

# Analyse de la régression robuste avec des hypothèses de concentration de la mesure

Cosme Louart, Romain Couillet

GIPSA-lab, INP Grenoble; List-CEA

23/10/2020



# Table of contents

## Introduction

I- Concentration of the Measure Phenomenon

II - Performances of the robust regression

# Robust Regression algorithm

- ▶ Data matrix  $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$
- ▶ labels :  $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$
- ▶ Robust regression problem with regularizing parameter:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho(y_i - x_i^T \beta) + \lambda \|\beta\|^2$$

with  $\rho : \mathbb{R} \rightarrow \mathbb{R}$  convex,  $\lambda > 0$ .

- ▶ Score of a new data  $x \in \mathbb{R}^p$  :  $\beta^T x$

**Performance:**  $\mathbb{E}_{X,x}[\rho(\beta^T x - y_x)]$

**Goal:** Understand the statistics of  $\beta = f(X)$ .

# Setting and conclusion

## Concentration hypotheses on the data $X$

- ▶ For all 1-Lipschitz maps  $f : \mathcal{M}_{p,n} \rightarrow \mathbb{R}$ :

$$\forall t > 0 : \mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \geq t) \leq Ce^{-ct^2}$$

- ▶  $x_1, \dots, x_n$  are independent

## Assets

- ▶ **(Representativity)** True if the columns are Lipschitz transformation of a Gaussian vector  $Z \sim \mathcal{N}(0, I_p)$ .  
→ dependence between entries of a column possibly complex
- ▶ **(Flexibility)** the inequality can be extended to the weight vector  $\beta = \beta(X)$

# Table of Contents

## Introduction

## I- Concentration of the Measure Phenomenon

## II - Performances of the robust regression

# Table of contents

## Introduction

## I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to “Concentrated equation”

## II - Performances of the robust regression

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

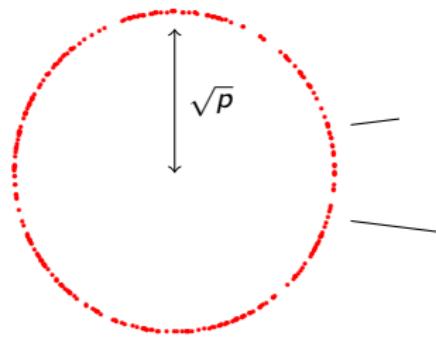
- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Concentration of Measure Phenomenon<sup>1</sup>

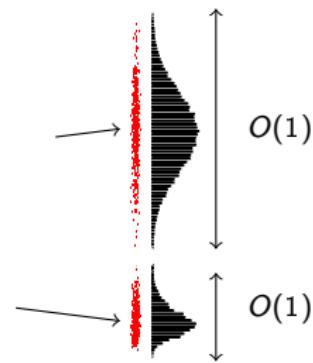
$$X = (X_1, \dots, X_p) \sim s_p$$



$$\frac{X_1 + \dots + X_p}{\sqrt{p}}$$

$$\|X\|_\infty$$

Observations



$$\begin{aligned} \text{Distribution diameter} &= \mathbb{E}[\|Z - \mathbb{E}Z\|] \\ &\stackrel{p \rightarrow \infty}{=} O(\sqrt{p}) \end{aligned}$$

$$\text{Observable diameter} \underset{p \rightarrow \infty}{=} O(1)$$

---

<sup>1</sup>Ledoux - 2001 : The concentration of measure phenomenon

# Fundamental example of the Theory

## Theorem

$Z \in \mathbb{R}^p$ , if  $Z$  uniformly distributed on  $\sqrt{p}\mathcal{S}^{p-1}$  or  $Z \sim \mathcal{N}(0, I_p)$ :  
 $\forall f : E \rightarrow \mathbb{R}$  1-Lipschitz :

$$\forall t > 0 : \mathbb{P}(|f(Z) - \mathbb{E}[f(Z')]| \geq t) \leq 2e^{-t^2/2},$$

we note (since  $2 \underset{p \rightarrow \infty}{=} O(1)$ ):

$$Z \propto \mathcal{E}_2(1) \quad \text{or, more simply,} \quad Z \propto \mathcal{E}_2$$

= Standard hypothesis

# Notations

$(E, \|\cdot\|)$ , normed vector space,  $Z \in E$ , random vector

- ▶  $\mathbb{R}^p$  endowed with:  $\|x\| = \sqrt{\sum_{i=1}^p x_i^2}$  or  $\|x\|_\infty = \sup_{1 \leq i \leq p} |x_i|$
- ▶  $\mathcal{M}_{p,n}$  endowed with:  $\|M\|_F = \sqrt{\text{Tr}(MM^T)} = \sqrt{\sum_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}} M_{i,j}^2}$   
or  $\|M\| = \sup_{\|x\| \leq 1} \|Mx\|$

Lipschitz concentration and linear concentration

- ▶ “ $Z \propto \mathcal{E}_2(\sigma)$ ”

$\exists C, c > 0 \mid \forall p, n \in \mathbb{N}, \forall f : E \rightarrow \mathbb{R}$  1-Lipschitz, :

$$\forall t > 0 : \mathbb{P}(|f(Z) - \mathbb{E}[f(Z)]| \geq t) \leq Ce^{-c(t/\sigma)^2},$$

$\sigma = \sigma_{p,n}$  : Observable Diameter of  $Z$ .

- ▶ “ $Z \in \tilde{Z} \pm \mathcal{E}_2(\sigma)$ ”

In particular, if  $\forall p, n \in \mathbb{N}, \forall u : E \rightarrow \mathbb{R}$  1-Lipschitz and linear :

$$\forall t > 0 : \mathbb{P}\left(\left|u(Z - \tilde{Z})\right| \geq t\right) \leq Ce^{-c(t/\sigma)^2},$$

$\tilde{Z}$  : Deterministic equivalent of  $Z$ . ( $Z \propto \mathcal{E}_2(\sigma) \implies Z \in \mathbb{E}[Z] \pm \mathcal{E}_2(\sigma)$ )

# How to build new concentrated random vectors ?

- ▶ If  $Z \in \mathcal{E}_2(\sigma)$  and  $f : E \rightarrow E$   $\lambda$ -Lipschitz,  $f(Z) \in \mathcal{E}_2(\lambda\sigma)$
- ▶ No simple way to set the concentration of  $(Z_1, \dots, Z_p)$  if  $Z_1 \in \mathcal{E}_2(\sigma), \dots, Z_p \in \mathcal{E}_2(\sigma)$  non independent
- ▶  $Z_1, Z_2 \in C\mathcal{E}_q(\sigma)$ , independent  $(Z_1, Z_2) \in \mathcal{E}_q(\sigma)$
- ▶  $(Z_1, Z_2) = f(Z)$  where  $Z \in \mathcal{E}_q(\sigma)$ , and  $f$  1-Lipschitz  $(Z_1, Z_2) \in \mathcal{E}_q(\sigma)$

