Gaussian concentration

[Borell - Tsirelson - Ibramigov - Sudakov]

Thm :- If $f: \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, then

$$\Pr_{X \sim N(0, I_n)} \left[f(X) - E[f(X)] > t \right] \leq \exp\left(-\frac{t^2}{2L^2}\right)$$

L-Lipschitz :- $|f(x) - f(y)| \leq L \, \|x - y\|_2$.

We will see two approaches to prove this theorem, one in outline, and the other in full.

Examples of Lipschitz functions

Consider a $n \times n$ symmetric matrix A. Let

$$f\left((A'_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq i}}\right) = \lambda_1(A)$$

[The largest eigenvalue of $\frac{1}{\sqrt{n}} A$, where $A_{ij} = A_{ji} = \frac{A'_{ij}}{\sqrt{n}}$]

Then, for any unit vector u, we have

$$\frac{1}{\sqrt{n}} \left| u^T (A + X) u - u^T A u \right| = \frac{1}{\sqrt{n}} |x^T X u|$$

so that f is $\frac{1}{\sqrt{n}} \sup_{u \in S^{n}} \sup_{\substack{X \in \mathbb{R}^{n \times n} \\ X_{ij} = X_{ji} \\ \sum_{1 \leq j \leq i \leq n} X_{ij}^2 = 1}} |u^T X u|$ Lipschitz.

But for such X, u.

$$|u^T X u| = \left| \sum_{i,j} u_i u_j X_{ij} \right|$$

$$\leq \sqrt{\sum_{i,j} X_{ij}^2 \sum_i u_i^2 \sum_j u_j^2} < \sqrt{2 \sum_{1 \leq j \leq i \leq n} X_{ij}^2} = \sqrt{2}$$

$\Rightarrow$ f is $\frac{\sqrt{2}}{\sqrt{n}}$ - Lipschitz.

Thus,

$$\Pr\left[\left| \lambda_1\left((X_{ij})_{1 \leq i \leq j \leq n}\right) - E[\lambda_1(X)] \right| > t \right] \leq 2 \exp\left(-\frac{nt^2}{4}\right)$$

[Not tight for small t; correct order is $t^{3/2}$ for upper tail]

Another application:

Let A be a fixed $m \times n$ matrix, and let

$$Y = AX \text{ where } X \sim N(0, I_n).$$

$$Z = \max_{1 \leq i \leq m} Y_i$$

Q:- How does Z concentrate?

Define:- $f(X) := \max_{1 \leq i \leq m} (AX)_i$

Now $$|f(x) - f(y)| = \left| \max_{1 \leq i \leq m} (Ax)_i - \max_{1 \leq j \leq m} (Ay)_j \right|$$

$$\leq \max_{1 \leq i \leq m} |(Ax)_i - (Ay)_i|$$

$$\leq \max_{1 \leq i \leq m} |(A(x-y))_i|$$

$$\leq \|x-y\| \max_{1 \leq i \leq m} \left(\sqrt{\sum_{1 \leq j \leq n} A_{ij}^2} \right)$$

$$= \|x-y\| \, \sigma \qquad \text{where } \sigma^2 := \max_{1 \leq i \leq m} E[Y_i^2].$$

$\Rightarrow$ f is σ-Lipschitz. Thus

$$\Pr\left[|Z - E[Z]| > t \right] \leq \exp\left(-\frac{t^2}{2\sigma^2} \right)$$

Note:- No dependence on m or n

"Dimension independent"

Q:- What is $E[\sup_{1 \leq i \leq n} Y_i]$? Interesting question with a lot of work behind it.

Note :- The result extends to "infinite" $\underline{\underline{m}}$. The framework is that we have a set $(Y_t)_{t \in T}$ of random variable s.t. every finite subset of the variables is "jointly Gaussian" [i.e., of the form described above.], and s.t. T forms a "totally bounded" metric space w.r.t. which $t \mapsto Y_t$ is continuous with probability one.

To reiterate :- This result has no dependence on m or n, but only on $\sup_{t \in T} E[Y_t^2]$.

