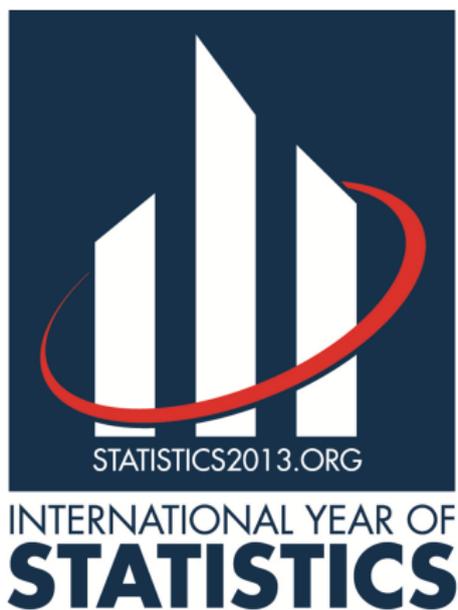


Profondeur statistique et quantile

Davy Paindaveine

Université libre de Bruxelles



Welcome to the International Year of Statistics

2013 is the International Year of Statistics. More than 2,000 organizations—professional statistical societies, colleges and universities, primary and secondary schools, businesses, government entities, and research institutes—are participating in this worldwide event. Why? Because statistics have powerful and far-reaching effects on everyone.

How Does Statistics Affect You?

You may not be aware of it, but statistics affects nearly every aspect of your life, including:

- Foods you eat
- Weather forecasts
- Emergency preparedness
- Assessing disease risks
- Protecting your pet's health
- Improving your health care
- Transportation systems you use
- Assessing your credit worthiness
- Pricing your insurance policies
- Ensuring national security
- Examining economic health
- Prosecuting criminals
- Ensuring the safety of medicine
- Rulemaking by governments
- Assessing teacher effectiveness
- Monitoring climate change

Statistics—An Excellent Career Choice

Since our world is becoming more quantitative and data-focused, job opportunities in statistics are plentiful and projected to increase worldwide. Many industries depend on statisticians to analyze data, which helps in making critical decisions. Statisticians work on important and challenging problems such as:

- Estimating the safety of nuclear power plants and alternative energy sources
- Evaluating the impact of air, water, and soil pollution
- Estimating the unemployment rate of a country
- Analyzing consumer demand for products and services
- Designing studies for and analyzing data from agricultural experiments to increase crop productivity and yields

- **September 2-3, 2013** — [2013 World Statistics Congress Statistics2013 Satellite Meeting](#), Standing Committee on Regional and Urban Statistics, Guangzhou, China
- **September 2-4, 2013** — [Asymptotic Statistics and Related Topics: Theories and Methodologies](#), Bernoulli Society Satellite Meeting to the ISI World Statistics Congress 2013, Tokyo, Japan
- **September 2-5, 2013** — [RSS 2013 International Conference](#), Royal Statistical Society, Newcastle, United Kingdom
- **September 2-6, 2013** — [15th Conference of the International Association for Mathematical Geosciences](#), Madrid, Spain
- **September 2-October 12, 2013** — [XI Concurs Student d'Estadística Aplicada](#); Almíral, Idescat, SCE; Cerdanyola, Spain
- **September 8-11, 2013** — [Workshop EVT2013-in honour of Ivette Gomes](#), Vimeiro, Portugal
- **September 9-11, 2013** — [S.Co.2013](#), Politecnico di Milano, Milano, Italy
- **September 9-12, 2013** — [Seminar of the Austro-Swiss Region of the International Biometric Society](#), Dornbirn, Austria
- **September 10-11, 2013** — [National Young Statisticians Symposium](#), Hacettepe University Department of Statistics, Ankara, Turkey
- **September 12, 2013** — [Statistics: The New Sexy?](#), University of Toronto Department of Statistical Sciences International Year of Statistics Public Lecture Series, Toronto, Canada
- **September 12, 2013** — [BarCamp S.Co.2013](#), Politecnico di Milano, Milano, Italy
- **September 12, 2013** — [China Statistics Open Day](#), National Bureau of Statistics of China, Beijing, China
- **September 12-13, 2013** — [Researchers and Statisticians Conference](#), Hacettepe University Department of Statistics, Ankara, Turkey
- **September 12-13, 2013** — [FernStat 2013: International Conference on Applied Statistics](#); Slovak Statistical and Demographical Society and Faculty of Economics, Matej Bel University Banská Bystrica; Šachtičky near Banská Bystrica, Slovakia
- **September 16-17, 2013** — [7th International Meeting Statistical Methods in Biopharmacy](#), "Emerging Topics for Statistical Methodology in Clinical Drug Development", French Society of Statistics, Paris, France
- **September 16-18, 2013** — [ENBIS Annual Conference](#), Ankara, Turkey
- **September 16-19, 2013** — [The International Conference on Applied Statistics 2013](#), Universitas Padjadjaran, Bandung, Indonesia
- **September 17-19, 2013** — [Structural Inference in Statistics](#), German Research Foundation (DFG) Research Unit 1735 and University of Potsdam, Universität Potsdam, Potsdam, Germany
- **September 17-21, 2013** — [Functional and Complex Structure Data Analysis](#), 14th Course in the ECAS Programme, ECAS 2013, Castro Urdiales, Spain
- **September 18-20, 2013** — [ENBIS – 13](#), the 13th Annual Conference of the European Network for Business and Industrial Statistics, Ankara, Turkey
- **September 18-20, 2013** — [9th Scientific Meeting of the Classification and Data Analysis Group](#) of the Italian Statistical Society, Italian Statistical Society, Modena, Italy
- **September 19-20, 2013** — [Statistical Mysteries](#), University of Wrocław Institute of Mathematics, Wrocław, Poland
- **September 19-21, 2013** — [Workshop on Multivariate Analysis and Random Matrices: New Tendencies](#), CIMAT, Guanajuato, Mexico
- **September 19-21, 2013** — [International Days of Statistics and Economics](#), Department of Statistics and Probability, Department of Microeconomics, and Faculty of Business Economics at the University of Economics; Prague, Czech Republic
- **September 19-21, 2013** — [y-BIS 2013: Joint Meeting of Young Business and Industrial Statisticians](#), ISBIS and ENBIS, Istanbul, Turkey
- **September 20, 2013** — [Statistics Data at MUSC](#), Department of Public Health Sciences at the Medical University of South Carolina (MUSC), Charleston, South Carolina, USA
- **September 21, 2013** — [New England Symposium on Statistics in Sports](#), American Statistical Association, Cambridge, Massachusetts, USA
- **September 22-25, 2013** — [10th Applied Statistics 2013 International Conference](#), Ribno(Bled), Slovenia
- **September 23, 2013** — [Statistical Lie—Truths & Myths](#), University of Wrocław Institute of Mathematics, Wrocław, Poland

1 Quantiles

2 Profondeur statistique

1 Quantiles

- Moyenne, médiane, quantiles
- Extension à la régression
- Extension à la position multivariée

Soit X une variable aléatoire, à valeurs dans \mathbb{R} .

Sa distribution est décrite par

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto F(x) = P[X \leq x], \end{aligned}$$

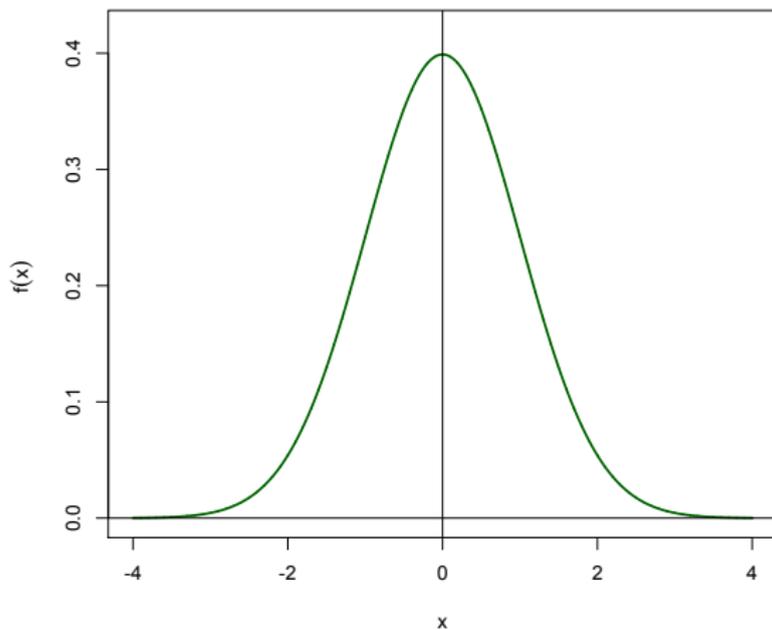
ou (si P ne charge aucun ensemble de mesure Lebesgue nulle) par

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R}^+ \\ x &\mapsto f(x), \end{aligned}$$

qui est telle que $P[X \in B] = \int_B f(x) dx$.

Clairement, $\int_{-\infty}^{\infty} f(x) dx = 1$ et $f \geq 0$.

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2).$$



Soit X une variable aléatoire, à valeurs dans \mathbb{R} .

Sa distribution est décrite par

$$\begin{aligned} F : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto F(x) = P[X \leq x], \end{aligned}$$

ou (si P ne charge aucun ensemble de mesure Lebesgue nulle) par

$$\begin{aligned} f : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto f(x), \end{aligned}$$

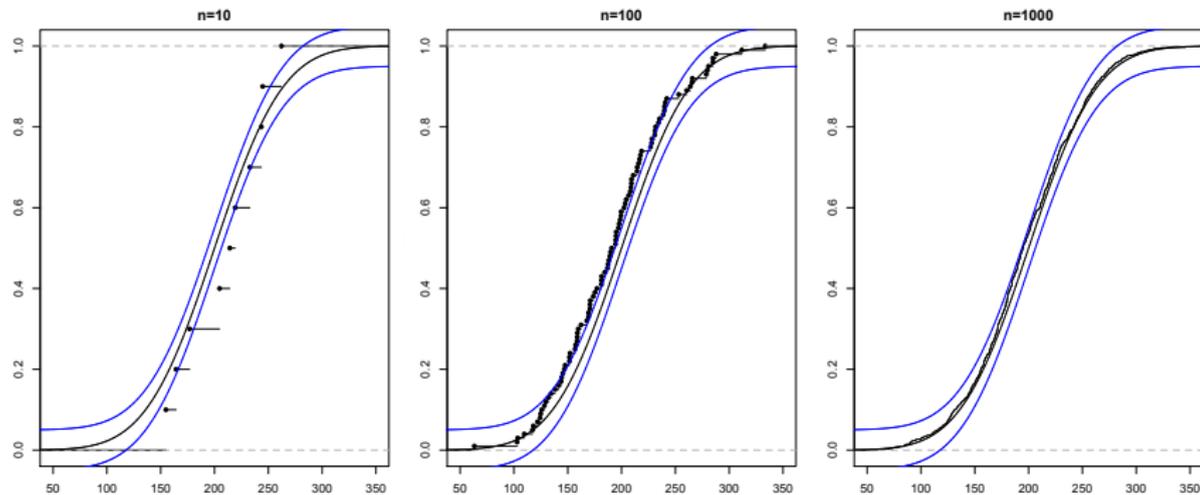
qui est telle que $P[X \in B] = \int_B f(x) dx$.

Si on observe X_1, \dots, X_n indépendantes et de loi F, f , on peut considérer

$$\begin{aligned} \hat{F}_n : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto \hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}[X_i \leq x], \end{aligned}$$

où $\mathbb{I}[A]$ est la fonction indicatrice de A .

X_1, \dots, X_n indépendants $\mathcal{N}(200, 50)$.



On s'intéresse souvent à des **fonctionnelles** du type

$$\begin{aligned} T : \mathcal{F} &\rightarrow \mathbb{R} \\ F &\mapsto T(F), \end{aligned}$$

qui vont chercher, dans F , la caractéristique de F à laquelle on s'intéresse.

Exemples :

Il est bien connu que la fonctionnelle "moyenne"

$$T_E(F) = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx = E[X]$$

est une mesure de **position**, et que la fonctionnelle "variance"

$$T_{\text{Var}}(F) = \int_{-\infty}^{\infty} (x - T_E(F))^2 dF(x) = \int_{-\infty}^{\infty} (x - T_E(F))^2 f(x) dx = \text{Var}[X]$$

est une mesure de **dispersion**. D'autres types de fonctionnelles mesurent l'asymétrie, le "poids des queues", ...

Un estimateur de $T(F)$, fondé sur nos observations X_1, \dots, X_n , peut naturellement être obtenu en considérant $T(\hat{F}_n)$.

Exemples :

La "moyenne empirique"

$$T_E(\hat{F}_n) = \int_{-\infty}^{\infty} x d\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n X_i \stackrel{\text{not}}{=} \bar{X}$$

et la "variance empirique"

$$T_{\text{Var}}(\hat{F}_n) = \int_{-\infty}^{\infty} (x - T_E(\hat{F}_n))^2 dF_n(x) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \stackrel{\text{not}}{=} s^2$$

Considérons les **M-fonctionnelles**, qui sont du type $T(F) = \arg \min_{a \in \mathbb{R}} S_a(F)$.

Exemple : $T_{L_2}(F) = \arg \min_{a \in \mathbb{R}} E[(X - a)^2]$

Considérons les **M-fonctionnelles**, qui sont du type $T(F) = \arg \min_{a \in \mathbb{R}} S_a(F)$.

Exemple : $T_{L_2}(F) = \arg \min_{a \in \mathbb{R}} E[(X - a)^2]$, pour laquelle on a

$$\begin{aligned} E[(X - a)^2] &= \int_{-\infty}^{\infty} (x - a)^2 dF(x) \\ &= \int_{-\infty}^{\infty} (x - a)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (x^2 + a^2 - 2ax) f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx + a^2 \int_{-\infty}^{\infty} f(x) dx - 2a \int_{-\infty}^{\infty} x f(x) dx \\ &= \text{Constante} + a^2 - 2aE[X] \quad \rightsquigarrow \text{dérivée : } 2(a - E[X]). \end{aligned}$$

Considérons les **M-fonctionnelles**, qui sont du type $T(F) = \arg \min_{a \in \mathbb{R}} S_a(F)$.

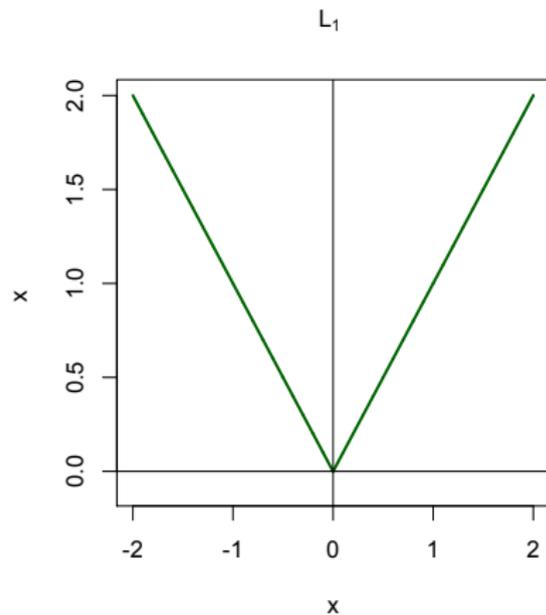
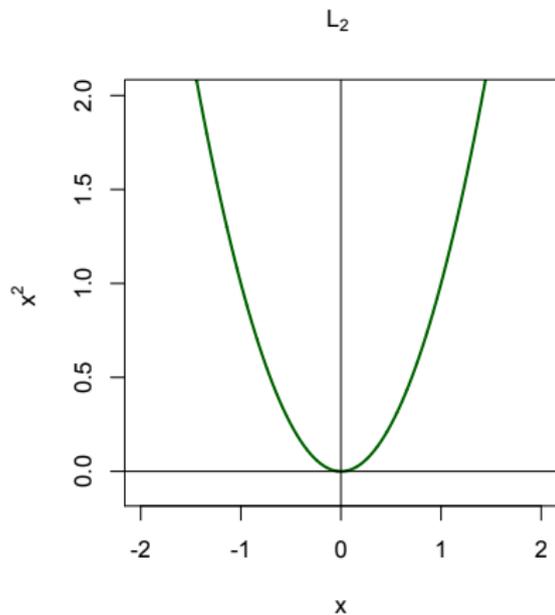
Exemple : $T_{L_2}(F) = \arg \min_{a \in \mathbb{R}} E[(X - a)^2]$, pour laquelle on a

$$\begin{aligned} E[(X - a)^2] &= \int_{-\infty}^{\infty} (x - a)^2 dF(x) \\ &= \int_{-\infty}^{\infty} (x - a)^2 f(x) dx \\ &= \int_{-\infty}^{\infty} (x^2 + a^2 - 2ax) f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx + a^2 \int_{-\infty}^{\infty} f(x) dx - 2a \int_{-\infty}^{\infty} x f(x) dx \\ &= \text{Constante} + a^2 - 2aE[X] \quad \rightsquigarrow \text{dérivée : } 2(a - E[X]). \end{aligned}$$

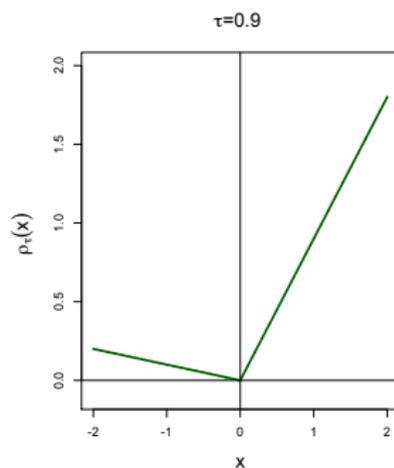
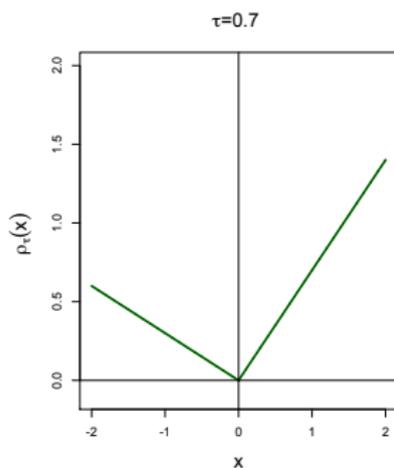
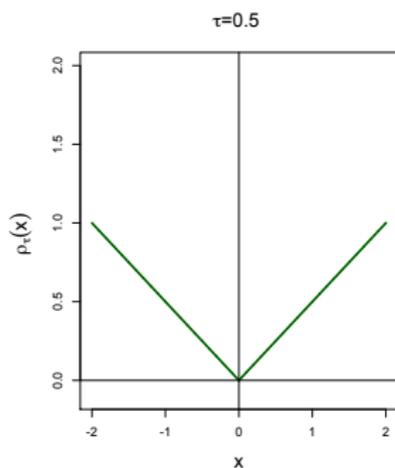
Théorème

$$T_{L_2}(F) = E[X].$$

Autre exemple : $T_{L_1}(F) = \arg \min_{a \in \mathbb{R}} E[|X - a|]$, où



Autre exemple : $T_{L_1, \tau}(F) = \arg \min_{a \in \mathbb{R}} E[\rho_\tau(X-a)]$, où $\rho(z) = \begin{cases} (1 - \tau)|z| & \text{si } z < 0 \\ \tau|z| & \text{si } z \geq 0. \end{cases}$



$$\begin{aligned} E[\rho_\tau(X - a)] &= \int_{-\infty}^{\infty} \rho_\tau(x - a) f(x) dx \\ &= \int_{-\infty}^a \rho_\tau(x - a) f(x) dx + \int_a^{\infty} \rho_\tau(x - a) f(x) dx \\ &= (1 - \tau) \int_{-\infty}^a (a - x) f(x) dx + \tau \int_a^{\infty} (x - a) f(x) dx \end{aligned}$$

La dérivée de

$$\begin{aligned} E[\rho_\tau(X - a)] &= \int_{-\infty}^{\infty} \rho_\tau(x - a) f(x) dx \\ &= \int_{-\infty}^a \rho_\tau(x - a) f(x) dx + \int_a^{\infty} \rho_\tau(x - a) f(x) dx \\ &= (1 - \tau) \int_{-\infty}^a (a - x) f(x) dx + \tau \int_a^{\infty} (x - a) f(x) dx \end{aligned}$$

en a vaut

$$\begin{aligned} (1 - \tau) \int_{-\infty}^a (1) \times f(x) dx + 0 + \tau \int_a^{\infty} (-1) \times f(x) dx - 0 \\ = (1 - \tau)P[X \leq a] - \tau P[X \geq a] = (1 - \tau)P[X \leq a] - \tau(1 - P[X \leq a]) \\ = P[X \leq a] - \tau = F(a) - \tau. \end{aligned}$$

La dérivée de

$$\begin{aligned} E[\rho_\tau(X - a)] &= \int_{-\infty}^{\infty} \rho_\tau(x - a) f(x) dx \\ &= \int_{-\infty}^a \rho_\tau(x - a) f(x) dx + \int_a^{\infty} \rho_\tau(x - a) f(x) dx \\ &= (1 - \tau) \int_{-\infty}^a (a - x) f(x) dx + \tau \int_a^{\infty} (x - a) f(x) dx \end{aligned}$$

en a vaut

$$\begin{aligned} (1 - \tau) \int_{-\infty}^a (1) \times f(x) dx + 0 + \tau \int_a^{\infty} (-1) \times f(x) dx - 0 \\ = (1 - \tau)P[X \leq a] - \tau P[X \geq a] = (1 - \tau)P[X \leq a] - \tau(1 - P[X \leq a]) \\ = P[X \leq a] - \tau = F(a) - \tau. \end{aligned}$$

Théorème

$T_{L_1, \tau}(F) = F^{-1}(\tau)$, le quantile d'ordre τ .

Donc, on a vu que

$$\arg \min E[|x - a|^2] = E[X]$$

$$\arg \min E[|x - a|] = \text{Med}[X]$$

De même, en remplaçant F par F_n ,

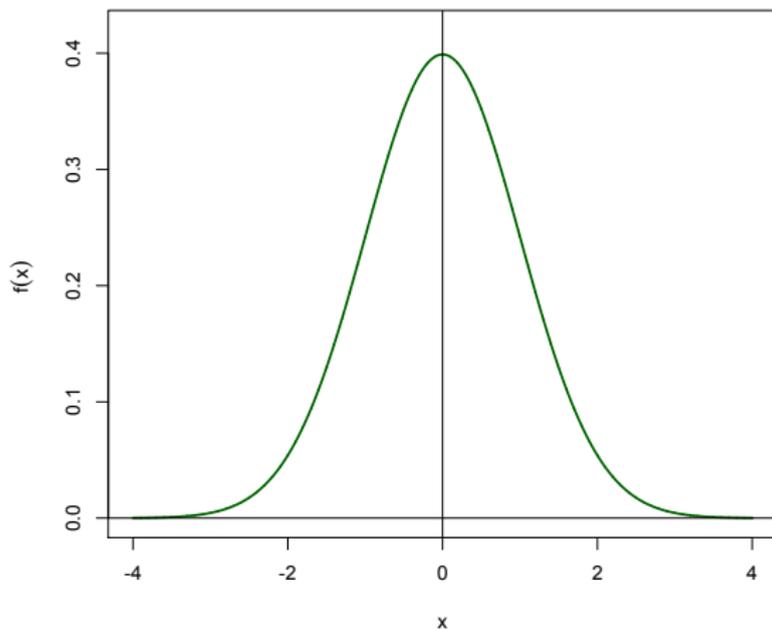
$$\arg \min \frac{1}{n} \sum_{i=1}^n |X_i - a|^2 = \bar{X}$$

$$\arg \min \frac{1}{n} \sum_{i=1}^n |X_i - a| = \text{Med}[X_1, \dots, X_n]$$

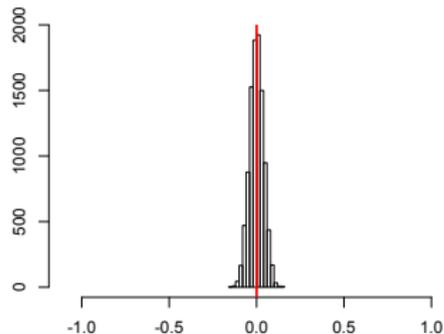
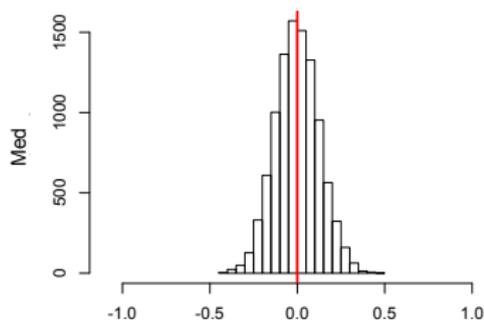
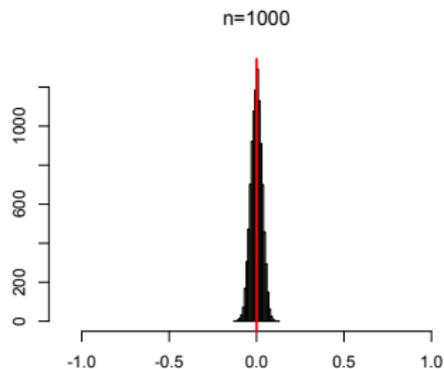
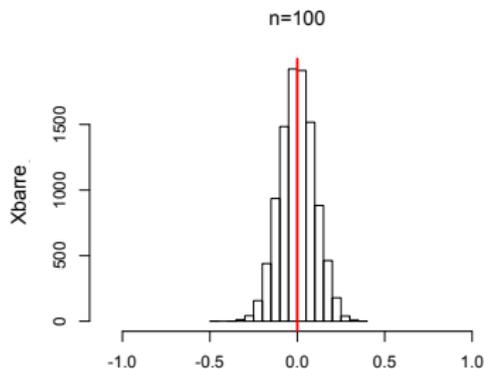
Si f est une fonction paire, $\text{Med}[X] = E[X]$, et $\bar{X} \neq \text{Med}[X_1, \dots, X_n]$ sont deux estimateurs de la même quantité.

Lequel choisir ?