# Realistic images built with GANS are concentrated

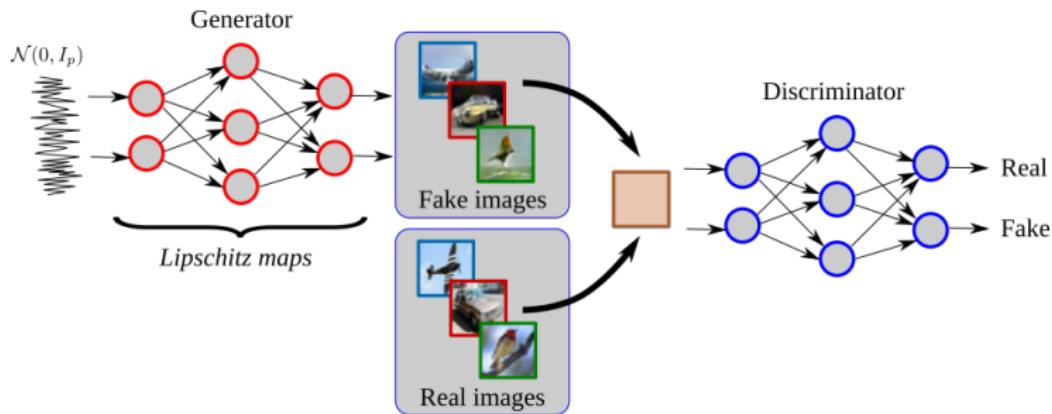


IMAGE =  $f(Z)$ , with  $f$  1 – Lipschitz and  $Z \sim \mathcal{N}(0, I_p)$



# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables**
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

## Same Notations:

### Lemma

Given a random  $Z \in \mathbb{R}$  (depending on  $p, n$ ):

$$Z \propto \mathcal{E}_2(\sigma) \iff Z \in \mathbb{E}[Z] \pm E_2(\sigma)$$

$$(\iff \forall t > 0 : \mathbb{P}(|Z - a| \geq t) \leq Ce^{-c(t/\sigma)^2})$$

### Lemma

$Z \in a \pm \mathcal{E}_2(\sigma)$  iff:

$$\iff \forall f : \mathbb{R} \rightarrow \mathbb{R}, \text{ } O(1)\text{-Lipschitz, } f(Z) \in f(a) \pm \mathcal{E}_2(\sigma)$$

### Example

$X \sim \mathcal{N}(0, I_p)$ ,  $f : \mathbb{R}^p \rightarrow \mathbb{R}$ , and  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  1-Lipschitz:

$$\phi(f(X)) \in \phi(\mathbb{E}[f(X)]) \pm \mathcal{E}_2$$

# Characterization with the centred moments

## Proposition

$$Z \in a \pm e^{-(\cdot/\sigma)^q}$$

$$\iff \exists c > 0 \mid \forall p, n \in \mathbb{N}, \forall r \geq 2 : \mathbb{E}[|Z - a|^r] \leq \left(\frac{r}{q}\right)^{\frac{r}{q}} c \sigma^r$$

Proof :

① Fubini:

$$\begin{aligned} \mathbb{E}[|Z - a|^r] &= \int_Z \left( \int_0^\infty \mathbb{1}_{t \leq |Z-a|^r} dt \right) dZ \\ &= \int_0^\infty \mathbb{P}(|Z - a|^r \geq t) dt \\ &\leq \int_0^\infty C e^{-t^{\frac{q}{r}}/\sigma^q} dt \dots \leq C \left(\frac{r}{q}\right)^{\frac{r}{q}} \sigma^r \end{aligned}$$

② Markov inequality:

$$\mathbb{P}(|Z - a| \geq t) \leq \frac{\mathbb{E}[|Z-a|^r]}{t^r} \leq C \left(\frac{r}{q}\right)^{\frac{r}{q}} \left(\frac{\sigma}{t}\right)^r,$$

$$\text{with } r = \frac{qt^q}{e\sigma^q} \geq q : \mathbb{P}(|Z - a| \geq t) \leq C e^{-(t/\sigma)^q/e}.$$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector**
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Control of the norm

## Lemma

Given  $(E, \|\cdot\|)$ , if  $Z \in \mathbb{E}[Z] \pm \mathcal{E}_2(\sigma)$ :

$$\mathbb{E}[\|Z\|] \leq \|\mathbb{E}[Z]\| + O(\sigma \sqrt{\eta_{\|\cdot\|}})$$

- ▶  $\eta(\mathbb{R}^p, \|\cdot\|_\infty) = \log(p)$
- ▶  $\eta(\mathbb{R}^p, \|\cdot\|_2) = p$
- ▶  $\eta(\mathcal{M}_{p,n}, \|\cdot\|) = n + p$
- ▶  $\eta(\mathcal{M}_{p,n}, \|\cdot\|_F) = np.$

## Example $Z \in \mathbb{R}^p$ , $X \in \mathcal{M}_{p,n}$

- ▶ if  $Z \in \tilde{Z} \pm \mathcal{E}_2$  :  $\mathbb{E} \|Z\|_\infty \leq \|\tilde{Z}\| + C\sqrt{\log p}$
- ▶ if  $Z \in \tilde{Z} \pm \mathcal{E}_2$  :  $\mathbb{E} \|Z\| \leq \|\tilde{Z}\| + C\sqrt{p}$
- ▶ if  $X \in \tilde{X} \pm \mathcal{E}_2$  :  $\mathbb{E} \|X\| \leq \|\tilde{X}\| + C\sqrt{p+n},$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors**
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Concentration of the sum and the product

## Proposition

If  $(X, Y) \in \mathcal{E}_2(\sigma)$  :  $X + Y \propto \mathcal{E}_q(\sigma)$

If  $\|\mathbb{E}[X]\|', \|\mathbb{E}[Y]\|' \leq \sigma \sqrt{\eta_{\|\cdot\|'}}$  where  $\forall x, y \in E$   $\|xy\| \leq \|x\|'\|y\|$ :

$$XY \propto \mathcal{E}_2\left(\sigma^2 \sqrt{\eta_{\|\cdot\|'}}\right) + \mathcal{E}_1(\sigma^2) \text{ in } (E, \|\cdot\|)^2$$

**Principal idea:**  $\|XY\| \leq \begin{cases} \|X\|\|Y\|' \\ \|X\|'\|Y\| \end{cases}$

## Example

$X \in \mathcal{M}_{p,n}$ ,  $Z \in \mathbb{R}^p$ ,  $Z, X \in \mathcal{E}_2$ ,  $\|\mathbb{E}[X]\| \leq O(1)$ ,  $\|\mathbb{E}[Z]\|_\infty \leq O(1)$ :

