Non-Asymptotic Theory of Random Matrices
Lecture 8: DUDLEY'S INTEGRAL INEQUALITY

Lecturer: Roman Vershynin Scribe: Igor Rumanov

Tuesday, January 30, 2007

Let A : $m \times n$ matrix with i.i.d. entries, $m > n$.

We want to estimate

$$\mathbb{E} \sup_{x \in S^{n-1}} | \, ||Ax||_2 - \sqrt{m} \, | \ \leq ?,$$

(we want $C\sqrt{n}$ in place of "?" here, it would be better estimate than in asymptotic theory: $C(\sqrt{m} + \sqrt{n})$). Under the absolute value sign here stands a random variable, even family of random variables indexed by points of sphere $x \in S^{n-1}$, i.e. a random process.

Random process:

$(X_t)_{t \in T}$ is a collection of random variables indexed by $t \in T$.

- Classical: $T = [a, b]$ - time interval. Examples of such processes are called Levy Processes.

(Ex.: Brownian motion)

- General: T is arbitrary, such as $T = S^{n-1}$.
- "Size of the random process":

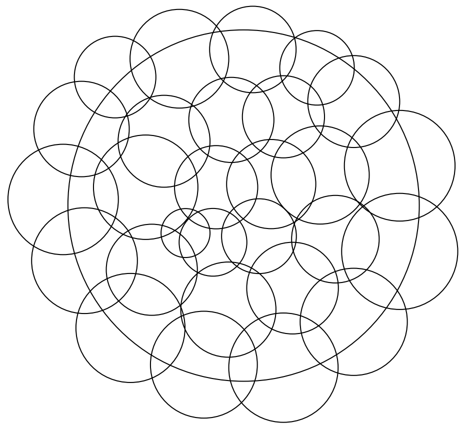
$$\mathbb{E} \sup_{t \in T} X_t \qquad \qquad (\textit{index set has to be compact})$$

How far a particle can get in time T ?

(Ex.: The highest level of water in a river in 10 years)

Previous approach: Discretization of T.

Consider an ε-net $\mathcal{N}$ of T: cover by ε-balls :



Compute $\mathbb{E} \sup_{t \in \mathcal{N}} X_t$,
approximate

Definition 1 (Covering Numbers)**.** *Let (T, d) be a compact metric space, $\varepsilon > 0$. Then covering number $N(T, \varepsilon)$ = minimal cardinality of an ε-net of T = minimum possible number of ε-balls to cover T.*

Measure of compactness of T:
$\log N(T, \varepsilon)$ is called metric entropy of T.

Sharper approach: Multiscale discretization.
Cover T progressively with radius ε^k-balls, $k = 1, 2, 3, ...$
The result will be

Dudley's Integral Inequality

Assumptions:
1) $\mathbb{E}X_t = 0$ for all t
2) Increments $|X_t - X_s|$ are proportional to the distance $d(t, s)$.

$\frac{|X_t - X_s|}{d(t,s)}$ is subgaussian for all t,s:

$$\mathbb{P}(|X_t - X_s| > u \cdot d(t, s)) \ \leq \ Ce^{-cu^2} \qquad \textit{for } u > 0,$$

"subgaussian increments". (Here C and c are some constants).

Theorem 2 (Dudley [1, 2]). *: For a process with subgaussian increments*

$$\mathbb{E}\sup_{t\in T} X_t \;\le\; C\int_0^\infty \sqrt{\log N(T,\varepsilon)}\,d\varepsilon$$

$\uparrow$ probabilistic $\qquad$ $\uparrow$ geometric (in T)

(one can replace the upper limit of ∞ in the integral with $diam(T)$).
Singularity here is at 0.

(For sphere $N(T,\varepsilon) \approx (\frac{1}{\varepsilon})^n$).

($\sqrt{logx}$ = inverse of e^{x^2}).

Proof: Let $diam(T) = 1$. (*Exercise:* general case)

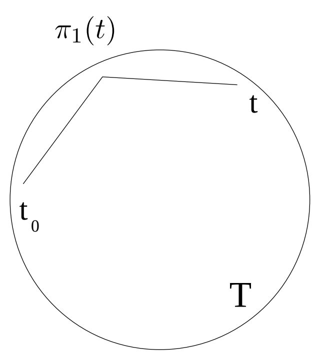
1) Let $t_0 \in T$ be arbitrary (reference point),

$$\mathbb{E}\sup_{t\in T} X_t \;=\; \mathbb{E}\sup_{t\in T}(X_t - \mathbb{E}X_{t_0}) \;\le\; \mathbb{E}\sup_{t\in T}(X_t - X_{t_0}),$$

by Jensen's inequality, because sup is convex function.

2) Multiscale discretization of T:

CHAINING :

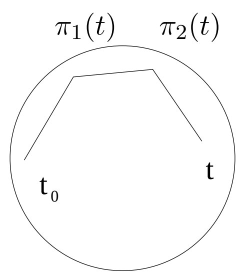


① Let $\mathcal{N}_1$ be a 1/2-net of T of size $N_1 = N(T, 1/2)$
Find $\pi_1(t) \in \mathcal{N}_1$ nearest to t

$$X_t - X_{t_0} = (X_t - X_{\pi_1(t)}) + (X_{\pi_1(t)} - X_{t_0})$$

↑ smaller than before (1/2)

↑ there are at most N_1 such r.v.'s (not too many)



② Let $\mathcal{N}_2$ be a 1/4-net of T of size $N_2 = N(T, 1/4)$
Find $\pi_2(t) \in \mathcal{N}_2$ nearest to t

$$X_t - X_{t_0} = (X_t - X_{\pi_2(t)}) + (X_{\pi_2(t)}) - X_{\pi_1(t)}) + (X_{\pi_1(t)} - X_{t_0})$$

↑ even smaller(1/4)

↑ there are (at most ?) $N_1 N_2 \le N_2^2$ such r.v.'s

....

