Mathematics of non-asymptotic statistics:

- typical quantities involved

  empirical processes: $X_1, \dots, X_n \in \mathbb{R}^d$, iid, $f : \mathbb{R}^d \to \mathbb{R}$

  $$R(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_1)]$$

  suprema of empirical processes:

  $$R(\mathcal{F}) = \sup_{f \in \mathcal{F}} \left\{ \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_1)] \right\}.$$

- tools of asymptotic statistics: Central Limit Theorem

  $$\sqrt{n}\, R(f) \xrightarrow{(d)} Z, \quad \text{where } Z \sim \mathcal{N}(0, \mathrm{Var}(f(X_1)))$$

  Ex: if f is L-Lipschitz and $X_1, \dots, X_n \in \mathbb{R}$, $\mathrm{Var}(X_i) = \sigma^2$, iid.

  $$\mathrm{Var}(f(X_1)) = \frac{1}{2} \mathbb{E}\left[(f(X_1) - f(X_2))^2\right]$$
  $$\leqslant \frac{L^2}{2} \mathbb{E}\left[(X_1 - X_2)^2\right] = L^2 \sigma^2$$

  So

  $$\lim_{n \to \infty} \mathbb{P}\left[\frac{1}{n} \sum_{i=1}^{n} f(X_i) \geqslant \mathbb{E}[f(X_1)] + \frac{L\sigma}{\sqrt{n}} x\right] \overset{\text{CLT}}{=} \mathbb{P}\left[Z \geqslant L\sigma x\right]$$
  $$\underset{\substack{\text{Lemma B.4} \\ \text{p 292}}}{\leqslant} e^{-x^2/2}$$

Can we get non-asymptotic versions of such statements?

Tools of non-asymptotics statistics: concentration inequalities.

Theorem B.7: Gaussian concentration inequality

If $X_1, \dots, X_n$ iid with $\mathcal{N}(0,\sigma^2)$ distribution, and $F : \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz, then there exists $\xi \overset{d}{\sim} \mathcal{E}xp(1)$ such that

$$F(X_1,\dots,X_n) \leq \mathbb{E}[F(X_1,\dots,X_n)] + L\sigma\sqrt{2\xi}$$

(equivalently $\mathbb{P}[F(X_1,\dots,X_n) \geq \mathbb{E}[F(X_1,\dots,X_n)] + L\sigma\sqrt{2x}] \leq e^{-x}$)

Ex: if $f : \mathbb{R} \to \mathbb{R}$ L-Lipschitz, then

$$\left| \frac{1}{n}\sum_{i=1}^n f(X_i) - \frac{1}{n}\sum_{i=1}^n f(Y_i) \right| \leq \frac{L}{n}\sum_{i=1}^n |X_i - Y_i|$$

$$\underset{\text{Cauchy-Schwartz}}{\leq} \frac{L}{\sqrt{n}}\sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Hence, $F(X_1,\dots,X_n) = \frac{1}{n}\sum_{i=1}^n f(X_i)$ is $\frac{L}{\sqrt{n}}$-Lipschitz, so Theorem B.7 gives

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X_1)] \geq \frac{L\sigma}{\sqrt{n}}x\right] \leq \mathbb{P}[\sqrt{2\xi} \geq x] = e^{-x^2/2}$$

Important example: $\varepsilon \sim \mathcal{N}(0, I_n)$, $F(\varepsilon) = \|\varepsilon\|$ is 1-Lipschitz so

$$\mathbb{P}[\|\varepsilon\| \geq \underbrace{\mathbb{E}[\|\varepsilon\|]}_{?} + \sqrt{2x}] \leq e^{-x}$$

$$\mathbb{E}[\|\varepsilon\|] \underset{\text{Jensen}}{\leq} \sqrt{\mathbb{E}(\|\varepsilon\|^2)} = \sqrt{\sum_{i=1}^n \mathbb{E}[\varepsilon_i^2]} = \sqrt{n}$$

So $\mathbb{P}[\|\varepsilon\|^2 \geq n + 2\sqrt{2nx} + 2x] \leq e^{-x}$

Exo 1.6.6: recover this bound from Markov inequality.