- ▶  $\frac{XX^T}{n} \in \mathcal{E}_2\left(\frac{\sqrt{p+n}}{n}\right) + \mathcal{E}_1\left(\frac{1}{n}\right)$  in  $(\mathcal{M}_{p,n}, \|\cdot\|_F)$
- ▶  $Z \odot Z \in \mathcal{E}_2(\sqrt{\log p}) + \mathcal{E}_1$  in  $(\mathbb{R}^p, \|\cdot\|)$

---

$^2 \iff \exists C, c > 0, \forall p, n, \forall f : E \rightarrow \mathbb{R}$ , 1-Lipschitz,  $\forall t > 0$ :

$$\mathbb{P}(|f(XY) - \mathbb{E}[f(XY)]| \geq t) \leq Ce^{-c(t/\sigma^2)^2/\eta_{\|\cdot\|'}} + Ce^{-ct/\sigma^2}$$

# Practical example: Hanson-Wright Theorem

## Theorem

Given random  $X, Y \in \mathbb{R}^p$ , and  $A \in \mathcal{M}_p$  deterministic, if  $(X, Y) \propto \mathcal{E}_2$  and  $\|\mathbb{E}[X]\|, \|\mathbb{E}[Y]\| \leq O(1)$ :

$$X^T A Y \propto \mathcal{E}_2(\sqrt{\log p} \|A\|_F) + \mathcal{E}_1(\|A\|_F)$$

## Proof:

- ▶ Decompose  $A = P \Lambda Q$ ,  $P, Q \in \mathcal{O}_p$ ,  $\Lambda \in \mathcal{D}_n$
- ▶ Note  $\check{X} \equiv PX$ ,  $\check{Y} \equiv QY$ ,  $\check{X}, \check{Y} \propto \mathcal{E}_2$
- ▶  $X^T A Y = \check{X}^T \Lambda \check{Y} = \lambda^T (\check{X} \odot \check{Y})$  where  $\Lambda = \text{Diag}(\lambda)$
- ▶  $\mathbb{E}[\|\check{X}\|_\infty] \leq \|\mathbb{E}[\check{X}]\|_\infty + O(\sqrt{\log p}) \leq \|\mathbb{E}[X]\| + O(\sqrt{\log p}) \leq O(\sqrt{\log p})$
- ▶  $\check{X} \odot \check{Y} \propto \mathcal{E}_2(\sqrt{\log n}) + \mathcal{E}_1$
- ▶  $\lambda^T (\check{X} \odot \check{Y}) \propto \mathcal{E}_2(\|\lambda\| \sqrt{\log n}) + \mathcal{E}_1(\|\lambda\|)$

# Hanson Wright Theorem

## Classical Theorem

If  $Z_1, \dots, Z_p \in C\mathcal{E}_2(\sigma)$  independent:

$$\mathbb{P} \left( |Z^T A Z - \mathbb{E} Z^T A Z| \geq t \right) \leq C \exp \left( -c \min \left( \frac{t^2}{\sigma^4 \|A\|_F^2}, \frac{t}{\sigma^2 \|A\|} \right) \right)$$

## With the Concentration of the measure phenomenon

If  $Z = (Z_1, \dots, Z_p) \in \mathcal{E}_2(\sigma)$ :

$$\begin{aligned} & \mathbb{P} \left( |Z^T A Z - \mathbb{E} Z^T A Z| \geq t \right) \\ & \leq C \exp \left( -c \min \left( \frac{t^2}{\sigma^4 \|A\|_F^2 \log p}, \frac{t}{\sigma^2 \|A\|_F} \right) \right) \end{aligned}$$

→ about the same result

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to “Concentrated equation”

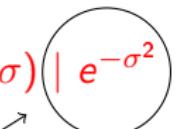
### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

## Real case

- ▶  $\sigma > 0$ , (changes with  $p, n$ ),
- ▶  $X \sim \mathcal{N}(0, \sigma^2)$  (then  $X \propto \mathcal{E}_2(\sigma)$ ),
- ▶  $Y \equiv \frac{1}{1-X}$  (solution to  $Y = 1 + XY$ )
- ▶  $f_Y(y) = \frac{e^{-(1-\frac{1}{y})^2/\sigma^2}}{\sqrt{2\pi}\sigma y^2},$
- ▶ Clearly :  $Y \not\propto \mathcal{E}_2(\sigma')$  because  $f_Y(y) \underset{y \rightarrow \infty}{\sim} \frac{e^{-1/\sigma^2}}{y^2},$
- ▶ Note  $\mathcal{A}_Y \equiv \{X \leq \frac{1}{2}\}$ ,  $\mathbb{P}(\mathcal{A}_Y^c) \leq 2e^{-\sigma^2/8},$
- ▶  $t \mapsto \frac{1}{1-t}$  4-Lipschitz on  $\mathcal{A}_Y,$

$\implies (Y|\mathcal{A}_Y) \propto \mathcal{E}_2(\sigma)$  and we note  $Y \overset{\mathcal{A}_Y}{\propto} \mathcal{E}_2(\sigma) \mid e^{-\sigma^2}$



(because  $\mathbb{P}(\mathcal{A}_Y^c) \leq Ce^{-c\sigma^2}$ )

# Concentration of solution to Concentrated equation

- ▶  $\mathcal{F}(E)$  : set of mapping  $E \rightarrow E$ ,
- ▶  $\|\phi\|_{\mathcal{B}(y_0, K)} = \sup_{\|x - y_0\| \leq K} \|\phi(x)\|$ ,
- ▶  $\|\phi\|_{\mathcal{L}} = \sup_{x, y \in E} \frac{\|\phi(x) - \phi(y)\|}{\|x - y\|}$ .

## Theorem

Given random  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  if  $\exists$  an event  $\mathcal{A}_\phi$  such that:

- ▶  $\mathcal{A}_\phi = \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$  and  $\mathbb{P}(\mathcal{A}_\phi^c) \leq Ce^{-cn}$  (for  $C, c > 0$ )
- ▶  $\exists! y_0 \in \mathbb{R}^n / y_0 = \mathbb{E}_{\mathcal{A}_\phi}[\phi(y_0)]$ .  $\forall K > 0$ , ( $K \leq O(1)$ ):

$$\phi \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right) \mid e^{-n} \quad \text{in } (\mathcal{F}(\mathbb{R}^n), \|\cdot\|_{\mathcal{B}(y_0, K)})$$

Then, under  $\mathcal{A}_\phi$  the equation  $Y = \phi(Y)$  admits a unique solution  $Y \in \mathcal{M}_{p,n}$  that satisfies:

$$Y \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right) \mid e^{-n}$$

# Heuristic of the proof

## Hypotheses

- ▶  $\mathcal{A}_\phi = \{\|\phi\|_{\mathcal{L}} \leq 1 - \varepsilon\}$  and  $\mathbb{P}(\mathcal{A}_\phi^c) \leq Ce^{-cn}$  (for  $C, c > 0$ )
- ▶  $\exists! y_0 \in \mathbb{R}^n \mid y_0 = \mathbb{E}_{\mathcal{A}_\phi}[\phi(y_0)]. \forall K > 0, (K \leq O(1)):$

$$\phi \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right) \mid e^{-n} \quad \text{in } (\mathcal{F}(\mathbb{R}^n), \|\cdot\|_{\mathcal{B}(y_0, K)})$$

