



HAL
open science

Réflexions sur l'estimation robuste

Francis Maisonneuve

► **To cite this version:**

| Francis Maisonneuve. Réflexions sur l'estimation robuste. 2014. hal-00995934

HAL Id: hal-00995934

<https://hal.science/hal-00995934>

Preprint submitted on 25 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Réflexions sur l'estimation robuste

Francis Maisonneuve

MINES ParisTech

mai 2014

Sommaire

1	Introduction	2
2	La médiane euclidienne comme “centre”	3
3	L'espérance τ-pondérée comme “rayon”	6
4	Construction d'une pondération robuste, phase 1	8
5	Aspect statistique, centre et rayon empiriques	12
6	Aspect statistique, la g-moyenne empirique	15
7	Exemples	20
8	Construction d'une pondération robuste, phase 2	25
9	Application et exemples	29

Réflexions sur l'estimation robuste

1 Introduction

Soit X une variable aléatoire réelle de loi μ . On dispose d'au moins deux définitions courantes et en partie concurrentes d'un indicateur de position (on dira en abrégé d'une "moyenne") de (la loi de) X :

• L'espérance de X , $E(X) = \int_{\mathbb{R}} x d\mu(x)$, qui a la propriété essentielle d'être une forme linéaire, a pour inconvénients :

- d'être parfois non définie ; par exemple pour la variable de Cauchy, ou pour $X = \frac{1}{2U-1}$ où U désigne la variable uniforme sur $]0, 1[$;
- d'être parfois infinie et en conséquence imprécise ; par exemple pour la variable positive $X = \frac{1}{U}$ qui a même espérance infinie que $2X = \frac{2}{U}$, etc. ;
- même quand $E(X)$ est bien définie et finie, de donner *trop de poids aux queues de distribution* (aux valeurs lointaines ou aberrantes) ; par exemple pour X discrète telle que $P(X = 0) = 0.999$ et $P(X = 10000) = 0.001$, on a $E(X) = 10$.

En résumé, $E(X)$ est *trop sensible aux valeurs extrêmes* de X .

• La médiane m_X de X , milieu (par convention) du *segment médian*

$$I_X \stackrel{\text{déf}}{=} \left\{ x \in \mathbb{R} : P(X < x) \leq \frac{1}{2} \text{ et } P(X > x) \leq \frac{1}{2} \right\},$$

dont les points intérieurs x (s'il y en a) vérifient $P(X = x) = 0$.

m_X a comme avantages sur $E(X)$ d'être *toujours définie* dans \mathbb{R} et de ne pas être sensible aux valeurs extrêmes de X ; elle a pour inconvénients :

- de ne pas être *toujours stable* vis à vis d'une petite variation de la loi de X ; par exemple pour X variable de Bernoulli de paramètre $P(X = 1) = p$ ($0 \leq p \leq 1$), $m_X = 0$ pour $p < \frac{1}{2}$, $I_X = [0, 1] \Rightarrow m_X = \frac{1}{2}$ pour $p = \frac{1}{2}$ et $m_X = 1$ pour $p > \frac{1}{2}$ (alors que $E(X) = p \in [0, 1]$, fonction continue de p au voisinage de $\frac{1}{2}$) ;
- de ne pas prendre en compte le "profil de la distribution" de X ; par exemple les variables Y et $\frac{Y}{U}$, où Y est une variable de Bernoulli de paramètre $\frac{1}{2}$ indépendante de la variable uniforme U sur $]0, 1[$, ont même segment médian $[0, 1]$ et donc même médiane (alors que $E(Y) = \frac{1}{2}$ et $E\left(\frac{Y}{U}\right) = E(Y) E\left(\frac{1}{U}\right) = +\infty$).

En résumé, m_X n'est sensible qu'aux valeurs centrales de X .

La construction d'un indicateur "intermédiaire", ne présentant pas (ou présentant moins) les inconvénients précédents, peut ainsi être fort utile. De multiples approches

ont été proposées ; on va s'intéresser ici au cas d'une variable X réelle ou **vectorielle**, selon une démarche qui pourrait être en partie originale.

L'idée principale consiste à déterminer dans un premier temps un "centre" et un "rayon" attachés à la loi de X qui même grossiers sont robustes et résistants, et sur lesquels on se base pour réévaluer la loi de X .

2 La médiane euclidienne comme "centre"

La donnée est une variable aléatoire réelle ou vectorielle $X = {}^t(X_1, \dots, X_d)$ de loi μ non concentrée en un seul point (t pour transposé). Lorsque $d \geq 2$, on se place dans le cas où il est pertinent de considérer ensemble les composantes X_ℓ : c'est le cas typiquement lorsqu'elles sont de même nature, comme des variables d'espace, de sorte qu'opérer une isométrie affine sur X a un sens. Plus généralement il est envisageable de procéder préalablement à une transformation affine bijective sur X , par exemple pour "normaliser" des échelles de mesure, et différentes méthodes sont proposées dans la littérature. Sinon, on forme les éventuels regroupements partiels qui s'imposent et on travaille séparément sur chacun d'eux.

On introduit pour commencer une notion de *médiane* de (la loi de) X : on va la définir comme le point $m_X \in \mathbb{R}^d$ tel que

$$E(|m_X - X|_2) = \min_{m \in \mathbb{R}^d} E(|m - X|_2),$$

où $|\cdot|_2$ désigne la **norme euclidienne** (ou *norme 2*) de \mathbb{R}^d pour assurer l'équivariance par isométrie — comme c'est le cas pour l'espérance $E(X) = {}^t(E(X_1), \dots, E(X_d))$.

On ne choisit donc pas la *norme 1*, $|x|_1 \stackrel{\text{déf}}{=} \sum_{\ell=1}^d |x_\ell|$, qui est aussi couramment utilisée pour définir la médiane vectorielle car elle ramène à la notion composante par composante, autre propriété vérifiée par l'espérance.

En fait comme on ne suppose pas que $E(X)$ est finie (c'est-à-dire définie dans \mathbb{R}^d), on va considérer la fonction sur \mathbb{R}^d :

$$\rho(m) = \underline{E(|m - X|_2 - |X|_2)} = \int_{\mathbb{R}^d} (|m - x|_2 - |x|_2) d\mu(x),$$

qui elle est toujours définie et vérifie $\forall m \in \mathbb{R}^d, |\rho(m)| \leq |m|_2$, car on a

$$\forall x \in \mathbb{R}^d, \left| |m - x|_2 - |x|_2 \right| \leq |m|_2.$$

Si $E(X)$ est finie, on a $\rho(m) = E(|m - X|_2) - E(|X|_2)$ où $E(|X|_2) < +\infty$ est une constante.

Proposition 2.1 La fonction ρ est **convexe** sur \mathbb{R}^d . Elle est même **strictement convexe**, sauf dans le cas particulier "monodimensionnel à trou" où X prend p.s. ses valeurs dans une droite de \mathbb{R}^d privée d'un intervalle ouvert non trivial.

Preuve Vérifions que $\forall x \in \mathbb{R}^d$ la fonction $m \in \mathbb{R}^d \mapsto |m - x|_2 - |x|_2$ est convexe :

Soient $m_1 \neq m_2$ dans \mathbb{R}^d ; on se place dans un plan contenant m_1, m_2 et x (plan déterminé de manière unique si le triangle m_1, m_2, x n'est pas dégénéré), et on choisit un repère orthonormé de ce plan tel que m_1, m_2 et x aient pour coordonnées respectives $(0, 0)$, $(a, 0)$ ($a > 0$) et (y, z) . Il s'agit de montrer que pour tout point $m \in]m_1, m_2[$ de coordonnées $(ta, 0)$ où $t \in]0, 1[$, on a

$$\sqrt{(y - at)^2 + z^2} \leq (1 - t) \sqrt{y^2 + z^2} + t \sqrt{(y - a)^2 + z^2};$$

il suffit donc de montrer que la fonction continue $t \in [0, 1] \xrightarrow{h} \sqrt{(y - at)^2 + z^2}$ est convexe.

Or pour $z \neq 0$, h est de classe C^2 sur $]0, 1[$ avec $h''(t) = \frac{a^2 z^2}{((y - at)^2 + z^2)^{\frac{3}{2}}} > 0$, donc h est strictement convexe ; et pour $z = 0$ (cas de triangle m_1, m_2, x dégénéré), on a $h(t) = |y - at|$, donc h est convexe, et elle est même strictement convexe — car non affine sur $[0, 1]$ — si et seulement si h s'annule sur $]0, 1[$, c'est-à-dire si et seulement si $0 < y < a$.

On conclut que la fonction d_x est convexe sur \mathbb{R}^d pour tout $x \in \mathbb{R}^d$, et que l'inégalité de convexité $d_x((1 - t)m_1 + tm_2) \leq (1 - t)d_x(m_1) + td_x(m_2)$ pour tout $t \in]0, 1[$ est stricte pour $m_1 \neq m_2$ si et seulement si $x \notin (m_1, m_2) \setminus]m_1, m_2[$, où (m_1, m_2) désigne la droite passant par m_1 et m_2 .

Comme $\rho(m) = \int d_x(m) d\mu(x)$, on voit en intégrant selon μ l'inégalité de convexité que $\forall m_1 \neq m_2$ dans \mathbb{R}^d et $\forall t \in]0, 1[$,

$$\rho((1 - t)m_1 + tm_2) \leq \int [(1 - t)d_x(m_1) + td_x(m_2)] d\mu(x) = (1 - t)\rho(m_1) + t\rho(m_2);$$

et l'inégalité est stricte si on n'a pas $x \in (m_1, m_2) \setminus]m_1, m_2[$ μ -presque partout, autrement dit si X ne prend pas p.s. ses valeurs dans $(m_1, m_2) \setminus]m_1, m_2[$. La fonction ρ est donc convexe, et elle est même strictement convexe si X ne prend pas p.s. ses valeurs dans une droite de \mathbb{R}^d privée d'un intervalle ouvert non trivial. \square

Intéressons-nous à présent à la régularité de la fonction convexe ρ , puis à l'existence et l'unicité du minimum. On a :

Proposition 2.2

- La fonction ρ est lipschitzienne de rapport 1 sur \mathbb{R}^d .
- Pour tout $a \geq 0$, on a $\forall m \in \mathbb{R}^d, \rho(m) \geq C_a |m|_2 - 2a$,
où $C_a = 2\mu(\{x \in \mathbb{R}^d : |x|_2 < a\}) - 1$.
- La fonction ρ est de classe C^1 sur $\mathbb{R}^d \setminus S_X$, où S_X désigne l'ensemble au plus dénombrable des points chargés par X , avec $\forall m = {}^t(m_1, \dots, m_d) \in \mathbb{R}^d \setminus S_X$:

$$1 \leq \ell \leq d, \frac{\partial \rho}{\partial m_\ell}(m) = E\left(\frac{m_\ell - X_\ell}{|m - X|_2}\right) \iff \nabla \rho(m) = E\left(\frac{m - X}{|m - X|_2}\right).$$

- La fonction ρ admet un minimum absolu, qui est atteint soit en un point de S_X , soit en un point m de $\mathbb{R}^d \setminus S_X$ caractérisé par $E\left(\frac{m - X}{|m - X|_2}\right) = 0$.
- Une condition nécessaire et suffisante pour que la fonction ρ atteigne son minimum absolu au point m de \mathbb{R}^d est :

$$\left|E\left(\frac{X - m}{|X - m|_2}\right)\right|_2 \leq P(X = m) \quad (1)$$

(avec la **convention** habituelle en intégration $\frac{0}{0} = 0 \cdot \infty = 0$ pour les valeurs de $\frac{X - m}{|X - m|_2}$).

- Ce point m , qu'on appelle *médiane euclidienne* ou *médiane spatiale* de X et qu'on notera m_X , est unique sauf dans le "cas exceptionnel", cas monodimensionnel à trou où ρ atteint son minimum sur tout un segment dit encore *segment médian* (dont l'intervalle intérieur est de mesure μ nulle); on convient alors comme d'habitude de prendre pour m_X le milieu dudit segment.

- La médiane euclidienne m_X est dans l'enveloppe convexe fermée du support $\text{supp } \mu$ de μ .

