Chap 2.

Class 4

Chap 2: I - LLN
II - Fluctua[tions]

II Fluctuat°

Now that we know that

$$\frac{L(Nx, Ny)}{N} \xrightarrow[N\to+\infty]{} s(x,y) \quad \text{if } \mathbb{E}|w|^{2+\varepsilon} < +\infty,$$

what about fluctuations?

That is to say, is there a limit:

$$\frac{L(Nx, Ny) - N s(x,y)}{\sigma(x,y)\, N^{\boxed{?}}} \longrightarrow ?$$

This turns out to be a surprisingly hard questo.

Conjecture: (Another KPZ conjecture, "Fixed point")

$$\boxed{\frac{L(Nx, Ny) - N s(x,y)}{\sigma(x,y)\, N^{1/3}} \xrightarrow{\mathcal{L}} TW}$$

(Tracy-Widom distributo)

As a first approximation to $N^{1/3}$ fluctuations, we will develop a general theory which works also for the Gaussian setting - this is the generic setting. It says that $N^{1/2}$ scaling is the worst case scenario

Thm: If ω bounded by $L>0$ a.s

Then $\exists C, c>0$ such that

$$\mathbb{P}\left(\left| L(N,M) - \mathbb{E} L(N,M) \right| \geq (N+M) \cancel{t} \right)$$

$$\leq C e^{-ct^2(N+M)}$$

⚠ a-dimensional.

The main tool is

II.$\boxed{1}$ Talagrand's inequality and its corollary

Thm [Talagrand]

Let $\begin{cases} \Omega = [0,1]^n & \text{be a product space with product measure} \\ \mathbb{P} = p^{\otimes n} \end{cases}$

Then for all $A \subset \Omega$, if $\mathcal{L}(X) = \mathbb{P}$,

$$\mathbb{E}\left[e^{\frac{1}{4} d_c(X,A)^2} \right] \leq \frac{1}{\mathbb{P}(A)}$$

⚠ No n A-dimensional

where $d_c(x,A)$ is the "convex distance" from x to A

Definition of the "convex distance"

$$d_c(x,A) := d_E(0, \operatorname{Conv} \mathcal{S}(A-x))$$

where $\operatorname{Conv} X :=$ Convex Hull of X

$$\mathcal{S}(A) := \{ s \in \{0,1\}^n \mid \exists y \in A, \ s_i = 0 \Rightarrow y_i = 0 \}$$

$$= \bigcup_{y \in A} \{ s \in \{0,1\}^n \mid s_i = 0 \Rightarrow y_i = 0 \}$$

$\mathcal{S}(A)$ measures complexity of A

Intuition: $\mathcal{S}(A)$ is the set of possible supports of elements in A.

why support? $s_i = 1 \Rightarrow y_i$ can be anything

$s_i = 0 \Rightarrow y_i = 0$ ie coordinate i is not used

Corollary: Let $X = (X_1, X_2, \dots, X_n)$ be a r.v with iid components in $[0,1]$.

$F : \mathbb{R}^n \to \mathbb{R}$ convex, 1- lipschitz.

Then (1) $\mathbb{P}(|F(X) - MF(X)| \geq t) \leq 4 e^{-t^2/4}$

(2) $\mathbb{P}(|F(X) - \mathbb{E} F(X)| \geq t) \leq C \cdot e^{-ct^2}$

where MY = Median of Y
= Any $a \in \mathbb{R}$ s.t $\mathbb{P}(Y \geq a) \wedge \mathbb{P}(Y \leq a) \geq 1/2$.

Proof of Corollary:

↳ [(1) ⇒ (2)] $|\mathbb{E} F(X) - MF(X)| \leq \mathbb{E} |F(X) - MF(X)|$

$$\leq \int_0^{+\infty} dt\, \mathbb{P}(|F(X) - MF(X)| \geq t) \leq \int_0^{+\infty} 4 dt\, e^{-t^2/4} = \underset{\text{Absolute}}{C}$$