Robust Johnson-Lindenstrauss

In the last lecture, we looked at JL matrices of the form

$$A = \left(\frac{1}{\sqrt{d}} A'_{ij} \right)_{\substack{k \le i \le d \\ 1 \le j \le m}}, \quad \begin{matrix} E[A'_{ij}] = 0 \\ E[A'^{2}_{ij}] = 1 \end{matrix}, \quad A'_{ij} \in \mathcal{G}(1)$$

We then saw that $\exists c$, constant such that if

$$d \geq \frac{c}{\varepsilon^2} \log\left(\frac{n}{\sqrt{\delta}}\right),$$ then for any set $S \subseteq \mathbb{R}^m$ of size n,

$$1 - \varepsilon \le \frac{\|Ax - Ay\|}{\|x - y\|} \le 1 + \varepsilon \quad \forall\, x \neq y \in S$$

with probability at least $1 - \delta$.

Our goal is now to remove the dependence on n in favour of more structural properties of S.

In the proof, we will need also the following fact [an analogue of "Spectral gap" in the discrete setting] that we will prove and strengthen later.

Gaussian Poincare inequality

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a differentiable fn. s.t. if $X \sim N(0, I_n)$ then both $E[f(X)^2]$ and $E[\|\nabla f(x)\|_2^2]$ exist.

Then:-

(1)

$$\mathrm{Var}[f(X)] \leq E[\|\nabla f(X)\|^2]$$

(2) (Corollary) If f is L-Lipschitz $[\,|f(x) - f(y)| \leq L\|x-y\|\,]$ and $E[f(x)^2]$ exists then

$$\mathrm{Var}[f(X)] \leq L^2 .$$

In particular,

$$E[f(X)]^2 \leq E[f(X)^2] \leq L^2 + E[f(X)]^2. \quad — \quad (*)$$

Now, as before, consider the set

$$T = \left\{ \frac{x-y}{\|x-y\|} \;\middle|\; x \neq y \in S \right\}.$$ We wish to show that both $Y = \inf_{u \in T} \|Au\|_2^2$ and $Z = \sup_{u \in T} \|Au\|_2^2$ concentrate.

The proof is based on the fact that if we define f as

$$f\left((A'_{ij})_{\substack{1\leq i\leq d \\ 1\leq j\leq m}}\right) = \left(\inf_{u\in T} \text{ or } \sup_{u\in T}\right) \|Au\|_2$$

$$\inf_{u\in T}\|Au\|_2 = \sqrt{Y}, \qquad \sup_{u\in T}\|Au\|_2 = \sqrt{Z}$$

then f is $\frac{1}{\sqrt{d}}$-Lipschitz.

From now on, we deal with the sup. The proof for inf is exactly the same.

Why is f $\frac{1}{\sqrt{d}}$ - Lipschitz ?

$$\left| f(A') - f(A'+X) \right| = \frac{1}{\sqrt{d}} \left| \sup_{u \in T} \|A'u\| - \sup_{u \in T} \|(A'+X)u\|_2 \right]$$

$$\leq \frac{1}{\sqrt{d}} \sup_{u \in T} \left| \|A'u\| - \|(A'+X)u\| \right|$$

$$\leq \frac{1}{\sqrt{d}} \sup_{u \in T} \|X'u\|$$

$$= \frac{1}{\sqrt{d}} \sup_{u \in T} \sqrt{|X_i^T u|^2} \leq \frac{1}{\sqrt{d}} \sqrt{\|X_i\|_2^2}$$

$$= \frac{1}{\sqrt{d}} \sqrt{\sum_{i,j} x_{ij}^2}.$$

So, we have, from Gaussian concentration

$$e^{-t} \geq \Pr\left(\sqrt{Z} > E\sqrt{Z} + \sqrt{\frac{2t}{d}}\right) = \Pr\left(Z > \left(E\sqrt{Z}\right)^2 + \frac{2t}{d} + 2\sqrt{\frac{2t}{d}}E[\sqrt{Z}]\right)$$

$$= \Pr\left(Z > EZ + \frac{2t}{d} + 2\sqrt{\frac{2tE[Z]}{d}}\right) \quad (\because t \geq 0, \sqrt{Z} \geq 0).$$

So, $\Pr\left(Z > E[Z] + \frac{2t}{d} + 2\sqrt{\frac{2t}{d}E[Z]}\right) \leq e^{-t} \quad \forall t > 0.$

where $Z = \sup_{u \in T} \|Au\|_2^2$.