Preuve

- On a $\forall m, m' \in \mathbb{R}^d$,

$$\begin{aligned} |\rho(m) - \rho(m')| &= \left| \int (|m - x|_2 - |m' - x|_2) d\mu(x) \right| \\ &\leq \int ||m - x|_2 - |m' - x|_2| d\mu(x) \leq \int |m - m'|_2 d\mu = |m - m'|_2. \end{aligned}$$

- Pour tout $a \geq 0$, $\rho(m) = I_1(a) + I_2(a)$ où
$$\begin{cases} I_1(a) \stackrel{\text{déf}}{=} \int_{\{|x|_2 < a\}} (|m - x|_2 - |x|_2) d\mu(x) \\ I_2(a) \stackrel{\text{déf}}{=} \int_{\{|x|_2 \geq a\}} (|m - x|_2 - |x|_2) d\mu(x) \end{cases};$$

Comme $|m - x|_2 + |x|_2 \geq |m|_2$, on a $|m - x|_2 - |x|_2 \geq |m|_2 - 2|x|_2$, donc

$$I_1(a) \geq \int_{\{|x|_2 < a\}} (|m|_2 - 2|x|_2) d\mu(x) \geq |m|_2 P(|X|_2 < a) - 2a;$$

et comme $|m - x|_2 - |x|_2 \geq -|m|_2$, on a

$$I_2(a) \geq - \int_{\{|x|_2 \geq a\}} |m|_2 d\mu(x) = -|m|_2 P(|X|_2 \geq a) = |m|_2 (P(|X|_2 < a) - 1).$$

D'où le résultat.

- On utilise le théorème de convergence dominée sur $\mathbb{R}^d \setminus \{m\}$ pour justifier les dérivations partielles sous le signe intégral et la continuité des dérivées partielles en $m \in \mathbb{R}^d \setminus S_X$, avec comme fonction dominatrice μ -intégrable la constante 1.

- Soit $a > 0$ strictement à droite du segment médian de la variable positive $|X|_2$; on a $C_a > 0$ (cf. le point 2), de sorte que $\lim_{|m|_2 \rightarrow \infty} \rho(m) = +\infty$. Comme ρ est continue sur \mathbb{R}^d (car lipschitzienne), elle admet donc bien un minimum absolu. Et comme ρ est convexe, si ce minimum est atteint en un point $m \in \mathbb{R}^d \setminus S_X$, où ρ est différentiable, m est caractérisé par l'équation $\nabla \rho(m) = 0$.

- Soit $m \in \mathbb{R}^d$ fixé; on considère la sous-probabilité $\nu = \mu - \mu(\{m\}) \delta_m$ portée par $\mathbb{R}^d \setminus \{m\}$, δ_m désignant la mesure de Dirac au point m .

On a $\nu(\{m\}) = 0$, donc (cf. le point 3) la fonction $\rho_\nu(m') \stackrel{\text{déf}}{=} \int (|m' - x|_2 - |x|_2) d\nu(x)$ est différentiable en $m' = m$, de gradient (avec toujours la convention $\frac{0}{0} = 0$)

$$\nabla \rho_\nu(m) = \int \frac{m - x}{|m - x|_2} d\nu(x) = \int \frac{m - x}{|m - x|_2} d\mu(x) = E\left(\frac{m - X}{|m - X|_2}\right).$$

On a par ailleurs $\forall m' \in \mathbb{R}^d$, $\rho_\nu(m') = \rho(m') - \mu(\{m\}) (|m' - m|_2 - |m|_2)$, et en particulier $\rho_\nu(m) = \rho(m) + \mu(\{m\}) |m|_2$.

Comme ρ_ν est convexe (proposition 2.1 en remplaçant μ par ν) et différentiable en m , on a

$$\forall m' \in \mathbb{R}^d, \rho_\nu(m') \geq \rho_\nu(m) + \nabla \rho_\nu(m) \cdot (m' - m),$$

d'où $\rho(m') - \mu(\{m\}) (|m' - m|_2 - |m|_2) \geq \rho(m) + \mu(\{m\}) |m|_2 + E\left(\frac{m - X}{|m - X|_2}\right) \cdot (m' - m)$

et donc $\rho(m') \geq \rho(m) + |m' - m|_2 \left[E\left(\frac{m - X}{|m - X|_2}\right) \cdot \frac{m' - m}{|m' - m|_2} + P(X = m) \right]$ pour tout $m' \in \mathbb{R}^d \setminus \{m\}$.

Or la condition (1) de l'énoncé équivaut au fait que le terme entre crochets ci-dessus est positif ou nul pour tout $m' \in \mathbb{R}^d \setminus \{m\}$, car $\frac{m' - m}{|m' - m|_2}$ est un vecteur unitaire arbitraire ; de sorte que (1) équivaut bien au fait que $\rho(m)$ est le minimum absolu de ρ .

- Comme vu à la proposition 2.1, ρ est strictement convexe sauf dans le cas monodimensionnel à trou : dans ce cas on peut éventuellement avoir ρ affine constante égale au minimum absolu sur tout un intervalle non trivial de la droite des valeurs de X , d'intérieur de mesure μ -nulle ; intervalle borné du fait que $\lim_{|m|_2 \rightarrow \infty} \rho(m) = +\infty$, et fermé du fait que ρ est continue sur \mathbb{R}^d .

- Soit m dans le complémentaire de l'enveloppe convexe fermée de $\text{supp } \mu$, de sorte que $P(X = m) = 0$; soit H un hyperplan séparant strictement m de cette enveloppe convexe fermée (théorème de Hahn – Banach, forme géométrique), et donc de $\text{supp } \mu$. On note $m' \neq m$ la projection orthogonale de m sur H .

On a $\forall x \in \text{supp } \mu$, $|m' - x|_2 - |x|_2 < |m - x|_2 - |x|_2$, donc $\rho(m') < \rho(m)$; de sorte que $m \neq m_X$. \square

Remarque 2.3 Pour $d = 1$, on a $\frac{m - X}{|m - X|_2} = \frac{m - X}{|m - X|} = \mathbb{1}_{\{X < m\}} - \mathbb{1}_{\{X > m\}}$; de sorte qu'on a $E\left(\frac{m - X}{|m - X|_2}\right) = P(X < m) - P(X > m)$ et le critère (1) s'écrit :

$$\begin{aligned} |P(X < m) - P(X > m)| \leq P(X = m) &\iff |2P(X \leq m) - 1 - P(X = m)| \leq P(X = m) \\ &\iff 0 \leq 2\left(P(X \leq m) - \frac{1}{2}\right) \leq 2P(X = m) \\ &\iff P(X \leq m) \geq \frac{1}{2} \text{ et } P(X < m) \leq \frac{1}{2}. \end{aligned}$$

On retombe bien sur la formulation banale de la médiane.

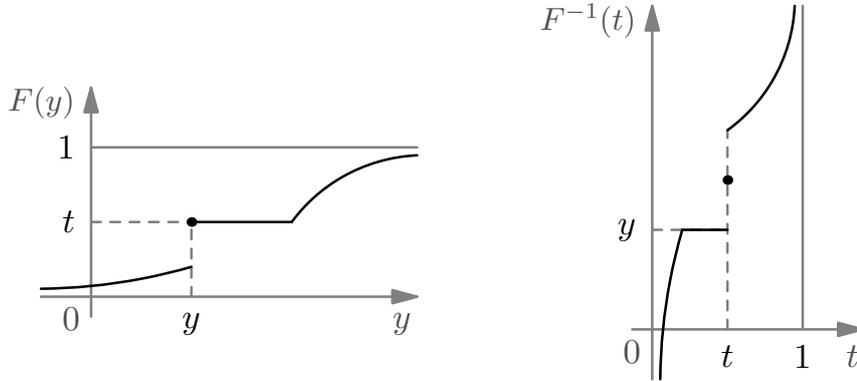
3 L'espérance τ -pondérée comme "rayon"

Soit $X = {}^t(X_1, \dots, X_d)$ de loi μ non concentrée eu un seul point, et soit la variable $R \stackrel{\text{déf}}{=} |X - m_X|_2 \geq 0$ distance "au centre" (c'est-à-dire à la médiane euclidienne). Par hypothèse, R n'est pas presque sûrement nulle. R a une espérance finie si et seulement si $E(X)$ est finie ; auquel cas on a $E(R) = \rho(m_X) + E(|X|_2)$ où $\rho(m_X) = \min_{m \in \mathbb{R}^d} \rho(m)$.

On veut définir un indicateur d'échelle (ou d'étendue) de R , et comme $R \geq 0$, un indicateur de position peut faire l'affaire ; on veut qu'il soit résistant, ce qui disqualifie

de toute façon l'espérance $E(R)$, et aussi que son éventuelle instabilité ne se surajoute pas à celle de m_X , ce qui tend à disqualifier la médiane m_R , qui est l'estimateur MAD bien connu (cf. l'introduction et le premier exemple du paragraphe 7).

De manière générale, soit Y une variable aléatoire **réelle** de fonction de répartition $F(y) \stackrel{\text{déf}}{=} P(Y \leq y)$; notons (abusivement) $F^{-1}:]0, 1[\rightarrow \mathbb{R}$ "l'inverse régularisée" de F (demi-somme de son inverse continu à gauche et de son inverse continu à droite).



F^{-1} est comme F une fonction croissante au sens large et on a la relation d'inclusion :

$$\forall y \in \mathbb{R},]0, F(y)[\subseteq \{t \in]0, 1[: F^{-1}(t) \leq y\} \subseteq]0, F(y)[\quad (2)$$

Soit U la variable uniforme sur $]0, 1[$; on déduit de cette relation (2) le résultat usuel

$$F^{-1}(U) \stackrel{\mathcal{L}}{=} Y \text{ (méthode d'inversion),}$$

car $\forall y \in \mathbb{R}, F(y) = P(U \in]0, F(y)[) \leq P(F^{-1}(U) \leq y) \leq P(U \in]0, F(y)[) = F(y)$;

$$\text{de sorte que } \begin{cases} E(Y) = E(F^{-1}(U)) = \int F^{-1} d\lambda_{]0,1[} & (\text{si } Y \text{ a une espérance}) \\ m_Y = F^{-1}\left(\frac{1}{2}\right) = \int F^{-1} d\delta_{\frac{1}{2}} & (\text{du fait que } F^{-1} \text{ est régularisée),} \end{cases}$$

où $\lambda_{]0,1[}$ désigne la mesure de Lebesgue sur $]0, 1[$ et $\delta_{\frac{1}{2}}$ celle de Dirac au point $\frac{1}{2}$.

Soit T une variable à valeurs dans $]0, 1[$ de loi notée τ ; je propose de considérer comme indicateur alternatif de position de Y l'espérance τ -pondérée

$$E_\tau(Y) \stackrel{\text{déf}}{=} \int F^{-1} d\tau = E(Y_T),$$

où $Y_T \stackrel{\text{déf}}{=} F^{-1}(T)$ est la V.A.R. composée de T par F^{-1} — sous réserve que l'espérance de cette *variable de simulation* Y_T soit bien définie. Y_T a même loi que Y si $T = U$ (auquel cas $\tau = \lambda_{]0,1[}$ et $E_\tau(Y) = E(Y)$); et Y_T coïncide avec la constante $F^{-1}\left(\frac{1}{2}\right) = m_Y$ si $T = \frac{1}{2}$ (auquel cas $\tau = \delta_{\frac{1}{2}}$ et $E_\tau(Y) = m_Y$).

Notons $\nu = F^{-1}(\tau)$ la loi de $Y_T = F^{-1}(T)$, et F^ν sa fonction de répartition. On a

$$\int_{\mathbb{R}} x d\nu(x) = \int_{[0, +\infty[} d\nu(x) \int_{[0, x]} dy - \int_{]-\infty, 0[} d\nu(x) \int_{]x, 0[} dy,$$

et donc d'après le théorème de Fubini

$$\int_{\mathbb{R}} x d\nu(x) = \int_{[0,+\infty[} dy \int_{[y,+\infty[} d\nu(x) - \int_{]-\infty,0[} dy \int_{]-\infty,y[} d\nu(x),$$

soit
$$E(Y_T) = \int_0^{+\infty} (1 - F^\nu(y)) dy - \int_{-\infty}^0 F^\nu(y) dy.$$

On se restreint dans la suite à considérer des **lois τ de T diffuses**, autrement dit des fonctions de répartition F^τ de T continues sur \mathbb{R} . Alors d'après la relation (2),

$$\forall y \in \mathbb{R}, P(T < F(y)) \leq P(F^{-1}(T) \leq y) \leq P(T \leq F(y)); \text{ soit } \underline{F^\nu} = F^\tau \circ F.$$

Ainsi
$$E_\tau(Y) = E(Y_T) = \int_0^{+\infty} [1 - F^\tau(F(y))] dy - \int_{-\infty}^0 F^\tau(F(y)) dy.$$

Revenons à la situation $Y = R \geq 0$; on a donc

$$E_\tau(R) = \int_0^{+\infty} [1 - F^\tau(F(r))] dr = \int_0^{+\infty} (1 - F^\nu(r)) dr \quad (3)$$

Je propose, pour assurer la pertinence et la résistance de l'indicateur d'échelle $E_\tau(R)$, de choisir une loi diffuse τ sur $]0, 1[$ portée par un sous-intervalle $]\frac{1}{2}, 1 - \varepsilon[$ (où $0 < \varepsilon < \frac{1}{2}$); en effet de cette manière on prend en compte les valeurs de R qui informent sur l'étendue, pour lesquelles $F \geq \frac{1}{2}$, mais pas ses valeurs extrêmes, pour lesquelles $F \geq 1 - \varepsilon$; cette dernière restriction assure aussi que l'espérance $E_\tau(R)$ est finie.

