Lecture 22

f is Lipchitz in a set distance d if

$$f(x) \le \sup_{y \in A} f(y) + d(x, A), \quad \forall A \subseteq \mathcal{X}.$$

Lemma $f(x) \le f(y) + d_{c(x)}(x, y)$ $\forall y, x$ where $\|c(x)\|_2 \le 1$ $\Rightarrow$ f is Lipchitz in d_T.

Corollary 3. f is Lipchitz in d_T. $X = (X_1, \ldots, X_n)$ indep.

$$\mathbb{P}(f(X) > \mathbb{M}f(X) + t) \le 2 . e^{-t^2/4}$$

$\boxed{\text{Ex 2}}$ Suprema of Rademacher Processes

$\mathcal{X}^n = \{-1, 1\}^n$, $X_i \overset{iid}{\sim} \text{unif}\{-1, 1\}$

Let A be an $m \times n$ matrix of reals.

Consider $f(x) = \max_{j \in [m]} (Ax)_j$, $\quad x \in \{-1, 1\}^n$.

For instance, for $A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ 0 & 0 & 1 \cdots & 1 \\ \vdots & & & \\ 0 & \cdots & 0 & 1 \end{pmatrix}$, $f(x) = \max_j \sum_{i=1}^{j} x_i$.

Thus, $f(X)$ denotes the max. deviation of a random walk from the origin.

→ Similarly, max of iid r.v.s with $|X_i| \le 1$ can be handled.

$$* \quad f(x) - f(y) \le A(x-y)_{j^*(x)}$$

$$= \sum_{i=1}^{n} A_{i, j^*(x)} (x_i - y_i)$$

$$(\#\#) \qquad \le \sum_{i=1}^{n} 2\,|A_{i,j^*(x)}|\, \mathbb{1}_{\{x_i \ne y_i\}}$$

$$\le \; d_c(x,y)$$

where $c_i = \max_j 2|A_{ij}|, \quad 1 \le i \le n.$

→ Thus, using the simpler bound

$$\mathbb{P}(X \in A)\,\mathbb{P}(d_c(X,A) > t) \le e^{-t^2/2\|c\|_2^2},$$

$$\mathbb{P}\big(f(X) > \mathbb{M}f(X) + t\big) \le 2e^{-t^2/8 \sum_{i=1}^{n} \max_j |A_{ij}|^2}$$

Also, using McDiarmid,

$$\mathbb{P}\big(f(X) > \mathbb{E}f(X) + t\big) \le e^{-t^2/2 \sum_{i=1}^{n} \max_j |A_{ij}|^2}.$$

* But we can improve the bound above using TCDI.

Specifically, by $(\#\#)$,

$$f(x) - f(y) \le \sum_{i=1}^{n} 2\, A_{i,j^*(x)}\, \mathbb{1}_{\{x_i \ne y_i\}}$$

Therefore, $f(x)/2\max_j \|A_j\|_2^2$ is Lipchitz in d_T.

$$\Rightarrow \mathbb{P}\big(f(X) > \mathbb{M}f(X) + t\big) \le e^{-t^2/16 \cdot \max_j \|A_j\|_2^2}.$$

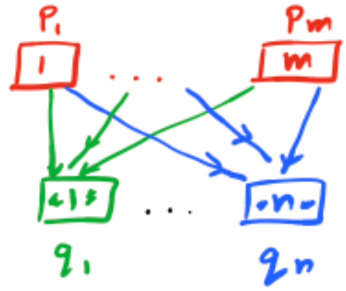
Another igenious perspective on concentration:

Measure Concentration & Transportation Cost Inequalities

$\boxed{A}$ Transportation cost inequalities and coupling

(wheat producers) T1

(wheat consumers) T2



$$\sum_{i=1}^{m} p_i = 1$$

$$\sum_{j=1}^{n} q_j = 1$$

* p_{ij} : fraction of traffic on path i-j
* c_{ij} = cost/unit of using the path i-j

Total cost $= \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}\, p_{ij}$

Optimal transport cost

For fixed p, q, what is the minimum cost $c(p,q)$ achieved by designing p_{ij} appropriately?

Definition (Coupling)

Let $(\mathcal{X}, P)$ and $(\mathcal{Y}, Q)$ be two probability spaces.

A coupling of (P, Q) is a pair of rvs (X, Y) s.t. $X \sim P$ and $Y \sim Q$. We say (X,Y) is a coupling of (P,Q), or, P_{XY} is a coupling of (P, Q).

Denote by $\mathcal{P}(P, Q)$ the set of all couplings of P and Q.

* A deterministic coupling $T: \mathcal{X} \to \mathcal{Y}$ is a change of variables from P to Q s.t. for all Q-integrable $\phi: \mathcal{Y} \to \mathbb{R}$,

$$\int_{\mathcal{Y}} \phi(y)\, Q(dy) = \int_{\mathcal{X}} \phi(T(x))\, P(dx).$$

* Optimal coupling or optimal transport is the solution (minimizer) for the Monge-Kantorovich minimization problem

$$c(P,Q) = \inf_{(X,Y) \in \mathcal{P}(P,Q)} \mathbb{E}[c(X,Y)].$$

Note that the minimization is over all joint distributions $P_{XY} \in \mathcal{P}(P,Q)$; such joint distributions are called transference plans and the ones attaining the infimum are called optimal transference plans.

Interesting Special Cases

* Deterministic optimal transference plan (Monge coupling)

In general, even a deterministic coupling need not exist. Here is an example where Monge coupling exists:

$$\left.\begin{array}{l} \mathcal{X} = \mathcal{Y} = \mathbb{R}; \quad P, Q \ll \mu \text{ (Lebesgue measure)} \\ c(x,y) = |x-y|^2 \end{array}\right\} P1$$

Let optimal cost for P1, $c(P,Q) = \inf_{(X,Y) \in \mathcal{P}(P,Q)} \mathbb{E}[c(X,Y)]$.

can be regarded as a measure of distance b/w P and Q. In the CS literature, this is called the Earthmover Distance. In fact,

$$c(P,Q) = \sup_{\substack{\text{s.t.} \\ h \text{ is Lipchitz}}} \mathbb{E}_Q[h(X)] - \mathbb{E}_P[h(X)]$$

$$\equiv \text{Wasserstein Distance } (W_1(P,Q))$$

* Hamming Cost

Another interesting setting where a comprehensive sol^n is avaible is the following:

$$\left.\mathcal{X} = \mathcal{Y}, \quad c(x,y) = \mathbb{1}_{\{x \neq y\}} \right\} \text{P2}$$

Lemma (Maximal coupling) For $c(x,y) = \mathbb{1}_{\{x \neq y\}}$, $c(P,Q) = d_{TV}(P,Q)$

$$\left(= \sup_{h : \mathcal{X} \to [-1,1], \text{ continuous}} \mathbb{E}_Q[h(X)] - \mathbb{E}_P[h(X)] \right)$$