Math690:Concentration(11/7/2017)

Guangshen Ma

November 2017

0.1 Review of last class

Last time we discuss the concentration problem of $F(x)$, where it satisfies:

$$F: \ R^n \ \rightarrow \ R$$

where $x_1, x_2, \cdots, x_n$ are independent.
Today's topic focus on the concentration of spectral edge (The largest or smallest eigenvalues), which includes the following sections:

- Widger matrix: The reason why the norm of the Widger matrixes is concentrated
- λ_2 of Erdos-Renyi random graph theory
- Randomized fast SVD(PCA)

Lemma 0.1 $\lambda_1(A)$ *(The largest eigenvalue) is a "*$1 - lipchitz$*" function*

- What the expectation of the spectral edge look like
- The fluctuation of the these eigenvalues (measure by high probability) would not be far away from its expectation (Concentration)

Widger's theorem:

$$W_{mxn}, \ W^T = W, \ W_{ij} \sim N(0,1), \ iid, \ i \leq j, \ (\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \ of \ W)$$

Wigner's semi-circle law:
Spectral distribution of $\frac{1}{\sqrt{n}} W \rightarrow \rho_{s.c}$ $(Weakly\ converge)$.

TW(Tracy-Widom distribution) law:
Spectral edge λ_1(similarly to λ_n) concentrates at 2 (-2 for λ_n). Based on the previous review, our goal is to show that:

$$\lambda_1(W) = 2\sqrt{n} + O(n)$$

We want to show that high order term $O(n)$ is going to be $O(1)$ with high probability.
Compared with the Tracy-widom result and we get:

$$\lambda_1 = 2\sqrt{n} + O(n^{-\frac{1}{6}})$$

This is like the true scaling of the concentration. But we are going to see our argument by Gaussian concentration equality to show the term $O(1)$ and how we control the deviation.
To show the above results, we have two steps:
Step 1: Show the expectation of λ_1

$$\mathbb{E}\,\lambda_1(W) \approx 2\sqrt{n}$$

Step 2: Use $1 - Lipchitz$ lemma to show the concentration

$$\lambda_1 \ is \ 1 - Lipchitz \ function: \ \ \mathbb{R}^{\frac{n(n-1)}{2}} \to \mathbb{R}$$

0.2 The first step

We start by using some theory from References.

Proposition[Ref[1],Tao]:

$$\mathbb{E}\,\lambda_1 \leq 2\sqrt{n}$$

Proof(Moment method[Ref[2]]:
We first consider the eigenvalues λ_i as the ith sample and take the moment of each sample and we get:

$$\sum_{i=1}^{n} \lambda_i^k = Tr(W^k)$$

This operation here is just the tool for the proof. Since we want to prove the upper bound and we relate λ_1 and get the upper bound argument(k can be considered as a very large even number):

$$\lambda_1^k \leq \sum_{i=1}^{n} \lambda_i^k = Tr(W^k), \ (k > 0)$$

Take expectation of both sides and we get:

$$\mathbb{E}\,\lambda_1^k \leq \mathbb{E}\,Tr(W^k)$$

By Jensen's inequality:

$$(\mathbb{E}\,\lambda_1)^k \leq (\lambda^k) \ \Rightarrow \ (\mathbb{E}\,\lambda_1)^k \leq Tr(W^k)$$

Recommend to take $k = 2, 4(or \ higher \ number)$ as an example for understanding the proof.
We use another *Lemma* without showing the details:

$$\mathbb{E}\,Tr(W^k) = (C_{\frac{k}{2}} + O_k(1))n^{\frac{k}{2}+1}$$

Where $k>0$ and k is fixed here.
We also know(Catolan number):

$$C_{\frac{k}{2}}=\frac{k!}{(\frac{k}{2}+1)!(\frac{k}{2})!}$$

If we take $(C_{\frac{k}{2}})^{\frac{1}{k}}{\rightarrow}2+O(1)$ when $k{\rightarrow}\infty$(This is true without providing details here). That is the reason we have the number 2 for the upper bound in the inequality.
Take both sides multiplied by order $\frac{1}{k}$ and we get:

$$\mathbb{E}\,Tr(W^k)^{\frac{1}{k}}=[(C_{\frac{k}{2}}+O_k(1))n^{\frac{k}{2}+1}]^{\frac{1}{k}}$$

The right hand side will be:

$$[C_{\frac{k}{2}}+O_k(1)]^{\frac{1}{k}}{\cdot}n^{\frac{1}{2}+\frac{1}{k}}$$

As $k\to\infty$ and we get:

$$\mathbb{E}\,\lambda_1{\leq}(2+O(1)){\cdot}\sqrt{n}$$

In practice, we usually take $k{\rightarrow}(lg^n)^2$(missing details here) and we have:

$$\mathbb{E}\,\lambda_1{\leq}(2+\epsilon)\sqrt{n}$$

This argument is corresponding to our purpose, i.e control λ_1 and its expectation value. That also explains why the semi-circle law we have the number 2.