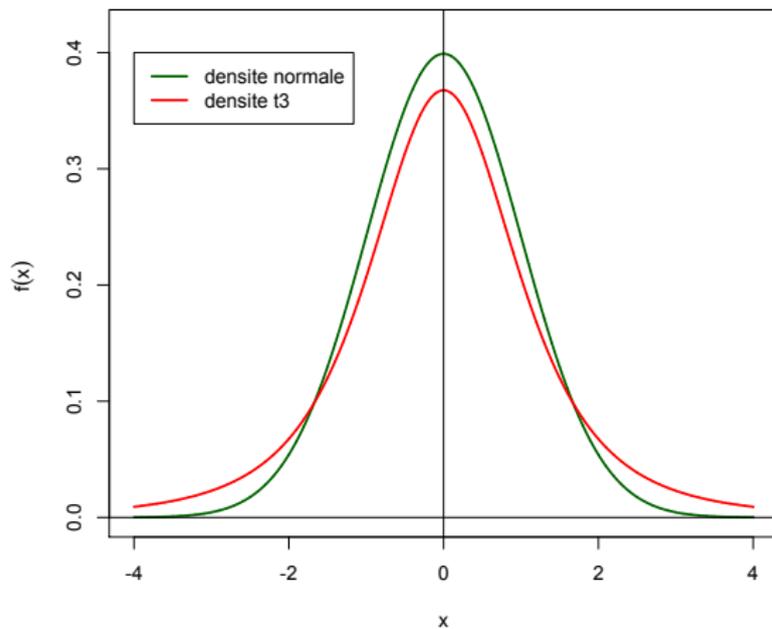
$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$



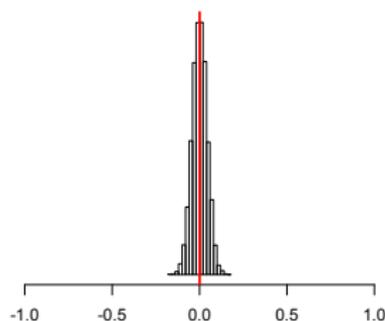
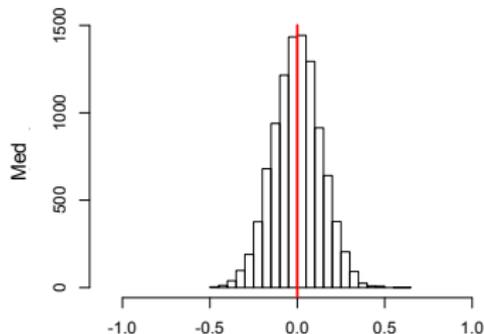
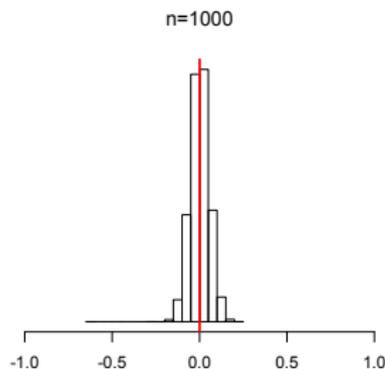
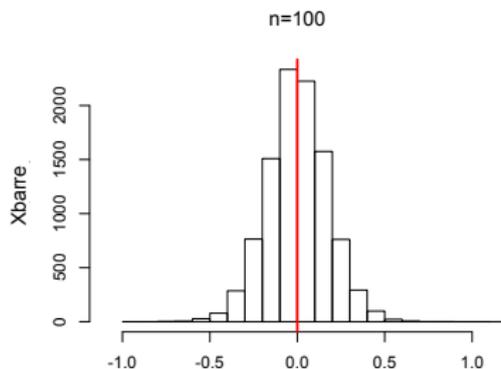
10^4 échantillons (indépendants) de type X_1, \dots, X_n i.i.d. $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$



$$f(x) = \frac{6\sqrt{3}}{\pi} (3 + x^2)^{-2}$$



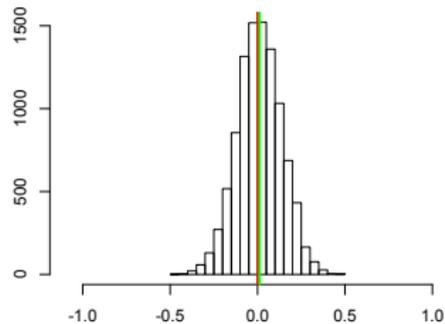
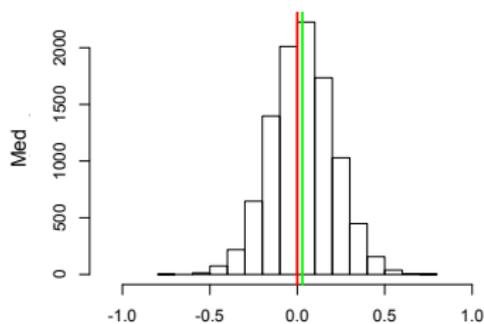
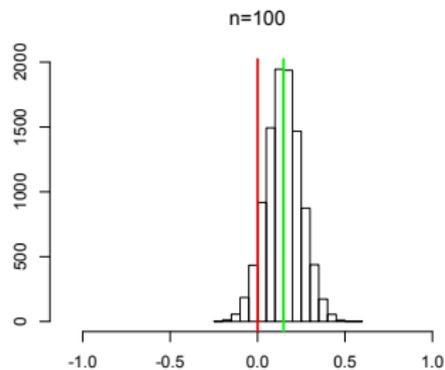
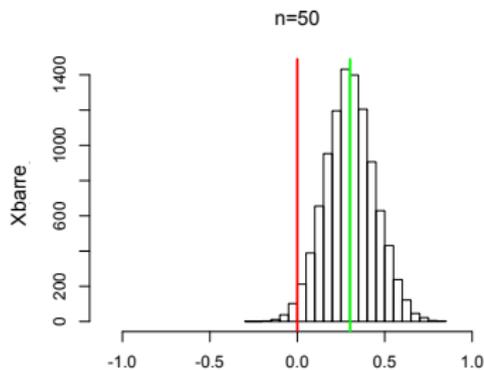
10^4 échantillons (indépendants) de type X_1, \dots, X_n i.i.d. $f(x) = \frac{6\sqrt{3}}{\pi}(3+x^2)^{-2}$



Quand tout va bien, la précision, c'est important...

Mais quand tout ne va pas si bien...

10^4 échantillons (indépendants) de type X_1, \dots, X_{n-1} i.i.d. $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \dots$
Et $X_n = 15$.

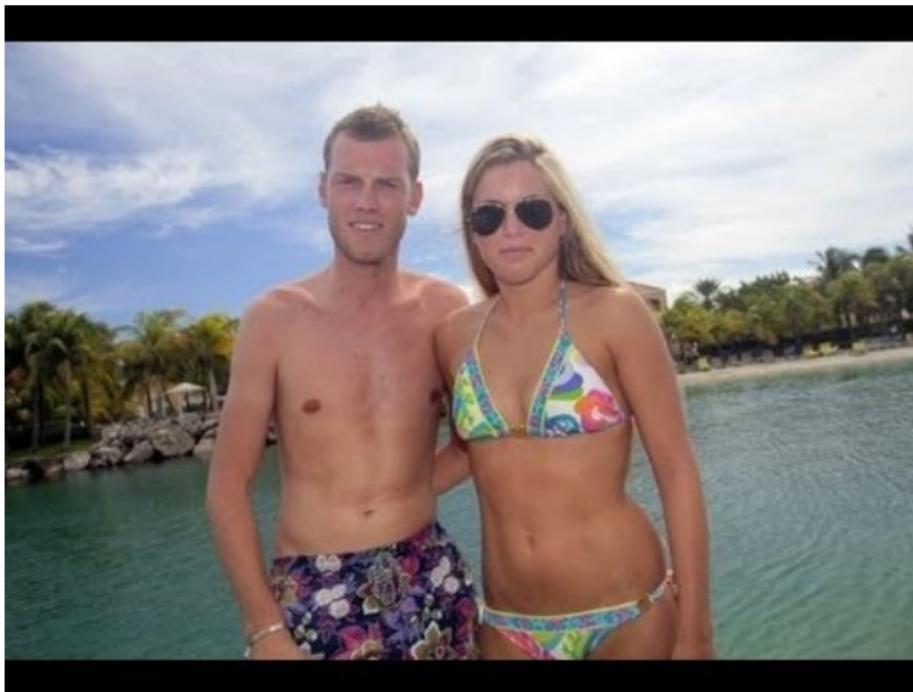


Ceci suggère que la médiane soit une "bonne" fonctionnelle

- comparable à la moyenne en termes d'efficacité
- bien meilleure que la moyenne en termes de robustesse

Les autres quantiles héritent des qualités de robustesse de la médiane (parce que ce sont aussi des quantités de type L_1).

Au delà de ça, **le concept de quantile joue un rôle crucial en statistique inférentielle.**





Question d'inférence statistique :

$$\begin{cases} H_0 : \text{il n'est pas dopé} \\ H_1 : \text{il est dopé} \end{cases}$$

\leadsto on va pour H_1 si et seulement si $x_{\text{Fr}} > F^{-1}(\tau)$
(τ grand choisi par l'organisme de contrôle).

Les quantiles sont aussi importants pour détecter des individus "à risque"...

Exemple : le surpoids chez les enfants.

Mais dans cet exemple, il est naturel d'utiliser une extension du concept de quantile.

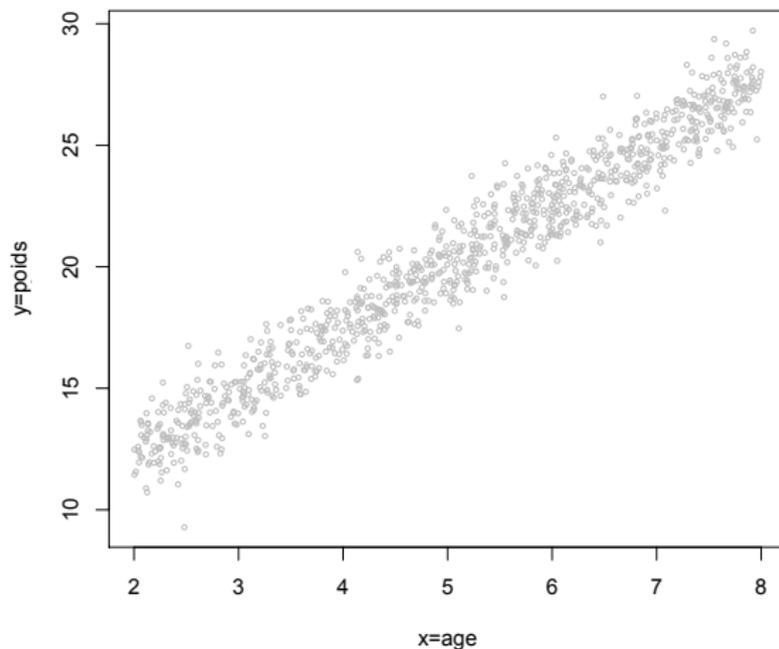
1 Quantiles

- Moyenne, médiane, quantiles
- Extension à la régression
- Extension à la position multivariée

Soit (X, Y) un vecteur aléatoire, où

- Y est la variable d'intérêt (le poids ?)
- X est une variable "explicative" (l'âge ?)

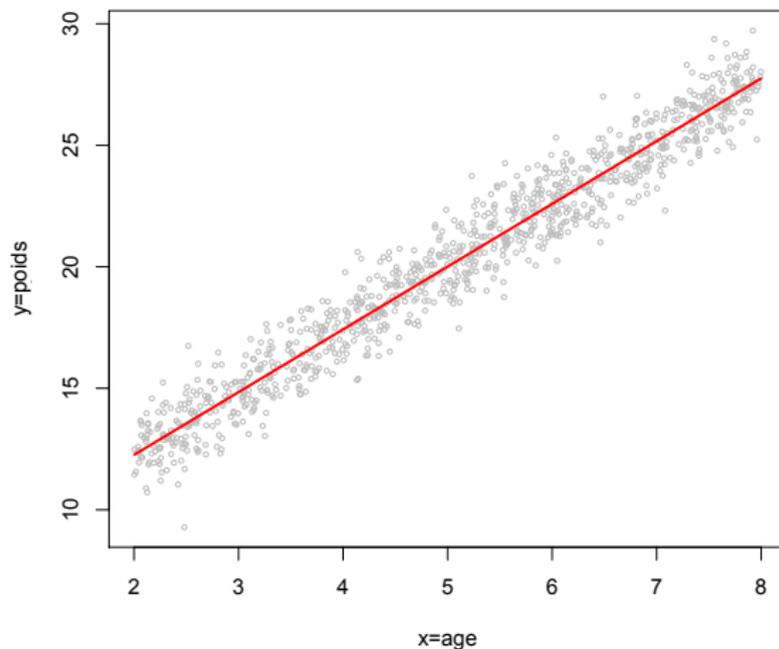
Supposons qu'on dispose d'observations indépendantes (X_i, Y_i) , $i = 1, \dots, n$.



$n = 1000$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$

Le pain quotidien : considérer la droite $y = \hat{a} + \hat{b}x$ des moindres carrés, où

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \frac{1}{n} \sum_{i=1}^n (Y_i - a - bX_i)^2.$$



$n = 1000$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$

Le pain quotidien : considérer la droite $y = \hat{a} + \hat{b}x$ des moindres carrés, où

$$(\hat{a}, \hat{b}) = \arg \min_{(a,b)} \frac{1}{n} \sum_{i=1}^n (Y_i - a - bX_i)^2.$$

La droite estime $y = a_{L_2} + b_{L_2}x$, où

$$(a_{L_2}, b_{L_2}) = \arg \min_{(a,b)} E[(Y - a - bX)^2].$$

Si le modèle sous-jacent est de la forme $Y = a_{L_2} + b_{L_2}X + \varepsilon$ pour une certaine variable ε de moyenne nulle et indépendante de X , on a, comme ci-dessus, que

$$a_{L_2} + b_{L_2}x = E[Y | X = x].$$

On conclut que la droite des moindres carrés estime la droite

$$y = E[Y | X = x].$$

Dans un souci de robustesse, on pourrait plutôt s'intéresser à la droite

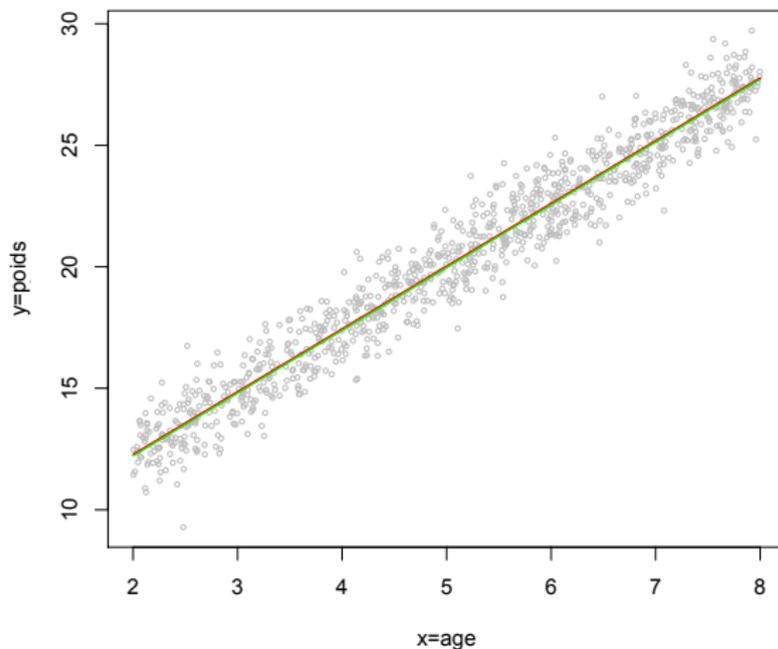
$$y = \text{Med}[Y | X = x].$$

Au vu de ce que nous avons vu, on peut deviner que cette droite est donnée par

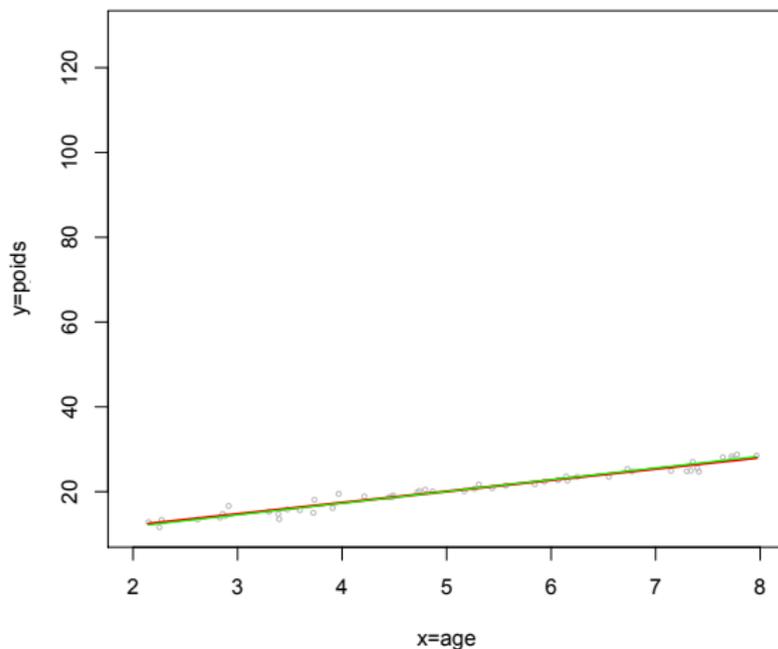
$$y = a_{L_1} + b_{L_1}x, \quad \text{où } (a_{L_1}, b_{L_1}) = \arg \min_{(a,b)} E[|Y - a - bX|].$$

Au niveau échantillon, ceci mène à **la droite des "moindres valeurs absolues"**

$$y = \hat{a} + \hat{b}x, \quad \text{où } (\hat{a}, \hat{b}) = \arg \min_{(a,b)} \frac{1}{n} \sum_{i=1}^n |Y_i - a - bX_i|.$$

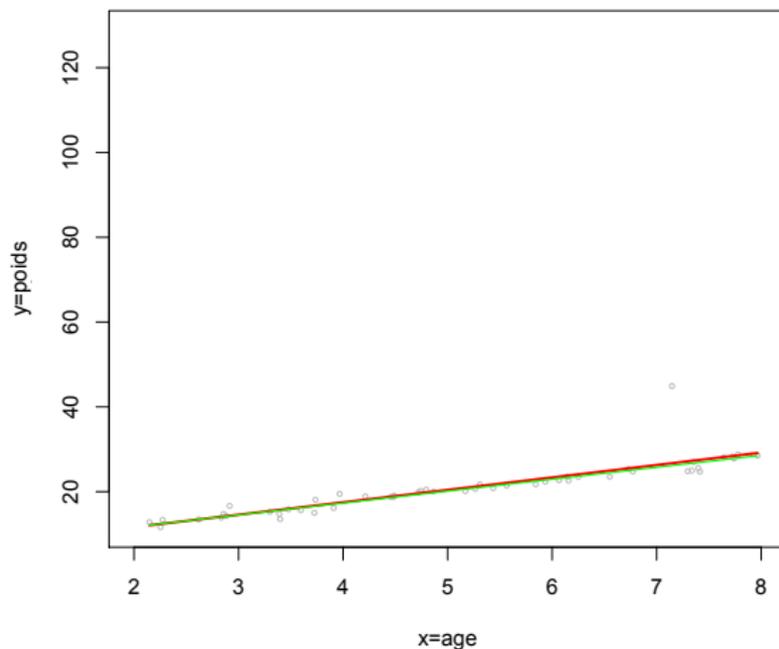


$n = 1000$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$



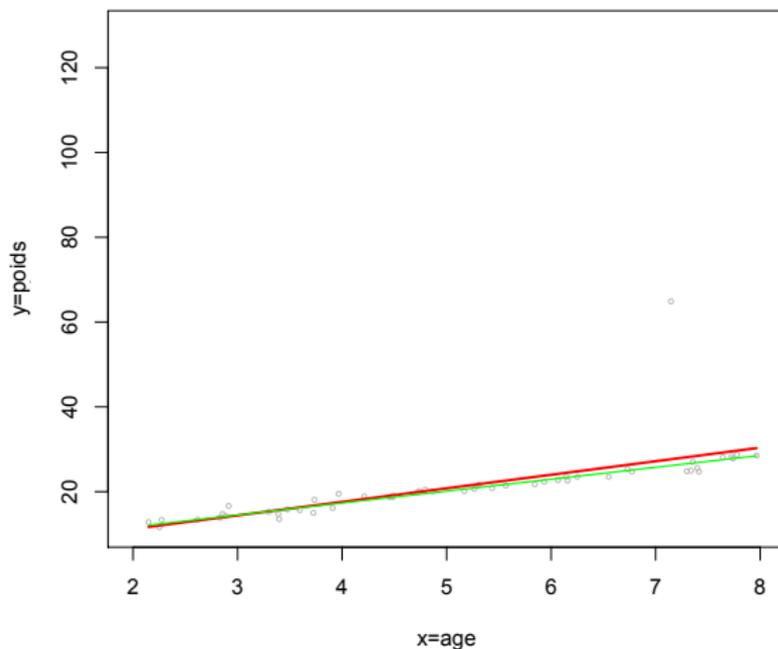
$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des $Y_i \dots$

Extension à la régression



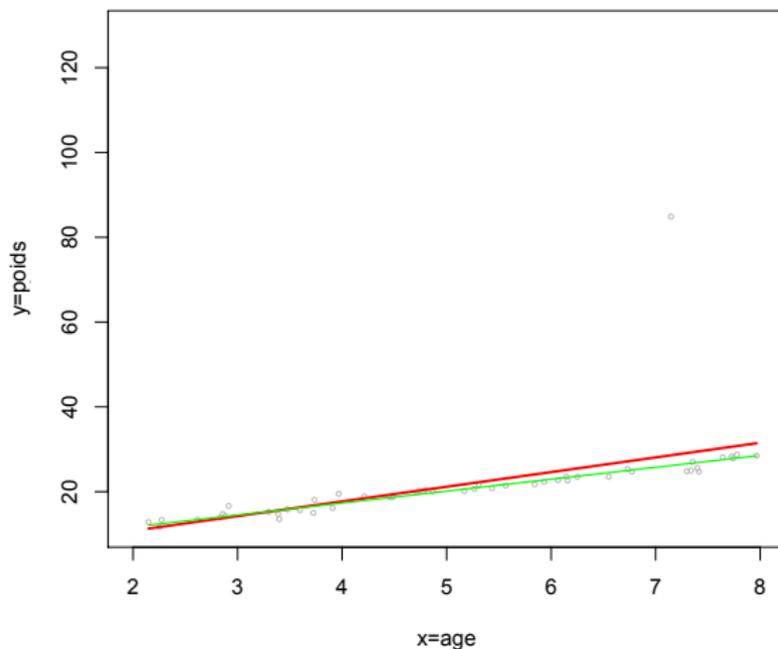
$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des Y_i ...

Extension à la régression



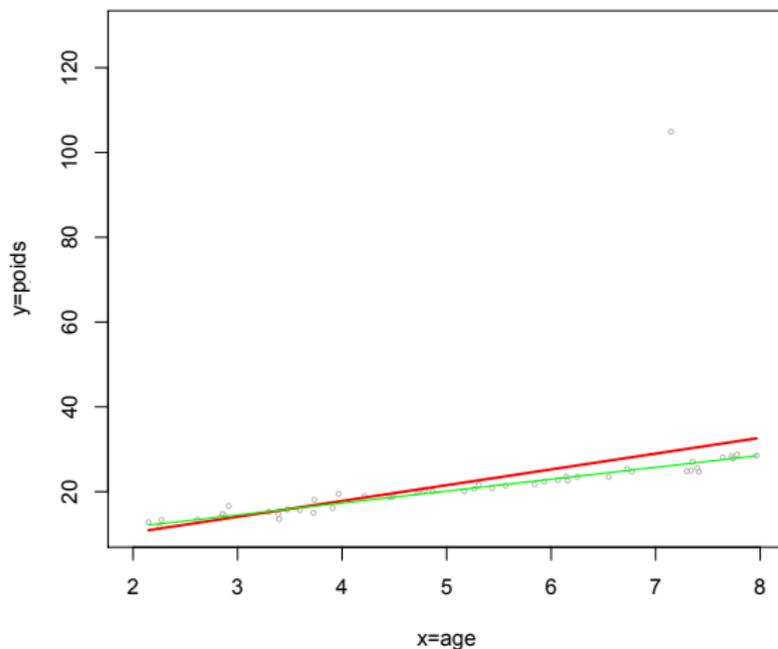
$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des Y_i ...

Extension à la régression



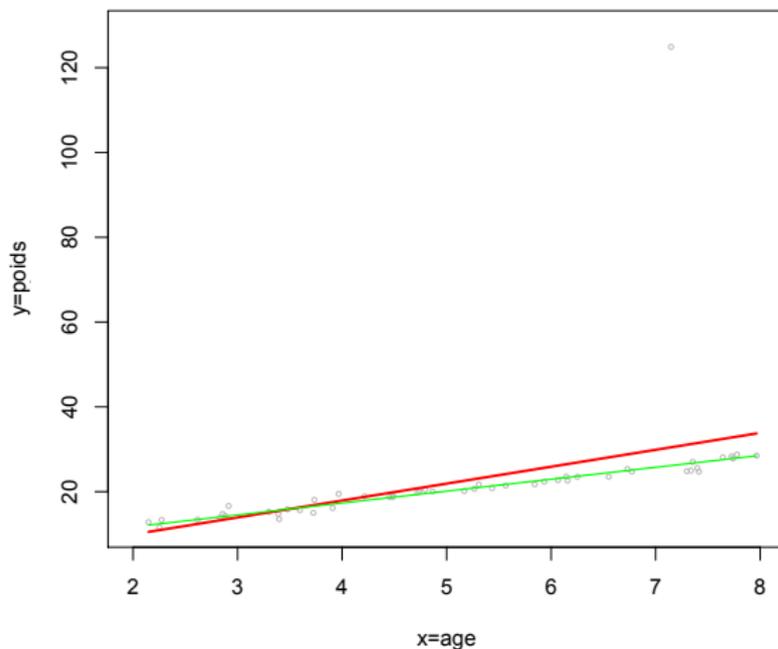
$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des $Y_i \dots$

Extension à la régression

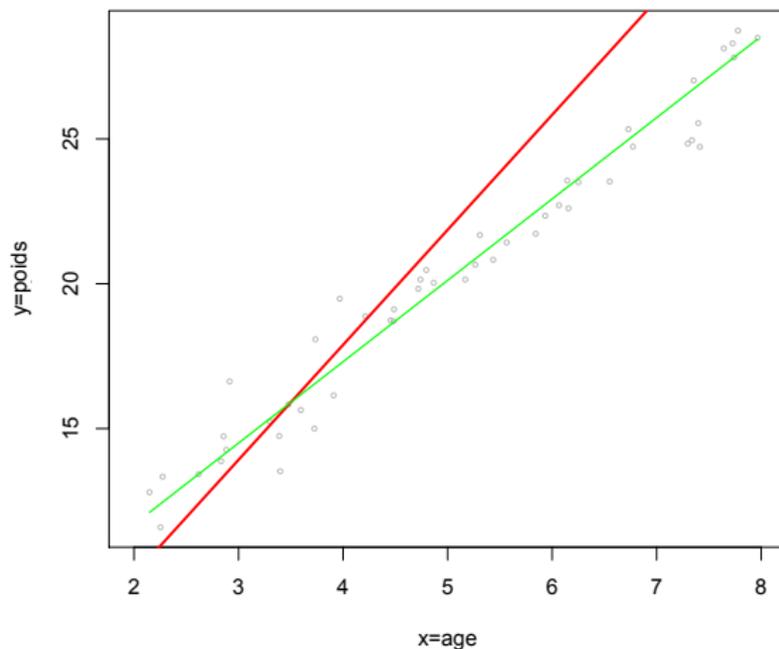


$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des Y_i ...

Extension à la régression



$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des Y_i ...