Le plus simple dans cette optique est de prendre τ de densité

$$f^\tau \stackrel{\text{déf}}{=} \frac{2}{1 - 2\varepsilon} \mathbb{1}_{]\frac{1}{2}, 1 - \varepsilon[};$$

avec ce choix F^τ est continue et affine par morceaux, telle que

$$F^\tau(t) = 0 \text{ si } t \leq \frac{1}{2}, \quad \underline{F^\tau(t) = \frac{2}{1 - 2\varepsilon} \left(t - \frac{1}{2}\right) \text{ sur }]\frac{1}{2}, 1 - \varepsilon]} \text{ et } F^\tau(t) = 1 \text{ si } t \geq 1 - \varepsilon.$$

Comme $F^\tau \circ F$ est croissante et continue à droite comme F , on note que d'après (3),

$$E_\tau(R) = 0 \iff F^\tau \circ F = 1 \text{ sur } \mathbb{R}_+ \iff P(X = m_X) = P(R = 0) = F(0) \geq 1 - \varepsilon.$$

4 Construction d'une pondération robuste, phase 1

Soit $X = {}^t(X_1, \dots, X_d)$ de loi μ non concentrée eu un seul point. On considère le changement de variables aléatoires de type hypersphérique obtenu en prenant comme **nouvelle origine le point m_X** :

$$R = |X - m_X|_2 \geq 0 \text{ et } \Theta \stackrel{\text{déf}}{=} \frac{X - m_X}{|X - m_X|_2} \in \mathbb{S}^{d-1} \cup \{0\}$$

où \mathbb{S}^{d-1} est l'hypersphère unité de \mathbb{R}^d ; de sorte que $\underline{X = m_X + R\Theta}$ avec

$$\Theta = 0 \iff X = m_X \iff R = 0 \text{ et } |\Theta|_2 = \mathbb{1}_{\{\Theta \neq 0\}}; \text{ d'où } R = R|\Theta|_2$$

et (proposition 2.2) $|E(\Theta)|_2 \leq P(R = 0) = P(\Theta = 0)$.

Bien sûr, les variables R et Θ ne sont pas indépendantes en général.

On souhaite définir une “moyenne” de X qui soit équivariante par similitude affine, comme l’est l’espérance $E(X) \in \mathbb{R}^d$ quand elle existe, et qui soit de plus résistante à un certain niveau de contamination. On va procéder, avant d’en prendre l’espérance, à une déformation graduelle de la loi μ en fonction de la distance r au point m_X de la manière suivante : étant donné une fonction g , définie à un facteur multiplicatif près, telle que

$$(*) \quad g : \overline{\mathbb{R}}_+ \rightarrow \mathbb{R}_+ \text{ **décroissante** (large), } g(1) > 0 \text{ et } \sup_{z \in \mathbb{R}_+} z g(z) < +\infty$$

(de sorte que $g(+\infty) = 0$), on considère la *variable de g -pondération de X* :

$$\mathcal{V}_X^g \stackrel{\text{déf}}{=} \frac{1}{E\left(g\left(\frac{R}{E_\tau(R)}\right)\right)} g\left(\frac{R}{E_\tau(R)}\right) = \phi_X^g(X)$$

où
$$\phi_X^g(x) \stackrel{\text{déf}}{=} \frac{1}{E\left(g\left(\frac{R}{E_\tau(R)}\right)\right)} g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) \geq 0.$$

La fonction g étant définie à un facteur multiplicatif près, on convient dans la suite de prendre $g(0) = 1$, de sorte que g prend ses valeurs dans $[0, 1]$. En accord avec la convention en intégration $\frac{0}{0} = 0 \cdot \infty = 0$, on a $\frac{R}{E_\tau(R)} = 0 \cdot \mathbb{1}_{\{R=0\}} + \frac{R}{E_\tau(R)} \mathbb{1}_{\{R>0\}}$; d’où ($g(0) = 1$ et $g(+\infty) = 0$) :

$$\begin{cases} \text{si } E_\tau(R) = 0, g\left(\frac{R}{E_\tau(R)}\right) = \mathbb{1}_{\{R=0\}} \text{ et } E\left(g\left(\frac{R}{E_\tau(R)}\right)\right) = P(R = 0) = P(X = m_X) \\ \text{si } E_\tau(R) > 0, \mathbb{1}_{\{R=0\}} \leq g\left(\frac{R}{E_\tau(R)}\right) \leq 1 \text{ et } P(X = m_X) \leq E\left(g\left(\frac{R}{E_\tau(R)}\right)\right) \leq 1. \end{cases}$$

On a donc en particulier (cf. la fin du paragraphe 3)

$$P(X = m_X) \geq 1 - \varepsilon \iff E_\tau(R) = 0 \implies \mathcal{V}_X^g = \frac{1}{P(R = 0)} \mathbb{1}_{\{R=0\}},$$

ce qui incite, pour éviter au maximum une variable \mathcal{V}_X^g aussi rudimentaire, à ne **pas** choisir $\varepsilon \in]0, \frac{1}{2}[$ trop voisin de $\frac{1}{2}$.

On note que $E\left(g\left(\frac{R}{E_\tau(R)}\right)\right) > 0$: en effet sinon on aurait $g\left(\frac{R}{E_\tau(R)}\right) \stackrel{\text{ps}}{=} 0$ et donc $R \stackrel{\text{ps}}{>} E_\tau(R)$ puisque g décroissante et $g(1) > 0$; on aurait donc $F(r) = P(R \leq r) = 0$ pour $0 \leq r \leq E_\tau(R)$. On en déduirait d’après la relation (3) du paragraphe 3,

$$E_\tau(R) = \int_0^{+\infty} [1 - F^\tau(F(r))] dr = E_\tau(R) + \int_{E_\tau(R)}^{+\infty} [1 - F^\tau(F(r))] dr.$$

On devrait donc avoir $F^\tau[F(E_\tau(R))] = 1$ du fait de la continuité à droite de $F^\tau \circ F$; d’où $P(R \leq E_\tau(R)) = F(E_\tau(R)) \geq 1 - \varepsilon$, en contradiction avec $F(E_\tau(R)) = 0$.

On en déduit $0 \leq \mathcal{V}_X^g \leq \frac{1}{E\left(g\left(\frac{R}{E_\tau(R)}\right)\right)} < \infty$ et $E(\mathcal{V}_X^g) = 1$ (d’où le terme *pondération*).

Une première idée en vue de définir l'indicateur robuste visé serait de remplacer la variable X par la variable virtuelle $\mathcal{V}_X^g X$; ou encore, vu de m_X , de remplacer $X - m_X$ par $\mathcal{V}_X^g (X - m_X)$ — ce qui revient à contracter les grandes distances à m_X sans considération angulaire (multiplication par $g(\frac{R}{E_\tau(R)}) \leq 1$), et à dilater parallèlement d'un même facteur toutes les distances à m_X “pour compenser” (division par $E(g(\frac{R}{E_\tau(R)})) \leq 1$).

En fait il semble plus intéressant de ne pas modifier X (“respect des données”) et de voir en $\mathcal{V}_X^g(\omega) = \phi_X^g(x)$ l'intensité de la confiance qu'on accorde à la donnée $x = X(\omega)$ du fait de la valeur de $R(\omega) = |x - m_X|_2$ comparée à $E_\tau(R)$, dans le “stock” de confiance à répartir $E(\mathcal{V}_X^g) = 1$. Ceci revient à considérer la probabilité $P_X^g \stackrel{\text{déf}}{=} \mathcal{V}_X^g P$ ayant \mathcal{V}_X^g comme dérivée de Radon–Nikodym par rapport à la probabilité initiale de référence P ; de sorte que X a vis à vis de P_X^g la nouvelle loi “robuste”

$$\mu^g = X(P_X^g) = X(\phi_X^g(X) P) = \phi_X^g X(P) = \phi_X^g \mu,$$

soit

$$\mu^g(dx) = \frac{1}{E(g(\frac{R}{E_\tau(R)}))} g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) \mu(dx).$$

On a alors pour toute variable Y (sous réserve d'existence des espérances), en notant E^g l'espérance relative à P_X^g ,

$$E^g(Y) = \int Y dP_X^g = \int Y \mathcal{V}_X^g dP = E(\mathcal{V}_X^g Y).$$

En particulier la variable X admet une espérance E^g finie car on a

$$\begin{aligned} E^g(|X - m_X|_2) &= \frac{1}{E(g(\frac{R}{E_\tau(R)}))} \int |x - m_X|_2 g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) \mu(dx) \\ &\leq \frac{M E_\tau(R)}{E(g(\frac{R}{E_\tau(R)}))} < +\infty \end{aligned}$$

où $M = \sup_{z \in \mathbb{R}_+} z g(z) < +\infty$ (cf. les propriétés (*)); on définit alors la g -moyenne de X par

$$E^g(X) = \int_{\mathbb{R}^d} x d\mu^g(x) = E(\mathcal{V}_X^g X) = m_X + E(\mathcal{V}_X^g (X - m_X)),$$

qui par construction est équivariante par similitude affine. En effet posons $\tilde{X} = \Phi(X)$ où Φ est une similitude affine de \mathbb{R}^d de rapport $\beta > 0$; on a $m_{\tilde{X}} = \Phi(m_X)$, de sorte que $\tilde{R} \stackrel{\text{déf}}{=} |\tilde{X} - m_{\tilde{X}}|_2 = \beta R$; on en déduit $\tilde{F}(r) = F(\frac{r}{\beta})$, donc $E_\tau(\tilde{R}) = \beta E_\tau(R)$: on a ainsi $\frac{\tilde{R}}{E_\tau(\tilde{R})} = \frac{R}{E_\tau(R)}$, d'où $\mathcal{V}_{\tilde{X}}^g = \mathcal{V}_X^g$. En conséquence $E^g(\Phi(X)) = \Phi(E^g(X))$.

Proposition 4.1 Soit $g : \overline{\mathbb{R}}_+ \rightarrow [0, 1]$ vérifiant les propriétés (*) et nulle à partir d'une valeur finie $c > 1$. Alors

- La probabilité μ^g est portée par l'hyperboule euclidienne ouverte $B(m_X, c E_\tau(R))$ de \mathbb{R}^d de centre m_X et de rayon $c E_\tau(R)$ (de sorte qu'on **rejette** purement et simplement les valeurs prises par X qui sont à distance de m_X supérieure ou égale à $c E_\tau(R)$); la variable X est donc P_X^g -ps bornée.

- Dans le cas où on choisirait $g = \mathbb{1}_{[0, c[}$ (où $1 < c < +\infty$), on aurait simplement

$$P_X^g = P(\cdot \mid R < c E_\tau(R)),$$

probabilité P conditionnelle à $R < c E_\tau(R)$. Ceci revient pour X à tenir compte sans changement des (seules) valeurs qui sont à distance de m_X strictement inférieure à $c E_\tau(R)$; de sorte que

$$E^g(X) = E(X \mid R < c E_\tau(R)),$$

espérance de X par rapport à la probabilité conditionnelle $P(\cdot \mid R < c E_\tau(R))$.

Preuve

• On a $\mu^g = \phi_X^g \mu$ avec $\phi_X^g(x) = 0$ pour $x \notin B(m_X, c E_\tau(R))$ d'après l'hypothèse sur g .

• Dans le cas où $g = \mathbb{1}_{[0, c[}$, on a $P_X^g = \frac{\mathbb{1}_{\{R < c E_\tau(R)\}}}{E(\mathbb{1}_{\{R < c E_\tau(R)\}})} P = P(\cdot \mid R < c E_\tau(R))$;

on en déduit $E^g(X) = \frac{1}{P(R < c E_\tau(R))} E(\mathbb{1}_{\{R < c E_\tau(R)\}} X) = E(X \mid R < c E_\tau(R))$. \square

Remarques 4.2

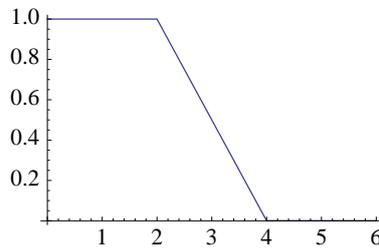
• L'inégalité $|\mathcal{V}_X^g(X - m_X)|_2 \leq \frac{M E_\tau(R)}{E(g(\frac{R}{E_\tau(R)})}$ garantira la **résistance** de l'indicateur empirique associé à $E^g(X) = E(\mathcal{V}_X^g X)$, même si g ne s'annule pas à partir d'une valeur finie c (cf. la fin du paragraphe 6).