## “Proof:”

- ▶  $Y \approx \phi^k(y_0)$  for  $k$  sufficiently big
- ▶ Under  $\mathcal{A}_\phi$ , for  $K \leq O(1)$ , sufficiently big  
 $\forall k \in \mathbb{N}, \phi^k(y_0) \in \mathcal{B}(y_0, K)$
- ▶ Since  $\phi$  concentrated in  $(\mathcal{F}(\mathbb{R}^n), \|\cdot\|_{\mathcal{B}(y_0, K)})$ ,

$$\forall k \in \mathbb{N}, \phi^k \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right) \mid e^{-n}$$

$$\implies \text{for } k \text{ sufficiently big, } Y \approx \phi^k(y_0) \stackrel{\mathcal{A}_\phi}{\propto} \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right) \mid e^{-n}$$

# Table of contents

## Introduction

## I- Concentration of the Measure Phenomenon

## II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Position of the problem

- ▶ Data matrix  $X = (x_1, \dots, x_n) \in \mathcal{M}_{p,n}$
- ▶ labels :  $Y = (y_1, \dots, y_n) \in \mathbb{R}^n$

Robust regression problem with regularizing parameter:

$$(P) : \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho(y_i - x_i^T \beta) + \lambda \|\beta\|^2$$

with  $\rho : \mathbb{R} \rightarrow \mathbb{R}$  convex,  $\lambda > 0$ .

Differentiation:

$$(P) \iff \beta = \frac{1}{n\lambda} \sum_{i=1}^n \rho'(y_i - x_i^T \beta) x_i \iff \beta = \frac{1}{n} X f(X^T \beta)$$

- ▶  $f_i \equiv \frac{1}{\lambda} \rho'(y_i - \cdot)$
- ▶  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n, f((z_i)_{1 \leq i \leq n}) = (f_i(z_i))_{1 \leq i \leq n}$

# Hypotheses

On  $X$ ,  $\forall i \in [n]$ ,  $\mu_i \equiv \mathbb{E}[x_i]$ ,  $\Sigma_i \equiv \mathbb{E}[x_i x_i^T]$ ,  $C_i \equiv \Sigma_i - \mu_i \mu_i^T$ :

- ▶  $p = O(n)$
- ▶  $x_1, \dots, x_n$  independent (with possibly different distributions)
- ▶  $X \propto \mathcal{E}_2$  (as if  $X \sim \mathcal{N}(0, I_{pn})$ )  $\implies \|C_i\| \leq O(1)$
- ▶  $\|\mu_i\| = O(1) \implies \mathbb{E}[\frac{1}{n} \|XX^T\|] \leq O(1)$

On  $f$ :

- ▶  $\|f\|_\infty \leq \infty (\leq O(1))$  (unnecessary)
- ▶  $\|f'\|_\infty, \|f''\|_\infty \leq \infty$

Contractivity of  $\beta = \frac{1}{n} X f(X^T \beta)$

- ▶  $\|f'\|_\infty \mathbb{E}[\|\frac{1}{n} XX^T\|] \leq 1 - 2\varepsilon$  with  $\varepsilon \geq O(1)$

# Goal

“Concentration of  $\beta$  and Estimation of first statistics”

$$\mu_\beta \equiv \mathbb{E}_{\mathcal{A}_\beta}[\beta] \quad C_\beta \equiv \mathbb{E}_{\mathcal{A}_\beta}[\beta\beta^T] - \mu_\beta\mu_\beta^T$$

**First approach:**  $\mu_\beta = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [f(x_i^T \beta) x_i]$

If we admit  $x_i$  behaves like a Gaussian vector,

→ **Issue:** dependence between  $x_i$  and  $\beta$

→ **Solution:** “Leave-one-out”:

- ▶ introduce  $\beta_{-i}$ :

$$\beta_{-i} = \frac{1}{n} \sum_{\substack{1 \leq j \leq n \\ j \neq i}} f(x_j^T \beta_{-i}) x_j$$

- ▶ Construct  $\zeta_i : \mathbb{R} \rightarrow \mathbb{R}$  deterministic |  $x_i^T \beta \approx \zeta_i(x_i^T \beta_{-i}^T)$

# Strategy of the study

1. Introduce event  $\mathcal{A}_\beta \equiv \{\|f'\|_\infty \|\frac{1}{n}XX^T\| \leq 1 - \varepsilon\}$  where  $\beta$  concentrate
2. disentangle  $\beta$  and  $x_i$  :

$$\beta_{-i}(t) = \frac{1}{n}X_{-i}f(X_{-i}^T\beta_{-i}(t)) + \frac{t}{n}f(x_i^T\beta_{-i}(t))x_i$$

where  $X_{-i} = (x_1, \dots, x_{i-1}, \mathbf{0}, x_{i+1}, \dots, x_n)$ .

- ▶ Differentiate  $\beta_{-i}(\cdot)$ ,
- ▶ Integrate approximation of  $\beta'_{-i}(t)$ .

3. Construct deterministic  $\zeta_i : \mathbb{R} \rightarrow \mathbb{R}$  st.  $\beta^T x_i = \zeta_i(\beta_{-i}^T x_i)$
4. Estimate  $\mu_\beta, C_\beta$  with Gaussian Hypotheses on  $x_1, \dots, x_n$ .

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$**
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

High probability of  $\mathcal{A}_\beta \equiv \{\|f'\|_\infty \|\frac{1}{n}XX^T\| \leq 1 - \varepsilon\}$

### Lemma

$$\|\frac{1}{n}XX^T\| \propto \mathcal{E}_2(1/\sqrt{n})$$

Contractivity of  $\beta = \frac{1}{n}Xf(X^T\beta)$

►  $\|f'\|_\infty \mathbb{E}[\|\frac{1}{n}XX^T\|] \leq 1 - 2\varepsilon$  with  $\varepsilon \geq O(1)$

### Lemma

$$\exists C, c > 0, \text{ constant } | \mathbb{P}(\mathcal{A}_\beta^c) \leq Ce^{-cn}$$

**Proof :**  $\mathbb{P}(\mathcal{A}_\beta^c) \leq \mathbb{P}(|\|\frac{1}{n}XX^T\| - \mathbb{E}[\|\frac{1}{n}XX^T\|]| \geq \frac{\varepsilon}{\|f'\|_\infty})$   
 $\leq Ce^{-cn\varepsilon^2/\|f'\|_\infty^2}$