$n = 50$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp \varepsilon \sim \mathcal{N}(0, 1)$, et un peu de frouchelage avec l'un des $Y_i \dots$

Pour obtenir une mesure de risque, on considère alors

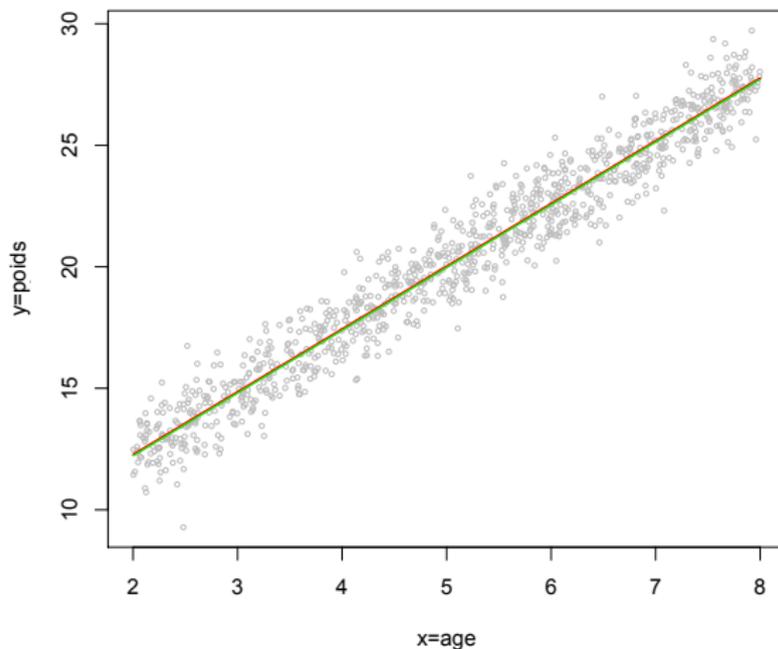
$$y = \text{quantile}_\tau[Y | X = x].$$

Au vu de ce que nous avons vu, on peut deviner que cette droite est donnée par

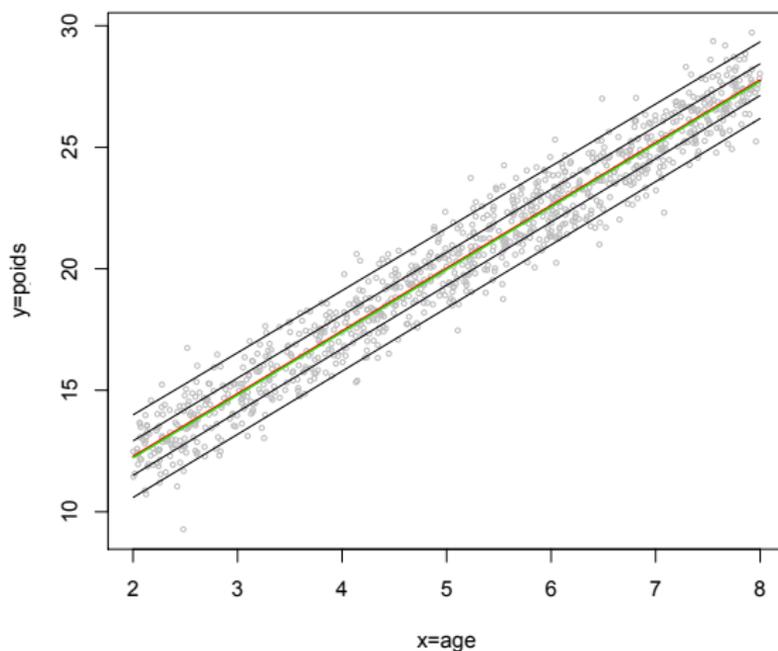
$$y = a_{L_1, \tau} + b_{L_1, \tau} x, \quad \text{où } (a_{L_1, \tau}, b_{L_1, \tau}) = \arg \min_{(a, b)} E[\rho_\tau(Y - a - bX)].$$

Au niveau échantillon, ceci mène à la droite

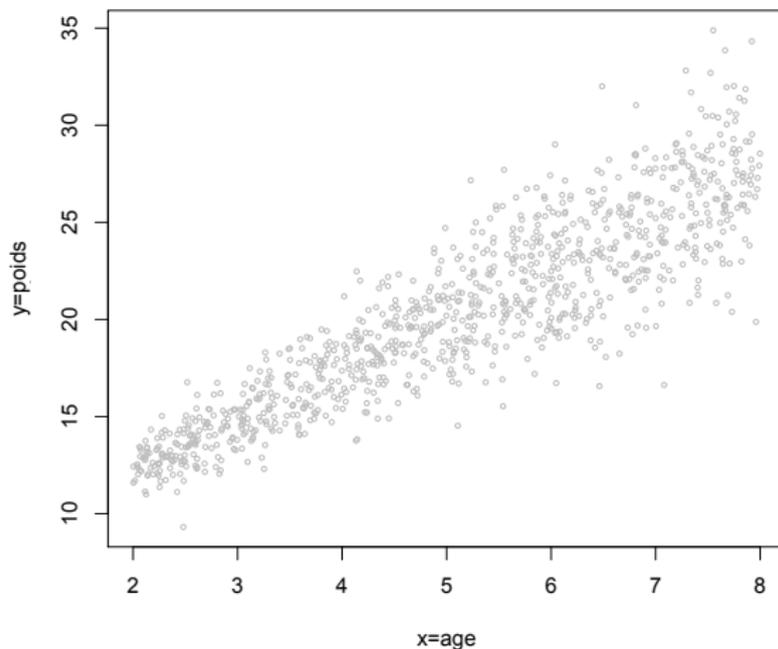
$$y = \hat{a} + \hat{b}x, \quad \text{où } (\hat{a}, \hat{b}) = \arg \min_{(a, b)} \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i - a - bX_i).$$



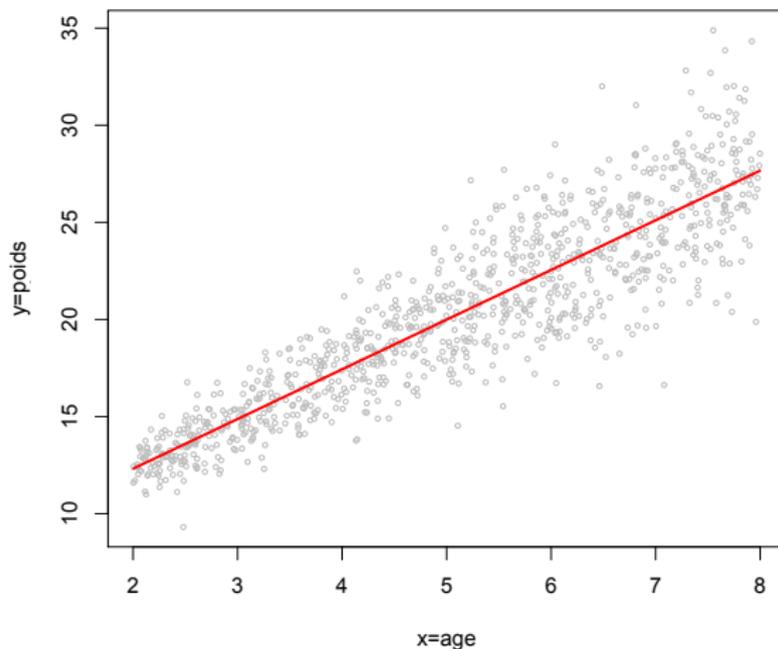
$n = 1000$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$



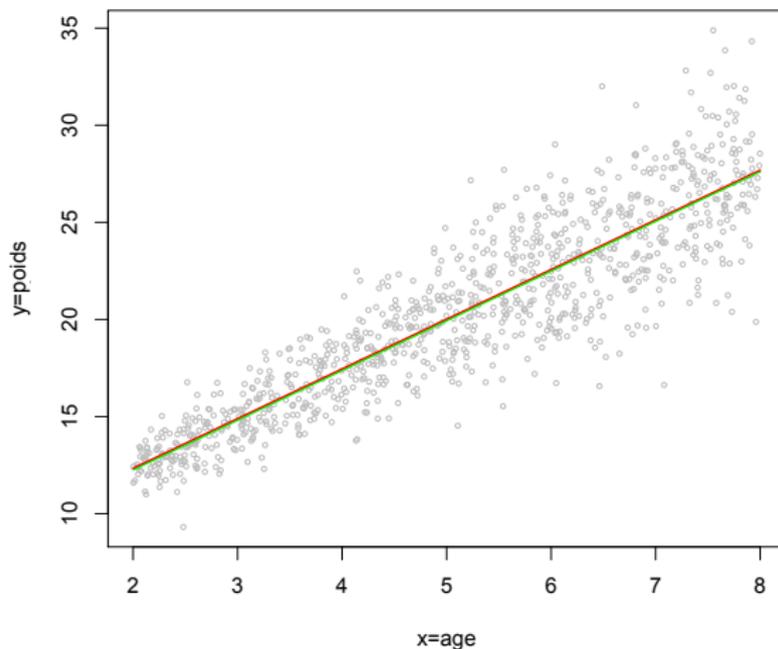
$n = 1000$ observations suivant $Y = (7 + 2.6X) + \varepsilon$
 $X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$



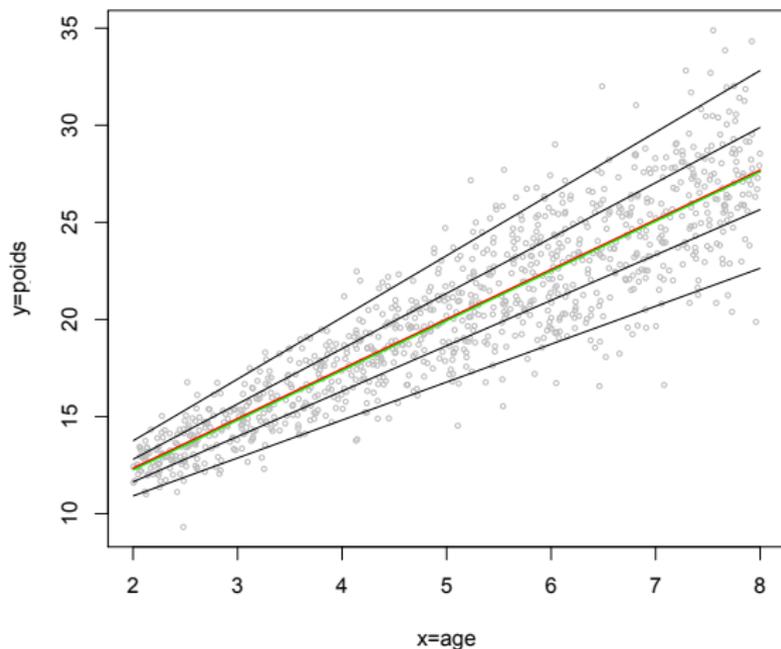
$$n = 1000 \text{ observations suivant } Y = (7 + 2.6X) + \frac{X}{2.5} \varepsilon$$
$$X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$$



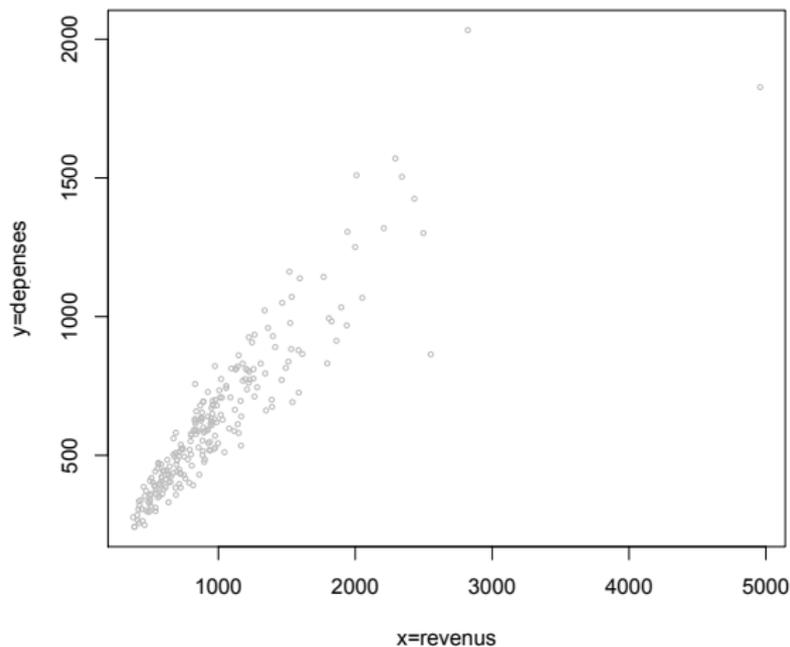
$$n = 1000 \text{ observations suivant } Y = (7 + 2.6X) + \frac{X}{2.5} \varepsilon$$
$$X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$$



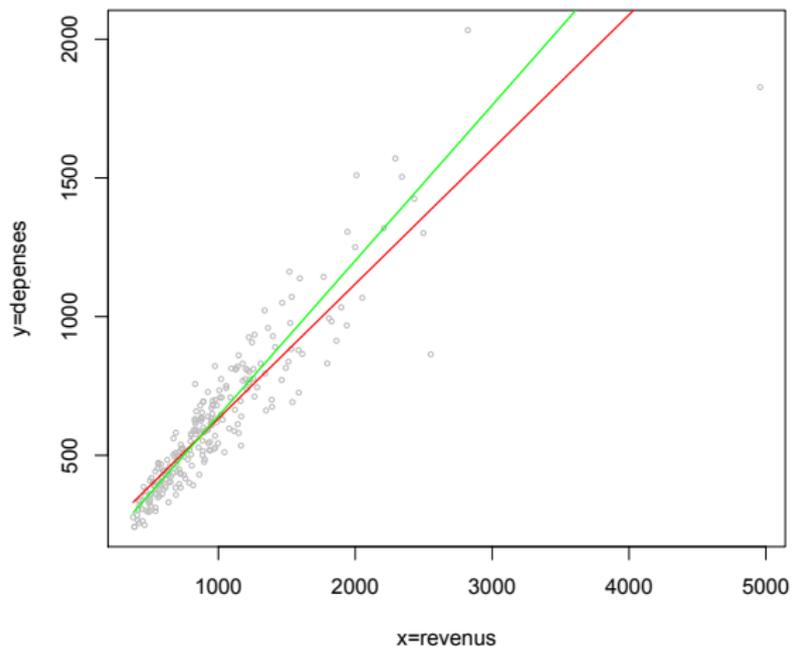
$$n = 1000 \text{ observations suivant } Y = (7 + 2.6X) + \frac{X}{2.5} \varepsilon$$
$$X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$$



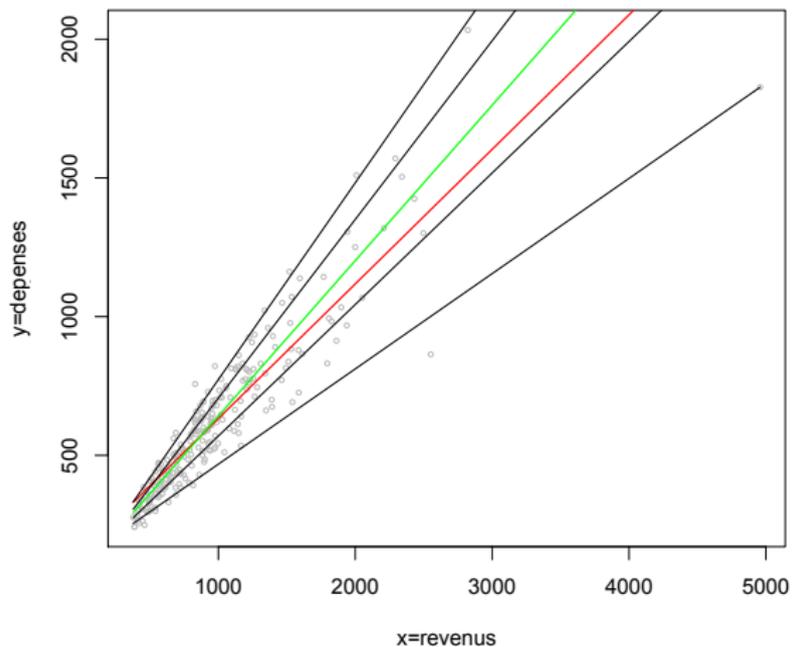
$$n = 1000 \text{ observations suivant } Y = (7 + 2.6X) + \frac{X}{2.5} \varepsilon$$
$$X \sim U([2, 8]) \perp\!\!\!\perp \varepsilon \sim \mathcal{N}(0, 1)$$



Ménages belges



Ménages belges



Ménages belges

Extension à la régression

Quantil Regression Shirt von Zazzle.de

http://www.zazzle.de/quantil_regression_tshirt-23580462208487819

ULB Info Sport Divers Webmail Publi Inception Proba Doodle

Modell Artikel Design



Quantil Regression Shirt

Von Robotern übersetzt. In Originalsprache ansehen

Größe:
Wähle eine [Größentabelle](#)

Menge:
1 Shirt. Nur **13,57 €** in Großmengen

So billig wie **13,75 €** für ein [Discount-T-Shirt](#)

In den Warenkorb **21,20 €** pro Shirt

Shirtarten und Farben

Basic 3/4-Arm Raglan: Weiß/Rot



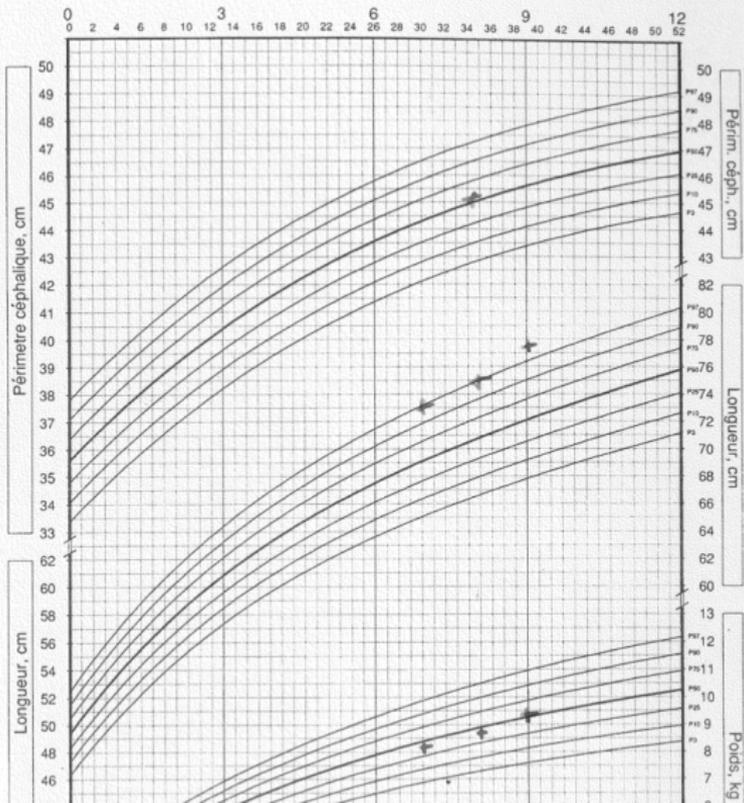
   

Discount-T-Shirt	Basic T-Shirt	Basic American Apparel T-Shirt	Damen Basic Doll (tailleur)
13,75 €	16,95 €	20,15 €	18,00 €

Gestalte es!



POIDS, TAILLE ET PÉRIMÈTRE CRANIE GARÇONS - 0/12 MOIS



1 Quantiles

- Moyenne, médiane, quantiles
- Extension à la régression
- Extension à la position multivariée

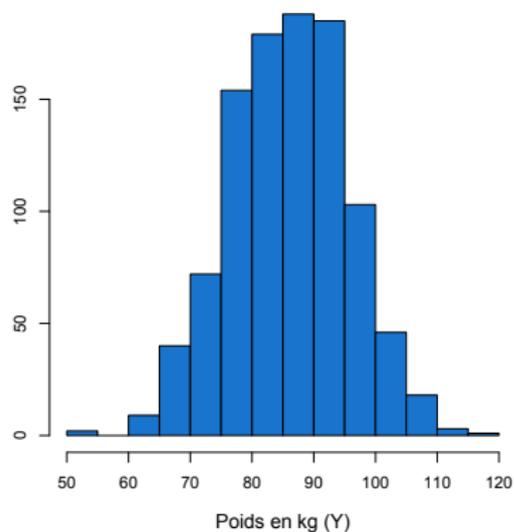
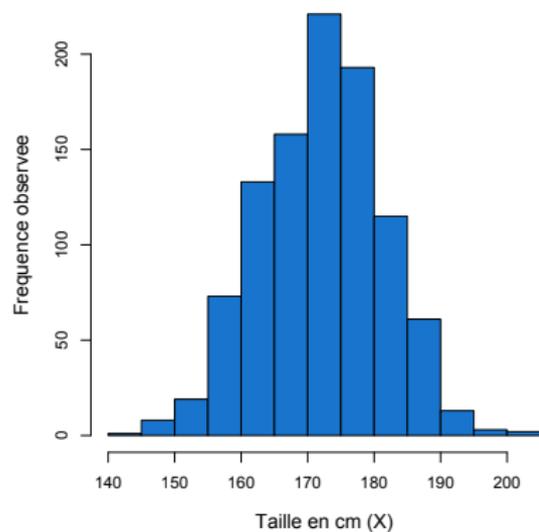
Au vu des courbes de croissance, le but ultime est de développer un concept de quantile pour le modèle de régression à réponse multivariée où le vecteur (X, \mathbf{Y}) comprend

- une variable explicative X
 - un d -vecteur de variables d'intérêt $\mathbf{Y} := (Y_1, \dots, Y_d)$.
-

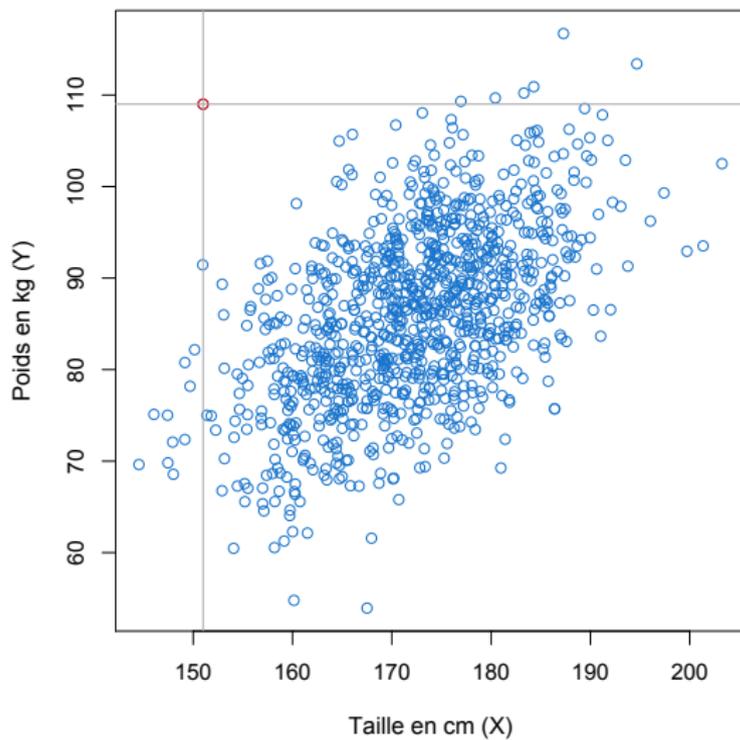
La difficulté principale :

Même dans le cas de la position (\gg regression), il n'y a pas de concept canonique de quantile.

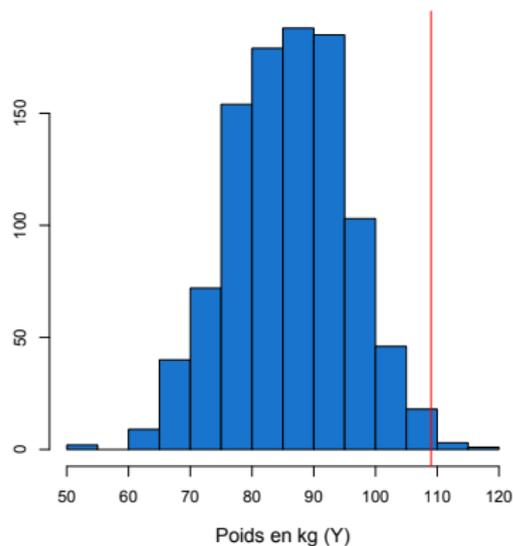
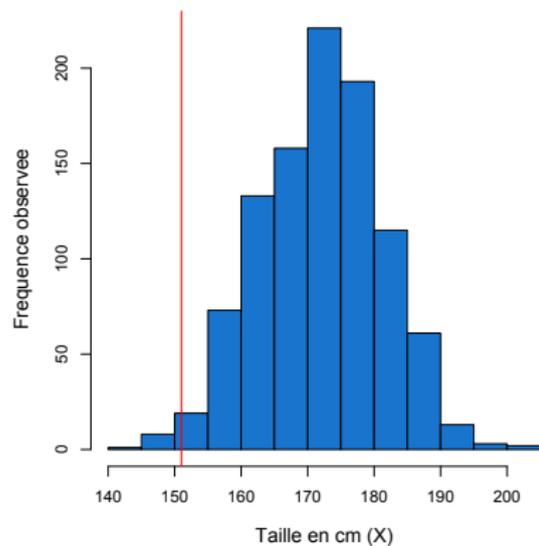
Extension à la position multivariée



Extension à la position multivariée



Extension à la position multivariée



Comment définir une médiane / un quantile pour $\mathbf{Y} := (Y_1, \dots, Y_d)$?

L'idée la plus naturelle consiste à prendre la médiane / le quantile composante par composante.

Cette idée souffre de nombreux défauts, le plus important étant le manque d'équivariance sous transformation affine.

Ici, nous allons considérer un concept de quantile **directionnel**.

Pourquoi ?

Parce que c'est déjà comme ça pour $d = 1$!



Ici, nous allons considérer un concept de quantile **directionnel**.

Pour

- un "ordre" $\tau \in (0, 1)$,
- une direction $\mathbf{u} \in \mathcal{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$,

nous définissons **la droite $(\tau\mathbf{u})$ -quantile**

$$\pi_{\tau\mathbf{u}} \quad (\pi_{\tau\mathbf{u}}^{(n)} \text{ au niveau échantillon})$$

comme la droite de regression τ -quantile obtenue quand \mathbf{u} a été choisi comme axe vertical (orienté).