• Si la fonction g vaut 1 sur un segment $[0, c'[$ avec $c' > 1$ et si $R \stackrel{\text{ps}}{<} c' E_\tau(R)$, on a $\mathcal{V}_X^g \stackrel{\text{ps}}{=} 1$ et $E^g(X) = E(X)$: ceci assure une forme de **robustesse** de la g -moyenne.

• Dans le cas $E_\tau(R) = 0$, on a $\mathcal{V}_X^g = \frac{1}{P(R=0)} \mathbb{1}_{\{R=0\}}$, donc $\mu^g = \frac{1}{\mu(\{m_X\})} \mathbb{1}_{\{m_X\}} \mu$ et $E^g(X) = m_X$.

Nous allons nous restreindre dans la suite à des fonctions g très simples, continues et affines par morceaux, qui dépendent d'un paramètre libre $c > 1$:

$$g_c(t) = 1 \text{ sur } \left[0, \frac{c}{2}\right[, \quad g_c(t) = 2 \left(1 - \frac{t}{c}\right) \text{ sur } \left[\frac{c}{2}, c\right] \text{ et } g_c(t) = 0 \text{ sur }]c, +\infty].$$



graphe de la fonction g_c pour $c = 4$

Avec de telles fonctions $g = g_c$, $E^g(X)$ apparaît comme une médiane corrigée “pour se rapprocher” de $E(X)$ lorsque cette espérance est finie, en rejetant les valeurs de R supérieures à $c E_\tau(R)$. La continuité de g_c évite des instabilités de comportement de $E^g(X)$ quand R prend des valeurs voisines de $\frac{c}{2} E_\tau(R)$, et sa décroissance entre $\frac{c}{2}$ et c exprime la diminution de la confiance sur les données à partir de $R \geq \frac{c}{2} E_\tau(R)$.

5 Aspect statistique, centre et rayon empiriques

Soit $(X^n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{R}^d , toutes de même loi non concentrée en un point ; et pour tout $n \in \mathbb{N}^*$, soit μ_n la répartition (aléatoire) de probabilité empirique portée par l'échantillon (X^1, \dots, X^n) :

$$\mu_n(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X^i(\omega)}.$$

Fixons ω . La médiane euclidienne empirique $M_n(\omega) \in \mathbb{R}^d$ est définie comme étant le M -estimateur

$$M_n(\omega) = \operatorname{argmin}_{m \in \mathbb{R}^d} \int (|m - x|_2 - |x|_2) d\mu_n(\omega)(x) = \operatorname{argmin}_{m \in \mathbb{R}^d} \sum_{i=1}^n |m - X^i(\omega)|_2,$$

Comme vu à la proposition 2.2 en remplaçant μ par la probabilité $\mu_n(\omega)$,

$$\left\{ \begin{array}{l} \cdot \text{ ou bien } M_n(\omega) \text{ coïncide avec l'un des } X^i(\omega) \\ \cdot \text{ ou bien } M_n(\omega) \in \mathbb{R}^d \setminus \{X^1(\omega), \dots, X^n(\omega)\} \text{ est la solution de l'équation en } m : \\ \quad \sum_{i=1}^n \frac{m - X^i(\omega)}{|m - X^i(\omega)|_2} = 0 \text{ (ou le milieu du segment médian si les } X^i(\omega) \text{ sont alignés).} \end{array} \right.$$

On peut évaluer en pratique $M_n(\omega)$ en appliquant par exemple un algorithme d'optimisation (non linéaire) à la fonction convexe $m \mapsto \sum_{i=1}^n |m - X^i(\omega)|_2$.

Pour $1 \leq i \leq n$, soit $R_n^i \stackrel{\text{déf}}{=} |X^i - M_n|_2$. On définit l'indicateur d'échelle empirique $\bar{R}_n = \bar{R}_n^\tau$ comme étant le L -estimateur

$$\bar{R}_n \stackrel{\text{déf}}{=} \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] R_n^{(i)}.$$

où $\{R_n^{(1)} \leq \dots \leq R_n^{(n)}\}$ est le pseudo-échantillon $\{R_n^1, \dots, R_n^n\}$ réordonné par valeurs croissantes. \bar{R}_n est une **moyenne pondérée des $R_n^{(i)}$** puisque la somme des n poids positifs $F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right)$ vaut $F^\tau(1) - F^\tau(0) = 1$.

La pertinence de cette définition de \bar{R}_n apparaîtra à la proposition 5.2.

Proposition 5.1 Si on n'est pas dans le "cas exceptionnel", on a convergence forte du M -estimateur médiane euclidienne empirique vers la médiane euclidienne de X :

$$\text{presque sûrement } \underline{M_n(\omega)} \xrightarrow[n \rightarrow \infty]{} m_X.$$

Preuve D'après la loi forte des grands nombres appliqué à la suite des variables réelles indépendantes et équidistribuées (d'espérances finies) $(|X^n - m_X|_2 - |X^n|_2)_{n \in \mathbb{N}^*}$, on a presque sûrement

$$\rho_n(\omega)(m_X) \stackrel{\text{déf}}{=} \frac{1}{n} \sum_{i=1}^n (|X^i(\omega) - m_X|_2 - |X^i(\omega)|_2) \xrightarrow[n \rightarrow \infty]{} E(|X - m_X|_2 - |X|_2) = \rho(m_X) \quad (\text{a})$$

On en déduit que presque sûrement, $\exists n_0(\omega) \in \mathbb{N}^*$ tel que $\forall n \geq n_0, \rho_n(\omega)(m_X) \leq \rho(m_X) + 1$.

Or d'après la proposition 2.2 en remplaçant μ par la probabilité $\mu_n(\omega)$, on a

$$\forall \omega, \forall a \geq 0 \text{ et } \forall m \in \mathbb{R}^d, \rho_n(\omega)(m) \geq C_n^a(\omega) |m|_2 - 2a,$$

$$\text{où } C_n^a(\omega) = 2 \mu_n(\omega)(\{x \in \mathbb{R}^d : |x|_2 < a\}) - 1 = \frac{2}{n} \sum_{i=1}^n \mathbb{1}_{\{|X^i|_2 < a\}}(\omega) - 1.$$

D'après la loi forte des grands nombres appliqué à la suite des variables réelles indépendantes et équidistribuées (d'espérances finies) $(\mathbb{1}_{\{|X^n|_2 < a\}})_{n \in \mathbb{N}^*}$

$$\text{presque sûrement } \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{|X^i|_2 < a\}}(\omega) \xrightarrow{n \rightarrow \infty} E(\mathbb{1}_{\{|X|_2 < a\}}) = P(|X|_2 < a),$$

donc presque sûrement $C_n^a(\omega) \xrightarrow{n \rightarrow \infty} C_a$.

Fixons $a > 0$ assez grand pour que $C_a = 2P(|X|_2 < a) - 1 > \frac{1}{2}$; on a donc

$$\text{presque sûrement } \exists n_1(\omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_1 \text{ et } \forall m \in \mathbb{R}^d, \rho_n(\omega)(m) \geq \frac{1}{2} |m|_2 - 2a.$$

Or $|m - m_X|_2 > \gamma \stackrel{\text{déf}}{=} |m_X|_2 + 2(\rho(m_X) + 1 + 2a) \implies \frac{1}{2} |m|_2 > \rho(m_X) + 1 + 2a$; on a donc pour cette valeur de γ , en posant $n_2 \stackrel{\text{déf}}{=} \max[n_0, n_1]$:

presque sûrement $\exists n_2(\omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_2,$

$$|m - m_X|_2 > \gamma \implies \rho_n(\omega)(m) > \rho(m_X) + 1 \geq \rho_n(\omega)(m_X).$$

Comme $\rho_n(\omega)$ est minimum en $M_n(\omega)$, ceci assure que

$$\text{presque sûrement } \exists n_2(\omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_2, M_n(\omega) \in B_f(m_X, \gamma) \quad (\text{b})$$

$B_f(m_X, \gamma)$ désignant l'hyperboule euclidienne fermée de \mathbb{R}^d de centre m_X et de rayon γ .

D'autre part, d'après le théorème fondamental de la statistique,

$$\text{presque sûrement la suite } (\mu_n(\omega))_{n \in \mathbb{N}^*} \text{ converge étroitement vers } \mu \quad (\text{c})$$

Montrons que pour tous les ω vérifiant les propriétés (a), (b) et (c) ci-dessus, et donc presque sûrement, on a $\lim_{n \rightarrow \infty} M_n(\omega) = m_X$; c'est-à-dire :

$$\forall \alpha \in]0, \gamma[, \exists n_3(\alpha, \omega) \geq n_2(\omega) \text{ tel que } \forall n \geq n_3, \text{ on a } |M_n(\omega) - m_X| < \alpha.$$

Par l'absurde, si pour un tel $\omega = \omega_0$ ce n'est pas le cas, il existe d'après (b) un $\alpha \in]0, \gamma[$ et une suite extraite $(M_{n^k}(\omega_0))_{k \in \mathbb{N}}$ de $(M_n(\omega_0))_{n \in \mathbb{N}}$ qui converge dans le sous-ensemble compact $B_f(m_X, \gamma) \cap B(m_X, \alpha)^c$, où $B(m_X, \alpha)$ est l'hyperboule euclidienne ouverte de centre m_X et de rayon α . Notons m_0 sa limite, qui vérifie $\rho(m_0) > \rho(m_X)$ puisqu'on n'est pas dans le "cas exceptionnel".

Posons $\delta = \rho(m_0) - \rho(m_X) > 0$. Comme $g_0(x) \stackrel{\text{déf}}{=} |m_0 - x|_2 - |x|_2$ est une fonction continue et bornée sur \mathbb{R}^d , on a $\rho_n(\omega_0)(m_0) = \int g_0(x) d\mu_n(\omega_0)(x) \xrightarrow{n \rightarrow \infty} \int g_0(x) d\mu(x) = \rho(m_0)$ par convergence étroite (c); donc pour n assez grand, $\rho_n(\omega_0)(m_0) \geq \rho(m_0) - \frac{\delta}{3}$.

Or pour tout ω les $\rho_n(\omega)$ sont 1-lipschitziens (remplacer μ par $\mu_n(\omega)$ dans la proposition 2.2). On a donc pour k assez grand, du fait que $M_{n^k}(\omega_0) \xrightarrow{k \rightarrow \infty} m_0$,

$$\min \rho_{n^k}(\omega_0) = \rho_{n^k}(\omega_0)(M_{n^k}(\omega_0)) \geq \rho_{n^k}(\omega_0)(m_0) - |(M_{n^k}(\omega_0)) - m_0|_2 \geq \rho(m_0) - \frac{2\delta}{3},$$

et a fortiori $\rho_{n^k}(\omega_0)(m_X) \geq \rho(m_0) - \frac{2\delta}{3} = \rho(m_X) + \frac{\delta}{3}$ par définition de δ : ce qui contredit (a). \square

Proposition 5.2 : Si on n'est pas dans le "cas exceptionnel", on a la convergence forte du L -estimateur d'échelle empirique vers l'indicateur d'échelle de X :

$$\text{presque sûrement } \underbrace{\bar{R}_n(\omega) \xrightarrow[n \rightarrow \infty]{} E_\tau(R)}.$$

Preuve

• Notons $1 \leq i \leq n$, $R_n^i = |X^i - M_n|_2$ et $Y^i = |X^i - m_X|_2$; on a par inégalité triangulaire

$$|R_n^i - Y^i| \leq |M_n - m_X|_2.$$

Or on sait (proposition 5.1) que presque sûrement $M_n(\omega) \xrightarrow[n \rightarrow \infty]{} m_X$; donc

$$\text{p.s. } \forall \alpha > 0, \exists n_0(\alpha, \omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_0 \text{ et } \forall i \in \{1, \dots, n\}, |R_n^i(\omega) - Y^i(\omega)| \leq \frac{\alpha}{2} \quad (\text{a})$$

Montrons que ceci implique que $\forall i \in \{1, \dots, n\}$, $|R_n^{(i)}(\omega) - Y^{(i)}(\omega)| \leq \frac{\alpha}{2}$: simplifions les notations en posant ici $1 \leq i \leq n$, $r^i = R_n^i(\omega)$ et $y^i = Y^i(\omega)$.

Par l'absurde, soit $i \in \{1, \dots, n\}$ tel que $|y^{(i)} - r^{(i)}| > \frac{\alpha}{2}$; on a $r^{(i)} = r^j$ et $y^{(i)} = y^k$ pour un certain couple $(j, k) \in \{1, \dots, n\}^2$ avec $j \neq k$ d'après (a).