# Concentration of $\beta$

## Lemma

Under  $\mathcal{A}_\beta$ ,  $\|\beta\| \leq O(1)$

**Proof:**  $\|\beta\| = \|\frac{1}{n}Xf(X^T\beta)\| \leq \frac{\|f\|_\infty}{n}\|X\|\|\mathbb{1}\| \leq O(1).$

Note  $\Psi$  such that  $\beta = \Psi(X)(\beta)$

Hypothesis for concentration of  $\beta$ :

1.  $\mathbb{P}(\mathcal{A}_\beta^c) \leq Ce^{-cn}$  (recall that  $\mathcal{A}_\beta \equiv \{\|\Psi(X)\| \leq 1 - \varepsilon\}$ )
2.  $\forall K > 0$ ,  $K \leq O(1)^3$ :  
 $(\Psi(X) \mid \mathcal{A}_\beta) \propto \mathcal{E}_2(1/\sqrt{n})$  in  $(\mathcal{F}(\mathbb{R}^p), \|\cdot\|_{\mathcal{B}(0,K)})$

---

<sup>3</sup>if  $y_0 = \mathbb{E}_{\mathcal{A}_\beta}[\Psi(X)(y_0)]$ ,  $\|y_0\| \leq O(1)$

# Concentration of $\beta$

## Proposition

$$\beta \mid \mathcal{A}_\beta \propto \mathcal{E}_2(1/\sqrt{n})$$

**Proof:** Recall that  $\Psi(A)(y) = Af(A^T y)$ , ( $\beta = \Psi(X)(\beta)$ )  
 $\Psi : \mathcal{M}_{p,n} \rightarrow (\mathcal{F}(\mathbb{R}^p), \|\cdot\|_{\mathcal{B}(0,K)})$  is  $O(1/\sqrt{n})$ -Lipschitz on  $\mathcal{A}_\beta$ .  
 $\forall \|y\| \leq K$ ,  $A, B \in \mathcal{A}_\beta$  ( $\|A\|, \|B\| \leq O(1)$ ):

$$\begin{aligned}\|\Psi(A)(y) - \Psi(B)(y)\| &\leq \frac{1}{n} \left\| (A - B)f(A^T y) \right\| + \frac{1}{n} \left\| B \left( f(A^T y) - f(B^T y) \right) \right\| \\ &\leq O\left(\frac{1}{\sqrt{n}}\right) \|A - B\|,\end{aligned}$$

$$\implies \Psi(X)(y) \propto \mathcal{E}_2(1/\sqrt{n}) \text{ in } (\mathcal{F}(\mathbb{R}^p), \|\cdot\|_\infty).$$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

## Differentiation of $\beta_{-i}(\cdot)$

- ▶  $\beta_{-i}(t) = \frac{1}{n} X_{-i} f(X_{-i}^T \beta_{-i}(t)) + \frac{t}{n} f(x_i^T \beta_{-i}(t)) x_i$
- ▶  $X_{-i} = (x_1, \dots, x_{i-1}, \mathbf{0}, x_{i+1}, \dots, x_n)$

### Proposition

$\beta_{-i}(\cdot)$  is differentiable and:

$$\beta'_{-i}(t) = \frac{1}{n} Q_{-i}(t) x_i \chi'(t)$$

where:

- ▶  $Q_{-i}(t) = (I_p - \frac{1}{n} X_{-i} D(t) X_{-i}^T)^{-1} \in \mathcal{M}_p$
- ▶  $D(t) = \text{Diag}(f'(x_j^T \beta_{-i}(t)))_{1 \leq j \leq n}$
- ▶  $\chi(t) = t f(x_i^T \beta_{-i}(t))$

→ Show that  $t \mapsto \frac{1}{n} Q_{-i}(\mathbf{t}) x_i$  is constant

# $\frac{1}{n} Q_{-i}(\cdot) x_i$ constant : Preliminary Lemmas

## Lemma

$$\|Q_{-i}(t)\| \leq \frac{1}{\varepsilon}$$

We note  $\beta_{-i} = \beta_{-i}(0)$ ,  $X_{-i} = X_{-i}(0)$  and  $Q_{-i} = Q_{-i}(0)$ .

## Lemma

$$x_i^T \beta_{-i}(t) \propto \mathcal{E}_2(1) \mid e^{-n}$$

$$\|x_i\| \leq O(2\sqrt{n}) \|\beta_{-i}(t)\| \leq O(1)\sqrt{n}$$

## Lemma

$$\frac{1}{\sqrt{n}} X_{-i}^T Q_{-i} x_i \propto \mathcal{E}_2(1) \mid e^{-n} \text{ and } \mathbb{E} \left[ \frac{1}{\sqrt{n}} \|X_{-i}^T Q_{-i} x_i\|_\infty \right] \leq O(1).$$

**Proof:**  $\|\frac{1}{\sqrt{n}} \mathbb{E}[X_{-i}^T Q_{-i} x_i]\|_\infty \leq \|\frac{1}{\sqrt{n}} \mathbb{E}[X_{-i}^T Q_{-i}] \mu_i\| \leq O(1)$

$$\mathbb{E} \left[ \frac{1}{\sqrt{n}} \|X_{-i}^T Q_{-i} x_i\|_\infty \right]$$

$\frac{1}{n} Q_{-i}(\cdot) x_i$  constant

## Proposition

$$\|Q_{-i}(t)x_i - Q_{-i}x_i\| \in O(1) \pm \mathcal{E}_2 \mid e^{-n}.$$

$$\begin{aligned}\textbf{Proof: } \| (Q_{-i}(t) - Q_{-i})x_i \| &\leq \frac{1}{n} \left\| Q_{-i}(t) X_{-i}(D_{-i} - D(t)) X_{-i}^T Q_{-i} x_i \right\| \\ &\leq O\left(\frac{1}{\sqrt{n}}\right) \|X_{-i}^T Q_{-i} x_i\|_\infty \|D_{-i} - D_{-i}(t)\|_F.\end{aligned}$$

Besides,  $D_{-i}(t) = \text{Diag}(f'(\mathbf{X}^T \beta_{-i}(t)))$  and:

$$\mathbf{X}^T \beta_{-i}(t) = \frac{1}{n} \mathbf{X}^T \mathbf{X}_{-i} f(\mathbf{X}^T \beta_{-i}(t)) + \frac{t}{n} \mathbf{X}^T x_i f(x_i^T \beta_{-i}(t)),$$

$$\begin{aligned}\|D_{-i} - D_{-i}(t)\|_F &\leq \|f''\|_\infty \|\mathbf{X}^T \beta_{-i}(t) - \mathbf{X}^T \beta_{-i}(0)\| \\ &\leq \frac{\|f''\|_\infty}{\varepsilon} \frac{t}{n} \left\| f(x_i^T \beta_{-i}(t)) \mathbf{X}^T x_i \right\| \\ &\leq O(\|f\|_\infty)\end{aligned}$$

