Session - 8 7/07/25

Supremum (subGaussian) :

Finite Maxima:

1. $X_t \sim sG(\sigma^2) \quad \forall\ t \in T$, then $\quad \boxed{\mathbb{E} X_t = 0}$

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \sqrt{2\sigma^2 \log|T|} \qquad X_t = \mu + Z_t$$

$\sqrt{\log|T|} \longrightarrow$ Entropic factor.

$$\mathbb{P}\left[\sup_{t\in T} X_t > \sqrt{2\sigma^2\log|T|} + u\right] \leq e^{-\frac{u^2}{2\sigma^2}}$$

2. When T is infinite :

$$X_i \text{ iid } X, \quad \mathbb{E}\left[\sup_{i\in\mathbb{N}} X_i\right] = +\infty, \quad X \sim N(0,1)$$

as long as $\nexists\, B$ s.t. $\mathbb{P}(X \leq B) = 1$

$$\mathbb{E}\left[\inf_{i\in\mathbb{N}} X_i\right] = -\infty$$

H.W. (RvH*) :

X_i iid $N(0,1)$

$$\frac{\sup_{i=1}^{n} X_i}{\sqrt{2\log N}} \xrightarrow{a.s.} 1$$

☆ Lipschitz process: $(X_t)_{t\in T}$ is called C - Lipschitz if

(T, d) – metric space

$$|X_s - X_t| \leq C\, d(s,t) \quad \text{a.s.},$$

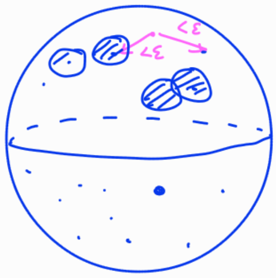
where C is a random variable.

Covering Method:

Lemma: Let (T,d) - compact metric space. Then, $\forall\ \varepsilon>0$, $\exists$ a set $\mathcal{N}_\varepsilon$ & $\pi: T \to \mathcal{N}_\varepsilon$ s.t. $\forall\ t \in T$, $\exists\ \pi(t) \in \mathcal{N}_\varepsilon$ s.t.

$$d(t, \pi(t)) < \varepsilon$$

and $|\mathcal{N}_\varepsilon| < \infty$.



Def$^{\underline{n}}$: $\mathcal{N}_\varepsilon$ as above is called an ε-net for (T,d)

H.W. Let (T,d) be a complete metric space.
ε-nets exist $\iff$ (T,d) is compact.

§. Let $(X_t)_t$ be a C-Lipschitz process. Then,

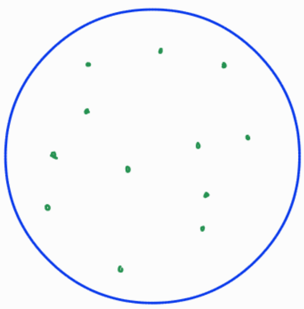
$$\sup_{t\in T} X_t = \sup_{t \in T} \left(X_{\pi(t)} + X_t - X_{\pi(t)} \right)$$

$$\leq \underbrace{\sup_{t\in T} X_{\pi(t)}}_{\sup_{s \in \mathcal{N}_\varepsilon} X_s} + \underbrace{\sup_{t \in T} \left(X_t - X_{\pi(t)} \right)}_{\leq C\varepsilon}$$

$$\boxed{\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq C\varepsilon + \sqrt{2\sigma^2 \log N(T,d,\varepsilon)}} \longrightarrow ☆$$

where $N(T,d,\varepsilon) = \inf\left\{ |\mathcal{N}_\varepsilon| \;\middle|\; \mathcal{N}_\varepsilon \text{ is an } \varepsilon\text{-net} \right\}$.

§. Example: $(B_2^n, \|\cdot\|_2)$ $\qquad B_2^n = \{ x \mid \|x\|_2 \leq 1 \}$



☆ Constructing an ε-net:

<u>ε- Packing:</u> A set $\mathcal{P}_\varepsilon \subseteq T$ is called a ε-packing if $\forall\, x, y \in \mathcal{P}_\varepsilon$, $|x-y| > \varepsilon$.

- Packing $\xrightarrow[\text{compact}]{(T,d) \text{ is}}$ maximal packing.

☆ Claim: Maximal ε-packing is an ε-net.