Si $r^j < y^k$, on a $y^j < y^k$ car $|y^j - r^j| \leq \frac{\alpha}{2}$; et si $y^k < r^j$, on a $r^k < r^j$ car $|y^k - r^k| \leq \frac{\alpha}{2}$; quitte à intervertir la liste des r et celle des y et l'indice j avec l'indice k , on peut donc supposer $r^j < y^k$ et $y^j < y^k$.

Soit $r^{j^1} = r^{(1)} \leq \dots \leq r^{j^{i-1}} = r^{(i-1)}$ qui sont $i-1$ points les plus petits de la liste des r , tous inférieurs ou égaux à $r^j = r^{(i)}$. Comme les indices j^1, \dots, j^{i-1} sont 2 à 2 distincts et distincts de j , les points $y^{j^1}, \dots, y^{j^{i-1}}$ ne peuvent pas être $i-1$ points les plus petits de la liste des y inférieurs ou égaux à $y^k = y^{(i)}$ car $y^j < y^k$ est l'un d'entre eux. Il existe donc $\ell \in \{1, \dots, i-1\}$ tel que $y^{j^\ell} \geq y^{(i)}$, de sorte que $|y^{j^\ell} - r^{j^\ell}| \geq |y^k - r^j| > \frac{\alpha}{2}$: contradiction avec $|y^{j^\ell} - r^{j^\ell}| \leq \frac{\alpha}{2}$.

Notons $\bar{Y}_n = \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] Y^{(i)}$. Comme $\bar{R}_n = \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] R_n^{(i)}$, on a

$$|\bar{R}_n - \bar{Y}_n| \leq \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] |R_n^{(i)} - Y^{(i)}|;$$

et comme $\sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] = 1$, on déduit de (a) que

$$\text{presque sûrement } \forall \alpha > 0, \exists n_0(\alpha, \omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_0, |\bar{R}_n(\omega) - \bar{Y}_n(\omega)| \leq \frac{\alpha}{2} \quad (\text{b})$$

• La suite (Y^1, \dots, Y^n) constitue un échantillon réel positif dont on note F_n la *fonction aléatoire de répartition empirique*

$$F_n(\omega) = \sum_{i=1}^n \frac{1}{n} \mathbb{1}_{]Y^i(\omega), +\infty[} = \sum_{i=1}^{n-1} \frac{i}{n} \mathbb{1}_{]Y^{(i)}(\omega), Y^{(i+1)}(\omega)]} + \mathbb{1}_{]Y^{(n)}(\omega), +\infty[}.$$

Posons $\hat{F}_n(\omega) \stackrel{\text{déf}}{=} (F^\tau \circ F_n)(\omega) = \sum_{i=1}^{n-1} F^\tau\left(\frac{i}{n}\right) \mathbb{1}_{]Y^{(i)}(\omega), Y^{(i+1)}(\omega)]} + \mathbb{1}_{]Y^{(n)}(\omega), +\infty[}$. On a

$$1 - \hat{F}_n(\omega) = \mathbb{1}_{]-\infty, Y^{(1)}(\omega)]} + \sum_{i=1}^{n-1} \left[1 - F^\tau\left(\frac{i}{n}\right) \right] \mathbb{1}_{]Y^{(i)}(\omega), Y^{(i+1)}(\omega)]};$$

donc par transformation d'Abel,

$$\begin{aligned} \int_0^{+\infty} (1 - \hat{F}_n(\omega)(r)) dr &= Y^{(1)}(\omega) + \sum_{i=1}^{n-1} \left[1 - F^\tau\left(\frac{i}{n}\right) \right] (Y^{(i+1)}(\omega) - Y^{(i)}(\omega)) \\ (\text{car } F^\tau(1) = 1) &= F^\tau\left(\frac{1}{n}\right) Y^{(1)}(\omega) + \sum_{i=1}^{n-1} \left[F^\tau\left(\frac{i+1}{n}\right) - F^\tau\left(\frac{i}{n}\right) \right] Y^{(i+1)}(\omega) \\ (\text{car } F^\tau(0) = 0) &= \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] Y^{(i)}(\omega) = \bar{Y}_n(\omega). \end{aligned}$$

• On rappelle que

$$F^\tau(t) = 0 \text{ si } t \leq \frac{1}{2}, F^\tau(t) = \frac{2}{1-2\varepsilon} \left(t - \frac{1}{2} \right) \text{ sur } \left[\frac{1}{2}, 1 - \varepsilon \right] \text{ et } F^\tau(t) = 1 \text{ si } t \geq 1 - \varepsilon$$

Comme $\lim_{r \rightarrow +\infty} F(r) = 1$ (où F est la fonction de répartition de $R = |X - m_X|_2$), on a

$$\exists A > 0 \text{ tel que } F \geq 1 - \frac{\varepsilon}{2} \text{ sur } [A, +\infty[, \text{ de sorte que } F^\nu = F^\tau \circ F = 1 \text{ sur } [A, +\infty[.$$

D'après le *théorème de Glivenko - Cantelli*, presque sûrement la suite $(F_n(\omega))_{n \in \mathbb{N}^*}$ converge uniformément vers F . On a donc

$$\text{presque sûrement } \exists n_1(\omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_1, |F_n(\omega) - F| \leq \frac{\varepsilon}{2};$$

donc presque sûrement $F_n(\omega) \geq 1 - \varepsilon$ sur $[A, +\infty[$, de sorte que $\hat{F}_n(\omega) = F^\tau(F_n(\omega)) = 1$ sur $[A, +\infty[$. En conséquence

$$\text{p.s. } \exists n_1(\omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_1, \int_0^{+\infty} (\hat{F}_n(\omega)(r) - F^\nu(r)) dr = \int_0^A (\hat{F}_n(\omega)(r) - F^\nu(r)) dr$$

Comme F^τ est (uniformément) continue sur $[0, 1]$, presque sûrement la suite de fonctions $(\hat{F}_n(\omega))_{n \in \mathbb{N}^*}$ converge uniformément vers $F^\tau \circ F = F^\nu$. On a donc

$$\text{p.s. } \forall \alpha > 0, \exists n_2(\alpha, \omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_2, \left| \int_0^A (\hat{F}_n(\omega)(r) - F^\nu(r)) dr \right| \leq \frac{\alpha}{2};$$

comme (cf. la relation (3) du paragraphe 3 et le point 2)

$$E_\tau(R) - \bar{Y}_n(\omega) = \int_0^{+\infty} (\hat{F}_n(\omega)(r) - F^\nu(r)) dr,$$

on déduit des deux relations précédentes, en posant $n_3 \stackrel{\text{déf}}{=} \max[n_1, n_2]$:

$$\text{presque sûrement } \forall \alpha > 0, \exists n_3(\alpha, \omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_3, |\bar{Y}_n(\omega) - E_\tau(R)| \leq \frac{\alpha}{2} \quad (\text{c})$$

En rapprochant les relations (b) et (c), on obtient finalement en posant $n_4 \stackrel{\text{déf}}{=} \max[n_0, n_3]$:

$$\text{presque sûrement } \forall \alpha > 0, \exists n_4(\alpha, \omega) \in \mathbb{N}^* \text{ tel que } \forall n \geq n_4, |\bar{R}_n(\omega) - E_\tau(R)| \leq \alpha,$$

autrement dit la convergence presque sûre de $\bar{R}_n(\omega)$ vers $E_\tau(R)$. \square

6 Aspect statistique, la g -moyenne empirique

Nous définissons la g -moyenne empirique \bar{X}_n^g comme étant la variable aléatoire

$$\bar{X}_n^g(\omega) \stackrel{\text{déf}}{=} \frac{\int g\left(\frac{|x - M_n(\omega)|_2}{\bar{R}_n(\omega)}\right) x d\mu_n(\omega)(x)}{\int g\left(\frac{|x - M_n(\omega)|_2}{\bar{R}_n(\omega)}\right) d\mu_n(\omega)(x)},$$

soit

$$\bar{X}_n^g = \frac{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) X^i}{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right)} = M_n + \frac{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) (X^i - M_n)}{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right)}.$$

Proposition 6.1 (*convergence de l'estimateur*) : Si on n'est pas dans le "cas exceptionnel", on a convergence forte de l'indicateur g -moyenne empirique vers la g -moyenne de X pour $g = g_c$:

$$\text{presque sûrement } \bar{X}_n^g(\omega) \xrightarrow[n \rightarrow \infty]{} E^g(X).$$

Preuve

• On reprend les notations $1 \leq i \leq n$, $R_n^i = |X^i - M_n|_2$ et $Y^i = |X^i - m_X|_2$. Montrons d'abord que $\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) > 0$, de sorte que \bar{X}_n^g est bien défini et admet la double expression donnée ci-dessus.

En effet on a $\sum_{i=1}^n g\left(\frac{R_n^i(\omega)}{\bar{R}_n(\omega)}\right) = 0 \iff \forall i \in \{1, \dots, n\}, \frac{R_n^i(\omega)}{\bar{R}_n(\omega)} \geq c$. Si $\bar{R}_n(\omega) = 0$, on aurait $\forall i \in \{1, \dots, n\}, R_n^i(\omega) > 0$; et si $\bar{R}_n(\omega) > 0$, on aurait $\forall i \in \{1, \dots, n\}, R_n^i(\omega) \geq c \bar{R}_n(\omega)$. Or ceci est absurde dans les deux cas, puisque $\bar{R}_n = \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] R_n^{(i)}$ est une moyenne pondérée des $R_n^{(i)}$ et puisque $c > 1$.

• On fait dans un premier temps l'hypothèse $P(X = m_X) < 1 - \varepsilon$, c'est-à-dire (cf. la fin du paragraphe 3) $E_\tau(R) > 0$; de sorte que $\bar{R}_n > 0$ pour n assez grand d'après la proposition 5.2, et on considère seulement de tels n . On pose

$$A_n^i = \left\{ \frac{1}{c} R_n^i \leq \bar{R}_n \right\} \quad \text{et} \quad B_n^i = \left\{ \frac{1}{c} Y^i \leq E_\tau(R) \right\}.$$

Comme $|R_n^i - Y^i| \leq |M_n - m_X|_2$, on a

$$\begin{cases} \frac{1}{c} R_n^i \mathbb{1}_{A_n^i \cup B_n^i} \leq \bar{R}_n \vee \left(E_\tau(R) + \frac{|M_n - m_X|_2}{c} \right) \stackrel{\text{d\u00e9f}}{=} \bar{R}_n a_n \\ \frac{1}{c} Y^i \mathbb{1}_{A_n^i \cup B_n^i} \leq E_\tau(R) \vee \left(\bar{R}_n + \frac{|M_n - m_X|_2}{c} \right) \stackrel{\text{d\u00e9f}}{=} E_\tau(R) b_n \end{cases},$$

où presque sûrement $a_n(\omega) \xrightarrow[n \rightarrow \infty]{} 1$ et $b_n(\omega) \xrightarrow[n \rightarrow \infty]{} 1$ d'après les propositions 5.1 et 5.2.

– Montrons d'abord que presque sûrement $\frac{1}{n} \sum_{i=1}^n g\left(\frac{R_n^i(\omega)}{\bar{R}_n(\omega)}\right) \xrightarrow[n \rightarrow \infty]{} E\left(g\left(\frac{R}{E_\tau(R)}\right)\right)$.

On a d'une part, du fait que g est nulle à partir de c et est $\frac{2}{c}$ -lipschitzienne,

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) - \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) \right| &\leq \frac{1}{n} \sum_{i=1}^n \left| g\left(\frac{R_n^i}{\bar{R}_n}\right) - g\left(\frac{Y^i}{E_\tau(R)}\right) \right| \mathbb{1}_{A_n^i \cup B_n^i} \\ &\leq \frac{1}{n} \sum_{i=1}^n \frac{2}{c} \left| \frac{R_n^i}{\bar{R}_n} - \frac{Y^i}{E_\tau(R)} \right| \mathbb{1}_{A_n^i \cup B_n^i} \\ &\leq \frac{2}{nc} \frac{1}{\bar{R}_n} \sum_{i=1}^n \left(|R_n^i - Y^i| + \left| 1 - \frac{\bar{R}_n}{E_\tau(R)} \right| Y^i \mathbb{1}_{A_n^i \cup B_n^i} \right) \\ &\leq \frac{2}{nc} \frac{1}{\bar{R}_n} \sum_{i=1}^n \left(|M_n - m_X|_2 + \left| 1 - \frac{\bar{R}_n}{E_\tau(R)} \right| c E_\tau(R) b_n \right) \end{aligned}$$

$$\leq \frac{2}{c} \frac{1}{\bar{R}_n} |M_n - m_X|_2 + 2b_n \frac{|E_\tau(R) - \bar{R}_n|}{\bar{R}_n};$$

et cette dernière expression tend presque sûrement vers 0 lorsque n tend vers l'infini d'après les propositions 5.1 et 5.2.

