MATH 690: Lecture Scribing

Xiuyuan Cheng, Scribing: Weiyao Wang

Nov.02, 2017

1 Summary

In last lecture, we finished the proof for JL lemma. In today's lecture, we hope to discuss the concentration of function $F(x)$ beyond the form of summation. We will also see two applications for concentration of measure: spectral edge, second largest eigenvalue $\lambda_2(G)$ where G is a random graph; randomized SVD (PCA).

2 McDiarmid Inequality

Theorem 1. *Let $x_1, ..., x_n$ be random variable, independent, and $x_i \in \Omega_i \subset \mathbb{R}$. Give function $F: \mathbb{R}^n \to \mathbb{R}$ there exists $c_i > 0$, such that $\forall i$, $\forall x_i, x_i' \in \Omega_i$*

$$|F(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n) - F(x_1, ..., x_{i-1}, x_i', x_{i+1}, ..., x_n)| \le c_i$$

Then $\forall \lambda > 0$, $Pr[|F(x) - \mathbb{E}\, F(x)| \ge \lambda(\sum_{i=1}^n c_i^2)^{\frac{1}{2}}] \le C_1 e^{-C_2 \lambda^2}$, where C_1 and C_2 are constants.

An exercise here would be to see how McDiarmid $\to$ Hoeffding.

Proof. We want to use a similar idea on exponential momentum as we proved lemmas previously. We will also use the idea of martingale difference here. We first consider $\mathbb{E}\, e^{t(F - \mathbb{E}\, F)}$.

We have $F(x) = \mathbb{E}[F|x_1, ..., x_n]$, $\mathbb{E}\, F = \mathbb{E}[F|\emptyset] := \mathbb{E}_0\, F$, which gives us the following equation of telescoping:

$$\begin{aligned} F(x) - \mathbb{E}\, F &= \sum_{k=1}^n \mathbb{E}[F|x_1, ..., x_k] - \mathbb{E}[F|x_1, ..., x_{k-1}] \\ &= \sum_{k=1}^n \mathbb{E}_k\, F - \mathbb{E}_{k-1}\, F \end{aligned}$$

Thus, plugging into $\mathbb{E}\, e^{t(F - \mathbb{E}\, F)}$, we have the following calculation:

$$\begin{aligned}\mathbb{E}\, e^{t(\sum_{k=1}^n \mathbb{E}_k F - \mathbb{E}_{k-1} F)} &= \mathbb{E}[e^{t(\mathbb{E}_n F - \mathbb{E}_{n-1} F)} e^{t(\sum_{k=1}^{n-1} \mathbb{E}_k F - \mathbb{E}_{k-1} F)}] \\ &= \mathbb{E}[\mathbb{E}[e^{t(\mathbb{E}_n F - \mathbb{E}_{n-1} F)} e^{t(\sum_{k=1}^{n-1} \mathbb{E}_k F - \mathbb{E}_{k-1} F)}]|x_1, ..., x_{n-1}] \\ &= \mathbb{E}[e^{t(\sum_{k=1}^{n-1} \mathbb{E}_k F - \mathbb{E}_{k-1} F)}\, \mathbb{E}[e^{t(\mathbb{E}_n F - \mathbb{E}_{n-1} F)}|x_1, ..., x_{n-1}]]\end{aligned}$$

From the equation above, we see that we can freeze $x_1, .., x_{n-1}$ and $\mathbb{E}_n F := G(x_n)$, $\mathbb{E}_{n-1} F = \mathbb{E}[G(x_n)]$ over x_n. By our assumption, we have $|G(x_n) - G(x_n')| \leq c_n$, $\forall x_n, x_n'$. Therefore, we have $\exists a_n, b_n$ such that $G(x_n) \in [a_n, b_n)$. This implies $e^{t(G(x_n) - \mathbb{E}[G(x_n)])} \leq e^{-\frac{t^2}{8} c_n^2}$, using Hoeffding's lemma.

Putting the calculation back and apply the argument of freezing variables recursively, we have $\mathbb{E}\, e^{t(F - \mathbb{E} F)} \leq \Pi_{k=1}^n e^{-\frac{t^2}{8} c_k^2} = e^{-\frac{t^2}{8} \sum_{k=1}^n c_k^2}$. Finally we use the similar argument to find the minimizer as in the proof for previous lemmas. $\square$

2.1 Remark on JL Lemma

To have a nontrivial dimension reduction (i.e. $d < D$), we will need $D \geq d \geq \frac{8}{\epsilon^2} ln(n)$, we have $n \leq e^{\frac{\epsilon^2}{8} D}$, this requires that we cannot have more than (constant)^D many points in $D-$dimensional space, where constant $= e^{\frac{\epsilon^2}{8}}$. When distortion tolerance is small (i.e. $\epsilon \to 0$), this is very restrictive.

3 Gaussian Concentration for Lipschitz Function

Recall

Definition 1. *Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}$ satisfies the property such that $\exists L > 0$ such that $|F(x) - F(y)| \leq L|x - y|_2$. We then say F is L-Lipschitz.*

Theorem 2. *Suppose $x_1, ..., x_n \sim N(0, 1)$, i.i.d; and $F : \mathbb{R}^n \to \mathbb{R}$ is 1-Lipschitz. Then $\forall \alpha > 0$, we have $Pr[|F(x_1, ..., x_n) - \mathbb{E}\, F(x_1, ..., x_n)| > \alpha] \leq 2e^{\frac{-\alpha^2}{2}}$*

Proof. We will only give the idea for the proof here. The idea is to sue rotation invariance of $X \sim N(0, I_n)$, and thena pply to spectral edge concentration. And we will also need a proposition that tells for $A = A^T$, $\lambda(A) \geq ... \geq \lambda_n(A)$; the map $\lambda_i : \mathbb{R}^{\frac{n(n-1)}{2}} \to \mathbb{R}$ is 1-Lipschitz. The proof for the proposition comes from the lemma $|\lambda_i(A) - \lambda_i(B)| \leq |A - B|_{Fro}$, which can be proved using Courant-Fischer Inequality we discussed at the beginning of the semester. $\square$