## Link between $\beta$ and $\beta_{-i}$

- ▶  $\beta'_{-i}(t) = \frac{1}{n} Q_{-i}(t) x_i \chi'(t)$
- ▶  $\chi(t) = t f(x_i^T \beta_{-i}(t))$
- ▶  $\left\| \frac{1}{n} Q_{-i}(\cdot) x_i - \frac{1}{n} Q_{-i} x_i \right\| \in O(1/n) \pm \mathcal{E}_2(1/n) \mid e^{-n}$
- ▶  $\chi'(t) \in O(1) \pm \mathcal{E}_2 \mid e^{-n}$

### Proposition

$$\left\| \beta - \beta_{-i} - \frac{1}{n} f(x_i^T \beta) Q_{-i} x_i \right\| \in 0 \pm \mathcal{E}_2\left(\frac{1}{n}\right) \mid e^{-n}$$

**Proof:**  $\beta_{-i}(1) = \beta$ ,  $\chi(0) = 0$ ,  $\chi(1) = \frac{1}{n} f(x_i^T \beta)$  so:

$$\beta - \beta_{-i} = \frac{1}{n} f(x_i^T \beta) Q_{-i} x_i + \frac{1}{n} \int_0^1 \chi'(u) (Q_{-i}(u) - Q_{-i}(0)) x_i du.$$

$$\implies x_i^\beta \approx x_i^T \beta_{-i} + \frac{1}{n} x_i Q_{-i} x_i f(x_i^T \beta).$$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

# Concentration of $\frac{1}{n}x_i^T Q_{-i} x_i$

Recall that:

- $Q_{-i} \equiv (I_p - \frac{1}{n}X_{-i}^T D(0) X_{-i})^{-1}$
- $D(0) \equiv \text{Diag}(f'(x_j^T \beta_{-i}))_{1 \leq j \leq n}$

## Lemma

$$\frac{1}{n}X_{-i}^T D(t) X_{-i} \asymp \mathcal{E}_2 \mid e^{-n}$$

$$\|X\| \lesssim \sqrt{n} \|D\| \leq \mathcal{O}(1) \|X\| \lesssim \sqrt{n}$$

## Lemma

$$\frac{1}{n}x_i^T Q_{-i} x_i \in \Delta_i \pm \mathcal{E}_2(1/\sqrt{n}) \mid e^{-n} \text{ with } \Delta_i \equiv \frac{1}{n} \text{Tr}(\Sigma_i \mathbb{E}[Q_{-i}])$$

**Proof:**  $\left| \frac{1}{n}x_i^T Q_{-i} x_i - \frac{1}{n} \text{Tr}(\Sigma_i \mathbb{E}[Q_{-i}]) \right|$

$$< \left| \frac{1}{n}x_i^T Q_{-i} x_i - \frac{1}{n} \text{Tr}(\Sigma_i Q_{-i}) \right| + \left| \frac{1}{n} \text{Tr}(\Sigma_i (Q_{-i} - \mathbb{E}[Q_{-i}])) \right|$$

## Deterministic mapping between $\beta$ and $\beta_{-i}$

From  $\|\beta - \beta_{-i} - \frac{1}{n}f(x_i^T \beta)Q_{-i}x_i\| \in 0 \pm \mathcal{E}_2\left(\frac{1}{n}\right) + e^{-n}$ , we deduce:  
►  $\|x_i^T \beta - x_i^T \beta_{-i} - \Delta_i f(x_i^T \beta)\| \in 0 \pm \mathcal{E}_2\left(\frac{1}{n}\right) + e^{-n}$

### Lemma

Given  $i \in [n]$ ,  $\exists! \zeta_i(t) \in \mathbb{R}$  /

$$\zeta_i(t) = t + \Delta_i f(\zeta_i(t))$$

**Proof:**  $\|f'\|_\infty \Delta_i = \mathbb{E}_{\mathcal{A}_Q} \left[ \frac{\|f'\|_\infty}{n} x_i^T Q_{-i} x_i \right]$   
 $\leq \mathbb{E}_{\mathcal{A}_Q} \left[ \frac{\|f'\|_\infty}{n} x_i^T \left( I_n - \frac{\|f\|_\infty}{n} X_{-i} X_{-i}^T \right)^{-1} x_i \right] < 1$

### Proposition

$$x_i^T Y \in \zeta_i(x_i^T Y_{-i}) \pm \mathcal{E}_2\left(\frac{1}{\sqrt{n}}\right) + e^{-n}$$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$**
- F - Application

## Integration on $x_i$ then on $\beta_{-i}$

Recall that

- ▶  $\mu_\beta = \frac{1}{n} \sum_{i=1}^n \mathbb{E} [f(x_i^T \beta) x_i]$ ,
- ▶ noting  $\xi_i = f \circ \zeta_i, \forall u \in \mathbb{R}^p, \|u\| \leq 1$ :

$$\left| \mathbb{E} [f(x_i^T \beta) u^T x_i] - \mathbb{E} [\xi_i(x_i^T \beta_{-i}) u^T x_i] \right| \leq O \left( \frac{1}{\sqrt{n}} \right)$$

### Assumption

$$x_i \sim \mathcal{N}(\mu_i, C_i)$$

(i) Stein formula, (ii) Concentration of  $\beta_{-i}$

$$\begin{aligned} \mathbb{E}_{-i, x_i} [\xi_i(x_i^T \beta_{-i}) u^T x_i] &\stackrel{(i)}{=} \mathbb{E}_{-i, z} [\xi_i(z_{-i})] u^T \mu_i + \mathbb{E}_{-i} [\mathbb{E}_z [\xi'_i(z_{-i})] u^T C_i \beta_{-i}] \\ &\stackrel{(ii)}{=} \mathbb{E} [\xi_i(z)] u^T \mu_i + \mathbb{E}_z [\xi'_i(z)] u^T C_i \mu_\beta + O \left( \frac{1}{\sqrt{n}} \right) \end{aligned}$$

with  $z_{-i} \sim \mathcal{N}(\beta_{-i}^T \mu_i, \beta_{-i}^T C_i \beta_{-i})$ , and  $z \sim \mathcal{N}(\mu_i^T \mu_\beta, \text{Tr}(\Sigma_\beta C_i))$

# Fixed point equation for $\mu_\beta$ and $C_\beta$

Noting:

- ▶  $z \sim \mathcal{N}(\mu_i^T \mu_\beta, \text{Tr}(C_\beta C_i))$
- ▶  $\tilde{\mu} \equiv \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi_i(z)] \mu_i$
- ▶  $\tilde{K} \equiv \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi'_i(z)] C_i$
- ▶  $\tilde{C} \equiv \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi_i^2(z)] C_i$

$$\color{red}\mu_\beta = \tilde{\mu} + \tilde{K} \color{red}\mu_\beta + O_{\|\cdot\|} \left( \frac{1}{\sqrt{n}} \right); \quad \color{red}C_\beta = \tilde{C} + \tilde{K} \color{red}C_\beta \tilde{K} + O_{\|\cdot\|_*} \left( \frac{1}{\sqrt{n}} \right)$$