D'autre part on a d'après la loi forte des grands nombres appliqué à la suite des variables réelles indépendantes et équidistribuées (d'espérances finies) $\left(g\left(\frac{Y^n}{E_\tau(R)}\right)\right)_{n \in \mathbb{N}^*}$

$$\text{presque sûrement } \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i(\omega)}{E_\tau(R)}\right) \xrightarrow{n \rightarrow \infty} E\left[g\left(\frac{|X - m_X|_2}{E_\tau(R)}\right)\right].$$

On en déduit finalement le résultat indiqué par addition.

– Montrons ensuite que

$$\text{presque sûrement } \frac{1}{n} \sum_{i=1}^n g\left(\frac{R_n^i(\omega)}{\bar{R}_n(\omega)}\right) (X^i(\omega) - M_n(\omega)) \xrightarrow{n \rightarrow \infty} E\left(g\left(\frac{R}{E_\tau(R)}\right) (X - m_X)\right).$$

$$\begin{aligned} \text{On a d'une part } & \left| \frac{1}{n} \sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) (X^i - M_n) - \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) (X^i - M_n) \right|_2 \\ & \text{(cf. le calcul précédent)} \leq \frac{1}{n} \sum_{i=1}^n \left| g\left(\frac{R_n^i}{\bar{R}_n}\right) - g\left(\frac{Y^i}{E_\tau(R)}\right) \right| R_n^i \mathbb{1}_{A_n^i \cup B_n^i} \\ & \leq c \bar{R}_n a_n \left(\frac{1}{n} \sum_{i=1}^n \left| g\left(\frac{R_n^i}{\bar{R}_n}\right) - g\left(\frac{Y^i}{E_\tau(R)}\right) \right| \mathbb{1}_{A_n^i \cup B_n^i} \right) \\ & \text{(cf. le calcul précédent)} \leq c \bar{R}_n a_n \left(\frac{2}{c} \frac{1}{\bar{R}_n} |M_n - m_X|_2 + 2b_n \frac{|E_\tau(R) - \bar{R}_n|}{\bar{R}_n} \right); \end{aligned}$$

et cette dernière expression tend presque sûrement vers 0 lorsque n tend vers l'infini d'après les propositions 5.1 et 5.2.

On a d'autre part, du fait que g est majorée par 1,

$$\begin{aligned} & \left| \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) (X^i - M_n) - \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) (X^i - m_X) \right|_2 \\ & = \frac{1}{n} \left| \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) (M_n - m_X) \right|_2 \\ & \leq \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i}{E_\tau(R)}\right) |M_n - m_X|_2 \leq |M_n - m_X|_2 \end{aligned}$$

qui tend aussi presque sûrement vers 0 lorsque n tend vers l'infini d'après la proposition 5.1.

De plus d'après la loi forte des grands nombres appliqué à la suite des variables réelles indépendantes et équidistribuées (d'espérances finies) $\left(g\left(\frac{Y^n}{E_\tau(R)}\right) (X^n - m_X)\right)_{n \in \mathbb{N}^*}$, on a

$$\text{presque sûrement } \frac{1}{n} \sum_{i=1}^n g\left(\frac{Y^i(\omega)}{E_\tau(R)}\right) (X^i(\omega) - m_X) \xrightarrow{n \rightarrow \infty} E\left[g\left(\frac{|X - m_X|_2}{E_\tau(R)}\right) (X - m_X)\right].$$

On en déduit finalement le résultat indiqué par addition.

– Comme $E\left(g\left(\frac{R}{E_\tau(R)}\right)\right) > 0$, on déduit des convergences presque sûres établies au deux sous-points précédents et en prenant leur quotient que

presque sûrement $\bar{X}_n^g(\omega) - M_n(\omega) \xrightarrow[n \rightarrow \infty]{} E(\mathcal{V}_X^g X) - m_X$;

de sorte que (proposition 5.1) presque sûrement $\bar{X}_n^g(\omega) \xrightarrow[n \rightarrow \infty]{} E(\mathcal{V}_X^g X) = E^g(X)$.

• Supposons pour terminer que $P(X = m_X) \geq 1 - \varepsilon$, de sorte que (cf. le paragraphe 4) $E_\tau(R) = 0$ et $\mathcal{V}_X^g = \frac{1}{P(R=0)} \mathbb{1}_{\{R=0\}}$, donc $E(\mathcal{V}_X^g X) = m_X$. Reprenant les considérations de début de preuve, notons

$$H^n \stackrel{\text{d\u00e9f}}{=} \sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) > 0 \text{ et } 1 \leq i \leq n, K_n^i \stackrel{\text{d\u00e9f}}{=} \frac{1}{H^n} g\left(\frac{R_n^i}{\bar{R}_n}\right) \mathbb{1}_{\{R_n^i < c \bar{R}_n\}} \geq 0.$$

$$\text{On a } \bar{X}_n^g = M_n + \frac{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right) (X^i - M_n)}{\sum_{i=1}^n g\left(\frac{R_n^i}{\bar{R}_n}\right)} = M_n + \sum_{i=1}^n K_n^i (X^i - M_n) \text{ avec } \sum_{i=1}^n K_n^i = 1 ; \text{ donc}$$

$$|\bar{X}_n^g - M_n|_2 \leq \sum_{i=1}^n K_n^i R_n^i \leq \sum_{i=1}^n K_n^i c \bar{R}_n = c \bar{R}_n.$$

On en d\u00e9duit que presque sûrement $\bar{X}_n^g(\omega) \xrightarrow[n \rightarrow \infty]{} m_X = E^g(X)$ d'apr\u00e8s les propositions 5.1 et 5.2 et puisque $E_\tau(R) = 0$. \square

En pratique, \u00e9tant donn\u00e9 un \u00e9chantillon $\{x^1, \dots, x^n\}$ de taille n de \mathbb{R}^d , on calcule

$$\bar{x}_n^g = \frac{1}{n} \sum_{i=1}^n v_n^i x^i$$

o\u00f9 m_n est la m\u00e9diane empirique de $\{x^1, \dots, x^n\}$ et o\u00f9 pour tout $i \in \{1, \dots, n\}$,

$$r_n^i \stackrel{\text{d\u00e9f}}{=} |x^i - m_n|_2, \text{ d'o\u00f9 } \bar{r}_n \stackrel{\text{d\u00e9f}}{=} \sum_{j=1}^n \left[F^\tau\left(\frac{j}{n}\right) - F^\tau\left(\frac{j-1}{n}\right) \right] r_n^{(j)}, \text{ et } v_n^i \stackrel{\text{d\u00e9f}}{=} \frac{g\left(\frac{r_n^i}{\bar{r}_n}\right)}{\frac{1}{n} \sum_{j=1}^n g\left(\frac{r_n^j}{\bar{r}_n}\right)}.$$

On peut voir dans l'expression de \bar{x}_n^g une forme de L -estimation en dimension d . \bar{x}_n^g est la *centre de gravit\u00e9* du nuage des points x^1, \dots, x^n affect\u00e9s des poids $\frac{v_n^1}{n}, \dots, \frac{v_n^n}{n}$.

On notera qu'\u00e0 la r\u00e9partition de probabilit\u00e9 empirique $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x^i}$ port\u00e9e par l'\u00e9chantillon $\{x^1, \dots, x^n\}$ correspond la probabilit\u00e9 $\mu_n^g \stackrel{\text{d\u00e9f}}{=} \frac{1}{n} \sum_{i=1}^n v_n^i \delta_{x^i}$ absolument continue par rapport \u00e0 μ_n . On a bien

$$\int_{\mathbb{R}^d} x d\mu_n^g(x) = \sum_{i=1}^n x^i \mu_n^g(\{x^i\}) = \frac{1}{n} \sum_{i=1}^n v_n^i x^i = \bar{x}_n^g.$$

Remarque 6.2 Dans une conception fr\u00e9quentiste forcen\u00e9e, on peut voir dans μ_n^g la r\u00e9partition de probabilit\u00e9 empirique d'un pseudo-\u00e9chantillon : en admettant que les poids $\frac{v_n^i}{n}$ sont assimilables \u00e0 des nombres rationnels, il existe un (plus petit) entier $N \in \mathbb{N}^*$ v\u00e9rifiant

$$\forall i \in \{1, \dots, n\}, \exists I(i) \in \mathbb{N} \text{ tel que } \frac{v_n^i}{n} = \frac{I}{N} ;$$

le pseudo-échantillon en question est formé à partir de $\{x^1, \dots, x^n\}$ en répétant chaque x_n^i exactement $I(i)$ fois — d'où la fréquence expérimentale (fictive!) $\frac{I}{N}$.

Point d'effondrement de la g -moyenne empirique :

Il est assez facile de voir que le point d'effondrement de la médiane euclidienne empirique M_n est (asymptotiquement) égal à 0.5, car les variations de M_n restent bornées si et seulement si moins de la moitié de l'échantillon est modifiée sans restriction.

On se convainc aussi facilement que le point d'effondrement de l'indicateur d'échelle empirique $\bar{R}_n = \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] R_n^{(i)}$ est au moins égal à ε ($0 < \varepsilon < \frac{1}{2}$), car les variations de \bar{R}_n restent bornées quand des points de l'échantillon s'éloignent sans restriction tant qu'ils sont en proportion inférieure à ε .

De là, du fait que $\sup_{z \in \mathbb{R}_+} z g(z) < +\infty$, il en est de même pour la g -moyenne empirique.

Montrons plus précisément que le **point d'effondrement ε^* de \bar{X}_n^g est égal à ε** , en prouvant qu'il est inférieur à tout $\varepsilon' \in]\varepsilon, \frac{1}{2}[$.

Pour tout entier $n > \frac{4}{1-2\varepsilon}$, il existe un entier k tel que $\varepsilon < \frac{k}{n} < \frac{k+1}{n} < \frac{1}{2}$: en effet si n est pair, $k = \frac{n}{2} - 2$ vérifie $\frac{k+1}{n} < \frac{1}{2}$ et $k > \frac{2}{1-2\varepsilon} - 2 = \frac{4\varepsilon}{1-2\varepsilon}$, donc $k > \varepsilon(2k+4) = \varepsilon n$; et si n est impair, $k = \frac{n-3}{2}$ vérifie $\frac{k+1}{n} < \frac{1}{2}$ et $k > \frac{\frac{3}{2}}{1-2\varepsilon} - \frac{3}{2} = \frac{3\varepsilon}{1-2\varepsilon}$, donc $k > \varepsilon(2k+3) = \varepsilon n$.

Il existe donc un plus petit entier, encore noté k , tel que $\varepsilon < \frac{k}{n} < \frac{k+1}{n} < \frac{1}{2}$; il vérifie $\frac{k}{n} - \varepsilon \leq \frac{1}{n}$ et donc $\frac{k+1}{n} \leq \varepsilon + \frac{2}{n}$. Pour tout $\varepsilon' \in]\varepsilon, \frac{1}{2}[$ on peut donc trouver $n > \frac{4}{1-2\varepsilon}$ tel que de plus $\frac{k+1}{n} < \varepsilon'$.

On considère alors l'échantillon monodimensionnel suivant :

$$x^1 = \dots = x^{n-k-1} = 0, \quad x^{n-k} = a \quad \text{et} \quad x^{n-k+1} = \dots = x^n = b \quad \text{avec} \quad 0 < a < b.$$

On a évidemment $m_n \in [0, b]$, et comme pour tout m dans cet intervalle,

$$\begin{aligned} \sum_{i=1}^n |m - x^i|_2 &= (n-k-1)m + |m-a| + k(b-m) \\ &= kb + (n-1-2k)m + |m-a| \end{aligned}$$

avec $2k+2 < n \implies n-1-2k > 1$, on a $m_n = 0$.

On a donc $(r_n^{(1)}, \dots, r_n^{(n)}) = (r_n^1, \dots, r_n^n) = (x^1, \dots, x^n)$, de sorte que

$$\bar{r}_n = \sum_{i=1}^n \left[F^\tau\left(\frac{i}{n}\right) - F^\tau\left(\frac{i-1}{n}\right) \right] r_n^{(i)} \geq \left[F^\tau\left(\frac{n-k+1}{n}\right) - F^\tau\left(\frac{n-k}{n}\right) \right] b \stackrel{\text{déf}}{=} Cb$$

car $x^{n-k+1} = b$, avec $C \leq 1$ et aussi $C > 0$ car $\frac{n-k}{n} < 1-\varepsilon \leq \frac{n-k+1}{n}$. On a donc $0 < \bar{r}_n \leq b$, et on a en choisissant $a = Cb$:

$$\sum_{i=1}^n g\left(\frac{r_n^i}{\bar{r}_n}\right) x^i \geq g\left(\frac{x^{n-k}}{\bar{r}_n}\right) x_{n-k} \geq g(1) C b > 0, \text{ et } \sum_{i=1}^n g\left(\frac{r_n^i}{\bar{r}_n}\right) \leq n \text{ car } g \leq 1;$$

de sorte que $\bar{x}_n^g = \frac{1}{\sum_{i=1}^n g\left(\frac{r_n^i}{\bar{r}_n}\right)} \sum_{i=1}^n g\left(\frac{r_n^i}{\bar{r}_n}\right) x^i \geq g(1) \frac{C}{n} b$, qui comme b est non borné.