En formules (pour le cas $d = 2$) :

Soit $\mathbf{u}^\perp \in \mathcal{S}^{d-1}$ orthogonal à \mathbf{u} . On peut alors décomposer \mathbf{y} en

$$\mathbf{y} = (\mathbf{y}'\mathbf{u}^\perp)\mathbf{u}^\perp + (\mathbf{y}'\mathbf{u})\mathbf{u}$$

$$(\approx (x, y) = x(1, 0) + y(0, 1)).$$

$\leadsto \pi_{\tau\mathbf{u}}$ est la droite d'équation

$$\mathbf{y}'\mathbf{u} = a_{\tau\mathbf{u}} + b_{\tau\mathbf{u}}(\mathbf{y}'\mathbf{u}^\perp)$$

$$(y = a_{L_{1,\tau}} + b_{L_{1,\tau}}x)$$

où $a_{\tau\mathbf{u}}, b_{\tau\mathbf{u}}$ sont donnés par

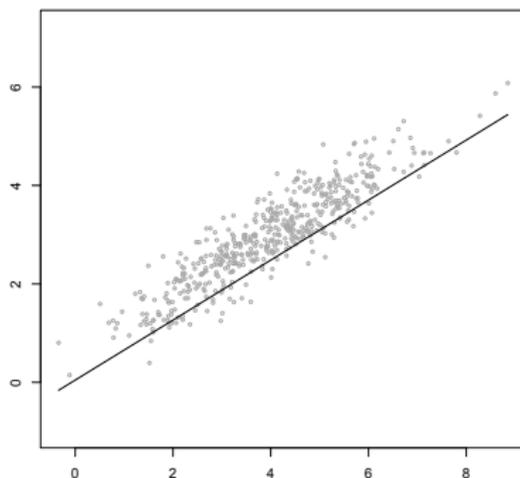
$$(a_{\tau\mathbf{u}}, b_{\tau\mathbf{u}}) = \arg \min_{(a,b)} E[\rho_\tau(\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^\perp))]$$

$$(a_{L_{1,\tau}}, b_{L_{1,\tau}}) = \arg \min_{(a,b)} E[\rho_\tau(Y - a - bX)].$$

Nous définissons la droite $(\tau\mathbf{u})$ -quantile $\pi_{\tau\mathbf{u}}^{(n)}$ comme la droite de régression τ -quantile obtenue quand \mathbf{u} a été choisi comme axe vertical (orienté).

$$\tau = .1$$

$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Conditions de premier ordre :

$$(a_{\tau u}, b_{\tau u}) = \arg \min_{(a,b)} E[\rho_{\tau}(\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^{\perp}))]$$

$$\text{où } \rho(z) = z(\tau - \mathbb{I}[z < 0]) = \begin{cases} (1 - \tau)|z| & \text{si } z < 0 \\ \tau|z| & \text{si } z \geq 0. \end{cases}$$

$$\begin{aligned} \Rightarrow \frac{\partial}{\partial a} E[\rho_{\tau}(\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^{\perp}))] &= \frac{\partial}{\partial a} E[\{\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^{\perp})\}(\tau - \mathbb{I}[\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^{\perp}) < 0])] \\ &= E[(-1) \times (\tau - \mathbb{I}[\mathbf{Y}'\mathbf{u} - a_{\tau u} - b_{\tau u}(\mathbf{Y}'\mathbf{u}^{\perp}) < 0])] = 0 \end{aligned}$$

$$\Rightarrow \tau = P[\mathbf{Y} \in H_{\tau u}^{(n)-}]$$

$$\begin{aligned} \frac{\partial}{\partial \mathbf{b}} E[\rho_\tau(\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^\perp))] &= \frac{\partial}{\partial \mathbf{b}} E[\{\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^\perp)\}(\tau - \mathbb{I}[\mathbf{Y}'\mathbf{u} - a - b(\mathbf{Y}'\mathbf{u}^\perp) < 0])] \\ &= E[-(\mathbf{Y}'\mathbf{u}^\perp) \times (\tau - \mathbb{I}[\mathbf{Y}'\mathbf{u} - a_{\tau\mathbf{u}} - b_{\tau\mathbf{u}}(\mathbf{Y}'\mathbf{u}^\perp) < 0])] = 0 \end{aligned}$$

$$\Rightarrow \tau E[\mathbf{Y}'\mathbf{u}^\perp] = E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]]$$

$$\Rightarrow \tau E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]] + \tau E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)+}]] = E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]]$$

$$\Rightarrow \tau E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)+}]] = (1 - \tau)E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]]$$

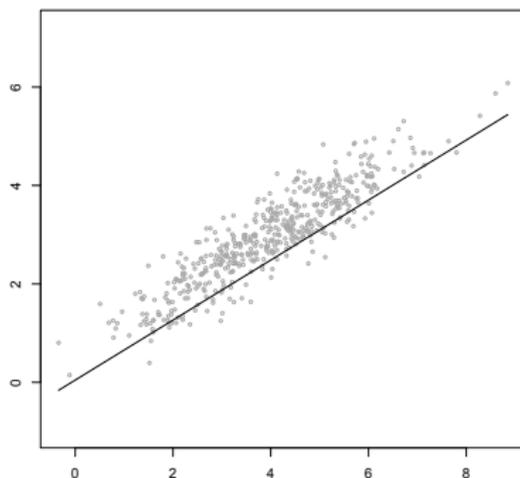
$$\Rightarrow \frac{1}{1 - \tau} E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)+}]] = \frac{1}{\tau} E[(\mathbf{Y}'\mathbf{u}^\perp)\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]]$$

$$\Rightarrow (\mathbf{u}^\perp)' \left[\frac{1}{1 - \tau} E[\mathbf{Y}\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)+}]] - \frac{1}{\tau} E[\mathbf{Y}\mathbb{I}[\mathbf{Y} \in H_{\tau\mathbf{u}}^{(n)-}]] \right] = 0.$$

Nous définissons la droite $(\tau\mathbf{u})$ -quantile $\pi_{\tau\mathbf{u}}^{(n)}$ comme la droite de régression τ -quantile obtenue quand \mathbf{u} a été choisi comme axe vertical (orienté).

$$\tau = .1$$

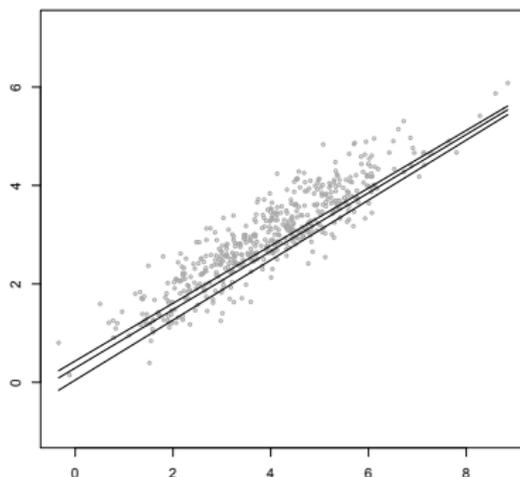
$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Nous définissons la droite $(\tau\mathbf{u})$ -quantile $\pi_{\tau\mathbf{u}}^{(n)}$ comme la droite de régression τ -quantile obtenue quand \mathbf{u} a été choisi comme axe vertical (orienté).

$$\tau = .1, .2, .3$$

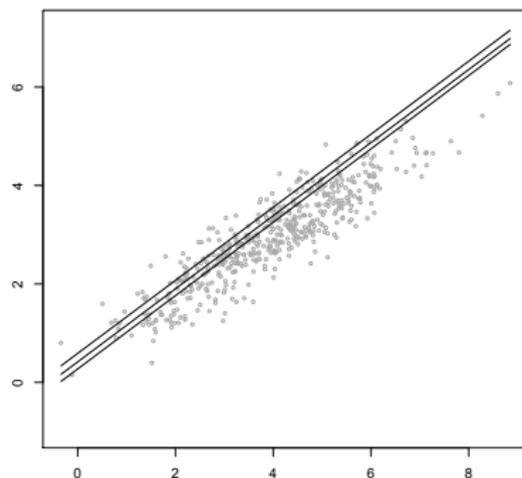
$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



Nous définissons la droite $(\tau\mathbf{u})$ -quantile $\pi_{\tau\mathbf{u}}^{(n)}$ comme la droite de regression τ -quantile obtenue quand \mathbf{u} a été choisi comme axe vertical (orienté).

$$\tau = .1, .2, .3$$

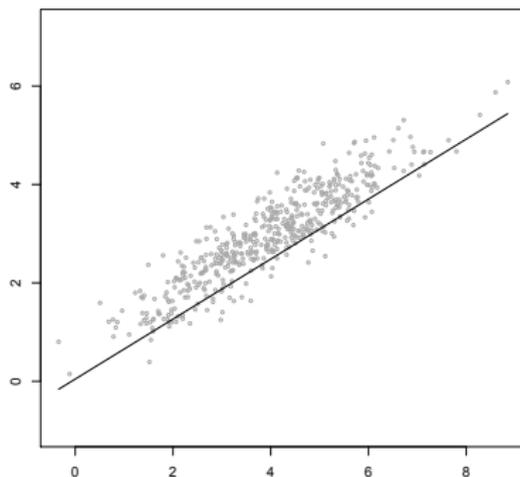
$$\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$



La dépendance en \mathbf{u} est importante.

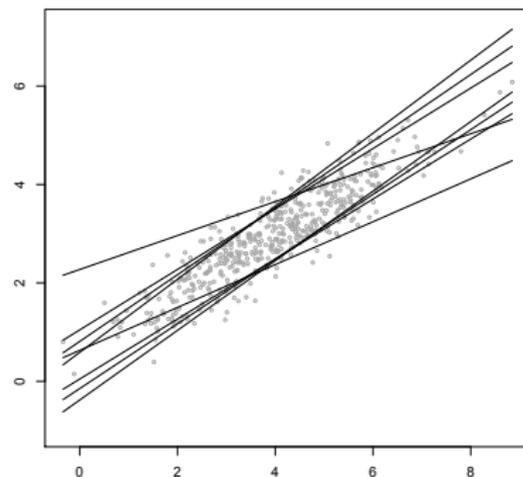
$$\tau = .1$$

$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$



La dépendance en \mathbf{u} est importante.

$\tau = .1$
 $\mathbf{u} = \begin{pmatrix} 0 \\ \pm 1 \end{pmatrix}, \begin{pmatrix} \pm 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \pm 1/\sqrt{2} \\ \pm 1/\sqrt{2} \end{pmatrix}$
i.e., 8 \mathbf{u} équiésacés



La dépendance en \mathbf{u} est importante.

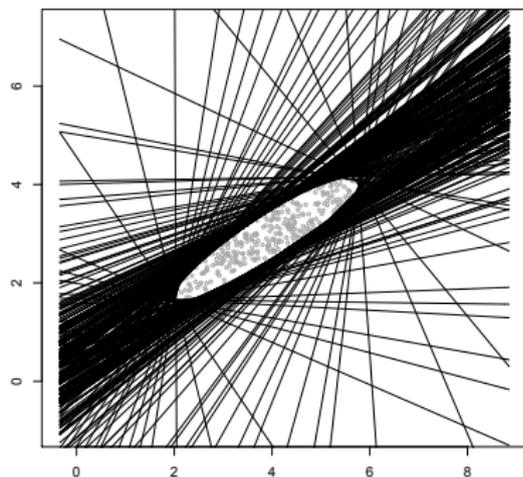
$$\tau = .1$$

512 \mathbf{u} équiespacés

La dépendance en \mathbf{u} est importante.

$$\tau = .1$$

512 \mathbf{u} équiésacés



Si on note $H_{\tau \mathbf{u}}^{(n)+}$ le demi-espace "au dessus de" $\pi_{\tau \mathbf{u}}^{(n)}$, nous pouvons obtenir, pour chaque $\tau \in (0, 1)$ fixé, une **région centrale d'ordre τ** en considérant

$$\bigcap_{\mathbf{u} \in S^{d-1}} H_{\tau \mathbf{u}}^{(n)+}.$$

Ces régions sont importantes pour détecter les individus à risque.

Si on note $H_{\tau\mathbf{u}}^{(n)+}$ le demi-espace "au dessus de" $\pi_{\tau\mathbf{u}}^{(n)}$, nous pouvons obtenir, pour chaque $\tau \in (0, 1)$ fixé, une **région centrale d'ordre τ** en considérant

$$\bigcap_{\mathbf{u} \in S^{d-1}} H_{\tau\mathbf{u}}^{(n)+}.$$

Ces régions sont importantes pour détecter les individus à risque.

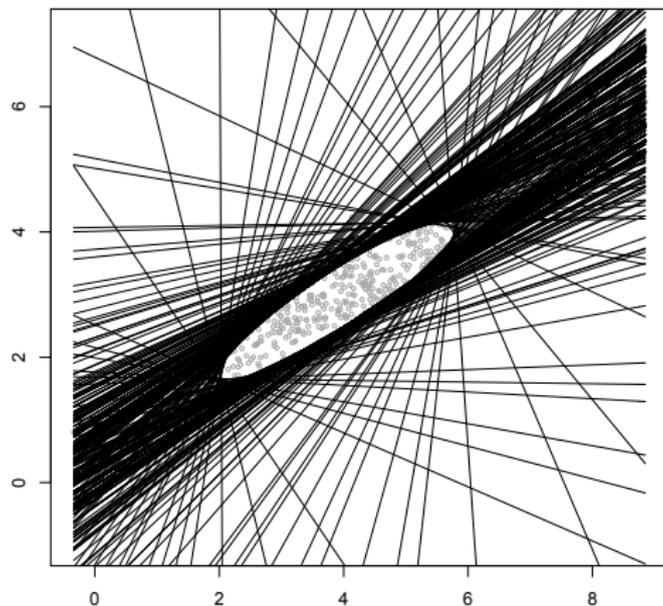
Evidemment, il ne semble pas facile de calculer ces régions car l'intersection est sur une infinité (non dénombrable) de directions \mathbf{u} ...



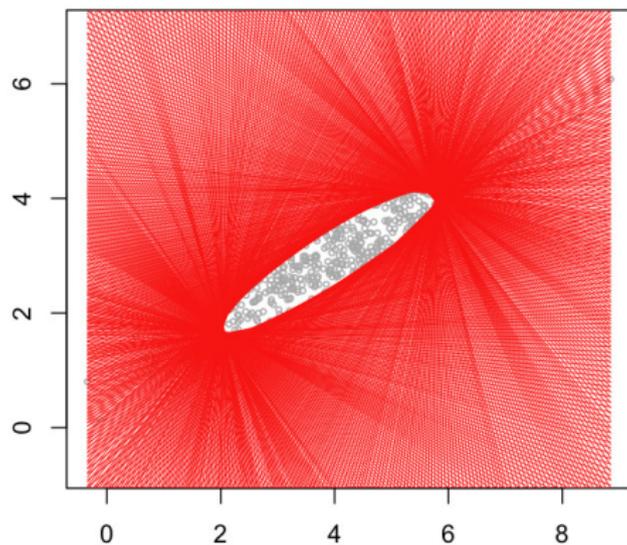
"Ça ressemble à l'alcool, c'est doré comme l'alcool... mais ce n'est pas de l'alcool"
(parfois, de façon éhontée, déformé en
"Ça a la couleur de l'alcool, le goût de l'alcool... mais ce n'est pas de l'alcool")

$$\bigcap_{\mathbf{u} \in \mathcal{S}^{d-1}} H_{\tau \mathbf{u}}^{(n)+}$$

"Ça ressemble à une intersection sur une infinité dénombrable de \mathbf{u} . . . mais ce n'est pas une intersection sur une infinité dénombrable de \mathbf{u} ")

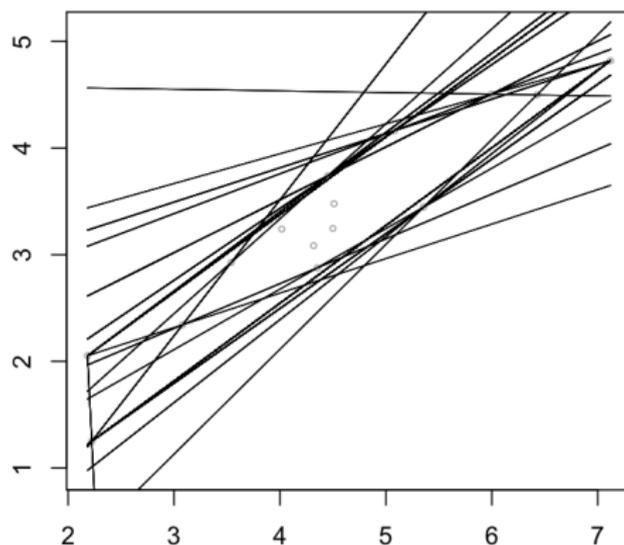


$\pi_{\tau \mathbf{u}}^{(n)}$, avec 512 $\mathbf{u} \in \mathcal{S}^1$, $\tau = .1$ ($n = 500$)



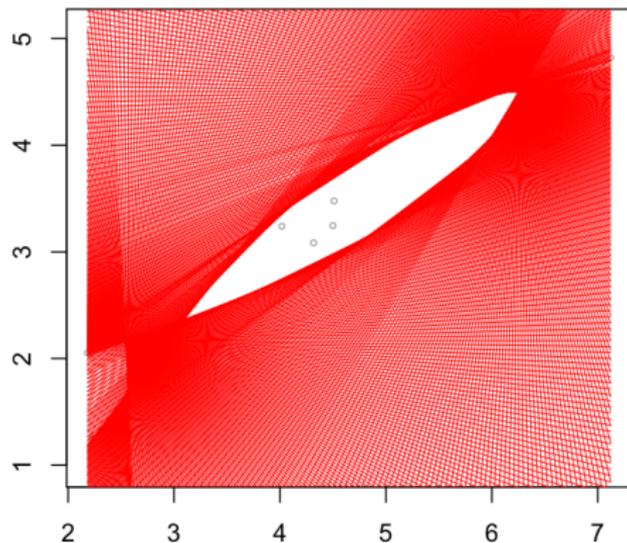
$\pi_{\tau \mathbf{u}; \text{proj}}^{(n)}$, avec 512 $\mathbf{u} \in \mathcal{S}^1$, $\tau = .1$ ($n = 500$)

Pour seulement $n = 15$ observations :



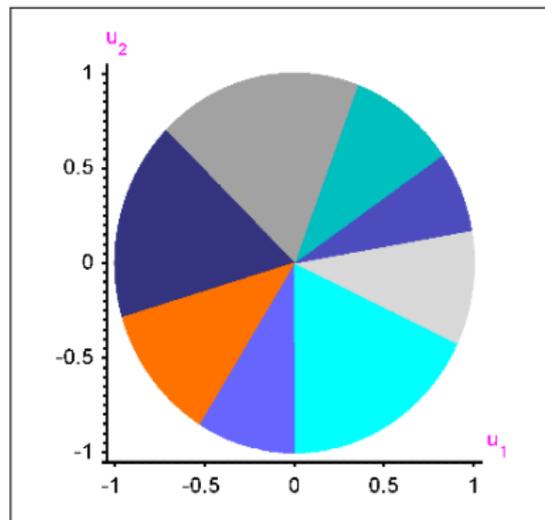
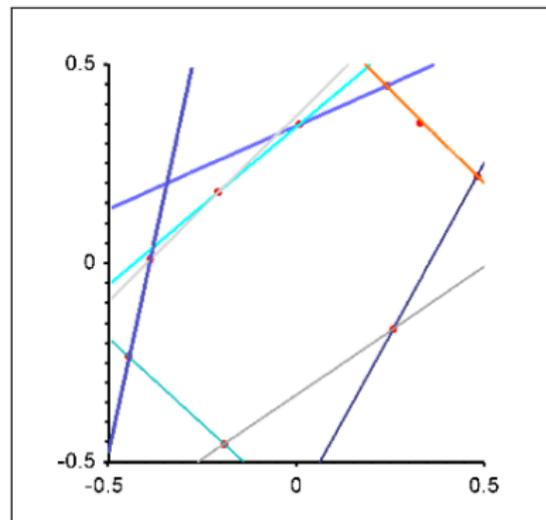
$\pi_{\tau \mathbf{u}}^{(n)}$, avec 512 $\mathbf{u} \in \mathcal{S}^1$, $\tau = .1$ ($n = 15$)

Pour seulement $n = 15$ observations :



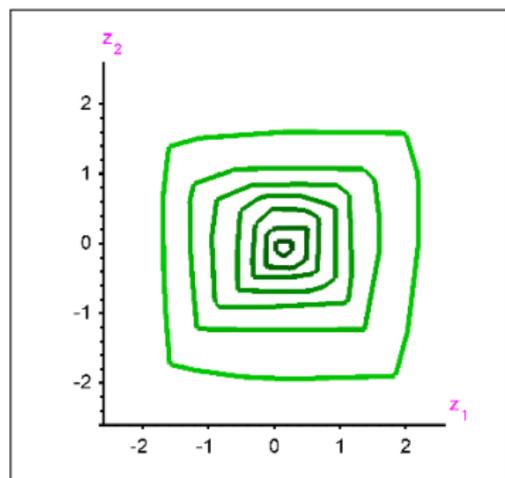
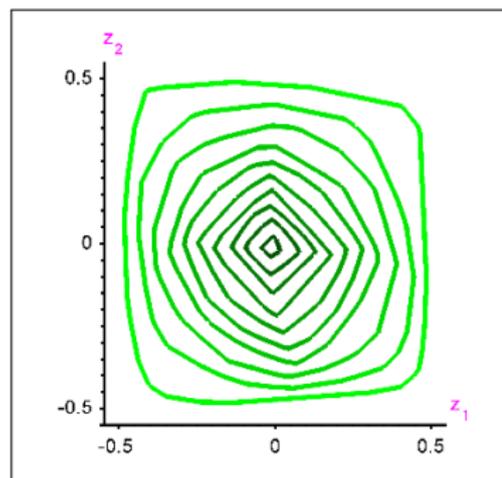
$$\pi_{\tau \mathbf{u}; \text{proj}}^{(n)}, \text{ avec } 512 \mathbf{u} \in \mathcal{S}^1, \tau = .1 (n = 15)$$

On peut identifier les sauts du processus $u \mapsto H_{\tau\mathbf{u}}^{(n)+}$ de façon efficace.

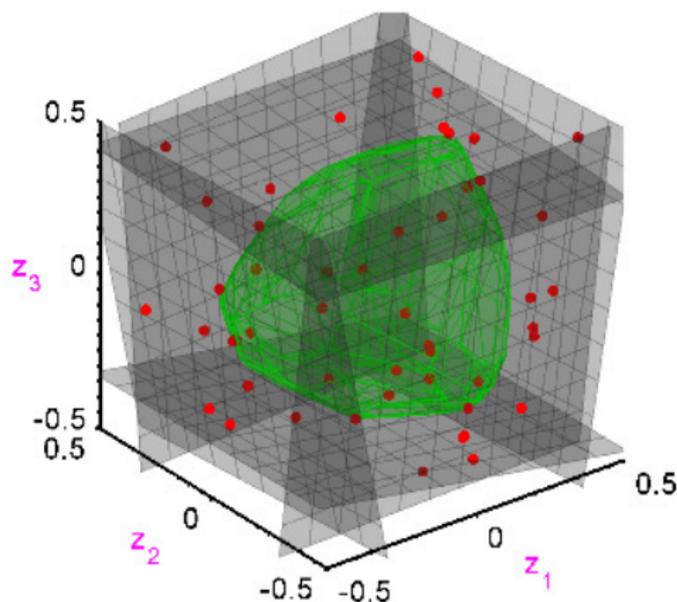


Donc ceci fournit un moyen efficace de calculer les régions

Extension à la position multivariée



Quelques régions centrales pour $n = 450$ observations indépendantes de loi uniforme sur $[-0.5, 0.5]^2$ (gauche) ou avec des marginales t_1 indépendantes (droite)



$\pi_{\tau \mathbf{u}}^{(n)}$, avec $\mathbf{u} = \{(\pm 1, 0, 0)', (0, \pm 1, 0)', (0, 0, \pm 1)'\}$, $\tau = .1$, et la région centrale qui en résulte pour 50 observations indépendantes de loi uniforme sur $[-.5, .5]^3$.

Au delà de $d = 3$, on ne peut plus faire de dessin, mais on a des indicatrices pour dire si un point donné de \mathbb{R}^d est ou n'est pas dans la région de centralité d'ordre τ ...

Ca l'air bien, les régions de centralité.

On va en regarder plus après la pause...

2 Profondeur statistique

- La profondeur de demi-espace
- Un zoo de profondeurs
- Classification, et extension locale de la profondeur

La profondeur de demi-espace

Soit $C_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i \in \mathbb{R}^d$.

Une **profondeur** mesure la **centralité** d'un quelconque $\boldsymbol{\theta} \in \mathbb{R}^d$ par rapport à C_n

$$D(\cdot, C_n) : \mathbb{R}^d \rightarrow [0, 1]$$

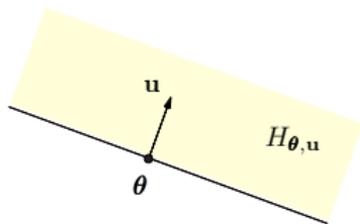
$$\boldsymbol{\theta} \mapsto D(\boldsymbol{\theta}, C_n)$$

Au plus grand $D(\boldsymbol{\theta}, C_n)$, au plus central $\boldsymbol{\theta}$ par rapport C_n .

La profondeur de demi-espace (1975)

$$D(\boldsymbol{\theta}, C_n) = \min_{\mathbf{u} \in S^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\boldsymbol{\theta}, \mathbf{u}}\},$$

où $S^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d
 $H_{\boldsymbol{\theta}, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \boldsymbol{\theta}) \geq 0\}$ est le demi-espace

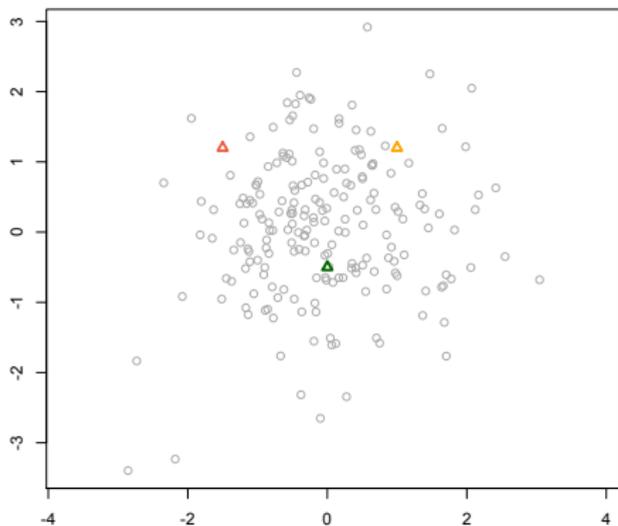
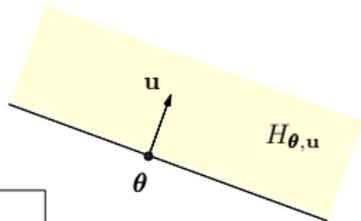


La profondeur de demi-espace (1975)

$$D(\theta, C_n) = \min_{\mathbf{u} \in S^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\theta, \mathbf{u}}\},$$

où $S^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d

$H_{\theta, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \theta) \geq 0\}$ est le demi-espace

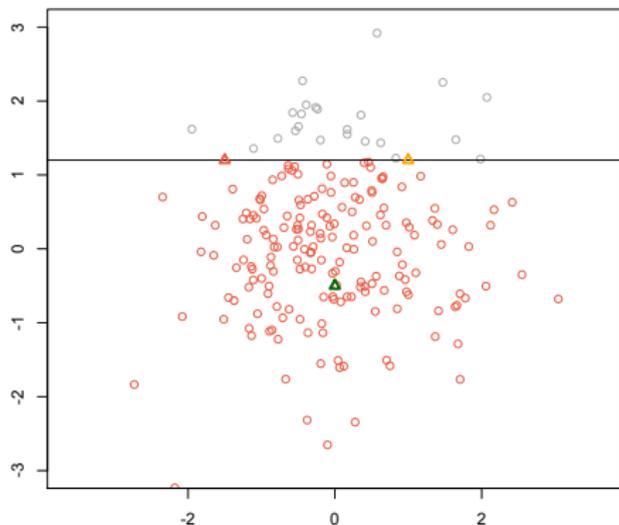
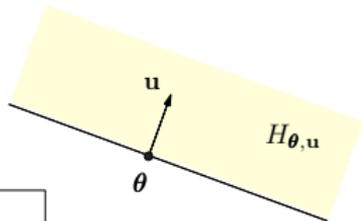


La profondeur de demi-espace (1975)

$$D(\theta, C_n) = \min_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\theta, \mathbf{u}}\},$$

où $\mathcal{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d

$H_{\theta, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \theta) \geq 0\}$ est le demi-espace

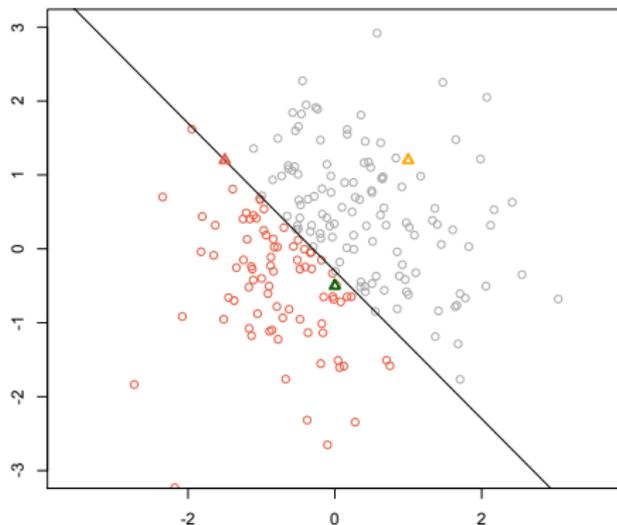
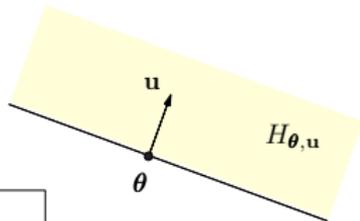


