1.2 Chaining argument: Dudley's integral $\neq$

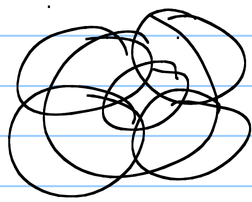
Note Title 2015-10-07

Goal: Bound $\mathbb{E}\sup_{x\in S^{n-1}} \|Ax-\sqrt{m}\| =: \mathbb{E}\sup_{t\in S^{n-1}} X_t$

= Expected supremum of a random process.

<u>Previous approach:</u>

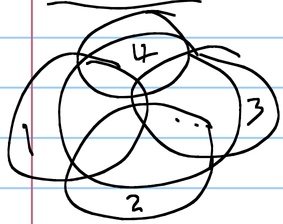
1-scale discretization.



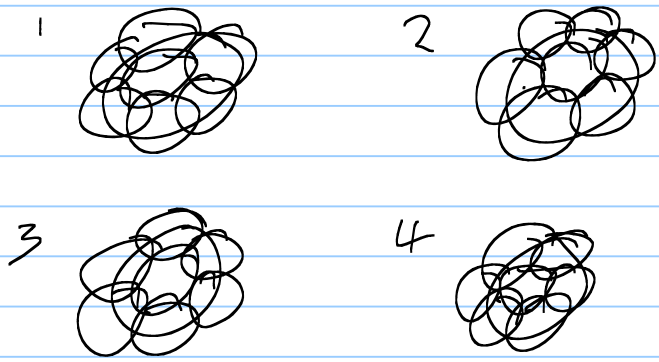
<u>Sharper approach:</u>

Multiscale discretization

<u>1-scale</u>



2-scales: discretize ①, ②, ③, ④.



3-scales: discretize ~30 little balls

⋮

This chaining argument gives:

<u>Dudley's $\neq$</u>

Assumptions on random process X_t, $t\in T$:

1. Centered: $\mathbb{E}X_t = 0$
2. Sub-Gaussian process:

$\|X_t - X_s\|_{\psi_2} \le C \cdot d(s,t)$ for some associated metric d.

Thm (Dudley's $\neq$) Let X_t, $t \in T$, be a centered sub-Gaussian process w/ associated metric d. Then,

$$\mathbb{E} \sup_{t \in T} |X_t| \le C \int_0^\infty \sqrt{\ln N(T,d,\epsilon)}\, d\epsilon$$

- $N(T,d,\epsilon)$ = covering number of T, with metric d, at scale ϵ.
- $\int_0^\infty$ can be replaced w/ $\int_0^{\mathrm{diam}(T)}$.

$\mathrm{diam}(T) := \inf\{t : \exists x \in T \text{ s.t. } \forall y \in T \ d(x,y) \le t\}$.

Proof: Dudley's $\neq$ is a special case of generic chaining (tbd)

Implication: sharper bounds on $\sigma_1(A)$, $\sigma_n(A)$

Thm Let $A \in \mathbb{R}^{m \times n}$ have iid $N(0,1)$ entries. Then,

$$\mathbb{E}\left| \sup_{x \in S^{n-1}} \|Ax\|_2 - \sqrt{m} \right| \le C\sqrt{n}$$

Proof: $X_t := \|At\|_2 - \mathbb{E}\|At\|_2 \qquad t \in S^{n-1}$

Note: By Gaussian concentration,

$$\left|\mathbb{E}\|At\|_2 - \sqrt{m}\right| \le C \qquad \text{(Exercise)}$$

Thus it suffices to bound centered process X_t. Sub-Gaussian increments?

Lemma $\|X_t - X_s\|_{\psi_2} \le C\|s-t\|_2$

Proof (of lemma): Let $y = \frac{t+s}{2}$, $z = \frac{t-s}{2}$

Observe:

- $t = y + z, \quad s = y - z$
- $\langle y, z\rangle = \frac{1}{4}\langle t+s, t-s\rangle = \frac{1}{4}\left(\|t\|_2^2 + \langle s,t\rangle - \langle s,t\rangle - \|s\|_2^2\right) = 0$

Useful Gaussian properties:

- $g_1 := Ay$ is independent of $g_2 := Az$ (since $\langle y,z\rangle = 0$)
- $g_2 = Az \sim \|z\|_2 \cdot N(0, I_m) = \frac{\|t-s\|_2}{2} N(0, I_m) =: \|z\|_2\, g$

- $X_t = \|g_1 + \|z\|_2 g\|, \quad X_s = \|g_1 - \|z\|_2 g\|$

Control moments of $X_t - X_s$:

$$\mathbb{E}|X_t - X_s|^p = \mathbb{E}\left|\|g_1 + \|z\|_2 g\|_2 - \|g_1 - \|z\|_2 g\|_2\right|^p \quad (*)$$

We will condition on g_1. Thus, for the moment, replace g_1 w/ const. vector x. Note that

$$f : g \mapsto \|x + \|z\|_2 g\|_2 - \|x - \|z\|_2 g\|_2$$

has Lipschitz constant $2\|z\|_2$, and $\mathbb{E}f(g) = 0$.

Thus, by Gaussian concentration,

$$\left\|\|x + \|z\|_2 g\|_2 - \|x - \|z\|_2 g\|_2\right\|_{\psi_2} \lesssim C \cdot \|z\|_2$$

$$\Rightarrow (*) = \mathbb{E}\left[\mathbb{E}\left|\|g_1 + \|z\|_2 g\|_2 - \|g_1 - \|z\|_2 g\|_2\right|^p \,\Big|\, g_1\right]$$

$$\leq \mathbb{E}\left[(C\|z\|_2)^p \cdot \sqrt{p}^{\,p} \,\Big|\, g_1\right] = C\|z\|_2^p \sqrt{p}^{\,p}$$

$$\Rightarrow \|X_t - X_s\|_p \leq C\|z\|_2 \sqrt{p}$$

$$\Rightarrow \|X_t - X_s\|_{\psi_2} \leq C\|z\|_2 = C\|x - y\|_2$$

QED!

Return to proof of thm:

By Dudley's $\neq$,

$$\mathbb{E} \sup_{t \in S^{n-1}} |X_t| \leq C \int_0^2 \sqrt{\ln N(S^{n-1}, \|\cdot\|_2, \epsilon)}\, d\epsilon$$

$$\overset{*}{=} C \int_0^2 \sqrt{\ln \left(\frac{c}{\epsilon}\right)^n}\, d\epsilon$$

$$= C\sqrt{n} \int_0^2 \sqrt{\ln\left(\frac{1}{\epsilon}\right)}\, d\epsilon$$

$$= C\sqrt{n}$$

(*): By covering number $\left(\frac{3}{\epsilon}\right)^n$ from previous lemma. 3 is replaced by c since $\epsilon \leq 2$ (not ≤ 1).

QED

Is $\sup_{x \in S^{n-1}} \left| \|Ax\|_2 - \sqrt{m} \right|$ concentrated around its expectation?

Yes, by Gaussian concentration:

Exercise: $f: A \to \sup_{x \in S^{n-1}} \left| \|Ax\|_2 - \sqrt{m} \right|$ has Lipschitz const. 1, i.e.,

$$|f(A) - f(B)| \leq \|A - B\|_F$$

Also, note: $\sup_{x \in S^{n-1}} \left| \|Ax\|_2 - \sqrt{m} \right| = \max\left(\sigma_1(A) - \sqrt{m},\ \sqrt{m} - \sigma_n(A)\right)$

Thus, by Gaussian concentration,

$$P\left(\sigma_1(A) \geq \sqrt{m} + C\sqrt{n} + t\right) \leq e \cdot e^{-ct^2}$$

$$P\left(\sigma_n(A) \leq \sqrt{m} - C\sqrt{n} - t\right) \leq e\, e^{-ct^2}$$

Optimal expectation? Almost. We will see that $C=1$.

Optimal concentration? For very large t, yes. Tails at least as heavy as $\underset{x,y \in S^{n-1}}{xAy} \sim N(0,1)$

Extend proof to sub-Gaussian case? Yes, but we would need a version of Dudley's ≠ allowing sub-Gauss & sub-exp. increments.

However, there is a way to control a sub-Gauss matrix (or even random process) by relating it to a Gaussian random matrix (or random process). This will be a consequence of the ultimate chaining argument: <u>Talagrand's generic chaining</u> (next lecture).