Ainsi le point d'effondrement est dépassé par la proportion $\frac{k+1}{n} < \varepsilon'$.

7 Exemples

• $d = 1$, X variable de Bernoulli de paramètre $P(X = 1) = p$ ($0 < p < 1$) et τ de densité $\frac{1}{1-2\varepsilon} \mathbb{1}_{] \varepsilon, 1-\varepsilon[}$ (d'où $F^\tau(t) = \frac{2}{1-2\varepsilon} (t - \frac{1}{2})$ sur $[\frac{1}{2}, 1 - \varepsilon]$) pour $0 < \varepsilon < \frac{1}{2}$.

– Pour $p = \frac{1}{2}$, on a $m_X = \frac{1}{2}$ et $R = \frac{1}{2}$; donc $E_\tau(R) = \frac{1}{2}$, $g\left(\frac{R}{E_\tau(R)}\right) = g(1) > 0$ et $\mathcal{V}_X^g \stackrel{\text{déf}}{=} \frac{1}{E(g(\frac{R}{E_\tau(R)})} g\left(\frac{R}{E_\tau(R)}\right) = 1$.

Ainsi $\mathcal{V}_X^g X = X$, donc $\mu^g = \mu = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1$ et $E^g(X) = E(X) = \frac{1}{2}$ (ouf!).

– Pour $0 < p < \frac{1}{2}$, on a $m_X = 0$ et $R = X$; $F(r) = 1 - p$ sur $[0, 1[$ et $F(r) = 1$ sur $[1, +\infty[$, donc $E_\tau(R) = \int_0^1 [1 - F^\tau(1-p)] dr$ d'après la relation (3) du paragraphe 3. Comme $\mu = (1-p) \delta_0 + p \delta_1$, on conclut que :

* Pour $0 < p \leq \varepsilon$, on a $E_\tau(R) = 0$, $\mu^g = \frac{1}{1-p} \mathbb{1}_{\{0\}} \mu = \delta_0$ et $E^g(X) = 0$ (cf. remarque 4.2).

* Pour $\varepsilon < p < \frac{1}{2}$, on a $E_\tau(R) = \frac{2(p-\varepsilon)}{1-2\varepsilon} > 0$ et

$$g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) = \begin{cases} g(0) = 1 \text{ pour } x = 0 \\ g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right) \text{ pour } x = 1 \end{cases};$$

donc $g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) \mu(dx) = (1-p) \delta_0(dx) + p g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right) \delta_1(dx)$

et pour raison de normalisation,

$$\mu^g = \frac{1}{1-p + p g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right)} \left[(1-p) \delta_0 + p g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right) \delta_1 \right];$$

soit μ^g loi de Bernoulli de paramètre

$$E^g(X) = \int x d\mu^g(x) = \frac{p g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right)}{1-p + p g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right)} \in \left[0, \frac{1}{2}\right].$$

On voit que $E^g(X)$ tend vers 0 quand p tend vers ε (car $g(+\infty) = 0$). On voit aussi que si g est continue sur $]1, +\infty[$, $E^g(X)$ est comme $E(X)$ une fonction continue de $p \in]\varepsilon, \frac{1}{2}[$; et que si $g(1^+) = 1$ (de sorte que g décroissante vaut 1 sur $[0, 1]$), $E^g(X)$ est aussi continue à gauche en $p = \frac{1}{2}$ comme $E(X) = p$.

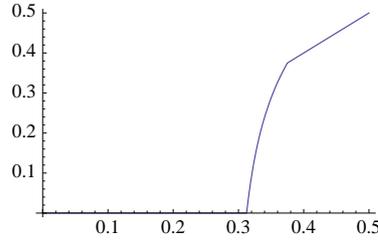
On peut trouver naturel d'imposer l'égalité $E^g(X) = E(X) = p$ par exemple sur un sous-intervalle de longueur $\frac{c-2}{c}$ fois celle de l'intervalle $]\varepsilon, \frac{1}{2}[$ où $c > 2$; autrement dit à partir de $p = \frac{2}{c}\frac{1}{2} + \frac{c-2}{c}\varepsilon = \varepsilon + \frac{1-2\varepsilon}{c}$. Ceci est obtenu si et seulement si de surcroît g vaut 1 sur $[0, \frac{c}{2}]$.

Pour $g = g_c$ (cf. le paragraphe 4), on a $g(t) = 2(1 - \frac{t}{c}) \mathbb{1}_{\{t < c\}}$ lorsque $t > \frac{c}{2}$, et les calculs donnent pour $p < \varepsilon + \frac{1-2\varepsilon}{c}$:

$$g\left(\frac{1-2\varepsilon}{2(p-\varepsilon)}\right) = \left(2 - \frac{1-2\varepsilon}{c(p-\varepsilon)}\right) \mathbb{1}_{\{p > \varepsilon + \frac{1-2\varepsilon}{2c}\}};$$

d'où
$$E^g(X) = \mathbb{1}_{\{p > \varepsilon + \frac{1-2\varepsilon}{2c}\}} \frac{2 - \frac{1-2\varepsilon}{c(p-\varepsilon)}}{\frac{1-p}{p} + 2 - \frac{1-2\varepsilon}{c(p-\varepsilon)}}.$$

Ainsi $E^g(X) = 0$ pour $p \leq \varepsilon + \frac{1-2\varepsilon}{2c}$ (et $E^g(X) = p$ pour $\varepsilon + \frac{1-2\varepsilon}{c} \leq p \leq \frac{1}{2}$).



$E^g(X)$ en fonction de $p \in]0, \frac{1}{2}[$ pour $\varepsilon = 0.25$ et $c = 4$

- Pour $\frac{1}{2} < p < 1$, on a $m_X = 1$ et $R = 1 - X$. Les calculs et les résultats (non détaillés ici) sont analogues aux précédents.

En conclusion l'instabilité de la médiane euclidienne au voisinage de $p = \frac{1}{2}$ est ici "absorbée" par la fonction élémentaire proposée $g = g_c$, la longueur 1 étant la longueur minimum du plateau initial de g pour que \mathcal{V}_X^g ne dépendent pas de la convention que m_X soit le milieu du segment médian dans le "cas exceptionnel"; une longueur égale ou supérieure à 2 pour ce plateau paraissant raisonnable en pratique.

- $d = 2$, échantillon de $n = 3$ points :

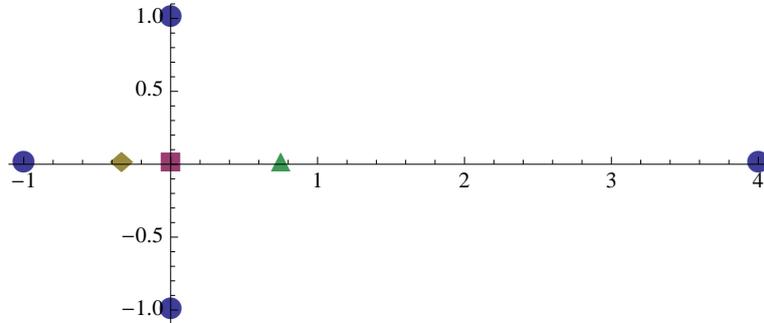
La médiane euclidienne m_3 n'est autre que le *point de Fermat* (dit aussi *point de Torricelli*) du triangle x^1, x^2, x^3 , lorsque les angles de ce triangle sont tous inférieurs à $\frac{\pi}{3}$; c'est-à-dire le point duquel on voit les 3 côtés sous le même diamètre (angulaire)

apparent $\frac{2\pi}{3}$ (de sorte que $\frac{m_3 - x^1}{r_3^1} + \frac{m_3 - x^2}{r_3^2} + \frac{m_3 - x^3}{r_3^3} = 0$). Sinon m_3 est le sommet du triangle dont l'angle est supérieur à $\frac{\pi}{3}$.

- $d = 2$, $\varepsilon = 0.25$, $c = 4$, échantillon de $n = 4$ points :

$$\{x^1 = (-1, 0), x^2 = (0, 1), x^3 = (0, -1), x^4 = (4, 0)\}.$$

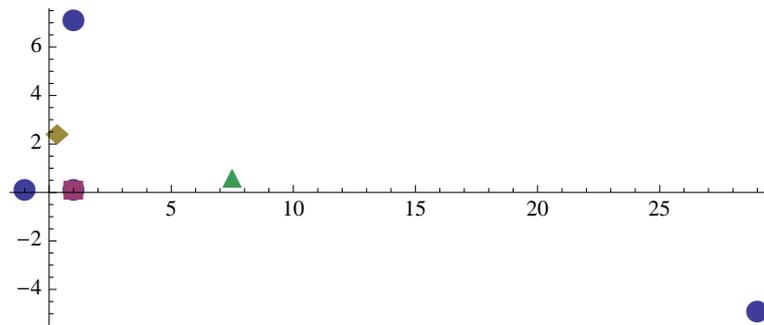
Les 4 points de l'échantillon sont les ronds bleus, la médiane euclidienne empirique est le carré rouge, la moyenne empirique est le triangle vert et la g_c -moyenne empirique est le losange jaune : ce dernier point est la moyenne empirique de $\{x^1, x^2, x^3\}$ par rejet du point lointain x^4 . De sorte que, bien que les 3 indicateurs de position soient alignés, la g_c -moyenne n'est pas entre la médiane et la moyenne (et il en serait exactement de même avec la configuration véritablement monodimensionnelle où $x^2 = x^3 = (0, 0)$).



• $d = 2$, $\varepsilon = 0.25$, $c = 4$, échantillon de $n = 4$ points :

$$\{x^1 = (-1, 0), x^2 = (1, 0), x^3 = (1, 7), x^4 = (29, -5)\}.$$

Dans le même esprit et avec le même code couleur, pour une configuration “vraiment bidimensionnelle”. La médiane euclidienne empirique est en x^2 .



• $d = 2$, $\varepsilon = 0.25$, $c = 4$, échantillons de $n = 5$ points :