La profondeur de demi-espace (1975)

$$D(\theta, C_n) = \min_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\theta, \mathbf{u}}\},$$

où $\mathcal{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d

$H_{\theta, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \theta) \geq 0\}$ est le demi-espace

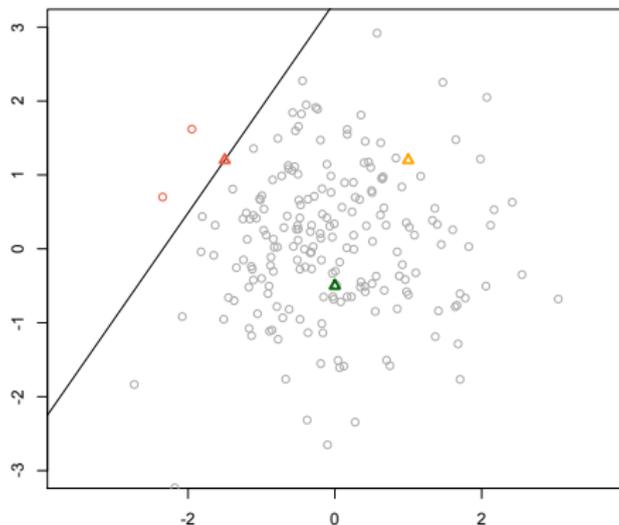
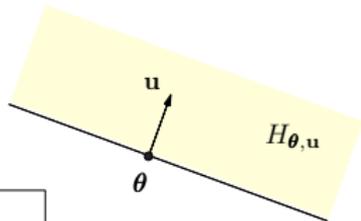


La profondeur de demi-espace (1975)

$$D(\theta, C_n) = \min_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\theta, \mathbf{u}}\},$$

où $\mathcal{S}^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d

$H_{\theta, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \theta) \geq 0\}$ est le demi-espace

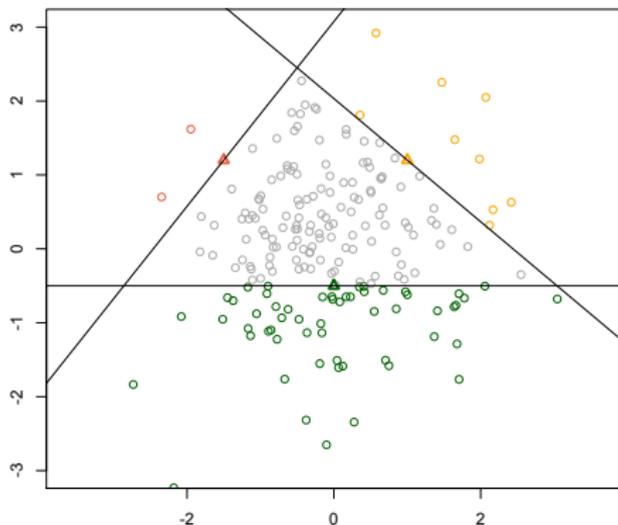
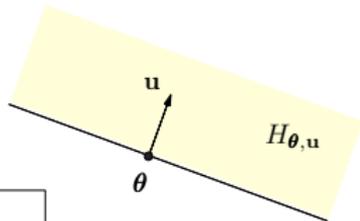


La profondeur de demi-espace (1975)

$$D(\theta, C_n) = \min_{\mathbf{u} \in S^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\theta, \mathbf{u}}\},$$

où $S^{d-1} = \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\| = 1\}$ est la sphère unité dans \mathbb{R}^d

$H_{\theta, \mathbf{u}} = \{\mathbf{y} \in \mathbb{R}^d : \mathbf{u}'(\mathbf{y} - \theta) \geq 0\}$ est le demi-espace



De $C_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ à une mesure de probabilité arbitraire P :

Si $P^{(n)} = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{y}_i}$ désigne la mesure de probabilité empirique associée à $C_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, on a

$$\begin{aligned} D(\boldsymbol{\theta}, C_n) &= \min_{\mathbf{u} \in \mathcal{S}^{d-1}} \frac{1}{n} \#\{\mathbf{y}_i \in H_{\boldsymbol{\theta}, \mathbf{u}}\} \\ &= \min_{\mathbf{u} \in \mathcal{S}^{d-1}} P^{(n)}[H_{\boldsymbol{\theta}, \mathbf{u}}]. \end{aligned}$$

↪ Extension à une P arbitraire :

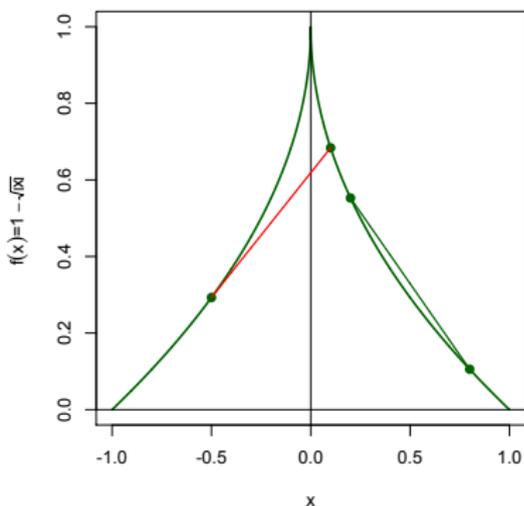
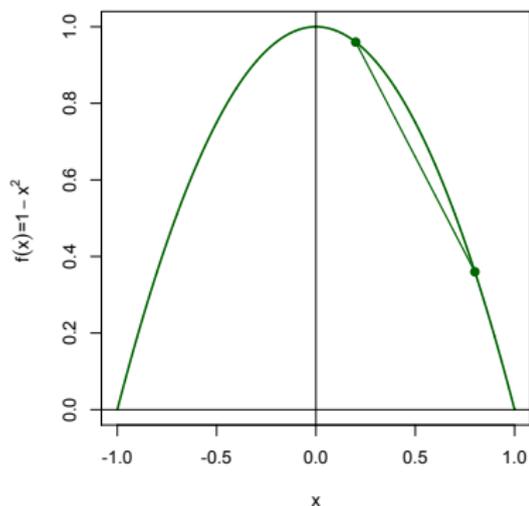
$$D(\boldsymbol{\theta}, P) = \inf_{\mathbf{u} \in \mathcal{S}^{d-1}} P[H_{\boldsymbol{\theta}, \mathbf{u}}].$$

Théorème

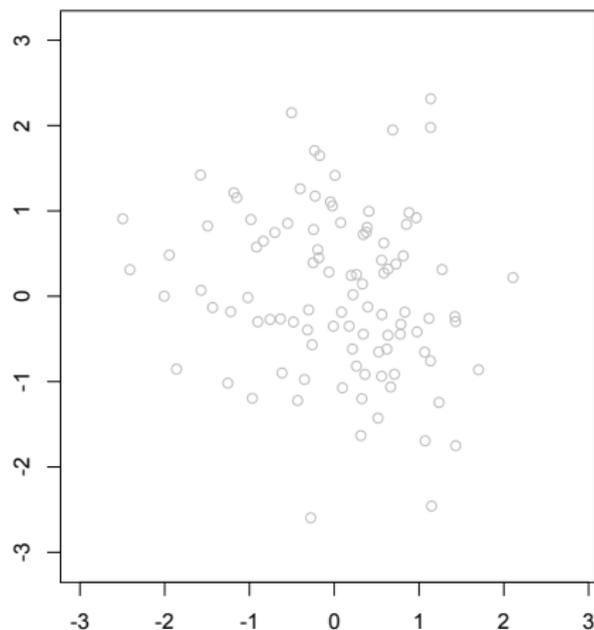
Soit P une mesure de proba. Alors $\theta \rightarrow D(\theta, P)$ est *quasi-concave*.

$f : \mathbb{R}^d \rightarrow \mathbb{R}$ concave $\stackrel{\text{def}}{\Leftrightarrow} \forall \theta_1, \theta_2 \in \mathbb{R}^d, \forall \lambda \in (0, 1), f((1 - \lambda)\theta_1 + \lambda\theta_2) \geq (1 - \lambda)f(\theta_1) + \lambda f(\theta_2)$

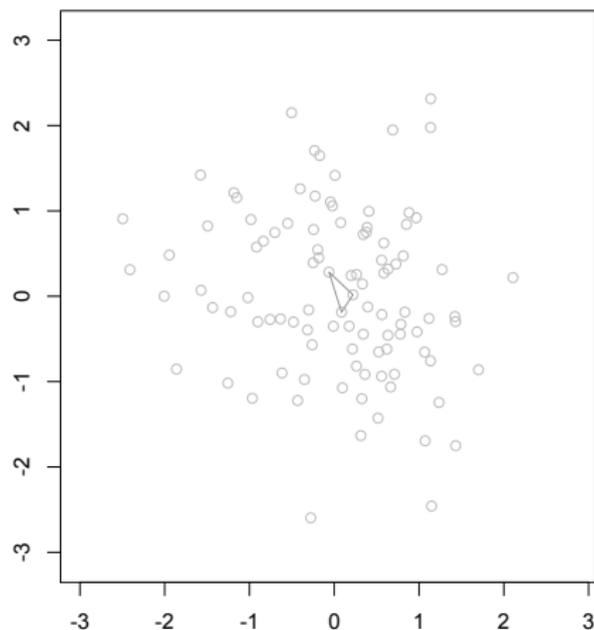
$f : \mathbb{R}^d \rightarrow \mathbb{R}$ quasi-concave $\stackrel{\text{def}}{\Leftrightarrow} \forall \theta_1, \theta_2 \in \mathbb{R}^d, \forall \lambda \in (0, 1), f((1 - \lambda)\theta_1 + \lambda\theta_2) \geq \min(f(\theta_1), f(\theta_2))$



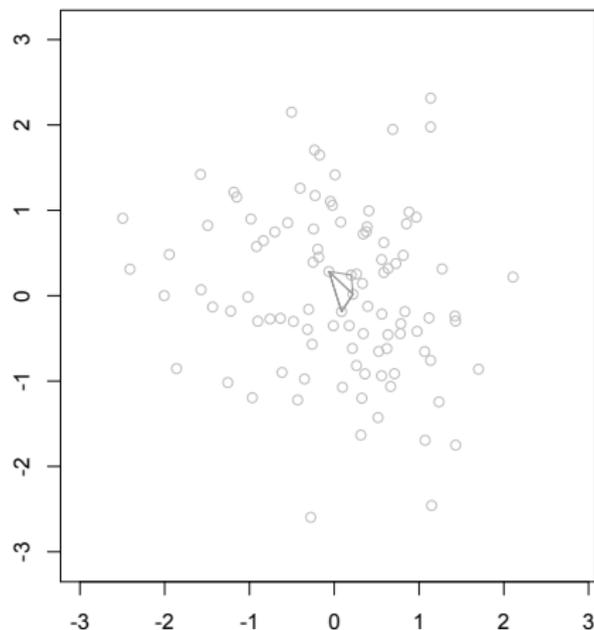
On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\boldsymbol{\theta} \in \mathbb{R}^d : D(\boldsymbol{\theta}, P) \geq \tau\}$



On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$

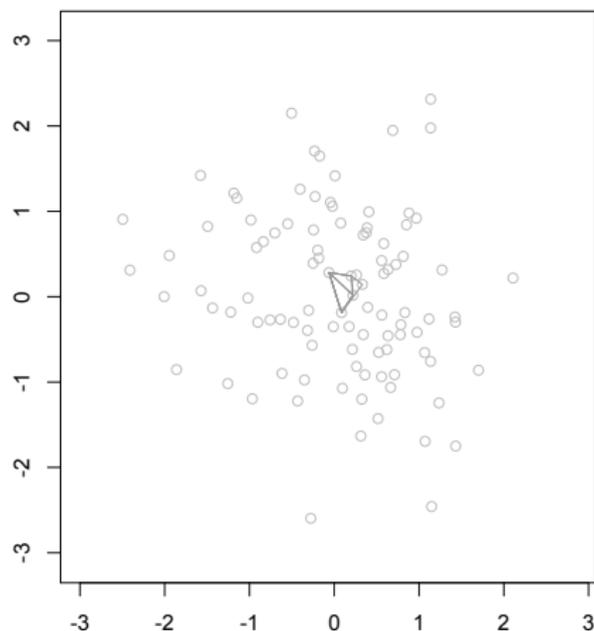


On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



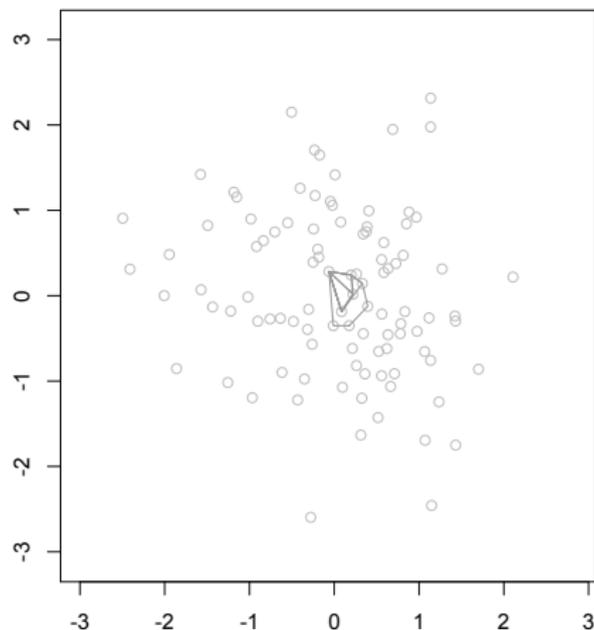
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



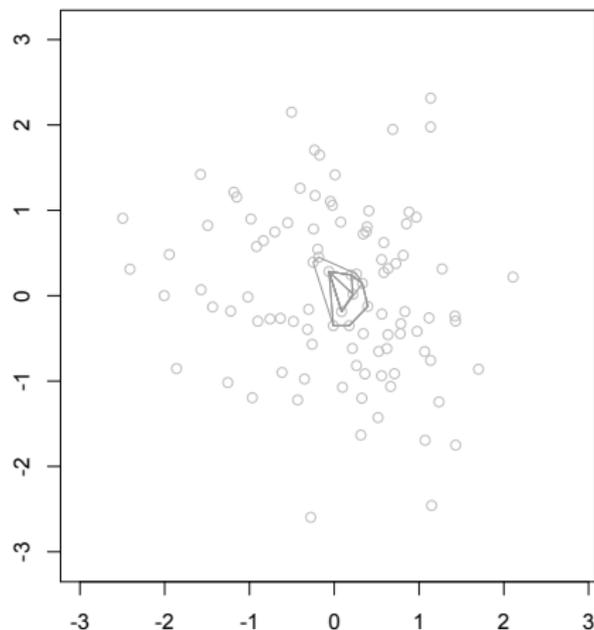
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



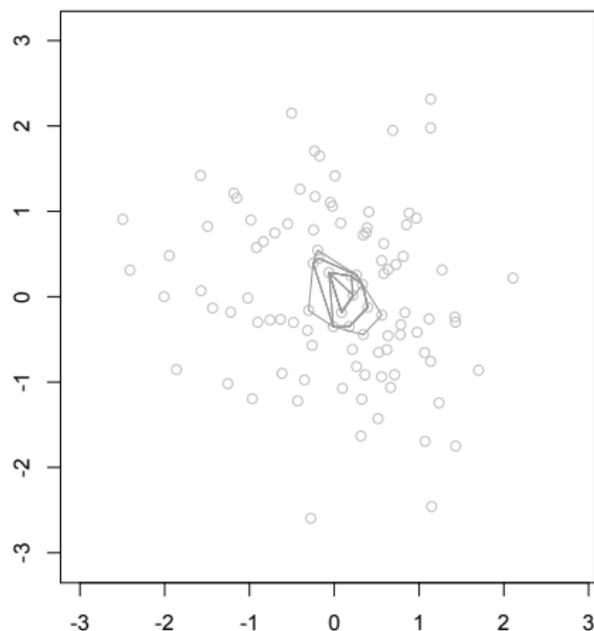
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



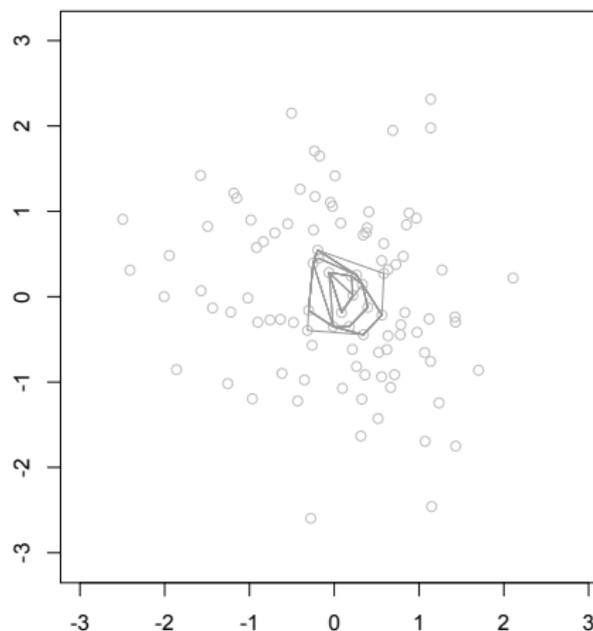
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



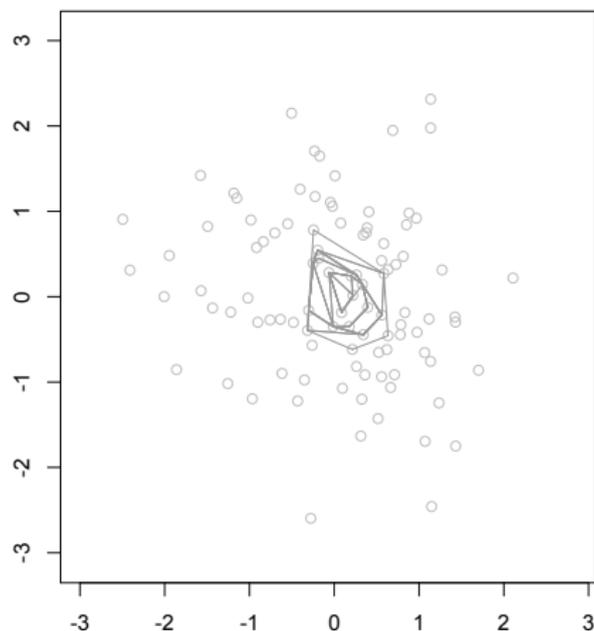
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



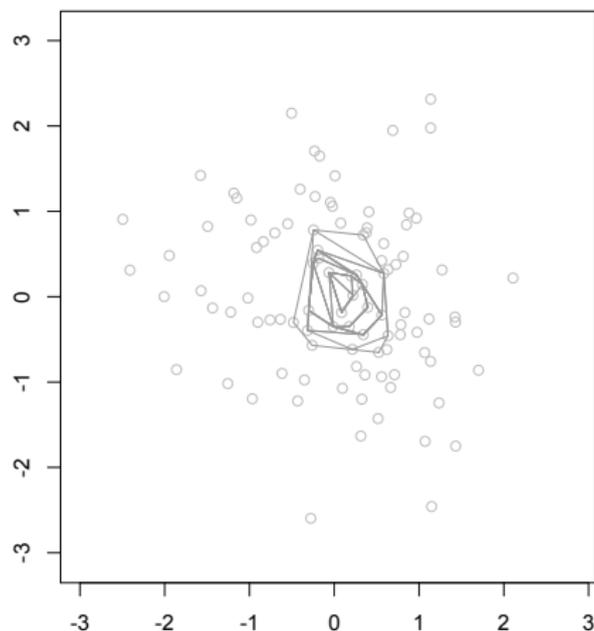
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



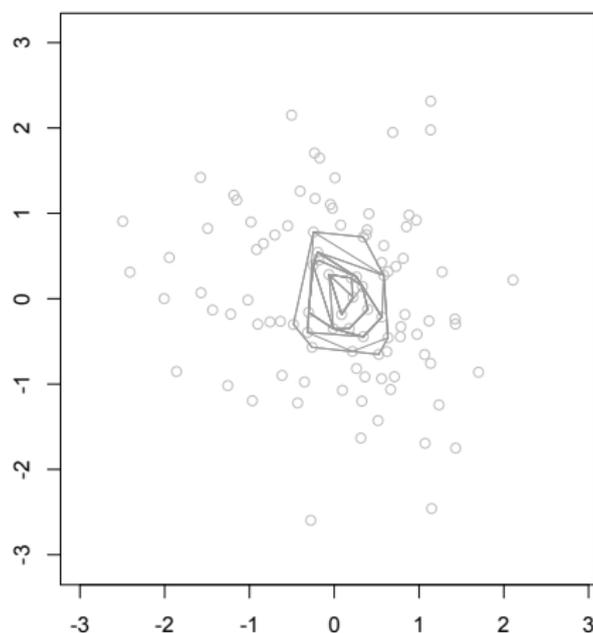
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



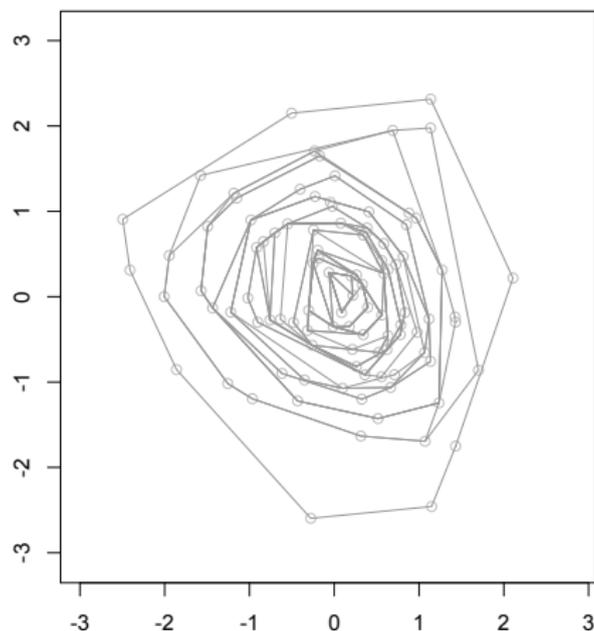
Les régions R_τ sont emboîtées.

On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



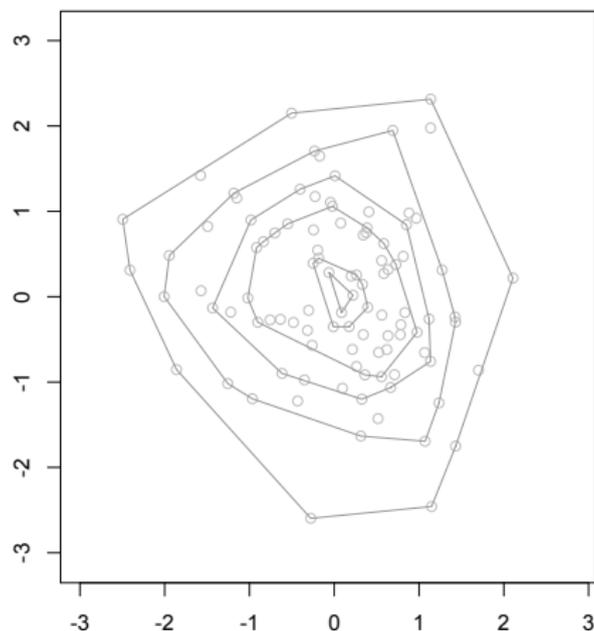
Les régions R_τ sont emboîtées.

On peut considérer les régions de profondeur $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$

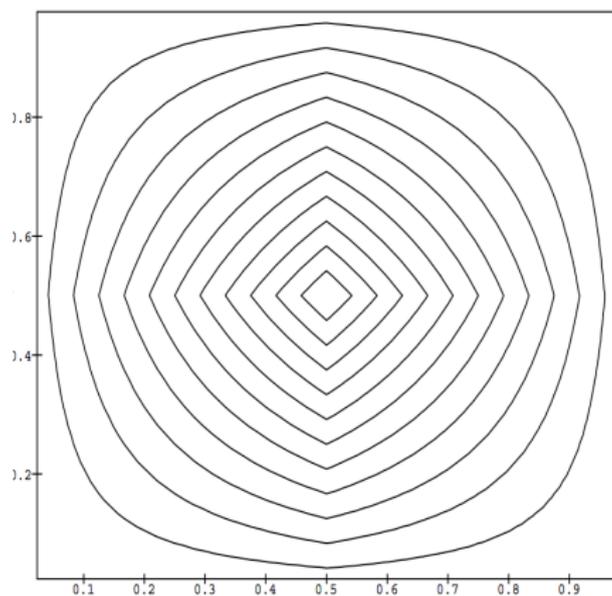


Les régions R_τ sont emboîtées.

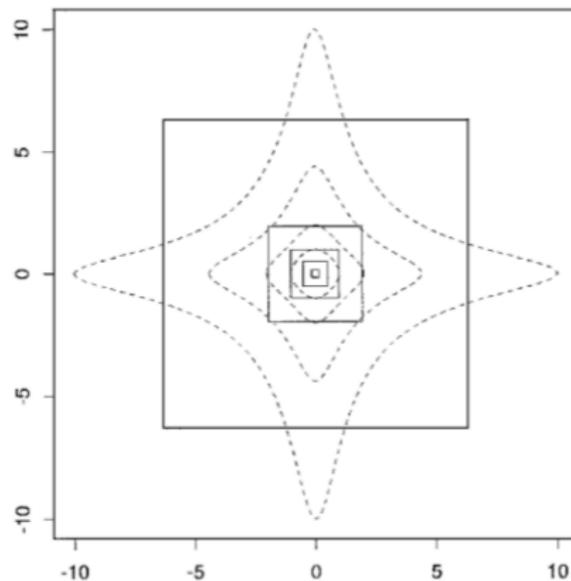
On peut considérer les **régions de profondeur** $R_\tau(P) \stackrel{\text{def}}{=} \{\theta \in \mathbb{R}^d : D(\theta, P) \geq \tau\}$



Les régions R_τ sont emboîtées.

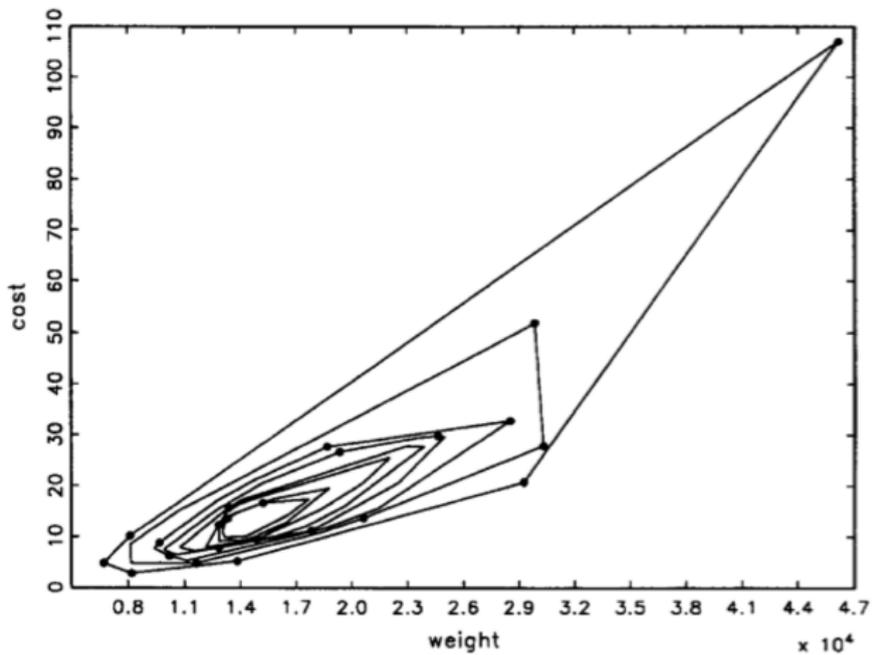


Quelques régions $R_\tau(P)$, pour $P =$ la distribution uniforme sur $[0, 1]^2$



Quelque régions $R_\tau(P)$, pour P = la distribution avec des marginales indépendantes de Cauchy

La profondeur de demi-espace



Théorème

Soit P une mesure de proba. Alors $\theta \rightarrow D(\theta, P)$ est *quasi-concave*.