Lower bound(similar argument with moment method):

$$Tr(W^k)=\sum_i\lambda_i^k\leq n{\cdot}\lambda_1^k$$

Similarly, take expectation on both sides and we can show that:

$$\mathbb{E}\,\lambda_1{\geq}(2-\epsilon)\sqrt{n}$$

So far we have finished the step 1.

0.3 The second step

One lipschitz function problem:

Lemma 0.2 $|\lambda_1(A)-\lambda_1(B)|{\leq}||A-B||_{Fros}$

Where λ_1 *is a one − lipschitz function of matrix A*). A_{nxn} and B_{nxn} are symmetric. We also have $\lambda_1:\mathbb{R}^{\frac{n(n-1)}{2}}\rightarrow\mathbb{R}$ is $1-Lipschitz$. The $1-Lipschitz$ mentioned here means the lipschitz constant C here is 1.

Proof of Lemma(Courant-Fisher):

$$|\lambda_1(A) - \lambda_1(B)| \leq ||A - B||_{op}$$

Where op symbol means operator norm.

$$\lambda_1(A) = \sup_{||v||_2=1} v^T A v$$

$$\lambda_1(B) = v_0^T B v_0 = v_0^T A v_0 + v_0^T (B - A) v_0$$

Where the first term $v_0^T A v_0 < \lambda_1(A)$.

Remark(Operator norm is bounded by Frobos norm):

$$\sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)|^2 \leq ||A - B||_{Fro}^2$$

Then we have(Citing the results from reference):

$$|\lambda_i(A) - \lambda_i(B)|^2 \leq ||A - B||_{Fro}^2$$

We use the above propositions to prove the lemma in step 2, and for $i = 1$ we have:

$$Pr[|\lambda_1 - \mathbb{E}\,\lambda_1| > \alpha] \leq 2 \cdot e^{-\frac{\alpha^2}{2}}$$

Therefore, the probability of the deviation from λ_1 and its expectation is bounded and also decay exponentially(The concept of concentration).

0.4 The Cheager's constant number

Remark:
λ_2 is the second eigenvalue of the normalized graph Laplacian.

$$L_{rw} = D^{-1}(D - W) = I - P$$

The symbol rw means random walk laplacian. And we have:

$$0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n (\leq 2)$$

The λ_2 here is strongly related to how much you can cut your graph.

Remark: Cheager's law

$$G = (V, E),\ W_{nxn},\ W_{ij} \geq 0,\ W = W^T$$

Define:

$$h(A) = \frac{|\partial A|}{min\{Vol(A), Vol(A^c)\}}$$

Where:

$$(\partial A) \stackrel{def}{\Longrightarrow} \sum_{i\in A, j\in A^c} W_{ij}$$

$$Vol(A) = \sum_{i\in A} d_i,\ d_i \stackrel{def}{\Longrightarrow} \sum_{j=1}^{n} W_{ij}$$

We define Cheager's constant as $h_G = \underset{A\subset v}{min}\, h(A)$. It measures how cutting a graph or manifold it is.

Theorem: Cheager Inequality

$$\frac{1}{2}h_G^2 \leq \lambda_2 \leq 2h_G$$

The hint of proving $\lambda_2 \leq 2h_G$:

$$\lambda_2 = \underset{f}{inf}\, \frac{f^T Lf}{f^T Df} \leq \frac{g^T Lg}{g^T Dg},\ L = D - W$$

Where g is: λ_2 gives implication on how cutting it is for the graph.

0.5 Application with fast PCA

Key ideas:

- PCA: We want to compute the low rank factorization of A_{nxn}
- Compute the first k singular values

Key idea: Tall gaussian matrix is well-conditioned

$$x_{ij} \sim N(0,1),\ iid\ (k \ll n)$$

We have the demo showing the follow properties:

$$\sigma_1(x) \geq \cdots \geq \sigma_k(x)$$

$$\sqrt{n} - \sqrt{k} - O(1) \leq \sigma_k(x) \leq \sigma_1(x) \leq \sqrt{n} + \sqrt{n} + O(1)$$

Based on the MP(mention in 9/7's note)density:

$$\begin{cases} \sigma_1^2 \approx (1 + \sqrt{\dfrac{k}{n}})^2 \\ \sigma_k^2 \approx (1 - \sqrt{\dfrac{n}{k}})^2 \end{cases} \tag{1}$$

0.6 Appendix

- Lipschitz continuty
- Courant-Fischer Minimax Theorem(9/5's note)
- L_{rw} random walk laplacian matrix
- Jensen's inequality
- Wigner's Semicircle Law
- Tracy–Widom distribution
- Method of moments
- MP law(9/7's note)