Proof: $\mathcal{P}_\varepsilon$ – max^{l} ε-packing

$\therefore$ $\mathcal{P}_\varepsilon \cup \{s\}$ is not an ε-packing $\forall\, s \in T$

$\Rightarrow \exists\, x_s \in \mathcal{P}_\varepsilon$ s.t. $|x_s - s| < \varepsilon \quad \forall\, s \in T$

$\therefore$ $\mathcal{P}_\varepsilon$ is also an ε-net. $\square$

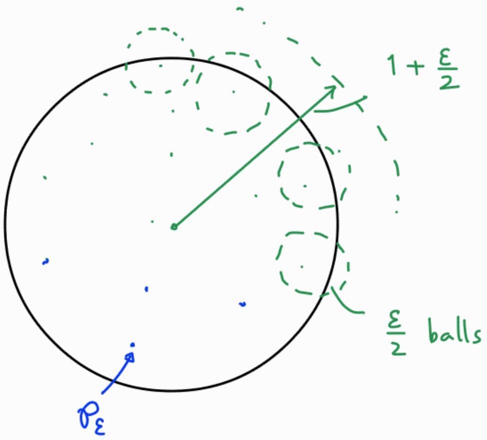
Def^{n}: $P(T, d, \varepsilon) = \max\{|\mathcal{P}_\varepsilon| \mid \mathcal{P}_\varepsilon$ is an ε-packing$\}$

We have shown $N(T, d, \varepsilon) \leq P(T, d, \varepsilon) \leq \cdot$

<u>Upper bound</u>:

$\mathbb{R}^n$

$(\varepsilon < 1)$

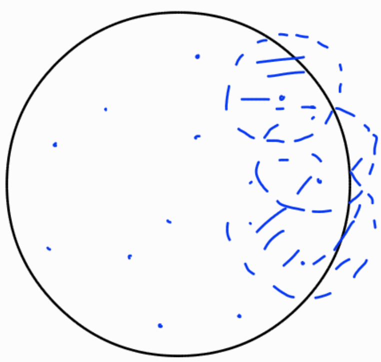


$$|\mathcal{P}_\varepsilon| \cdot \cancel{\pi_n} \left(\frac{\varepsilon}{2}\right)^n \leq \left(1 + \frac{\varepsilon}{2}\right)^n \cdot \cancel{\pi_n}$$

$$\Rightarrow |\mathcal{P}_\varepsilon| \leq \left(1 + \frac{2}{\varepsilon}\right)^n$$

$$\Rightarrow \quad N(T, d, \varepsilon) \leq \left(1 + \frac{2}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n$$

Lower bound:



$$N \cdot \cancel{\pi_n} \varepsilon^n \geq \cancel{\pi_n}$$

$$\Rightarrow N \geq \left(\frac{1}{\varepsilon}\right)^n$$

$$\boxed{\left(\frac{1}{\varepsilon}\right)^n \leq N(T, d, \varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^n}$$

Lemma: $$\left(\frac{1}{\varepsilon}\right)^n \leq N\left(B_2^n, \|\cdot\|_2, \varepsilon\right) \leq \left(\frac{3}{\varepsilon}\right)^n$$

☆ **Remark:** This leads to another notion of dimension for compact metric spaces.

☆ **Theorem:** For any compact m.s. (T,d), we have

$$\underset{\left(\frac{1}{2\varepsilon}\right)^n}{\underset{\geq}{P(T,d,2\varepsilon)}} \leq N(T,d,\varepsilon) \leq P(T,d,\varepsilon) \leq \left(\frac{3}{\varepsilon}\right)^n$$

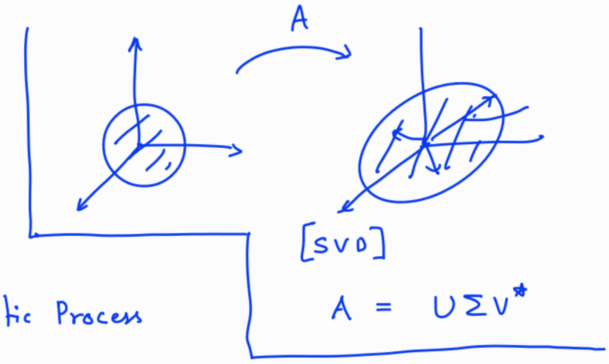
Proof: (H.W.)