$f : \mathbb{R}^d \rightarrow \mathbb{R}$ quasi-concave $\stackrel{\text{def}}{\Leftrightarrow} \forall \theta_1, \theta_2 \in \mathbb{R}^d, \forall \lambda \in (0, 1), f((1 - \lambda)\theta_1 + \lambda\theta_2) \geq \min(f(\theta_1), f(\theta_2))$

Corollaire

Soit P une mesure de proba. Alors, pour tout $\tau > 0$, $R_\tau(P)$ est *convexe*.

Preuve : soient $\theta_1, \theta_2 \in R_\tau(P)$ (donc $D(\theta_i, P) \geq \tau, i = 1, 2$).

Alors on obtient directement

$$D((1 - \lambda)\theta_1 + \lambda\theta_2, P) \geq \min(D(\theta_1, P), D(\theta_2, P)) \geq \tau,$$

ce signifie que $\theta \in R_\tau(P)$.

□

Les régions de profondeur sont donc toujours convexes (et emboîtées).

Le résultat suivant est la clé vers une meilleure compréhension de ces régions.

Théorème

Soit P une mesure de proba.

Alors, pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] > 1 - \tau\}$.

Les régions de profondeur sont donc toujours convexes (et emboîtées).

Le résultat suivant est la clé vers une meilleure compréhension de ces régions.

Théorème

Soit P une mesure de proba.

Alors, pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] > 1 - \tau\}$.

On obtient alors directement

Corollaire

Soit P une mesure de proba.

Alors, pour tout $\tau > 0$, $R_\tau(P)$ est fermée.

Théorème

Soit P une mesure de proba.

Alors, pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] > 1 - \tau\}$.

Corollaire

Soit P une mesure de proba admettant une densité. Alors,

(i) pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] \geq 1 - \tau\}$.

(ii) pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] = 1 - \tau\}$.

Preuve : (i) si P admet une densité, $P[\partial H] = 0$ pour tout demi-espace H . Donc les deux intersections sont sur les mêmes demi-espaces...

(ii) Clair !

□

Théorème

Soit P une mesure de proba admettant une densité. Alors,
Alors, pour tout $\tau > 0$, $R_\tau(P) = \cap \{H : H \text{ demi-espace fermé avec } P[H] = 1 - \tau\}$.

Pour $d = 1$ (avec P étant de loi F, f) :

$$D(\theta, P) = \min(F(\theta), 1 - F(\theta))$$



$$R_\tau(P) = [F^{-1}(\tau), F^{-1}(1 - \tau)]$$



Théorème

Soit P une mesure de proba admettant une densité. Alors,
Alors, pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H : H \text{ demi-espace fermé avec } P[H] = 1 - \tau\}$.

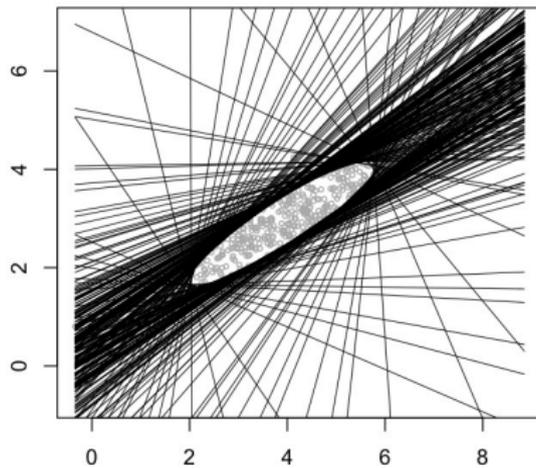
Corollaire (2010)

Soit P une mesure de proba admettant une densité. Alors,
Alors, pour tout $\tau > 0$, $R_\tau(P) = \bigcap \{H_{\tau\mathbf{u}}^+ : \mathbf{u} \in S^{d-1}\}$

Preuve :

(\subset) trivial car $R_\tau(P)$ est une intersection sur plus de demi-espaces fermés.

(\supset) Tout demi-espace fermé avec $P[H] = 1 - \tau$ est un $H_{\tau\mathbf{u}}^+$ (il suffit de prendre \mathbf{u} dans la direction joignant le barycentre de $H_{\tau\mathbf{u}}^+$ à celui de $(H_{\tau\mathbf{u}}^+)^c$. \square



Quelques propriétés supplémentaires, dans le désordre :

- Pour tout $\tau > 0$, les régions $R_\tau(P)$ sont aussi bornées, donc compactes.
- Le point le plus profond $\theta^* = \arg \max_{\theta \in \mathbb{R}^d} D(\theta, P)$ joue le rôle de médiane multivariée (puisque pour $d = 1$, c'est la médiane usuelle).
- Pour tout P , $\frac{1}{d+1} \leq D(\theta^*, P) \leq \frac{1}{2} + \frac{1}{2} \max_{\theta \in \mathbb{R}^d} P[\{\theta\}]$.
- la profondeur de demi-espace caractérise les $P^{(n)}$ (1999), les distributions P ayant une densité à support compact (2003), les distributions P ayant des contours lisses (2010). Toutes les distributions ???
- $D(\mathbf{A}\theta + \mathbf{b}, P^{\mathbf{A}\mathbf{Y} + \mathbf{b}}) = D(\theta, P^{\mathbf{Y}})$ pour toute matrice $d \times d$ inversible \mathbf{A} , tout d -vecteur \mathbf{b} et toute distribution $P = P^{\mathbf{Y}}$ (invariance affine).

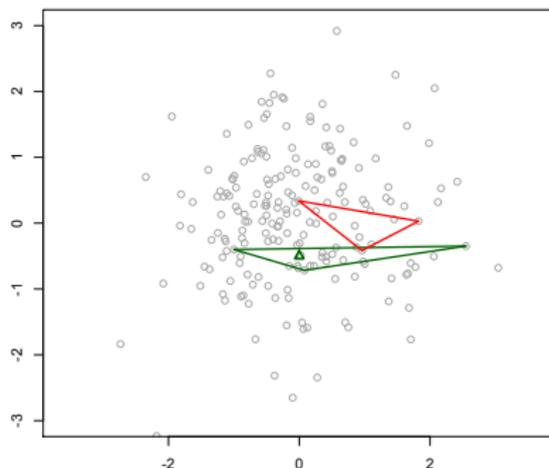
2 Profondeur statistique

- La profondeur de demi-espace
- Un zoo de profondeurs
- Classification, et extension locale de la profondeur

En notant de nouveau $C_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$, $\mathbf{y}_i \in \mathbb{R}^d$, la profondeur simpliciale (1990)

$$D_S(\boldsymbol{\theta}, C_n) := \frac{1}{\binom{n}{d+1}} \sum_{i_1 < i_2 < \dots < i_{d+1}} \mathbb{I}[\boldsymbol{\theta} \in \text{Simpl}(\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \dots, \mathbf{y}_{i_{d+1}})],$$

où $\text{Simpl}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{d+1})$ désigne le simplexe de sommets $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{d+1}$.



La profondeur spatiale (2000)

$$D_{Sp}(\boldsymbol{\theta}, C_n) = 1 - O_{Sp}(\boldsymbol{\theta}, C_n), \text{ avec } O_{Sp}(\boldsymbol{\theta}, C_n) = \left\| \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{y}_i - \boldsymbol{\theta}}{\|\mathbf{y}_i - \boldsymbol{\theta}\|} \right\|.$$

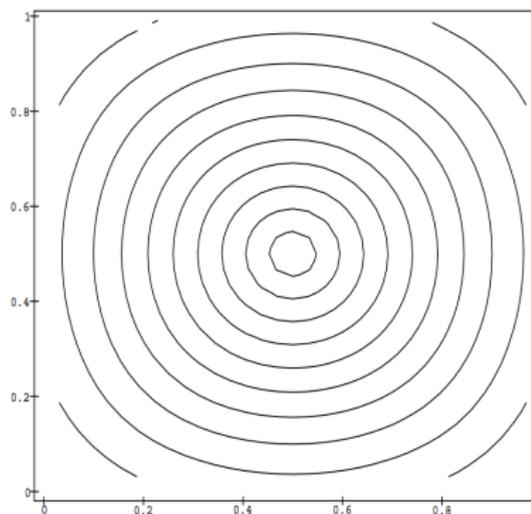
La quantité $O_{Sp}(\boldsymbol{\theta}, C_n)$ est une mesure "d'outlyingness" de $\boldsymbol{\theta}$ par rapport à C_n .

Le point le plus profond est le $\boldsymbol{\theta}_0$ tel que $O_{Sp}(\boldsymbol{\theta}_0, C_n) = 0$. Il est appelé "médiane spatiale" (il généralise encore la médiane univariée).

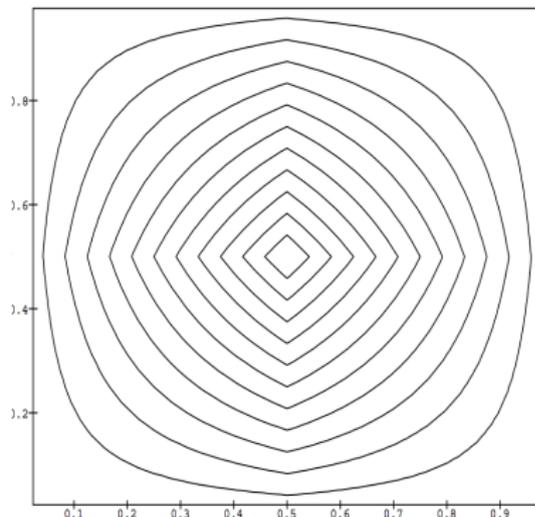
L'unicité en dimension $d > 1$ a été prouvée en 1987 dans un article d'une page et demi dans la meilleure revue de statistique mathématique.

Cette profondeur n'est pas invariante sous transformations affines.

Profondeur spatiale



Profondeur de demi-espace

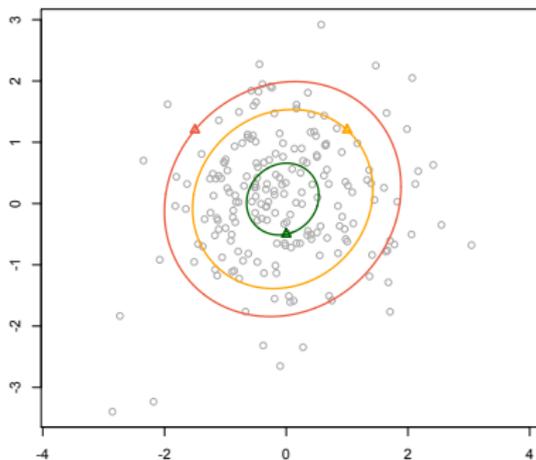


Quelques régions $R_\tau(P)$, pour $P =$ la distribution uniforme sur $[0, 1]^2$

La profondeur de Mahalanobis

$$D_M(\boldsymbol{\theta}, C_n) = \frac{1}{1 + dist}, \text{ avec } dist = \sqrt{(\boldsymbol{\theta} - \bar{\mathbf{y}}_n)' S_n^{-1} (\boldsymbol{\theta} - \bar{\mathbf{y}}_n)},$$

$$\text{où } \bar{\mathbf{y}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i \text{ et } S_n = \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}}_n)(\mathbf{y}_i - \bar{\mathbf{y}}_n)'$$



Une approche axiomatique

Une profondeur devrait satisfaire les propriétés suivantes :

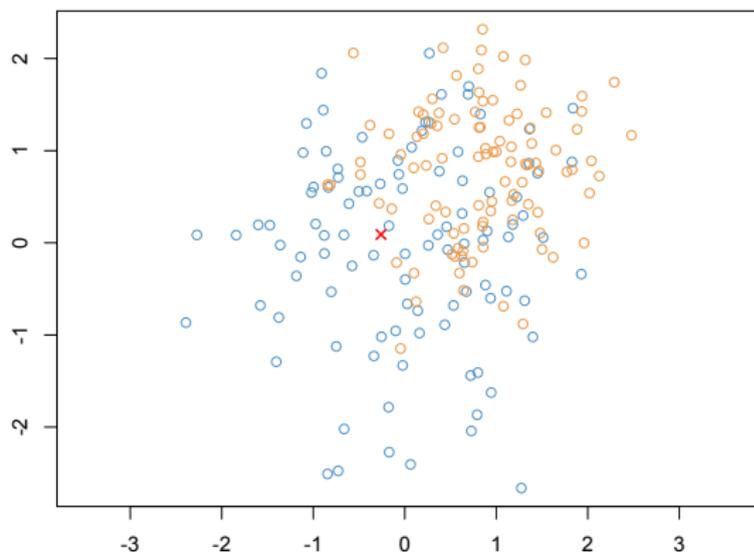
- (P1) *affine-invariance* : $D(\mathbf{A}\boldsymbol{\theta} + \mathbf{b}, P^{\mathbf{A}\mathbf{Y} + \mathbf{b}}) = D(\boldsymbol{\theta}, P^{\mathbf{Y}})$ pour toute matrice $d \times d$ inversible \mathbf{A} , tout d -vecteur \mathbf{b} et toute distribution $P = P^{\mathbf{Y}}$;
- (P2) *maximalité au centre* : $D(\boldsymbol{\theta}_0, P) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^d} D(\boldsymbol{\theta}, P)$ pour toute P symétrique par rapport à $\boldsymbol{\theta}_0$;
- (P3) *monotonicity relative to the deepest point* : pour toute P ayant $\boldsymbol{\theta}_0$ comme point le plus profond et tout $\mathbf{u} \in \mathbb{R}^d$, $r \mapsto D(\boldsymbol{\theta}_0 + r\mathbf{u}, P)$ est décroissante en $r \geq 0$;
- (P4) *"vanishing at infinity"* : pour toute P , $D(\boldsymbol{\theta}, P) \rightarrow 0$ si $\|\boldsymbol{\theta}\| \rightarrow \infty$.

2 Profondeur statistique

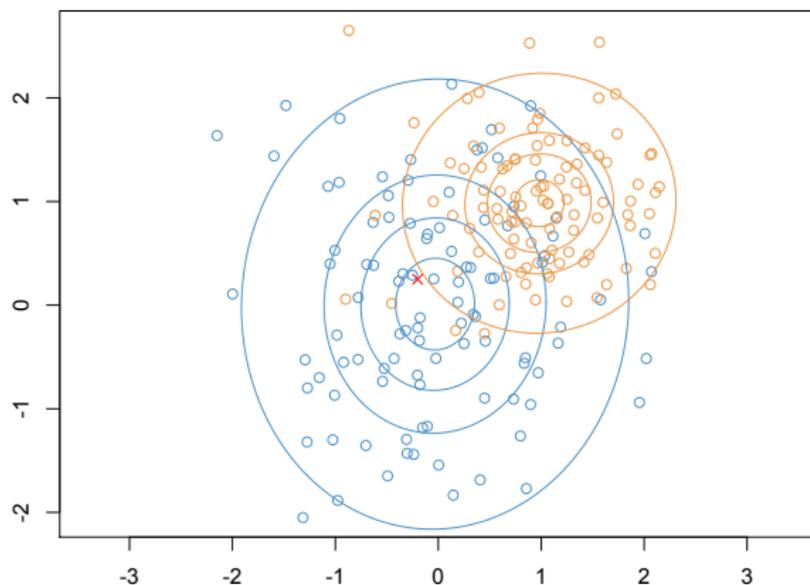
- La profondeur de demi-espace
- Un zoo de profondeurs
- Classification, et extension locale de la profondeur

Classification, et extension locale de la profondeur

Soit (\mathbf{X}, Y) , où $\mathbf{X} \in \mathbb{R}^d$ et $Y \in \{0, 1\}$. Si on observe que la partie \mathbf{x} d'une réalisation (\mathbf{x}, y) de (\mathbf{X}, Y) , comment faire une "prédiction" $m(\mathbf{x})$ du y correspondant ?



Classification, et extension locale de la profondeur



La procédure classique : $m(x) = \mathbb{I}[d_{\Sigma_1}(\mathbf{x}, \mu_1) < d_{\Sigma_0}(\mathbf{x}, \mu_0)]$ (\leadsto le point rouge est plus proche du μ bleu que du μ orange \Rightarrow classification en bleu)

Sous des conditions "elliptiques", cette procédure classique se réécrit

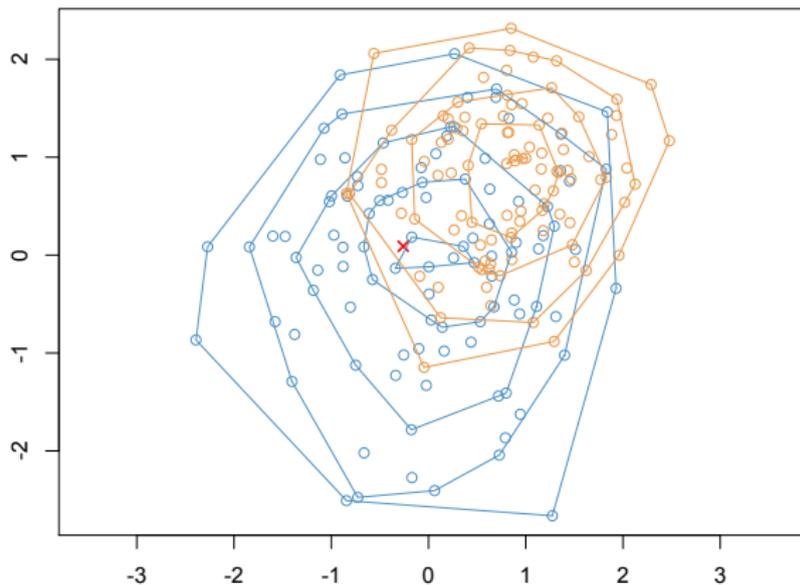
$$\begin{aligned} m(\mathbf{x}) &= \mathbb{I}\left[d_{\Sigma_1}(\mathbf{x}, \boldsymbol{\mu}_1) < d_{\Sigma_0}(\mathbf{x}, \boldsymbol{\mu}_0)\right] \\ &= \mathbb{I}\left[D_M(\mathbf{x}, P_1) > D_M(\mathbf{x}, P_0)\right], \end{aligned}$$

ce qui suggère de considérer une procédure (2005→ 2013)

$$m(\mathbf{x}) = \mathbb{I}\left[D(\mathbf{x}, P_1) > D(\mathbf{x}, P_0)\right],$$

fondée sur une profondeur D quelconque ("max-depth approach").

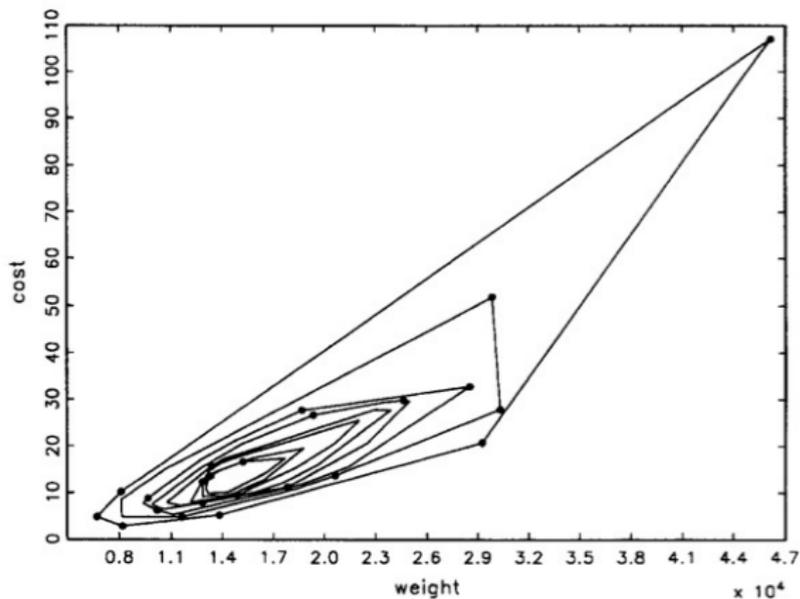
Classification, et extension locale de la profondeur



Le point rouge est plus profond dans la population bleue que dans la population orange \Rightarrow classification en bleu

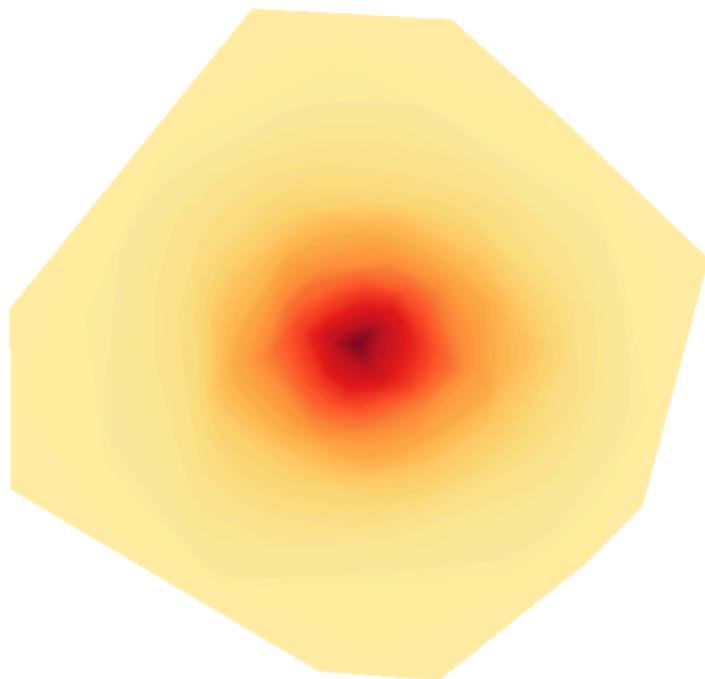
Cette approche est cependant peu efficace dans les situations où la profondeur échoue à refléter la centralité correctement...

Les régions de profondeur ne fournissent pas un résumé de P .
Elles "révèlent" la structure de P .

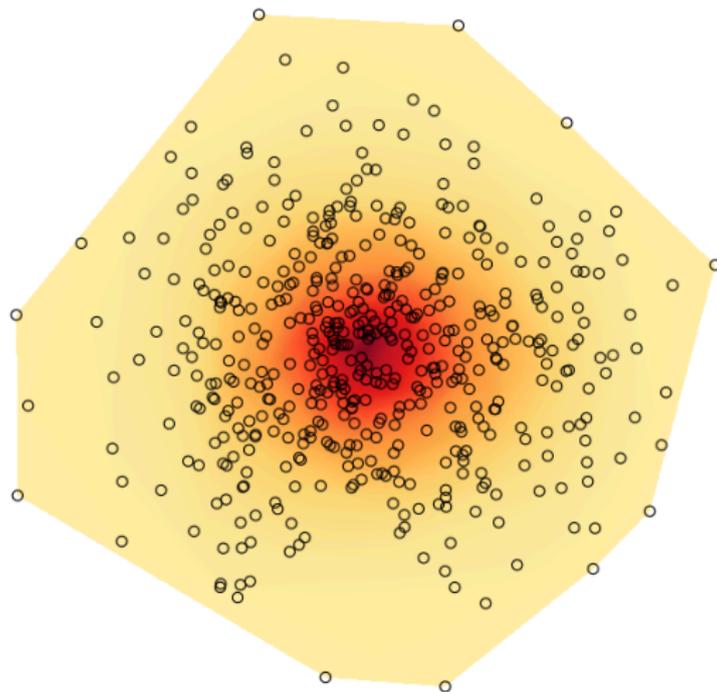


Contours of depths 1 to 8, in a plot of the weight and the cost of 23 single-engine aircraft.

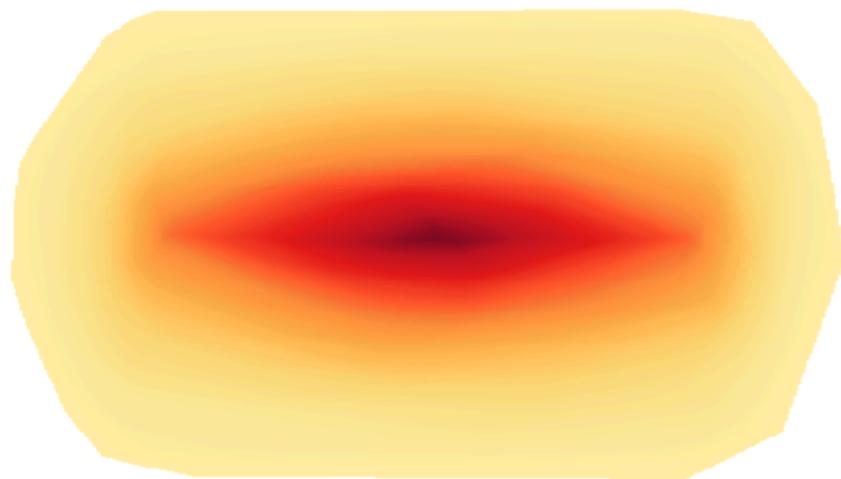
Les régions de profondeur révèlent la structure de P



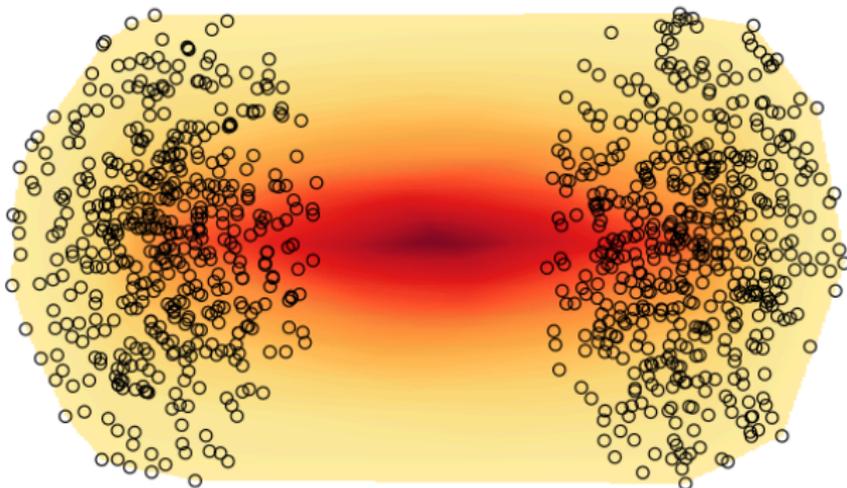
Les régions de profondeur révèlent la structure de P



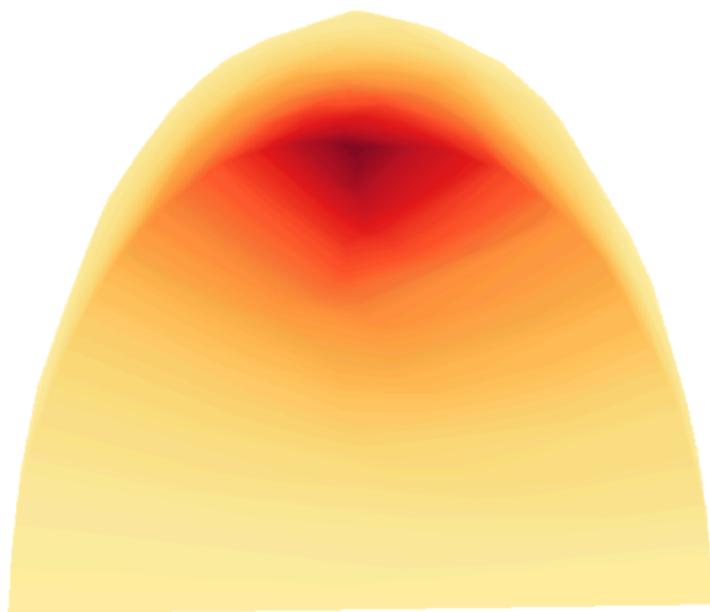
Mais il y a des cas importants où ceci n'a pas lieu...