Remark: $F(\varepsilon) = -\|\varepsilon\|$ is also 1-Lipschitz so

$$\mathbb{P}\left[\|\varepsilon\| \leq \mathbb{E}[\|\varepsilon\|] - \sqrt{2x}\right] \leq e^{-x}$$

lower bound on $\mathbb{E}[\|\varepsilon\|]$?

as $\|\varepsilon\| \leq \mathbb{E}[\|\varepsilon\|] + \sqrt{2\zeta}$ with $\zeta \overset{d}{\lesssim} \mathcal{E}xp(1)$

then
$$\underbrace{\mathbb{E}[\|\varepsilon\|^2]}_{=m} \leq \mathbb{E}\left[\left(\mathbb{E}[\|\varepsilon\|] + \sqrt{2\zeta}\right)^2\right]$$
$$\underset{\text{Jensen}}{\leq} \left(\mathbb{E}[\|\varepsilon\|] + \sqrt{2\underbrace{\mathbb{E}[\zeta]}_{=1}}\right)^2$$

as $x \to (a + \sqrt{2x})^2$ is concave

So we have proved

$$\sqrt{m} - \sqrt{2} \leq \mathbb{E}[\|\varepsilon\|] \leq \sqrt{m}$$

Bounding expectations of supremum

- Deviations: Assume that $X_1, \dots, X_p$ are independent

$$\mathbb{P}\left[\max_{j=1,\dots,p} X_j > t\right] = 1 - \mathbb{P}\left[\forall j:\ X_j \leq t\right]$$
$$\overset{\perp\!\!\!\perp}{=} 1 - \prod_{j=1}^{p} \mathbb{P}\left[X_j \leq t\right]$$
$$= 1 - \prod_{j=1}^{p} \left(1 - \mathbb{P}[X_j > t]\right)$$
$$\underset{t \to \infty}{\sim} \sum_{j=1}^{p} \mathbb{P}[X_j > t]$$

In addition, for __any__ $X_1, \dots, X_p$, we always have

$$\mathbb{P}\left[\max_{j=1,\dots,p} X_j > t\right] \leq \sum_{j=1}^{p} \mathbb{P}[X_j > t] \quad \text{(union bound)}$$

Expectations:

- $X_j \geq 0$: then $\mathbb{E}[\max_{j=1,\dots,p} X_j] \leq \sum_{j=1}^{p} \mathbb{E}[X_j]$
- without assumptions: $\forall \alpha \in \mathbb{R}$

$$\mathbb{E}[\max_{j=1,\dots,p} X_j] \underset{\substack{\uparrow \\ \max_j X_j \leq \alpha + \max_j (X_j-\alpha)_+}}{\leq} \alpha + \sum_{j=1}^{p} \mathbb{E}[(X_j-\alpha)_+]$$

Ex: if $X_j \in \mathrm{SubG}(\sigma^2)$, ie $\mathbb{E}[X_j]=0$ and $\mathbb{E}[e^{sX_j}] \leq e^{\sigma^2 s^2/2}$;

. we have

$$\mathbb{P}[X_j \geq t] \underset{\substack{\text{Markov} \\ \text{(Lemma B.1.)}}}{\leq} e^{-st}\,\mathbb{E}[e^{sX_j}] \leq e^{-st} e^{\sigma^2 s^2/2} \underset{\substack{\uparrow \\ s = t/\sigma^2}}{=} e^{-t^2/2\sigma^2}$$