Application: Random Matrices

- A – $m\times n$ random matrix, $a_{ij} \sim sG(\sigma^2)$ and a_{ij}'s are independent.

$$\|A\|_{op} = \sup_{x \in B^n} \|Ax\|_2$$

$$= \sup_{x\in B^m, y\in B^n} |\langle x, Ay\rangle|$$



Define $X_{x,y} = |\langle x, Ay\rangle|$ — Stochastic Process

$$|X_{x,y} - X_{x',y'}| = \big|\,|\langle x, Ay\rangle| - |\langle x', Ay'\rangle|\,\big|$$

$\langle x, Ay'\rangle$

$$\leq |\langle x, Ay\rangle - \langle x', Ay'\rangle|$$

$$\leq \|A\|_{op}\,\underbrace{\left(\|x-x'\| + \|y-y'\|\right)}_{d_{B^n\times B^n}((x,y),(x',y'))} \quad \text{a.s.}$$

- Also, $X_{x,y} \in sG(\sigma^2)$, $\langle x, Ay\rangle = \sum_{i,j} x_i a_{ij} y_j \in sG\left(\sigma^2 \cdot \left(\sum x_i^2\right)\left(\sum y_j^2\right)\right)$
$\subseteq sG(\sigma^2)$.

$$\mathbb{E}\,\|A\|_{op} = \mathbb{E}\left[\sup_{x,y\in B^n} X_{x,y}\right]$$

$$= \varepsilon\,\mathbb{E}[\|A\|_{op}] + \sqrt{2\sigma^2 \log N(B^m\times B^n, d, \varepsilon)}$$

ε-net for $B^n\times B^n \longrightarrow \mathcal{N}^m_{\varepsilon/2}\times \mathcal{N}^n_{\varepsilon/2}$

Cardinality $\leq \left(\frac{6}{\varepsilon}\right)^m\cdot\left(\frac{6}{\varepsilon}\right)^n$

$$\therefore \quad (1-\varepsilon)\,\mathbb{E}\,\|A\|_{op} \leq \sqrt{2\sigma^2 (m+n)\log\left(\frac{6}{\varepsilon}\right)} \qquad \varepsilon = \frac{1}{2}$$

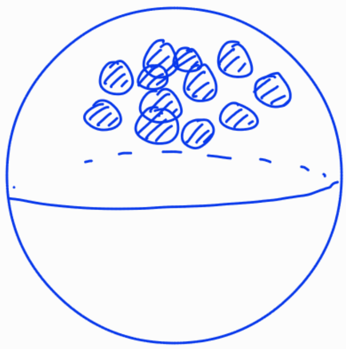
$$\boxed{\mathbb{E}\,\|A\|_{op} \lesssim \sigma\left(\sqrt{m+n}\right).}$$

∎

$\oslash$ $$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \varepsilon\,\mathbb{E}[C] + \mathbb{E}\left[\sup_{t\in\mathcal{N}_\varepsilon} X_t\right]$$

Controlling deviations.

Union bound.



$\oslash$ $\sup_{x\in B^n} \|Ax\|$

1. Split T into patches which intersect minimally (minimal ε-net)

2. Bound the deviation from a point on each patch.

☆ Chaining:

$\pi_0 \longrightarrow \varepsilon$-net. , $\pi_1 \longrightarrow \varepsilon/2$ net

finite

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \leq \mathbb{E}\sup_{t\in T}\left[X_{\pi_0(t)}\right] + \mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi_0(t)}\right)\right]$$

Supremum of r.v. $\longrightarrow$ Involving the geometry of the 'dependence metric'.

$$\leq \mathbb{E}\sup_{t\in T} X_{\pi_0(t)} + \mathbb{E}\left[\sup_{t\in T}\left(-X_{\pi_0(t)} + X_{\pi_1(t)}\right)\right] + \mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi_1(t)}\right)\right]$$

finite

$$\boxed{\mathbb{E}[\sup X_t] \leq \sum_k 2^{-k}\sqrt{\log N(T, d, 2^{-k})}\cdot\sigma}$$

⋮

$$\leq \mathbb{E}\left[\sup_{t\in T} X_{\pi_0(t)}\right] + \sum_{k=1}^{n}\mathbb{E}\left[\sup_{t\in T}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right)\right] + \underbrace{\mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi_n(t)}\right)\right]}_{\to 0 \text{ as } n\to\infty}$$

$$\boxed{\pi_k \rightsquigarrow \varepsilon = 2^{-k}}$$

finite maximum

If, $\mathbb{E}\left[\sup_{t\in T}\left(X_t - X_{\pi_n(t)}\right)\right] \longrightarrow 0$ as $n \to \infty$,

then we do not a Lipschitz bound!