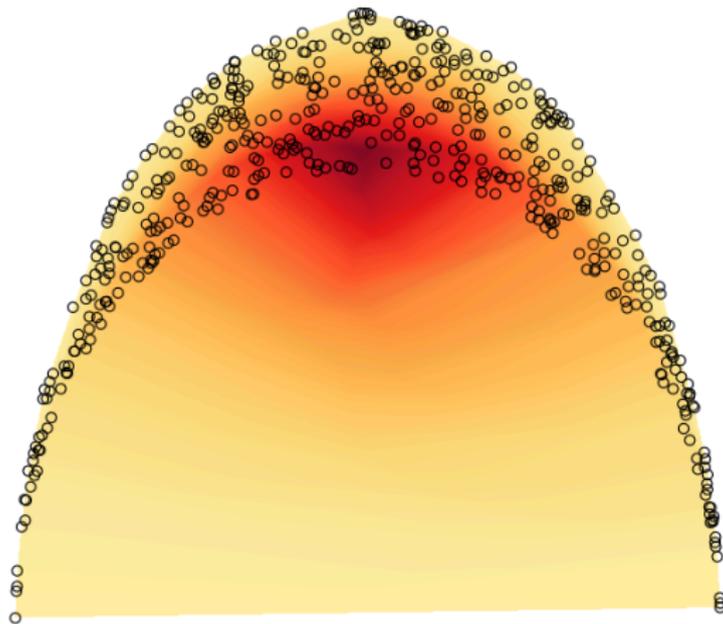
Mais il y a des cas importants où ceci n'a pas lieu...



Mais il y a des cas importants où ceci n'a pas lieu...



Mais il y a des cas importants où ceci n'a pas lieu...



Ceci motive qu'on étende la notion de profondeur pour pouvoir traiter

- les distributions multi-modales
- les distributions à support non-convexe.

L'idée est de mesurer une **centralité locale** au lieu de la centralité globale

~> **Profondeur locale !**

Définition (2011)

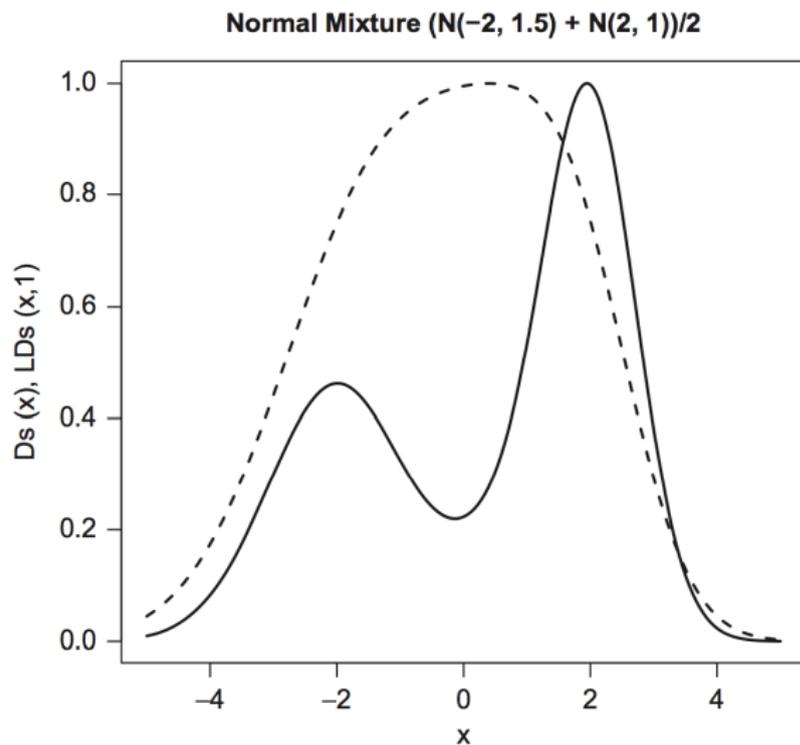
Pour tout $\alpha > 0$, la profondeur de demi-espace α -locale $LD_H(\cdot; P, \alpha)$ envoie θ sur

$$LD_H(\theta; P^{\mathbf{Y}}, \alpha) = \inf_{\mathbf{u} \in S^{d-1}} P[0 \leq \mathbf{u}'(\mathbf{Y} - \theta) \leq \alpha].$$

Définition (2011)

Pour tout $\alpha > 0$, la profondeur simpliciale α -locale $LD_S(\cdot; P, \alpha)$ envoie θ sur la probabilité qu'un simplexe aléatoire de volume au plus α recouvre θ .

Clairement, on retrouve les concepts globaux pour $\alpha \rightarrow \infty$.



Ces profondeurs locales sont d'une nature locale :

Si P admet une densité f , on a

$$\frac{1}{\alpha} LD_H(\boldsymbol{\theta}; P, \alpha) \rightarrow f(\boldsymbol{\theta}),$$

et

$$\frac{1}{\alpha} \sqrt{LD_S(\boldsymbol{\theta}; P, \alpha)} \rightarrow f(\boldsymbol{\theta}),$$

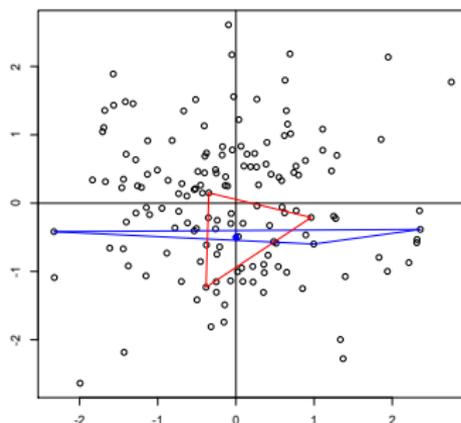
quand $\alpha \rightarrow 0$.

Donc, ces profondeurs fournissent un continuum entre

- des mesures de centralité globales (profondeurs), obtenues pour $\alpha \rightarrow \infty$, et
- des mesures de masses de proba (densités), obtenues pour $\alpha \rightarrow 0$.

Or la profondeur et la densité ne sont pas de la même nature.

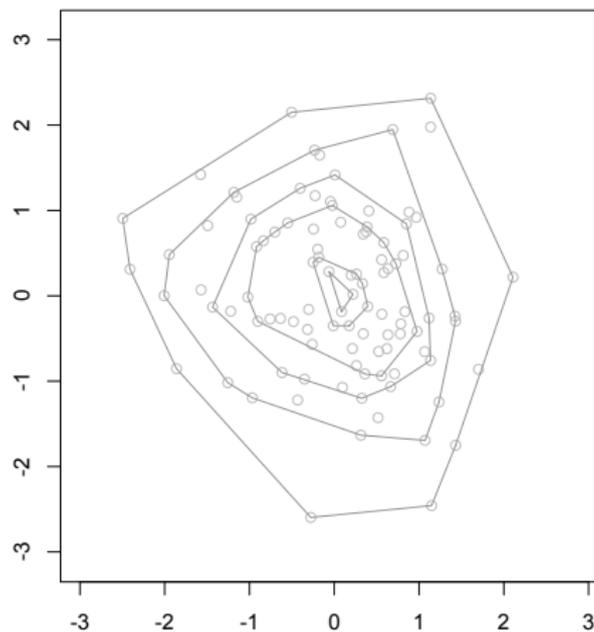
- Ce lien non désirable avec la densité
- La localisation pas si évidente



- Le manque d'affine-invariance
- Le fait que chaque profondeur D requiert une localisation ad hoc
- ... donc des implémentations algorithmiques séparées

motivent une autre approche...

Notre approche s'appuie sur des voisinages construits en exploitant l'approche axiomatique

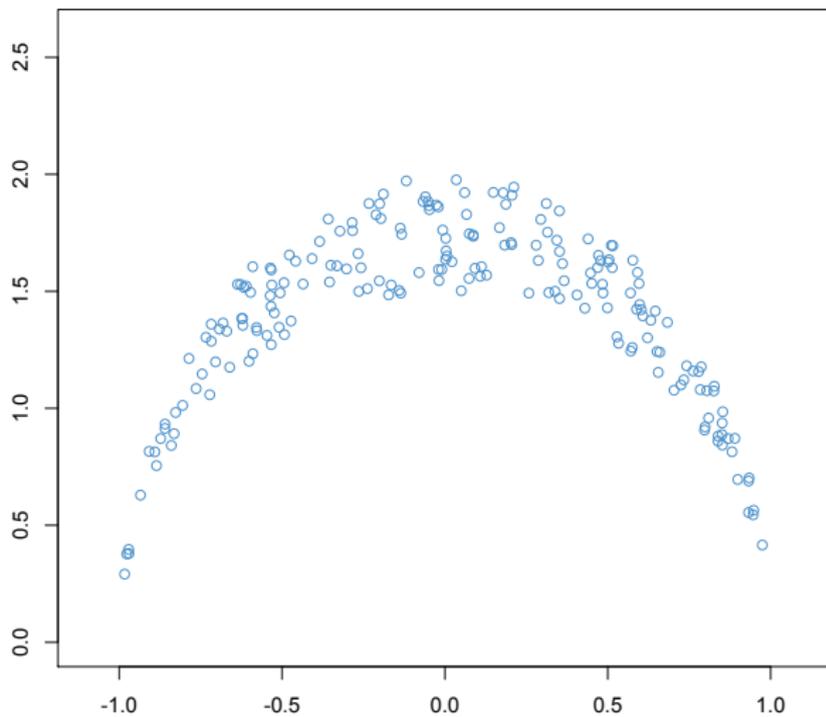


Cette autre approche s'appuie sur des voisinages construits en exploitant l'approche axomatique

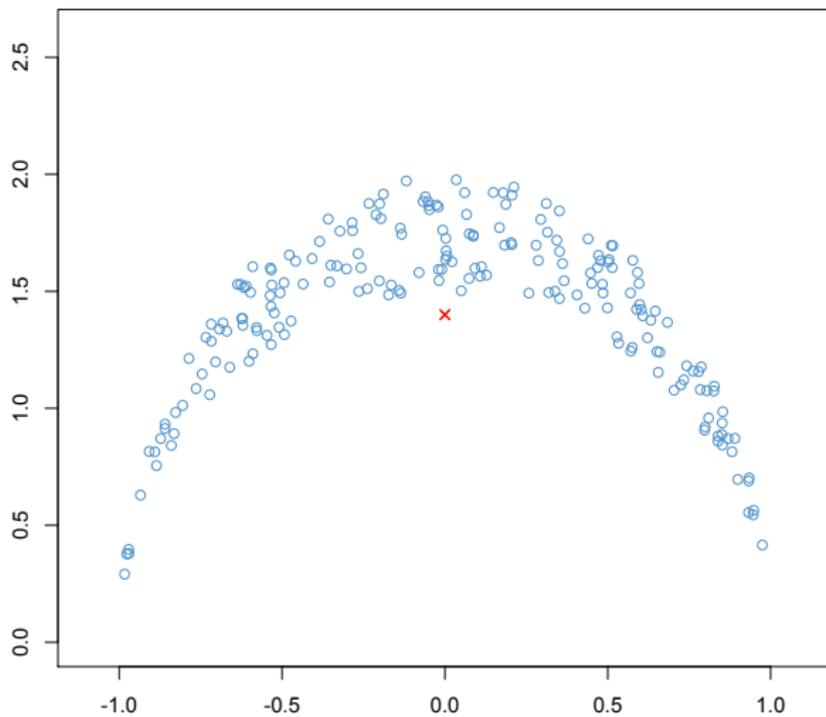
Une profondeur devrait satisfaire les propriétés suivantes :

- (P1) *affine-invariance* : $D(\mathbf{A}\boldsymbol{\theta} + \mathbf{b}, P^{\mathbf{A}\mathbf{Y}+\mathbf{b}}) = D(\boldsymbol{\theta}, P^{\mathbf{Y}})$ pour toute matrice $d \times d$ inversible \mathbf{A} , tout d -vecteur \mathbf{b} et toute distribution $P = P^{\mathbf{Y}}$;
- (P2) *maximalité au centre* : $D(\boldsymbol{\theta}_0, P) = \sup_{\boldsymbol{\theta} \in \mathbb{R}^d} D(\boldsymbol{\theta}, P)$ pour toute P symétrique par rapport à $\boldsymbol{\theta}$;
- (P3) *monotonicity relative to the deepest point* : pour toute P ayant $\boldsymbol{\theta}_0$ comme point le plus profond et tout $\mathbf{u} \in \mathbb{R}^d$, $r \mapsto D(\boldsymbol{\theta}_0 + r\mathbf{u}, P)$ est décroissante en $r \geq 0$;
- (P4) *"vanishing at infinity"* : pour toute P , $D(\boldsymbol{\theta}, P) \rightarrow 0$ si $\|\boldsymbol{\theta}\| \rightarrow \infty$.

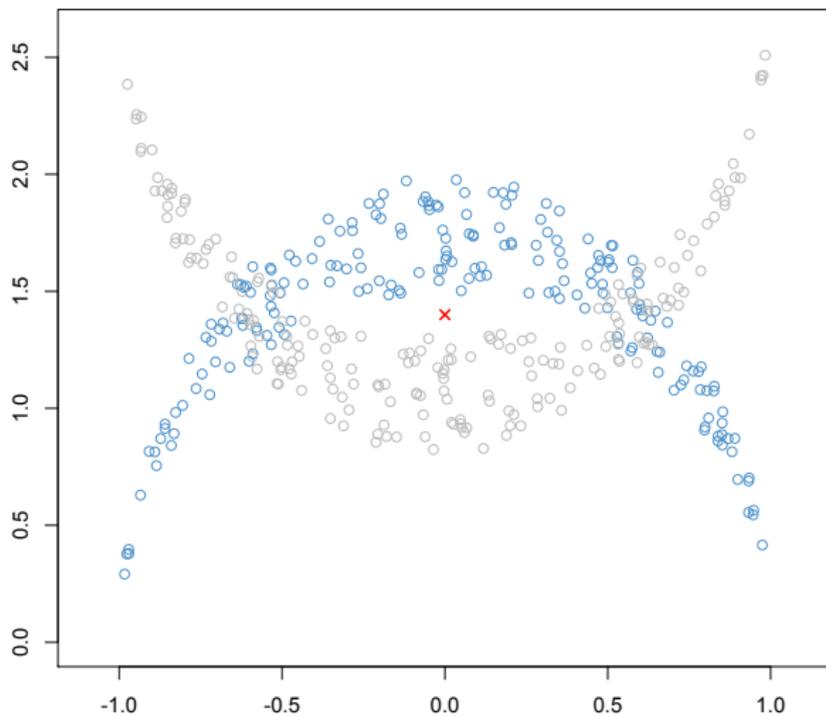
Classification, et extension locale de la profondeur



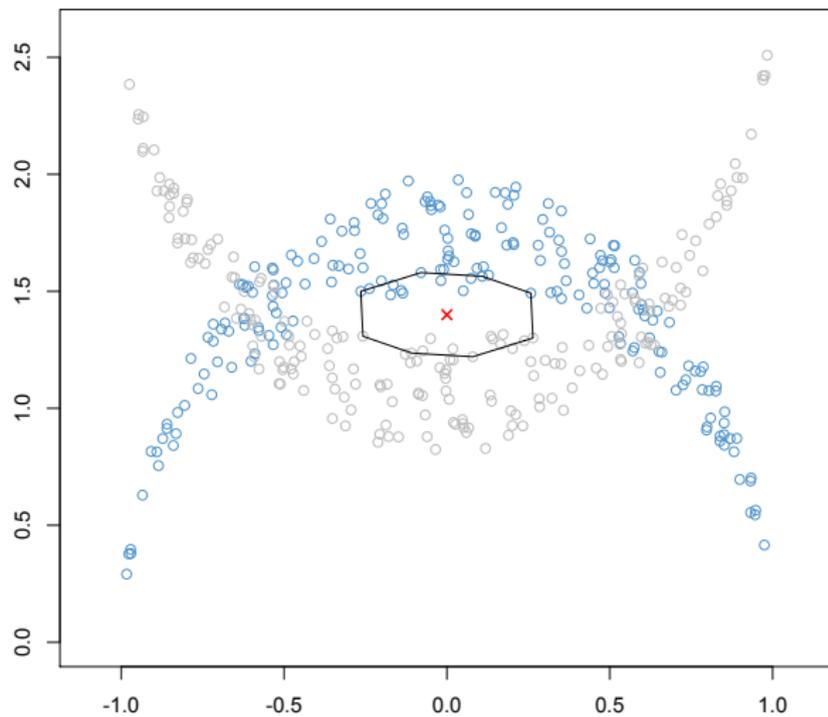
Classification, et extension locale de la profondeur



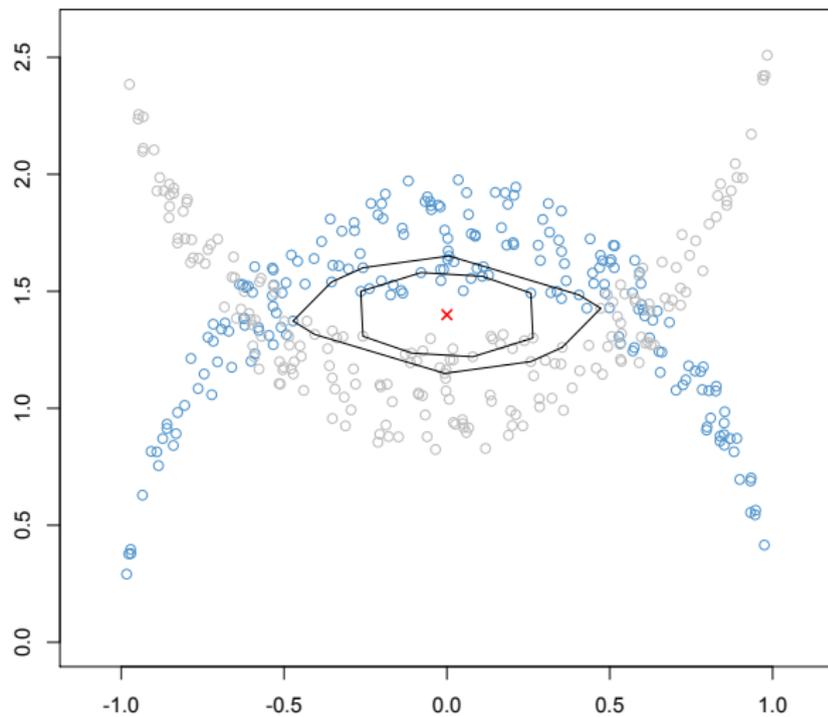
Classification, et extension locale de la profondeur



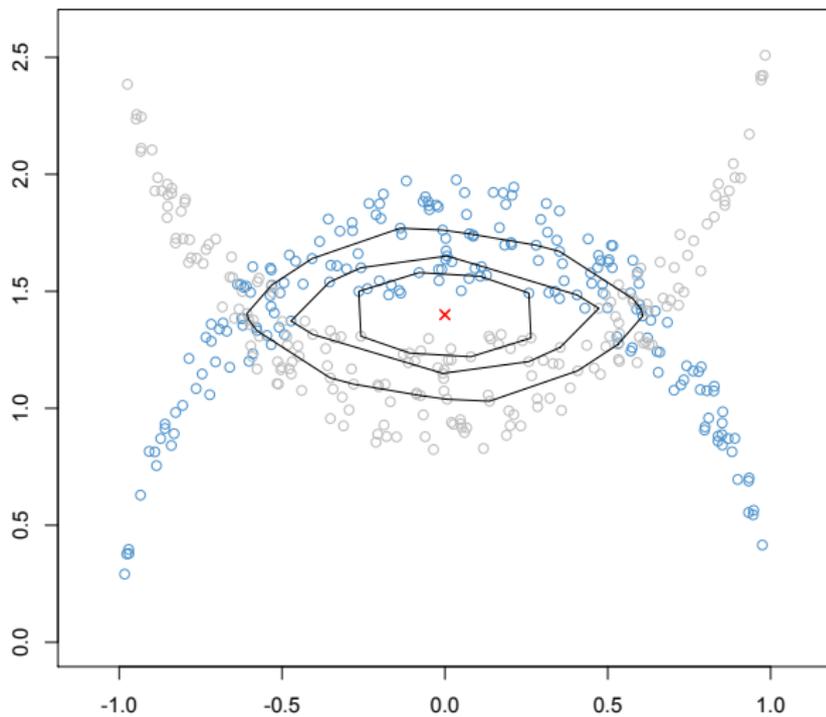
Classification, et extension locale de la profondeur



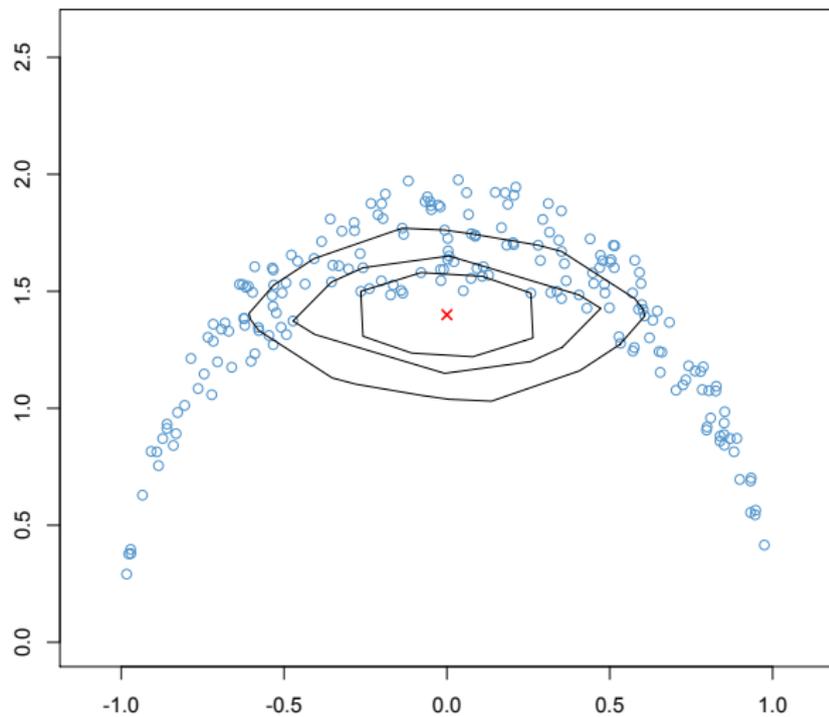
Classification, et extension locale de la profondeur



Classification, et extension locale de la profondeur



Classification, et extension locale de la profondeur



Soit D une fonction de profondeur. Soit $\beta \in (0, 1]$.

Définition

La profondeur locale associée à D , au niveau de localité β , est

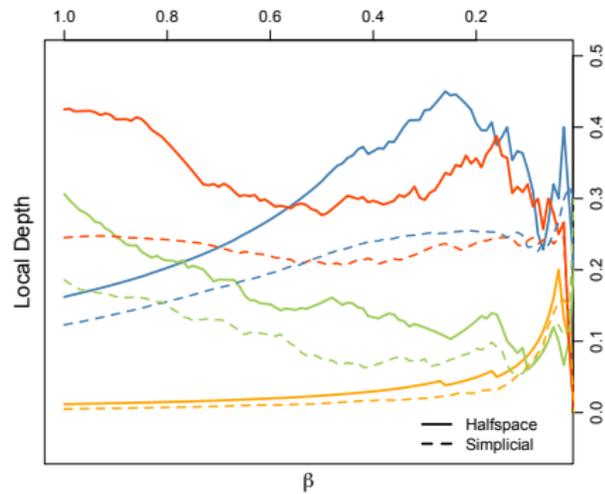
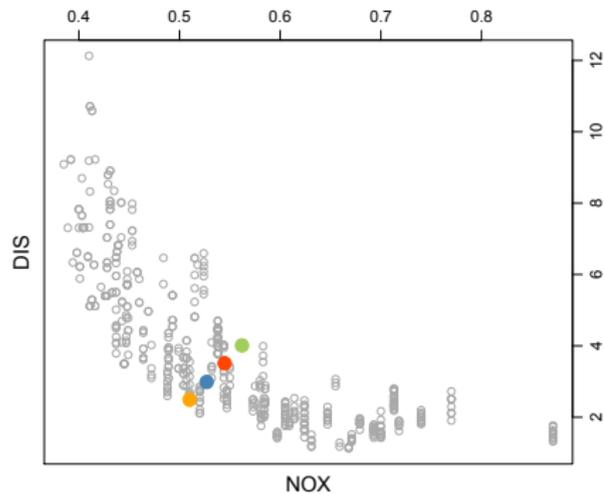
$$\begin{aligned} LD^\beta(\cdot, P) : \mathbb{R}^d &\rightarrow [0, 1] \\ \theta &\mapsto LD^\beta(\theta, P) = D(\theta, P_\theta^\beta), \end{aligned}$$

où $P_\theta^\beta : B \mapsto P_\theta^\beta[B | R_\theta^\beta(P)]$ est la distribution conditionnelle de P , par rapport à la plus petite région de profondeur, dans la symétrisée P_θ de P , contenant une masse de proba $\geq \beta$.