Then $\mathbb{P}(|F(X) - \mathbb{E} F(X)| \geq t)$

$$= \mathbb{P}(|F(X) - \mathbb{E} F(X)| - C \geq t - C)$$

Triangular Ineq. $\leq \mathbb{P}(|F(X) - MF(X)| + \underbrace{|\mathbb{E} F(X) - MF(X)| - C}_{\leq 0} \geq t - C)$

$$\leq \mathbb{P}(|F(X) - MF(X)| \geq t - C) \leq 4 e^{-\frac{1}{4}(t-C)^2}$$

$$\leq C' e^{-c' t^2}.$$

↳ [Talagrand inequality ⇒ (1)]

① First, the convex distance controls the convex distance

$\forall A \subset [0,1]^n$ convex, $x \in [0,1]^n$, $d_E(x, A) \leq d_C(x, A)$

Indeed $d_C(x, A) \leq t \iff \exists\, w \in \mathbb{R}^m$ s.t $\begin{cases} |w|_E \leq t \\ w \in \operatorname{Conv} \mathcal{S}(A - x) \end{cases}$

$\iff \exists\, w \in \mathbb{R}^n$ s.t $\begin{cases} |w|_E \leq t \\ w = \sum_{j=1}^{m} \lambda_j s_j \end{cases}$

$s_j \in \mathcal{S}(A - x)$; $\sum \lambda_j = 1$

IV

By definition of $\mathcal{S}(A-x)$,

$\forall s_j \in U_A(x)$, $\exists \tilde{z}_j \in A-x$, $s_j^i = 0 \implies \tilde{z}_j^i = 0$

$\exists z_j \in A$, $s_j^i = 0 \implies (z_j - x)^i = 0$ / $(z_j - x)^i \in [-1,1]$

$\hookrightarrow \quad |(z_j^i - x^i)| \leqslant s_j^i$ as in $[-1,1]$ & support conditions

Let $z = \sum_{j=1}^{m} \lambda_j z_j \in A$ by convexity of A.

$$\implies d_E(x, A) \leqslant |z - x|_E = \sqrt{\sum_{i=1}^{n} |z^i - x^i|^2} \leqslant \sqrt{\sum_{i=1}^{n} (\omega^i)^2} = \|\omega\|_E \leqslant t$$

But $|(z-x)^i| \leqslant \sum_{j=1}^{m} \lambda_j |(z_j - x)^i| \leqslant \sum_{j=1}^{m} \lambda_j s_j^i = \omega^i$

In conclusion $d_c(x,A) \leqslant t \implies d_E(x,A) \leqslant t$ $\square$

② Let us now prove that for all $a \in \mathbb{R}$, $t \geqslant 0$ we have:

$$(\star) \quad \mathbb{P}(F(X) \geqslant a+t)\, \mathbb{P}(F(X) \leqslant a) \leqslant e^{-t^2/4}$$

By writing $A = \{F \leqslant a\}$, we have

$$(\star) \iff \mathbb{P}(F(X) \geqslant a+t) \leqslant \frac{e^{-t^2/4}}{\mathbb{P}(A)}$$

By Talagrand's inequality:

$$\mathbb{P}(d_c(X,A) \geqslant t) = \mathbb{E}\left[\mathbb{1}_{\{d_c(X,A) \geqslant t\}} e^{t^2/4}\right] e^{-t^2/4} \overset{\text{Chebychev}}{\leqslant} \mathbb{E}\left(e^{d_c(X,A)^2/4}\right) e^{-t^2/4} \overset{\text{Talagrand}}{\leqslant} \frac{e^{-t^2/4}}{\mathbb{P}(A)}$$

Thus $(\star) \Leftarrow \mathbb{P}(F(X) \geqslant a+t) \leqslant \mathbb{P}(d_c(X,A) \geqslant t)$

$\Leftarrow \{d_c(X,A) \geqslant t\} \supset \{d_E(X,A) \geqslant t\}$ (step 1)

Use F 1-lip $\supseteq \{F(X) \geqslant t+a\}$

For the last implicato:

$$F(x) \geq t+a \Rightarrow \forall y \in A, \quad |x-y| \geq |F(x)-F(y)|$$
$$\geq F(x) - F(y)$$
$$\geq t+a - F(y) \geq t$$
$$\Rightarrow d_E(x,A) \geq t \quad \boxed{}$$

(3) • Apply (*) with $a = MF(X)$: $\mathbb{P}(F(X)-MF(X) \geq t) \frac{1}{2} \leq e^{-t^2/4}$
and $a = t - MF(X)$: $\frac{1}{2}\mathbb{P}(MF(X)-F(X) \geq t) \leq e^{-t^2/4}$