ⓚ Let $\mathcal{N}_k$ be a 2^{-k}-net of T of size $N_k = N(T, 2^{-k})$
Find $\pi_k(t) \in \mathcal{N}_k$ nearest to t

....

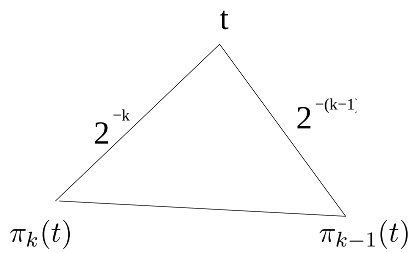
$$\boxed{X_t - X_{t_0} = \sum_{k=1}^{\infty} X_{\pi_k(t)} - X_{\pi_{k-1}(t)}}$$

↑ chaining identity, because $X_t - X_{\pi_k(t)} \to 0$ *a.s.*

$(\pi_0(t) = t_0)$, (*<u>Exercise:</u>* use $\pi_k(t) \to t$).

Nice properties of multiscale discretization:

1)Increments are small:



$$d(\pi_k(t), \pi_{k-1}(t)) \le d(\pi_k(t), t) + d(\pi_{k-1}(t), t) \le 2^{-k} + 2^{-(k-1)} = 3 \cdot 2^{-k}.$$

2) There are at most $N_k N_{k-1} \le N_k^2$ pairs of $(\pi_k(t), \pi_{k-1}(t))$, whatever t is.

Increments:

$$\mathbb{P}\left(|X_{\pi_k(t)} - X_{\pi_{k-1}}(t)| > u \cdot a_k\right) \le C \exp\left(-\frac{cu^2 a_k^2}{d(\pi_k(t), \pi_{k-1}(t))^2}\right) = C \exp\left(-c' \cdot 2^{2k} u^2 a_k^2\right)$$

(holds for $\forall a_k > 0$).
Thus we can bound every increment in the Chaining Identity:
the failure (to bound) probability is

$$p = \mathbb{P}(\exists k, \exists t \in T : |X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| > u \cdot a_k) \le \sum_{k=1}^{\infty} N_k^2 \cdot C \exp(-c \cdot 2^{2k} u^2 a_k^2).$$

In case of success: if $\forall k$, $\forall t \in T$:

$$|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \le u a_k,$$

then $|X_t - X_{t_0}| \le u \sum a_k$.
Hence

$$\mathbb{P}\left(\sup_t |X_t - X_{t_0}| > u \sum_k a_k\right) \le p. \qquad (*)$$

It remains to choose weights a_k. We have tradeoff here: we want $\sum a_k$ to be small, but for decreasing failure probability a_k have to be large.
How large ? Say, for $u \geq 1$ we want the summands in p be $\sim 2^{-k}$. Therefore

$$a_k = c' \cdot 2^{-k} \sqrt{\log 2^k N_k^2} \qquad\qquad (for\ u \geq 1).$$

Then

$$p \;\leq\; \sum_{k=1}^{\infty} C N_k^2 \cdot (2^k N_k^2)^{-u^2} \;\leq\; C \sum_{k=1}^{\infty} 2^{-ku^2}.$$

So subgaussian failure probability obeys the bound $\boxed{p \;\leq\; C \cdot 2^{-u^2}}$.

This way we get an estimate for the sum of weights which appears in (*):

$$\sum a_k = c' \sum 2^{-k} \sqrt{\log 2^k N_k^2} \;\leq \qquad (use\ \sqrt{a+b} \leq 2(\sqrt{a} + \sqrt{b}))$$

$$\leq c'' (\underbrace{\sum 2^{-k} \sqrt{\log 2^k}}_{\leq\, const} + \underbrace{\sum 2^{-k} \sqrt{\log N_k}}_{\substack{\geq\, const \quad \text{because} \\ diam T = 1,\ N_1 \geq 2}}) \leq$$

$$\leq\; C''' \sum_k 2^{-k} \sqrt{\log N_k} \;=\; C''' \sum_k 2^{-k} \sqrt{\log N(T, 2^{-k})} \;\leq\; C^{IV} \int_0^1 \sqrt{\log N(T,\varepsilon)} d\varepsilon \;:=\; S$$

(compare series with integrals in the last inequality)
We have

$$\boxed{\mathbb{P}\left(\sup_t |X_t - X_{t_0}| > uS\right) \;\leq\; Ce^{-u^2} \qquad for\ u \geq 1.}$$

Thus, the random variable $\frac{1}{S} \sup_t |X_t - X_s|$ is subgaussian and Dudley's inequality follows immediately.

<u>Problem:</u> to find sharp estimate (a function better than S - will be done next time - see Lecture 9).

References

[1] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.

[2] Michel Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.