$\longrightarrow$ Dudley's Inequality.

Wasserstein LLN:

Measures: Ω is a set, $\mu : \mathcal{B} \longrightarrow [0,\infty)$, where $\mathcal{B} \subseteq \mathcal{P}(\Omega)$ s.t

i) $\mu(\emptyset) = 0$

ii) $\mu\left(\coprod_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty}\mu(A_i)$

iii) $A \subseteq B \in \mathcal{B}$, then $\mu(A) \leq \mu(B)$

- On $\mathbb{R}$ $\longrightarrow$ Lebesgue measure
  $\longrightarrow$ Dirac measure

$$\delta_a(A) = \mathbb{1}_{[a\in A]}$$

$$\int f\,d\mu = \mu(f) = \mu f$$

- $\int f\,\delta_a := f(a)$

⊘ Another look at weak law:

X_i iid μ, $\quad \mathrm{Supp}[X_i] \subseteq [0,1]$.

$\rightarrow \mathcal{F} = \{ f \mid |f(x) - f(y)| \leq |x-y|, \ 0 \leq f \leq 1 \}$

Define $\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$ $\rightsquigarrow$ Random measure

- $\int f \, d\mu_n = \frac{1}{n} \left(f(X_1) + \dots + f(X_n) \right)$, $\quad f \in \mathcal{F}$

$$\xrightarrow{p} \int f \, d\mu \quad \text{(weak law)}$$
$$= \mathbb{E} \, f(X).$$

- $\mathbb{E} \left| \int f d\mu_n - \int f d\mu \right| \leq \frac{\|f\|_\infty}{\sqrt{n}}$ (Chebychev)

Wasserstein metric:

$$\boxed{W(\mu_n, \mu) = \sup_{f \in \mathcal{F}} \left| \int f d\mu_n - \int f d\mu \right|}$$

As $n \to \infty$, $W(\mu_n, \mu) \longrightarrow 0$ as each individual term

⊘ What is the rate of convergence ??

Thm: $W(\mu_n, \mu) = O\left(\frac{1}{\sqrt{n}}\right)$.

1st attempt: $W(\mu_n, \mu) = \sup_{f \in \mathcal{F}} X_f$, $\quad X_f = \left| \int f d\mu_n - \int f d\mu \right|$

Lipschitz bound

Size of an ε-net here

$$|X_f - X_g| \leq \left| \left| \int f d\mu_n - \int f d\mu \right| - \left| \int g d\mu_n - \int g d\mu \right| \right|$$
$$\leq \underline{2 \|f - g\|_\infty}$$

$N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon) \leq e^{C/\varepsilon}$

Cannot be improved.

$$\mathbb{E}\left[W(\mu_n,\mu) \right] \leq \inf_{\varepsilon>0} \left[2\varepsilon + \sqrt{2\sigma^2 \log e^{C/\varepsilon}} \right],$$

where $\sigma^2 = \sup_f \operatorname{Var}(X_f)$

$$= \sup_{f\in\mathcal{F}} \mathbb{E}\left[\Big| \underbrace{\int f\,d\mu_n}_{= \frac{\sum f(X_i)}{n}} - \int f\,d\mu \Big|^2 \right]$$

$$\operatorname{Var}\left(\frac{\sum X_i}{n}\right) = \frac{1}{n}\operatorname{Var}(X)$$

$$= \frac{\operatorname{Var}(f(X))}{n}$$

$$\mathbb{E}\left[W(\mu_n,\mu) \right] \leq \inf_{\varepsilon>0} \left[2\varepsilon + \frac{C_0}{\sqrt{\varepsilon n}} \right]$$

$$\lesssim n^{-1/3}. \qquad ▨$$