Similarly or $Z = Y = \inf_{u \in T} \|Au\|_2^2$

$$e^{-t} \geq \Pr\left(\sqrt{Z} < E[\sqrt{Z}] - \sqrt{\frac{2t}{d}}\right)$$ [Again using the Lipschitz property of $\sqrt{Z}$]

$$\geq \Pr\left(Z < E[\sqrt{Z}]^2 + \frac{2t}{d} - 2\sqrt{\frac{2t}{d}}E[\sqrt{Z}]\right) \quad \left(\text{if } t \leq \frac{d}{2}E[\sqrt{Z}]^2\right)$$

$$\geq \Pr\left(Z < E[Z] - \frac{1}{d} + \frac{2t}{d} - 2\sqrt{\frac{2tE[Z]}{d}}\right)$$

$\boxed{\text{Using } E[Z] \leq E[\sqrt{Z}]^2 + \frac{1}{d} \text{ from } \circledast \text{ and } E[Z] \geq E[\sqrt{Z}]^2 \text{ when } Z \geq 0 \text{ (convexity of } x \mapsto x^2)}$

$$\geq \Pr\left(Z < E[Z] - 2\sqrt{\frac{2tE[Z]}{d}}\right)$$ when $t \geq \frac{1}{2}$ and $t \leq \frac{d}{2}E[\sqrt{Z}]^2$

Note that this inequality is trivially true if $t \geq \frac{d}{2}\left(E[Z] - \frac{1}{d}\right)$ and $E[Z] \geq \frac{4}{3d}$

when $Z = \sup_{u \in T} \|Au\|_2^2$

or $Z = Y = \inf_{u \in T} \|Au\|_2^2$.

We are interested in $\sup_{u\in T} \left| |Au|^2 - 1 \right| = \max(Z-1, 1-Y)$

So,
$$\sup_{u\in T} \left| \|Au\|^2 - 1 \right| \leq \max\left(E[Z]-1, 1-E[Y]\right) + \frac{2t}{d} + 2\sqrt{\frac{2E[Z]t}{d}}$$

with prob $\geq 1-2e^{-t}$ when $t > \frac{1}{2}$ and $t \leq \frac{d}{2} E[\sqrt{Y}]^2$

$$\Delta(d) = d \max\left(E(Z)-1, 1-E[Y]\right)^2.$$

[Note: $E[Y] \leq 1 \leq E[Z]$ (Why?)]

Then the R.H.S. is at most

$$\sqrt{\frac{\Delta}{d}} + \frac{2t}{d} + 2\sqrt{\frac{2t}{d}\left(1+\sqrt{\Delta}\right)}$$

$$\leq \sqrt{\frac{\Delta}{d}} + \frac{2t}{d} + 2\sqrt{\frac{2t}{d}} + 2\sqrt{\frac{2t}{d}\cdot\sqrt{\Delta}}$$

$$\leq 2\sqrt{\frac{\Delta}{d}} + \frac{4t}{d} + 2\sqrt{\frac{2t}{d}}$$

Thus $d \geq \frac{K}{\varepsilon^2}\left(\Delta(d) + \log\left(\frac{2}{\delta}\right)\right)$ suffices for a large enough constant K. [along with $E[Y] \geq \frac{4}{3d}$]

Note:- no dependence of d on $|S|$.
$|S|$ comes in only through

$$\Delta = \Delta(d) := d \max\left(E\sup_{u\in T}\|Au\|^2 - 1,\ 1 - E\inf_{u\in T}\|Au\|^2\right)^2.$$

where $T = \left\{ \frac{x-y}{\|x-y\|} \;\middle|\; x \neq y \in S \right\}$

<u>One example</u>: if $S = \{ u \in \mathbb{R}^m \mid$ at most k entries of u are non-zero$\}$

Then $\Delta(d)$ can be tightly estimated, and it turns out that

$d = \Theta\left(\frac{1}{\varepsilon^2}\left[k \log\left(\frac{m}{k} \right) + \log\left(\frac{1}{\delta} \right) \right] \right)$ suffices.