So

$$\mathbb{E}[\max_{j=1,\dots,p} X_j^2] \leq \alpha + \sum_{j=1}^{p} \underbrace{\mathbb{E}[(X_j^2-\alpha)_+]}$$

$$= \int_0^\infty \mathbb{P}[X_j^2-\alpha > t]\,dt$$

$$= 2\int_0^\infty e^{-(t+\alpha)/2\sigma^2}\,dt = 4\sigma^2 e^{-\alpha/2\sigma^2}$$

for $\alpha = 2\sigma^2 \log 2p$, we get

$$\mathbb{E}[\max_{j=1,\dots,p} X_j^2] \leq 2\sigma^2 \log(2ep)$$

Remark: we also get

$$\mathbb{E}[\max_{j=1,\dots,p} X_j] \underset{\text{Jensen}}{\leq} \sqrt{\mathbb{E}[(\max_{j=1..p} X_j)^2]}$$

$$\leq \sqrt{\mathbb{E}[\max_{j=1..p} X_j^2]}$$

$$\leq \sigma\sqrt{2\log(2ep)}$$

We can generalize this last approach:

<u>Lemma</u>: for any $\varphi : \underset{\text{interval}}{I} \longrightarrow \mathbb{R}^+$ convex, we have

$$\boxed{\varphi\left(\mathbb{E}\left[\max_{j=1,\dots,p} X_j\right]\right) \leqslant \sum_{j=1}^{p} \mathbb{E}\left[\varphi(X_j)\right]}$$

<u>Proof</u>: $\varphi\left(\mathbb{E}\left[\max_j X_j\right]\right) \overset{\text{Jensen}}{\leqslant} \mathbb{E}\left[\varphi(\max_j X_j)\right]$

$$\leqslant \mathbb{E}\left[\max_j \varphi(X_j)\right]$$

$$\underset{\varphi \geqslant 0}{\leqslant} \sum_j \mathbb{E}\left[\varphi(X_j)\right]$$

$\square$

<u>Ex</u>: if $X_j \in \mathrm{SubG}(\sigma^2)$: with $\varphi(x) = e^{sx}$

$$\mathbb{E}\left[\max_{j=1..p} X_j\right] \leqslant \varphi^{-1}\left(\mathbb{E}\left[\max_{j=1..p} \varphi(X_j)\right]\right)$$

$$\leqslant \frac{1}{s} \log\left(\sum_{j=1}^{p} \underbrace{\mathbb{E}\left[e^{sX_j}\right]}_{\leqslant e^{s^2\sigma^2/2}}\right)$$

$$= \frac{1}{s}\log p + \frac{\sigma^2}{2} s$$

$$s = \sqrt{\frac{2\log p}{\sigma^2}} \longrightarrow \; = \sigma\sqrt{2\log p}$$

- <u>Supremum of an infinite number of variables?</u>

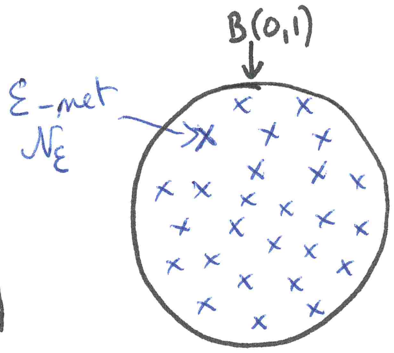
For example, for $X \in \mathbb{R}^{m \times p}$ with $X_{ij} \overset{iid}{\sim} \mathcal{N}(0,1)$, then

$$|X|_{op} = \max_{u \in B(0,1)} \|Xu\|$$

$\rightsquigarrow$ we cannot bound with $\sum_{u \in B(0,1)} \mathbb{E}\left[\|Xu\|\right]$ $(= +\infty)$

For an ε-net $\mathcal{N}_\varepsilon$:

$$\max_{u\in B(0,1)} \|Xu\| = \max_{u\in\mathcal{N}_\varepsilon}\|Xu\| + \left(\max_{u\in B(0,1)}\|Xu\| - \max_{u\in\mathcal{N}_\varepsilon}\|Xu\|\right)$$



$$\leq \underbrace{\max_{u\in\mathcal{N}_\varepsilon}\|Xu\|}_{\text{can be handled as before.}} + \underbrace{\max_{\|u\|\leq\varepsilon}\|Xu\|}_{\leq |X|_{op}\,\varepsilon}$$

Sometimes, some different levels of approximation are needed

$\leadsto$ chaining.

Matthieu Lerasle will explain all this –