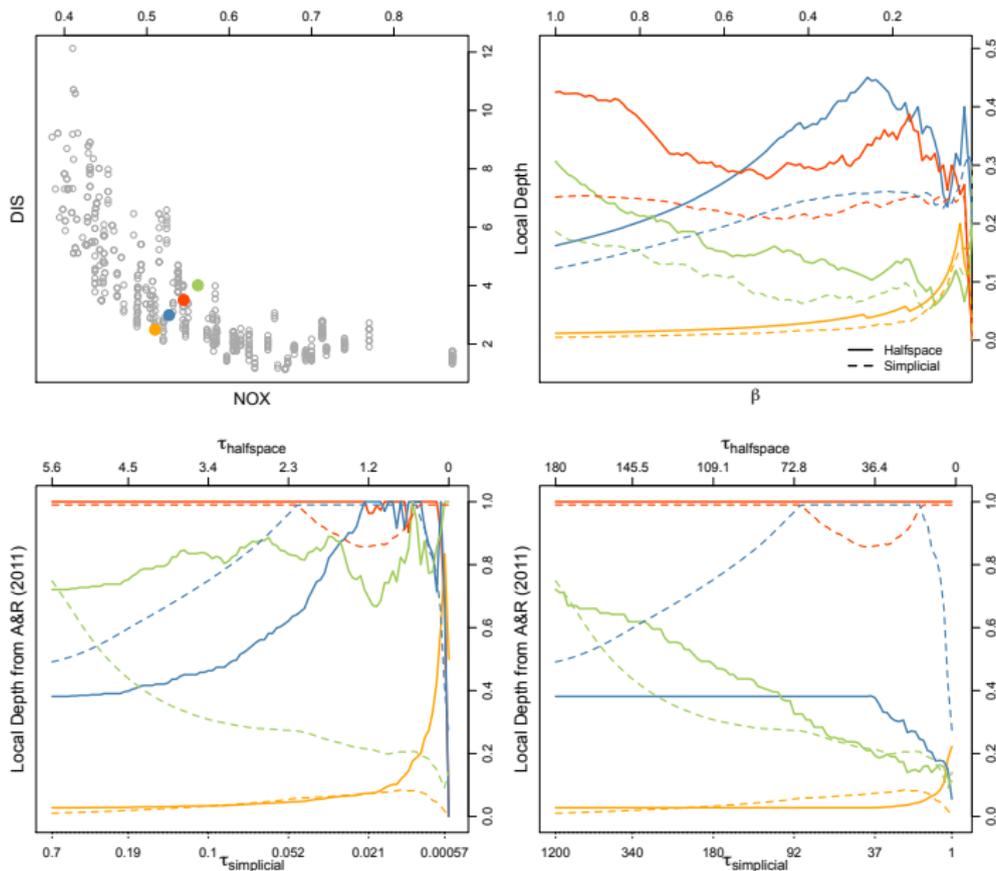
Remarques :

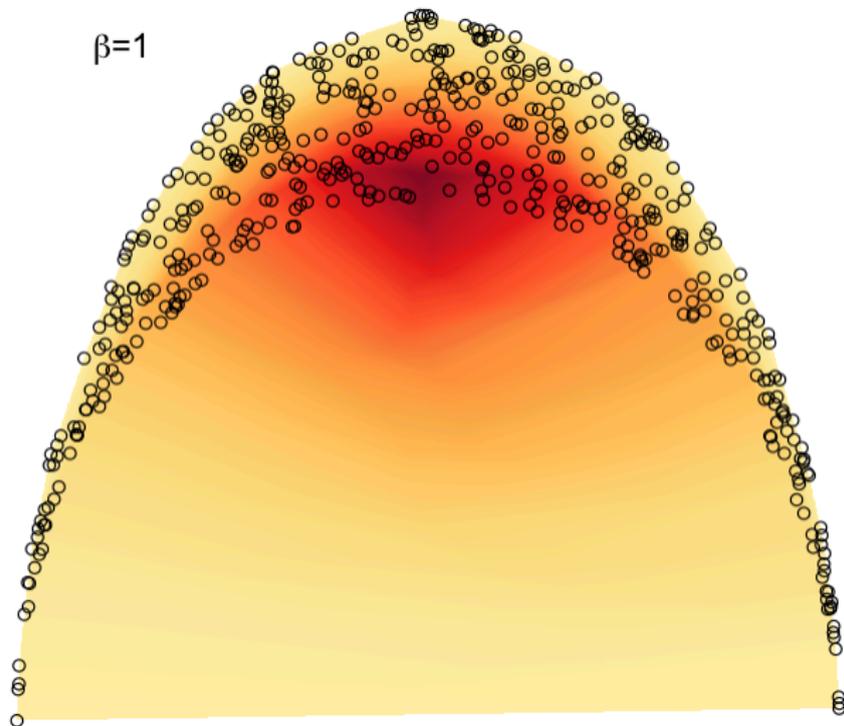
- Pour $\beta = 1$, $LD^\beta(\theta, P) = D(\theta, P)$ (pas de localisation)
- Pour tout β , $LD^\beta(\theta, P) = D(\theta, P_\theta^\beta)$ reste une mesure de centralité
- Pour tout β , $LD^\beta(\cdot, P)$ est affine-invariante
- La localisation s'applique à chaque D de façon homogène

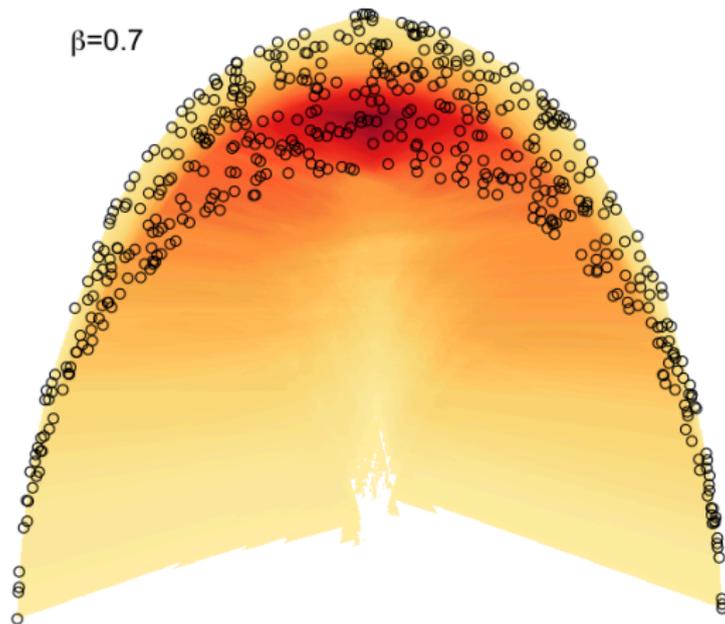
Classification, et extension locale de la profondeur

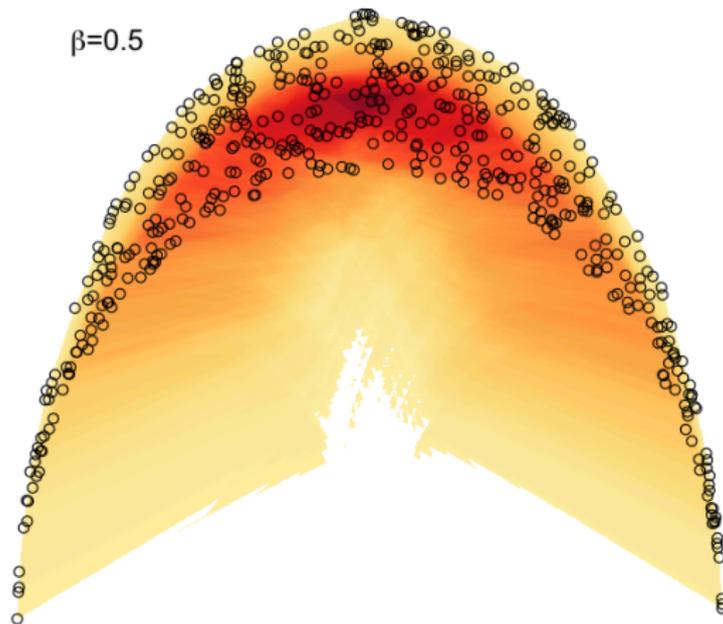


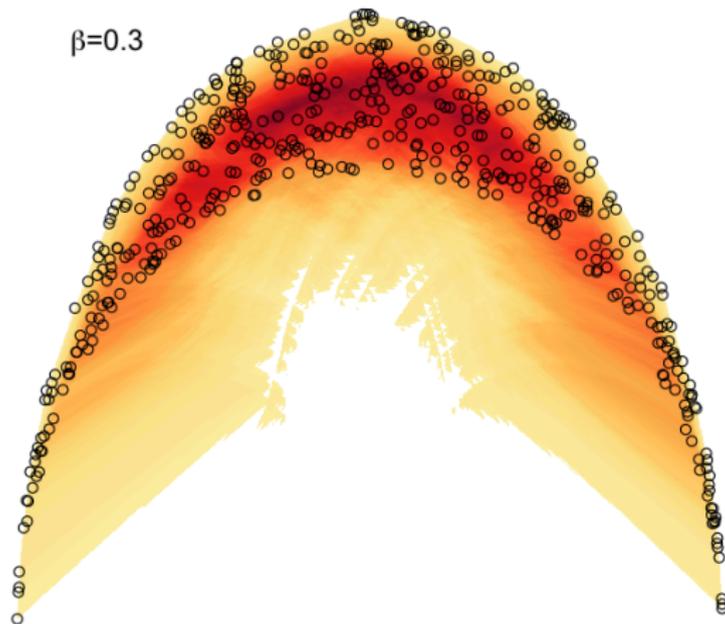
Classification, et extension locale de la profondeur

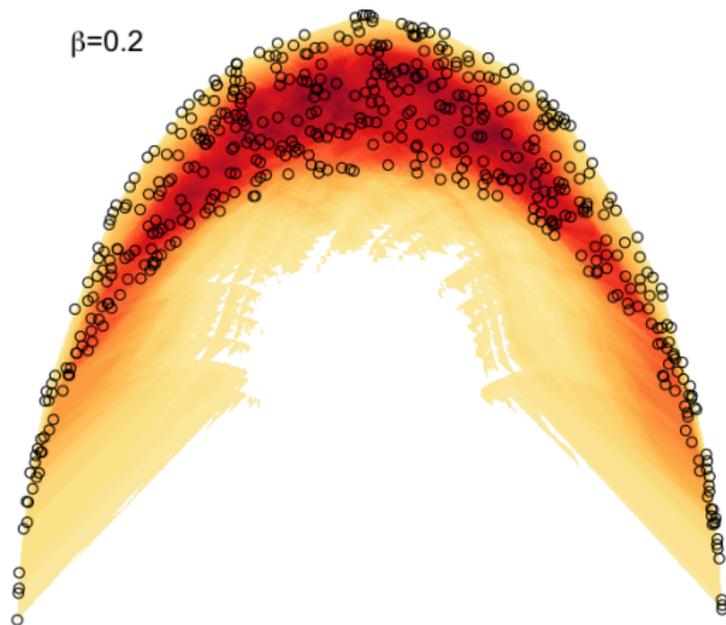


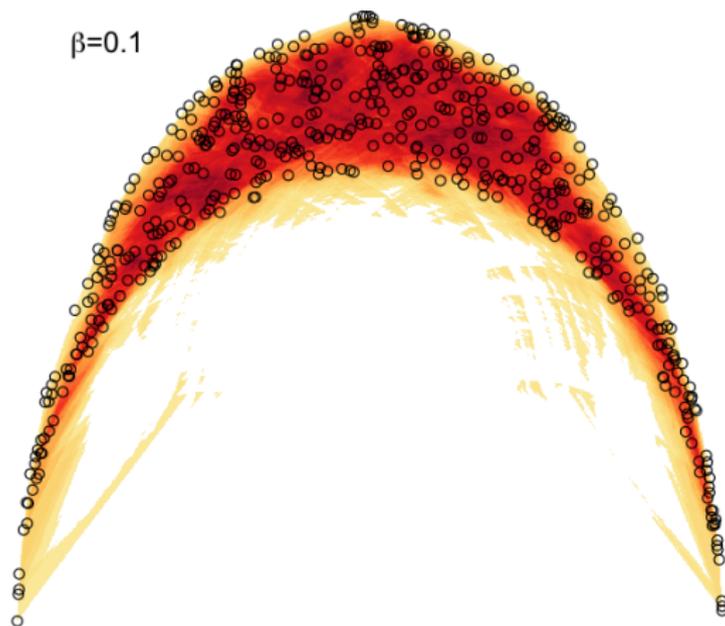






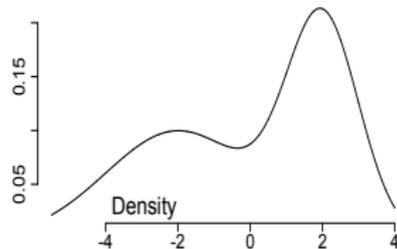
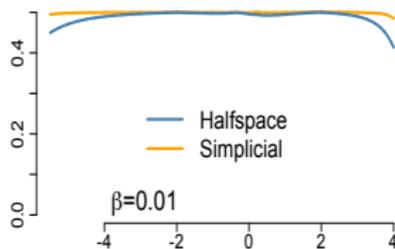
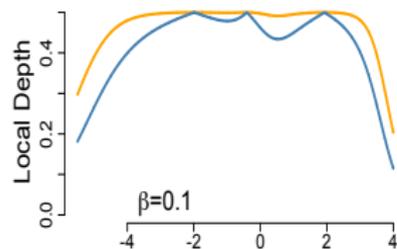
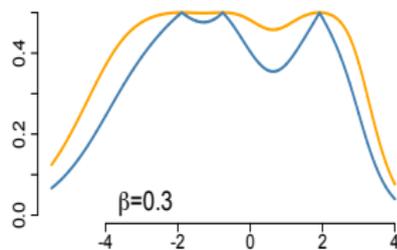
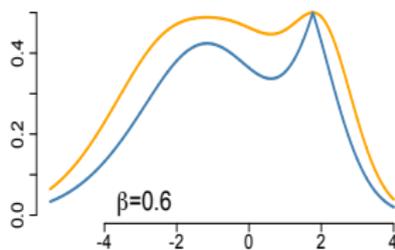
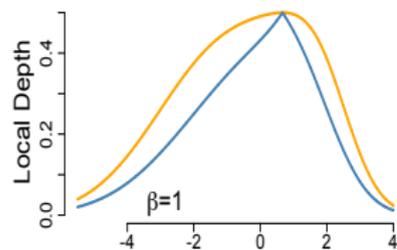




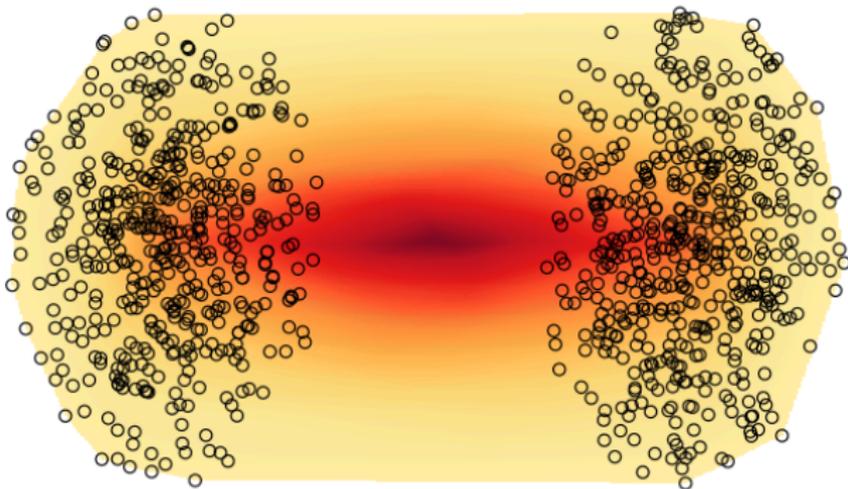


Pour θ dans le support de P , $\lim_{\beta \rightarrow 0} LD^\beta(\theta, P) = c$

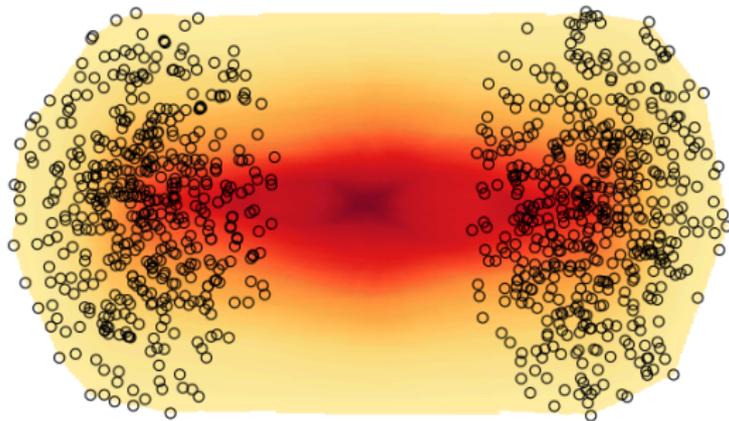
$$Y \sim \frac{1}{2}\mathcal{N}(-2, 2) + \frac{1}{2}\mathcal{N}(2, 1)$$



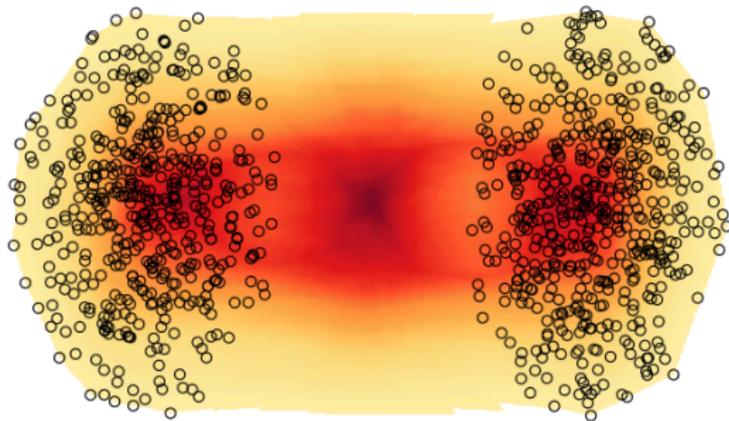
$\beta=1$



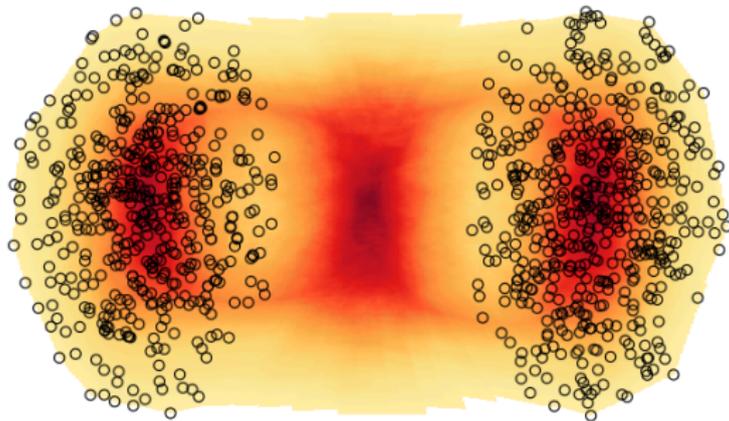
$\beta=0.7$



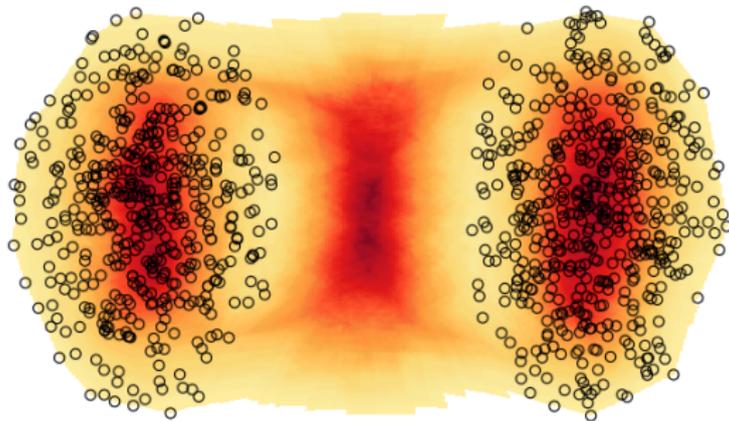
$\beta=0.5$



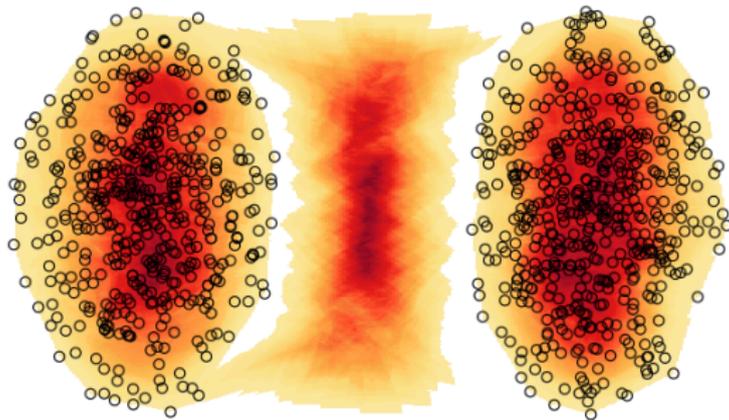
$\beta=0.3$



$\beta=0.2$

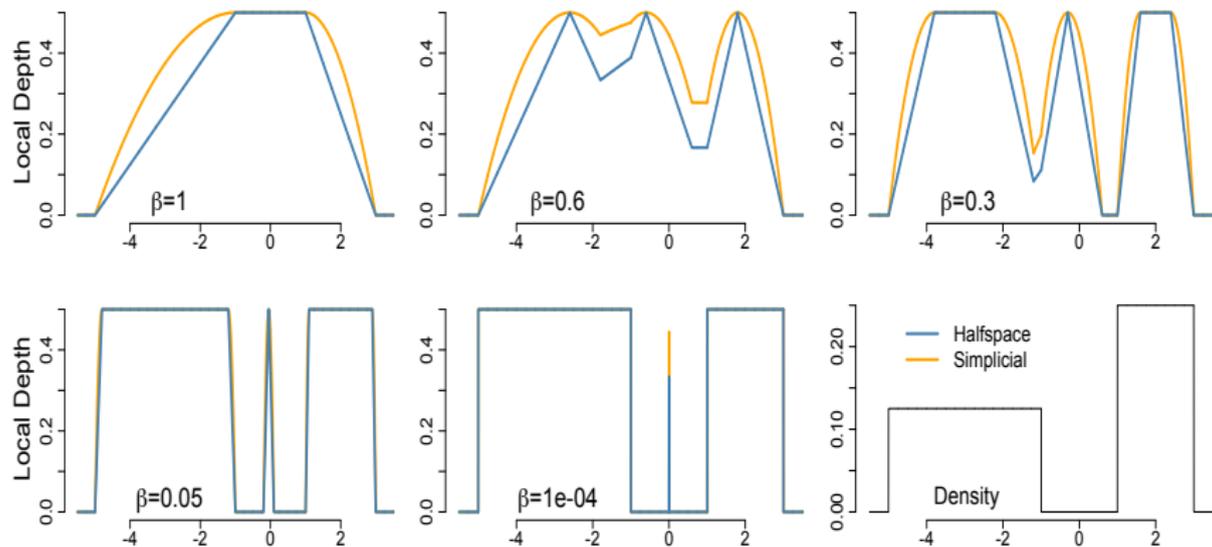


$\beta=0.1$



Classification, et extension locale de la profondeur

$$Y \sim \frac{1}{2}\text{Unif}(-5, -1) + \frac{1}{2}\text{Unif}(1, 3)$$



Application en classification

Puisque la profondeur locale est robuste à la non-convexité, on peut penser à remplacer les **max-depth classifiers**

$$m_D(\mathbf{x}) = \mathbb{I}[D(\mathbf{x}, P_1) > D(\mathbf{x}, P_0)]$$

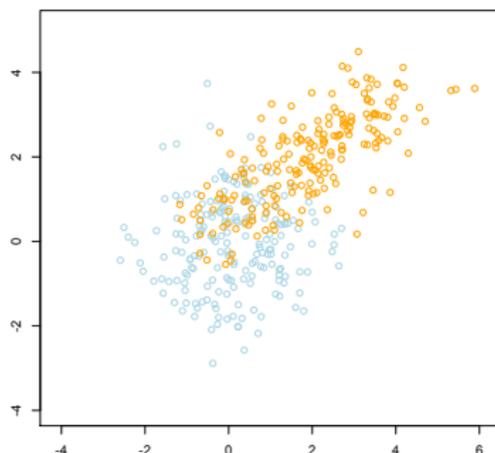
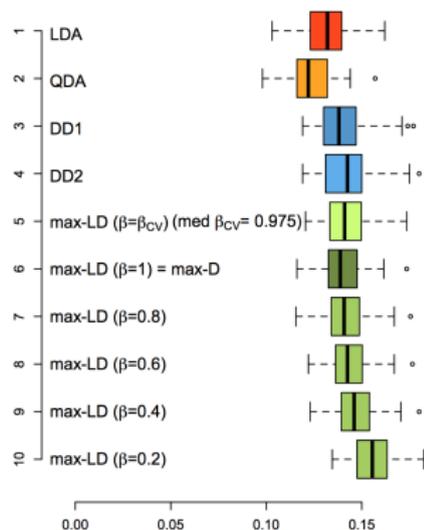
par les **max-local-depth classifiers**

$$m_D^\beta(\mathbf{x}) = \mathbb{I}[LD^\beta(\mathbf{x}, P_1) > LD^\beta(\mathbf{x}, P_0)],$$

pour un certain β bien choisi.

Classification, et extension locale de la profondeur

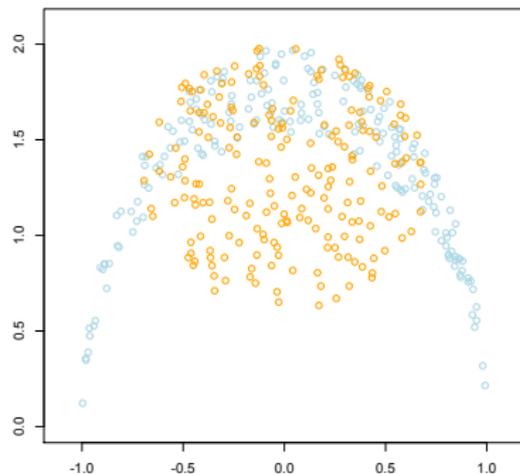
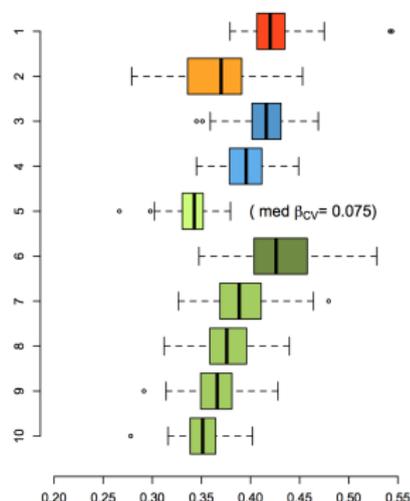
Fréquences de misclassification, dans trois configurations différentes, des LDA/QDA et de divers classifieurs fondés sur la profondeur (100 échantillons d'entraînement de taille $400 = 200 + 200$, et 100 échantillons tests de taille $= 1000 = 500 + 500$)



Configuration 1 : multinormalité avec $\Sigma_0 \neq \Sigma_1$

Classification, et extension locale de la profondeur

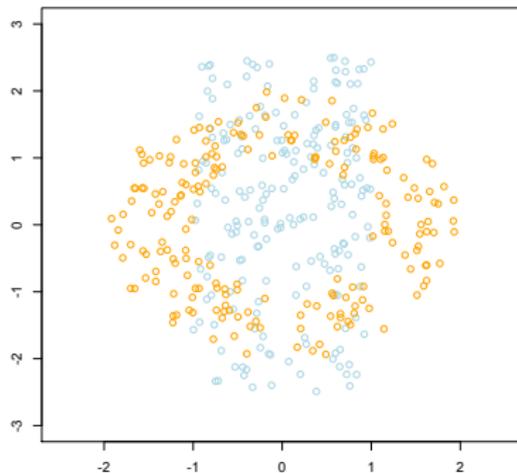
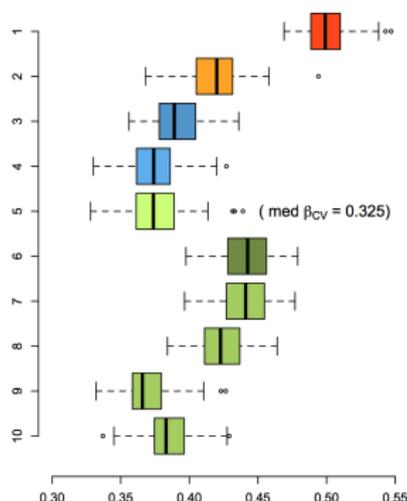
Fréquences de misclassification, dans trois configurations différentes, des LDA/QDA et de divers classifieurs fondés sur la profondeur (100 échantillons d'entraînement de taille $400 = 200 + 200$, et 100 échantillons tests de taille $= 1000 = 500 + 500$)



Configuration 2 : distributions à support lunaire et sphérique

Classification, et extension locale de la profondeur

Fréquences de misclassification, dans trois configurations différentes, des LDA/QDA et de divers classifieurs fondés sur la profondeur (100 échantillons d'entraînement de taille $400 = 200 + 200$, et 100 échantillons tests de taille $1000 = 500 + 500$)



Configuration 3 : distributions à support sur un anneau et un rectangle