## Lemma

$$|\Delta_i - \frac{1}{n} \text{Tr}(\Sigma_i(1 - \tilde{K})^{-1})| \leq O\left(\frac{1}{\sqrt{n}}\right) \text{ and } \|\tilde{K}\| \leq 1 - \varepsilon.$$

“Proof:”  $\xi'_i(t) = \frac{f'(t + \Delta_i \xi_i(t))}{1 - \Delta_i f'(t + \Delta_i \xi_i(t))}$  and:

$$\begin{aligned} \Delta_i &= \mathbb{E} \left[ \frac{1}{n} \text{Tr} \left( \Sigma_i \left( I_p - \frac{1}{n} X f_d'(X^T \beta) X \right)^{-1} \right) \right] \\ &= \frac{1}{n} \text{Tr} \left( \Sigma_i \left( I_p - \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \frac{f'(x_j^T \beta)}{1 - \Delta_j f'(x_j^T \beta)} \right] C_j \right)^{-1} \right) + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

# Fixed point equation for $\mu_\beta$ , $C_\beta$ , $\Delta$

## Proposition (Unproven)

$\exists! (\Delta, m, \sigma) \in (\mathbb{R}^n)^3$  satisfying:

- ▶  $\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi_i(z_i)] \mu_i$
- ▶  $\tilde{C} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi_i(z_i)^2] C_i$
- ▶  $\tilde{K} = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\xi'_i(z_i)] C_i$
- ▶  $\tilde{Q} = (I_p - \tilde{K})^{-1}$
- ▶  $\tilde{\mathcal{Q}} : \mathcal{M}_p \rightarrow \mathcal{M}_p, \forall M :$   
$$\tilde{\mathcal{Q}}(M) = M + \tilde{K} \tilde{\mathcal{Q}}(M) \tilde{K}$$
- ▶  $m_i = \mu_i^\top \tilde{Q} \tilde{\mu}$
- ▶  $\sigma_i^2 = \frac{1}{n} \text{Tr}(C_i \tilde{\mathcal{Q}}(\tilde{C})) + \tilde{\mu}^\top \tilde{Q} C_i \tilde{Q} \tilde{\mu}.$
- ▶  $z_i \sim \mathcal{N}(m_i, \sigma_i^2)$
- ▶  $\Delta_i = \frac{1}{n} \text{Tr} \left( C_i \left( I_p - \tilde{K} \right)^{-1} \right)$
- ▶  $\xi_i(z) = f(z + \Delta_i \xi_i(z))$

With these definitions,

$$\left\| \mu_\beta - \tilde{Q} \tilde{\mu} \right\| \leq \mathcal{O} \left( n^{-\frac{1}{2}} \right) \quad \left\| C_\beta - \frac{1}{n} \tilde{\mathcal{Q}}(\tilde{C}) \right\|_* \leq \mathcal{O} \left( n^{-\frac{1}{2}} \right),$$

# Sommaire

## Introduction

### I- Concentration of the Measure Phenomenon

- A - Description of the phenomenon
- B - Concentration of random variables
- C - Concentration of the norm of a random vector
- D - Concentration of the sum and the product of random vectors
- E - Concentration of fixed point solution to "Concentrated equation"

### II - Performances of the robust regression

- A - Position of the problem
- B - Concentration of  $\beta$
- C - Leave-one-out
- D - From an approximation to a deterministic fixed point equation
- E - Estimation of  $\mu_\beta$
- F - Application

## Softmax classification

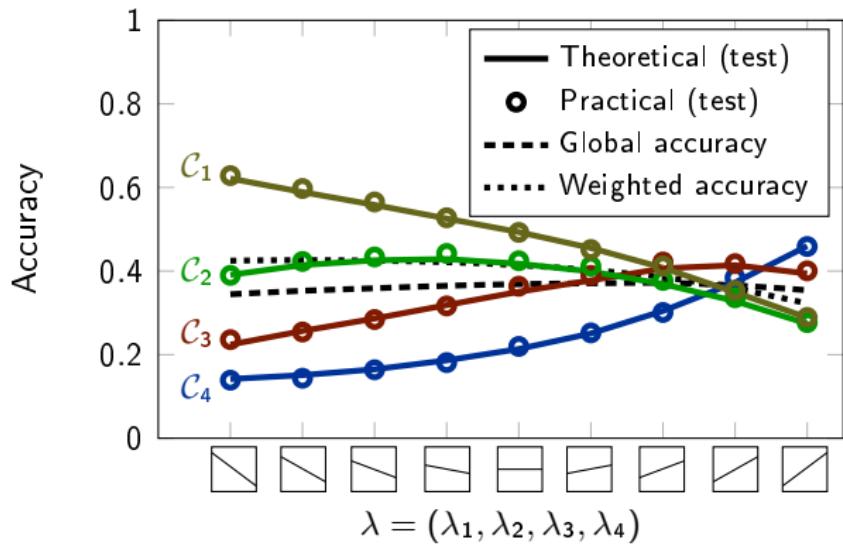
- ▶  $(x_i)_{1 \leq i \leq n}$  belong to  $k$  possible classes  $\mathcal{C}_1, \dots, \mathcal{C}_k$ ,
- ▶ Labels  $y_1, \dots, y_n \in \mathbb{R}^k$ , if  $x_i \in \mathcal{C}_\ell$ ,  $y_i = e_\ell$
- ▶ Knowing  $(x_1, y_1), \dots, (x_n, y_n)$ :  
Learning Procedure = Attribute a weight  $w_\ell$  to each class  $\mathcal{C}_\ell$ ,
- ▶ Given  $x \in \mathbb{R}^P$ , score to be in  $\mathcal{C}_\ell$  :  $p_\ell(x) = \frac{\exp(w_\ell^T x)}{\sum_{j=1}^k \exp(w_j^T x)}$
- ▶ Choose the weights  $w_1, \dots, w_k \in \mathbb{R}^P$  that minimize:

$$\begin{aligned}\mathcal{L}(w_1, \dots, w_k) &= -\frac{1}{n} \sum_{i=1}^n \sum_{\ell=1}^k y_{i,\ell} \log(p_\ell(x_i)) + \sum_{\ell=1}^k \lambda_\ell \|w_\ell\|^2 \\ &= \frac{1}{n} \sum_{i=1}^n y_i^T \log(\text{Softmax}(W^T x_i)) + \|W\Lambda\|_F^2\end{aligned}$$

⇒ If  $\lambda$  is big enough, the weights concentrate and we can estimate their statistics.