Together $\mathbb{P}(|F(X)-MF(X)| \geq t) \leq 4e^{-t^2/4}$ $\boxed{}$

II-[2] Application to DLPP:
The concentration of measure.

$$L(N,M) = \max_{\substack{\pi : (1,1) \to (N,M)}} \sum_{(i,j)\in\pi} w_{ij} = F\Big(\underbrace{w_{ij} ; \begin{array}{l} 1\leq i \leq N \\ 1\leq j \leq M \end{array}}_{X \in \mathbb{R}^n,\ n = N\times M}\Big)$$

F is

- convex as max of linear functions

$$F(X) = \max_{\pi \in \Pi^{N,M}} \mathcal{L}_\pi(X)$$

Dual canonical basis

where $\mathcal{L}_\pi = \sum_{(i,j)\in\pi} e_{ij}^* \in (\mathbb{R}^{N\times M})^*$ linear forms

- L- Lipschitz for $L = \sqrt{N+M}$

Notice that if $X \in \mathbb{R}^n$ is s.t $\exists! \pi^*$ optimal:

$$L(N,M) = \mathcal{L}_{\pi^*}(X) > \max_{\pi \neq \pi^*} \mathcal{L}_\pi(X)$$

open condition!

Then $\begin{cases} F(X) = \mathcal{L}_{\pi^*}(X) \\ \& \ \nabla F(X) = \mathcal{L}_{\pi^*} \end{cases} \Rightarrow \|\nabla F(X)\|_2 = \|\mathcal{L}_{\pi^*}\|_2 = \sqrt{N+M}$

Is It always true? No ... The complement set

$$\{ X \in \mathbb{R}^n \mid \exists \, \pi_1 \neq \pi_2 \text{ s.t } L(N,M) = \mathcal{L}_{\pi_1}(X) = \mathcal{L}_{\pi_2}(X) \}$$

$$\subset \bigcup_{\pi_1 \neq \pi_2} \ker \left(\mathcal{L}_{\pi_1} - \mathcal{L}_{\pi_2} \right)$$

is contained in a closed subset of zero Lebesgue dimension $\leadsto$

$$\begin{cases} \|\nabla F(X)\|_2 = \sqrt{N+M} & \text{for a.e } X \\ F \text{ differentiable a.e} \end{cases}$$

Therefore

$$|F(X) - F(Y)| \underset{(*)}{=} \left| \int_0^1 \nabla F(\underbrace{X + t(Y-X)}_{\xi_t}) \cdot (Y - X) dt \right|$$

$$\leq \int_0^1 | \nabla F(\xi_t) \cdot (Y - X) dt |$$

$$\overset{C.S}{\leq} \sqrt{N+M} \, \|X - Y\|_2$$

⚠ We are not exactly done

If $\lambda(\mathbb{R}^k)$ is the Lebesgue measure on $\mathbb{R}^k$,

$\| \nabla F(X) \|_2 = \sqrt{N+M}$ for $\underline{\lambda(\mathbb{R}^k)}$ - a.e

And $(*)$ needs a.e equality for the Lebesgue measure on lines.

How to fix it? One needs "for almost every lines"

+ (X, Y) are random, not falling on lines.

We will not discuss these details.

We mention them for the sake of completeness.

If $\begin{cases} |w_{ij}| \leqslant L \quad \text{for } L \leqslant 1 \\ \tilde{F} = \dfrac{F}{L\sqrt{N+M}} \quad \text{is convex \& } \dfrac{1}{L}\text{-Lipschitz} \\ \qquad\qquad\qquad\qquad \rightsquigarrow 1\text{-Lipschitz} \end{cases}$

Thus $\exists\ C, c > 0$,

$$\mathbb{P}\left(\left| \tilde{F}(w_{ij}) - \mathbb{E}\,\tilde{F}(w_{ij}) \right| \geqslant t \right) \leqslant C \cdot e^{-ct^2}.$$

$$\Rightarrow \mathbb{P}\left(\left| L(N,M) - \mathbb{E}\, L(N,M) \right| \geqslant t L \sqrt{N+M} \right) \leqslant C \cdot e^{-ct^2}$$

$\boxtimes$