} (f - f_i) \leq f$$

In the case of the longest increasing subsequence function $f_i$ can be taken as the length of the longest subsequence of $X^{(i)}$. Another example of such a function is the VC-dimension.

In this subsection we will prove a concentration inequality for self-bounding functions.

**Theorem 20.5.** Let $Z = f(X)$ be a self-bounding function of independent random variables. Then

$$\varphi_Z \leq \mathbb{E}Z \cdot \phi(\lambda), \forall \lambda > 0.$$ 

**Proof.** Recall the inequality

$$\text{Ent}(Z) \leq \mathbb{E} \sum_{i=1}^{n} \text{Ent}^{(i)}(Z).$$
Recall also,

\[
\text{Ent}^{(i)}(e^{\lambda Z}) \leq \mathbb{E}(e^{\lambda Z_i} - 1 - (\lambda(Z_i - Z)))
\]

\[
= \mathbb{E}(e^{\lambda Z}e^{\lambda(Z_i-Z)} - 1 - (\lambda(Z_i - Z)))
\]

\[
= \mathbb{E}(e^{\lambda Z} \phi(\lambda(Z_i - Z)))
\]

where we take \(Z_i = f_i(x^{(i)})\) and \(\phi(x) = e^x - 1 - x\) (note \(Z_i\) does not depend on \(X_i\)).

Note that \(\frac{\phi(-x)}{x}\) increases as \(x > 0\) increases. Thus, (since \(0 \leq Z - Z_i \leq 1\) by the first self-bounding property) for \(\lambda > 0\),

\[
\frac{\phi(-\lambda(Z - Z_i))}{\lambda(Z - Z_i)} \leq \frac{\phi(-\lambda)}{\lambda}
\]

\[
\phi(-\lambda(Z - Z_i)) \leq \phi(-\lambda)(Z - Z_i).
\]

Putting these all together, we get

\[
\text{Ent}(e^{\lambda Z}) \leq \mathbb{E}\sum_{i=1}^{n}(e^{\lambda Z} \phi(-\lambda)(Z - Z_i)) \leq \mathbb{E}(e^{\lambda Z} \phi(-\lambda)\sum_{i=1}^{n}(Z - Z_i)).
\]

Applying the second part of the self-bounding definition

\[
\text{Ent}(e^{\lambda Z}) \leq \mathbb{E}(e^{\lambda Z} \phi(-\lambda)Z).
\]

Thus, in terms of the moment generating function \(m\) of \(Z\), we get

\[
\text{Ent}(e^{\lambda Z}) = \lambda m'(\lambda) - m(\lambda) \log m(\lambda) \leq \phi(-\lambda)m'(\lambda).
\]

Subtracting the right hand side

\[
(1 - e^{-\lambda})m'(\lambda) - m(\lambda) \log m(\lambda) \leq 0.
\]

For the centered \(Z\) distribution, we have \(m_Z(\lambda) = \mathbb{E}e^{\lambda(Z-\mathbb{E}Z)} = e^{-\lambda \mathbb{E}Z} \cdot m(\lambda)\), so substituting, we have

\[
(1 - e^{-\lambda}) (m_Z'(\lambda) + \mathbb{E}Z \cdot m_Z(\lambda)) - m_Z(\lambda) \log m_Z(\lambda) e^{-\lambda \mathbb{E}Z} \leq 0.
\]

In terms of the log MGF

\[
\frac{e^\lambda - 1}{e^\lambda - 1} \varphi_Z' - \varphi_Z \leq \phi(-\lambda)\mathbb{E}Z.
\]

Manipulating,

\[
\frac{1}{e^\lambda - 1} \varphi_Z' + \left(\frac{1}{e^\lambda - 1}\right)' \varphi_Z \leq \frac{e^\lambda}{(e^\lambda - 1)^2} (e^{-\lambda} - 1 + \lambda)\mathbb{E}Z.
\]

\[
\left(\frac{1}{e^\lambda - 1} \varphi_Z\right)' \leq \mathbb{E}Z \cdot \left(\frac{-\lambda}{e^\lambda - 1}\right)'.
\]

This implies

\[
\varphi_Z \leq \mathbb{E}Z \cdot \phi(\lambda), \forall \lambda > 0.
\]

\[\blacksquare\]
This is similar to arguments for the Poissonian we have seen before. A Poisson can be thought of as the limit of a sum of Bernoullis:

\[ d_{TV}(\text{Bin}(n,p), \text{Poisson}(np)) \leq np^2 \]

In general, we can always bound repulsive systems like this with an i.i.d. counterpart. For self-bounding \( f \)

\[ \mathbb{P}(f(X) \geq t) \leq \mathbb{P}(\text{Poisson}(\mathbb{E}Z) \geq t) \]

In summary, the Entropy method bound the log MGF.

### 20.2 The Mass Transportation Principle

Recall the following duality equivalence relating a KL-divergence bound on similarity in expectation to a bound on the centered log MGF:

\[ \mathbb{E}_Q Z - \mathbb{E}_P Z \leq (\phi^*)^{-1}(D(Q||P)), \forall Q << P \]

if and only if

\[ \log \mathbb{E} e^{\lambda Z} \leq \phi(x), \forall \lambda \in (0, b) \]

where \( \phi^*(x) = \sup_{0 \leq t \leq b} tx - \phi(t) \).

Pinsker’s inequality is the case where \( Z \) is an indicator random variable

**Theorem 20.6.** **Pinsker’s inequality**

\[ d_{TV}(P, Q) = \sup_A |P(A) - Q(A)| = \inf_{\pi(P, Q)} \mathbb{P}(X \neq Y) \leq \sqrt{\frac{1}{2}D(Q||P)} \]

Here \( \pi \) represents the set of couplings. Specifically, this is equivalent to

\[ \inf_{\pi(P, Q)} \mathbb{P}(X \neq Y)^2 \leq \frac{1}{2}D(Q||P) \]

We now see the Marton Transportation inequality, which strengthens Pinsker by bounding an strictly larger distance in the case where \( P \) is a vector of independent random variables.

**Theorem 20.7.** If \( P = P_1 \otimes \cdots \otimes P_n \), then

\[ \inf_{\pi} \sum_{i=1}^n \mathbb{P}(X_i \neq Y_i)^2 \leq \frac{1}{2}D(Q||P) \]

In fact, there is an even stronger version: the Conditional Transportation inequality.

**Theorem 20.8.**

\[ \inf_{\pi} \sum_{i=1}^n \mathbb{E} \Pr(X_i \neq Y_i|X)^2 \leq 2D(Q||P) \]
These are useful when $Z = f(X)$, $X = (X_1, \ldots, X_n)$ are independent, and $f(x) - f(y) \leq \sum_i C_i(x) \mathbf{1}_{X_i \neq Y_i}$, in which case

$$E_Q f - E_P f = \inf_{\pi} \mathbb{E}(f(Y) - f(X))$$

$$\leq \inf \mathbb{E} \sum_i C_i(x) \mathbf{1}_{X_i \neq Y_i}$$

$$\leq \inf \mathbb{E} \left( \sqrt{\sum_i C_i(x)} \sqrt{\sum \Pr(X_i \neq Y_i)} \right)$$

$$\leq \inf \mathbb{E} \left( \sqrt{\sum_i C_i(x)} \sqrt{2D(Q||P)} \right)$$

Next time, we see proofs of these Transportation Inequalities.