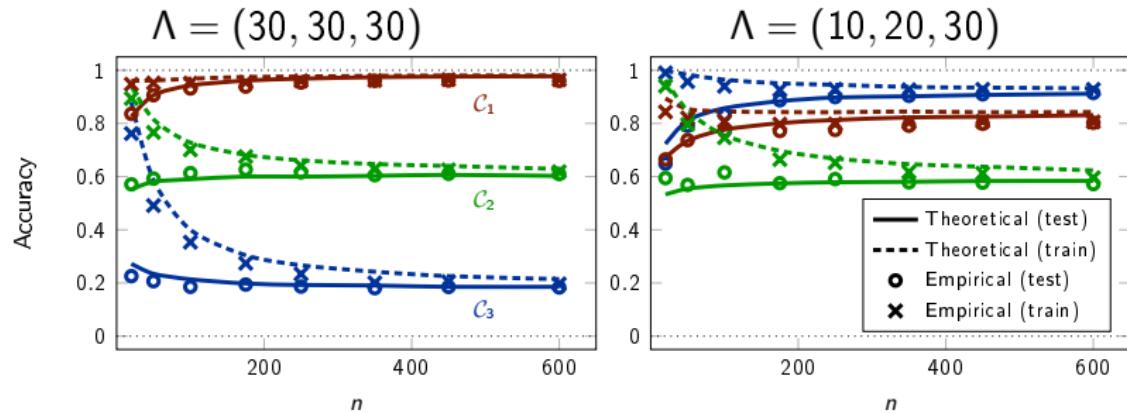
# Prediction of performances on Gaussian data

With Gaussian data,  $n = p = 200$ ,  
4 classes  $\#\mathcal{C}_1 > \#\mathcal{C}_2 > \#\mathcal{C}_3 > \#\mathcal{C}_4$



# Prediction with GAN-generated MNIST data

With GAN-generated data,  $p = 784$ , 3 classes  $\#\mathcal{C}_1 > \#\mathcal{C}_2 > \#\mathcal{C}_3$ .



THANK YOU!!

# Integration on $\beta$

## Lemma

For any  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  such that  $\|\phi'\|_\infty \leq O(1)$ :

$$\mathbb{E}_{\beta, z} \left[ \phi \left( \mu_i^T \beta + \sqrt{\beta^T C_i \beta} z \right) \right] = \mathbb{E}_z \left[ \phi \left( \mu_i^T \mu_\beta + \sqrt{\text{Tr}(\Sigma_\beta C_i)} z \right) \right] + O \left( \frac{1}{\sqrt{n}} \right)$$

where  $z \sim \mathcal{N}(0, 1)$  independent with  $\beta$  and  $\Sigma_\beta = \mu_\beta \mu_\beta^T + C_\beta$

**Proof:**  $\mathbb{E}_z \left[ \phi \left( \mu_i^T \beta + \sqrt{\beta^T C_i \beta} z \right) \right] = \psi(\mu_i^T \beta, \beta^T C_i \beta)$  where  
 $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$   $O(1)$ -Lipschitz, thus:

$$\mathbb{E}_z \left[ \phi \left( \mu_i^T \beta + \sqrt{\beta^T C_i \beta} z \right) \right] \in \psi \left( \mathbb{E}_\beta[\mu_i^T \beta], \mathbb{E}_\beta[\beta^T C_i \beta] \right) \pm \mathcal{E}_2 \left( \frac{1}{\sqrt{n}} \right)$$

## Control of the norm

- ▶ Infinite norm ( $Z \in \mathbb{R}^p$ ,  $Z \propto \mathcal{E}_2(\sigma)$ ) :

$$\begin{aligned}\mathbb{P}\left(\|Z - \tilde{Z}\|_\infty \geq t\right) &= \mathbb{P}\left(\sup_{1 \leq i \leq p} e_i^T (Z - \tilde{Z}) \geq t\right) \\ &\leq p \sup_{1 \leq i \leq p} \mathbb{P}\left(e_i^T (Z - \tilde{Z}) \geq t\right) \\ &\leq p C e^{-(t/\sigma)^q} \leq C' e^{-(t/\sigma \sqrt{\log(p)})^q},\end{aligned}$$

- ▶ For the general case, use of “ $\varepsilon$ -nets”.

If  $\exists H \subset (E^*, \|\cdot\|_*)$  |  $\forall z \in E : \|z\| = \sup_{f \in H} f(z)$ .<sup>4</sup>

$$Z \in \tilde{Z} \pm C \mathcal{E}_2(\sigma) \implies \|Z - \tilde{Z}\| \in 0 \pm \mathcal{E}_2(\sigma \sqrt{\dim(\text{Vect}(H))})$$

---

<sup>4</sup>on  $(\mathbb{R}^p, \|\cdot\|)$ ,  $H = \mathbb{R}^p$ , and  $\dim(\text{Vect}(H)) = p$

## Norm degree

### Degree of a subset $H \subset E^*$ and of a norm

- ▶  $\eta_H = \log(\#H)$  if  $H$  is finite
- ▶  $\eta_H = \dim(\text{Vect}(H))$  if  $H$  is infinite

### Degree of a norm

- ▶  $\eta_{\|\cdot\|} = \inf \left\{ \eta_H, H \subset E^* \mid \forall x \in E, \|x\| = \sup_{f \in H} f(x) \right\}$

### Example

- ▶  $\eta(\mathbb{R}^p, \|\cdot\|_\infty) = \log(p)$
- ▶  $\eta(\mathcal{M}_{p,n}, \|\cdot\|) = n + p$
- ▶  $\eta(\mathbb{R}^p, \|\cdot\|_r) = p$  for  $r \geq 1$
- ▶  $\eta(\mathcal{M}_{p,n}, \|\cdot\|_F) = np.$

## Concentration of the norm

If  $Z \in \tilde{Z} \pm C\mathcal{E}_2(\sigma)$ :

$$\|Z - \tilde{Z}\| \in 0 \pm C'\mathcal{E}_2(c'\sigma\eta_{\|\cdot\|}^{1/q}) \quad \text{and} \quad \mathbb{E} \|Z - \tilde{Z}\| \leq C'\sigma\eta_{\|\cdot\|}^{1/q}$$

Example  $Z \in \mathbb{R}^p$ ,  $X \in \mathcal{M}_{p,n}$

- ▶ if  $Z \in \tilde{Z} \pm 2\mathcal{E}_2(\sqrt{2})$  :  $\mathbb{E} \|Z\| \leq \|\tilde{Z}\| + C\sqrt{p}$
- ▶ if  $X \in \tilde{X} \pm 2\mathcal{E}_2(\sqrt{2})$  :  $\mathbb{E} \|X\| \leq \|\tilde{X}\| + C\sqrt{p+n}$ ,
- ▶ if  $X \in \tilde{X} \pm 2\mathcal{E}_2(\sqrt{2})$  :  $\mathbb{E} \|X\|_F \leq \|\tilde{X}\|_F + C\sqrt{pn}$ .