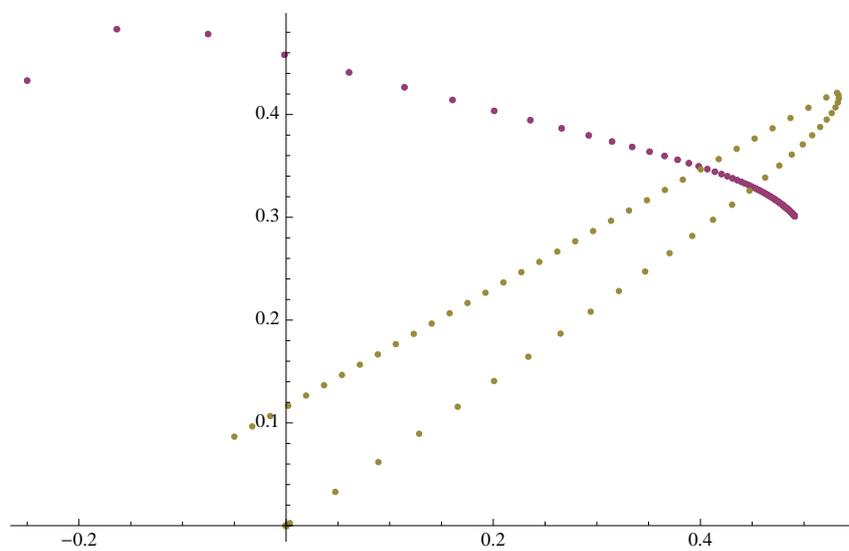
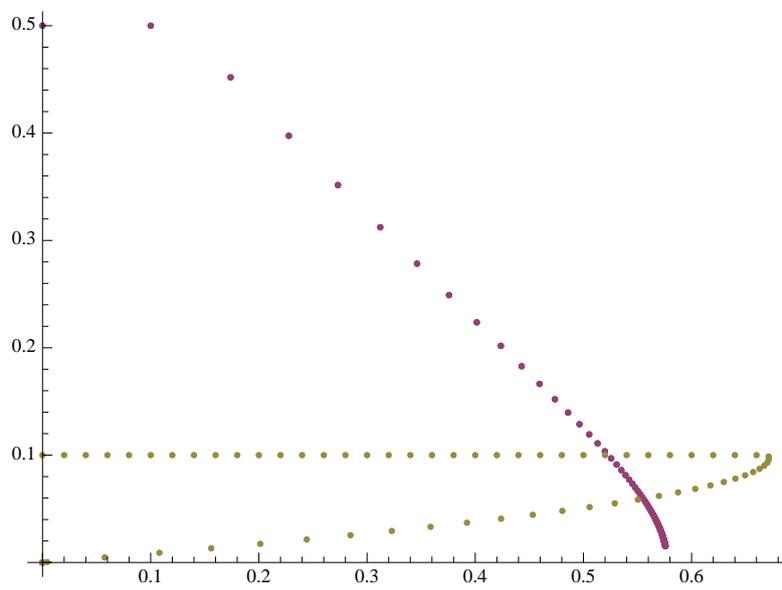
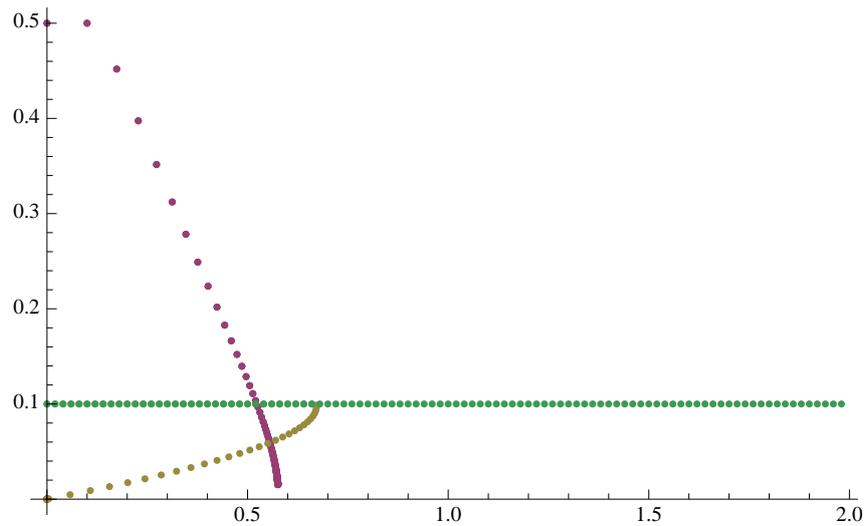
On considère la suite d'échantillons de 5 points de \mathbb{R}^2 dépendant de $k \in \{0, 100\}$:

$$\{x^1 = (-1, 0), x^2 = (1, 0), x^3 = (0, -1), x^4 = (0, 1), x^5 = (\frac{k}{10}, 0.5)\}.$$

Sur la première figure apparaissent les médianes euclidiennes empiriques (points rouges approximativement alignés en biais) ; les moyennes empiriques (points verts alignés horizontalement) ; et les g_c -moyennes empiriques (points jaunes). On voit que pour les petites valeurs de k , moyennes et g_c -moyennes empiriques coïncident parfaitement.

Sur la seconde figure en repère orthonormé, on a omis les moyennes empiriques, pour voir plus en détail le processus de retour des g -moyennes empiriques vers $(0, 0)$ — qui est la moyenne des 4 premiers points de l'échantillon — au fur et à mesure que le cinquième point devient plus clairement un outlier.

La troisième figure — toujours en repère orthonormé — illustre l'équivariance de l'estimateur par rotation (de $\frac{\pi}{6}$).



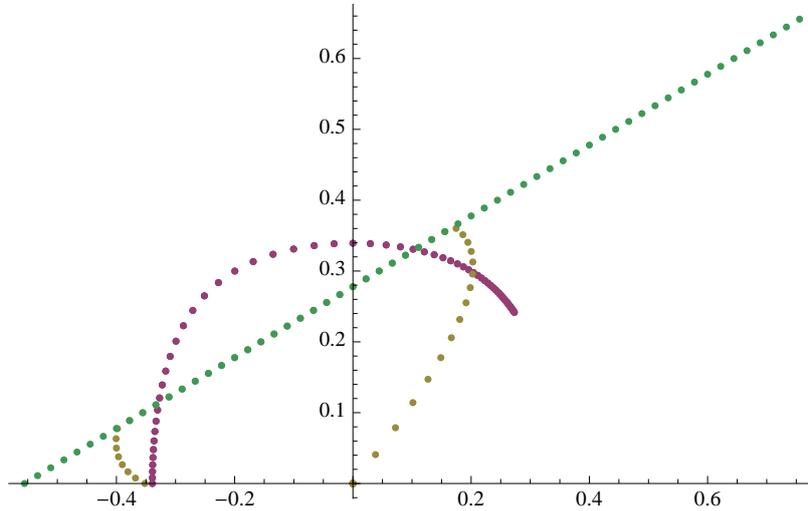
- $d = 2$, $\varepsilon = 0.25$, $c = 4$, échantillons de $n = 9$ points :

On considère la suite d'échantillons de 9 points de \mathbb{R}^2 dépendant de $k \in \{0, 60\}$:

$$\{x^1 = x^2 = (1, 1), x^3 = x^4 = (-1, 1), x^5 = x^6 = (-1, -1),$$

$$x^7 = x^8 = (1, -1), x^9 = (\frac{k}{5} - 5, \frac{k}{10})\}.$$

Les codes couleurs sont les mêmes que précédemment.



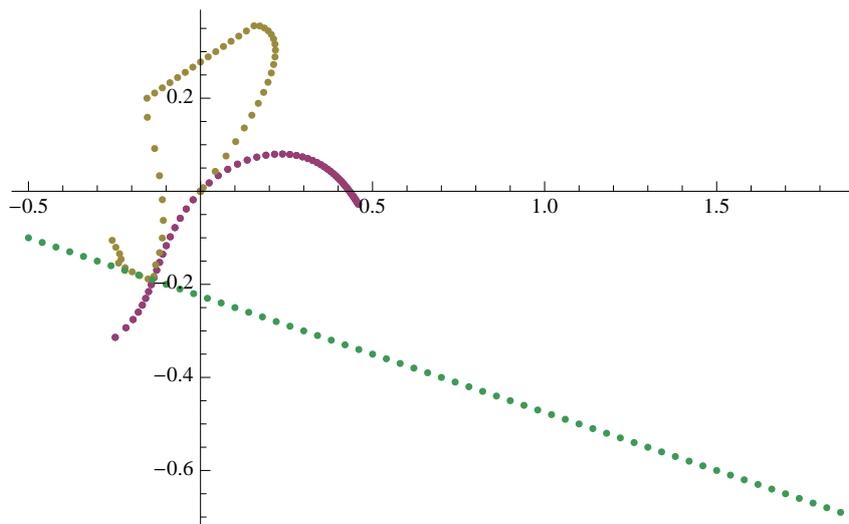
- $d = 2$, $\varepsilon = 0.25$, $c = 4$, échantillons de $n = 10$ points :

On considère la suite d'échantillons de 10 points de \mathbb{R}^2 dépendant de $k \in \{0, 60\}$:

$$\{x^1 = x^2 = (1, 1), x^3 = x^4 = (-1, 1), x^5 = x^6 = (-1, -1),$$

$$x^7 = x^8 = (1, -1), x^9 = (\frac{k}{5} - 5, \frac{k}{10}), x^{10} = (\frac{k}{5}, -\frac{k}{5} - 1)\}.$$

Les codes couleurs sont les mêmes que précédemment.



8 Construction d'une pondération robuste, phase 2

On reprend le fil du paragraphe 4. Le fait de ne pas considérer la distribution angulaire des points aléatoires par rapport à la médiane m_X est regrettable, même si ce n'est pas essentiel lorsqu'on se focalise sur l'estimation robuste de la moyenne.

On a vu au passage qu'une façon profonde d'analyser le problème consiste à affecter à chaque valeur $X(\omega)$ de X un coefficient de confiance $\mathcal{V}_X(\omega)$, qui dépend de la variable X , tel que la variable aléatoire \mathcal{V}_X ainsi définie vérifie $E(\mathcal{V}_X) = 1$. La variable \mathcal{V}_X^g construite au paragraphe 4 exprime la seule "confiance radiale", mais rien n'interdit de chercher à en tirer parti pour définir **dans un second temps** une nouvelle variable \mathcal{W}_X de "confiance radio-angulaire", permettant de rejeter ou d'atténuer plus finement le poids de certaines valeurs de X .

Voici une proposition en ce sens, en supposant que g vérifie les propriétés (*) du paragraphe 4 où la dernière est renforcée en $\sup_{z \in \mathbb{R}_+} z^2 g(z) \stackrel{\text{déf}}{=} M_2 < +\infty$ (propriété trivialement vraie si g est nulle à partir d'une valeur finie $c > 1$) : en effet X de loi μ^g est alors de carré intégrable car

$$\begin{aligned} E^g(|X - m_X|_2^2) &= \frac{1}{E(g(\frac{R}{E_\tau(R)})})} \int |x - m_X|_2^2 g\left(\frac{|x - m_X|_2}{E_\tau(R)}\right) \mu(dx) \\ &\leq \frac{M_2 E_\tau(R)^2}{E(g(\frac{R}{E_\tau(R)})})} < +\infty. \end{aligned}$$

On peut ainsi considérer légitimement non seulement l'espérance $E^g(X)$, mais aussi la matrice de covariance de X de taille d , qu'on notera K_X^g pour indiquer qu'on se réfère aussi à la loi μ^g et non à μ .

Pour tout vecteur unitaire u de \mathbb{R}^d déterminant une direction, l'indicateur de position dans cette direction u est

$$E^g(X \cdot u) = \int x \cdot u d\mu^g(x) = \int x d\mu^g(x) \cdot u = E^g(X) \cdot u;$$

et on peut prendre comme indicateur d'échelle dans la direction u , en introduisant la variable g -centrée $Y \stackrel{\text{déf}}{=} X - E^g(X)$:

$$\text{var}^g(X \cdot u) = E^g((Y \cdot u)^2) = {}^t u E^g(Y {}^t Y) u = {}^t u K_X^g u.$$

On suppose pour simplifier que la matrice K_X^g est **inversible**, de sorte qu'on a ${}^t u K_X^g u > 0$ pour toute direction u . On va caractériser la confiance en la valeur $X(\omega)$ dans la direction u par le rapport

$$S_u(\omega) \stackrel{\text{déf}}{=} \frac{(Y(\omega) \cdot u)^2}{E^g((Y \cdot u)^2)} = \frac{({}^t u Y(\omega))^2}{{}^t u K_X^g u} \geq 0,$$

la confiance étant d'autant plus grande que le rapport est petit ; et la confiance globale en $X(\omega)$ sera fondée sur le nombre $S(\omega) \stackrel{\text{déf}}{=} \sup_{|u|_2=1} S_u(\omega) \geq 0$.

En fait il est clair que la condition $|u|_2 = 1$ n'intervient pas vraiment dans la définition de $S(\omega)$, du fait que le numérateur et le dénominateur de $S_u(\omega)$ sont des fonctions

homogènes de même degré 2 en u : on peut donc remplacer cette condition $|u|_2 = 1$ par $u \neq 0$ ou encore par $Y(\omega) \cdot u = c$, c étant une constante réelle arbitraire non nulle (car le cas $S_u(\omega) = 0$ peut être ignoré). On est alors ramené au *problème standard de minimisation sous contrainte*

$$(\mathcal{P}) \quad \left| \begin{array}{l} \min \frac{1}{2} {}^t u K_X^g u \\ {}^t u Y(\omega) = c \end{array} \right. .$$

Du fait de la convexité du problème, on atteint bien un minimum absolu sous ladite contrainte en un point $u \in \mathbb{R}^d$; et d'après la théorie des multiplicateurs de Lagrange, u vérifie par dérivation la relation $K_X^g u + \lambda Y(\omega) = 0$ pour une valeur réelle convenable du multiplicateur λ . La matrice K_X^g étant supposée inversible, ceci équivaut à

$$u = -\lambda (K_X^g)^{-1} Y(\omega), \text{ avec } -\lambda {}^t Y(\omega) (K_X^g)^{-1} Y(\omega) = c.$$

Ainsi $-\lambda = \frac{c}{{}^t Y(\omega) (K_X^g)^{-1} Y(\omega)}$, d'où $u = c \frac{(K_X^g)^{-1} Y(\omega)}{{}^t Y(\omega) (K_X^g)^{-1} Y(\omega)}$; et pour cette valeur de u , on a ${}^t u K_X^g u = c^2 \frac{{}^t Y(\omega) (K_X^g)^{-1} K_X^g (K_X^g)^{-1} Y(\omega)}{({}^t Y(\omega) (K_X^g)^{-1} Y(\omega))^2} = \frac{c^2}{{}^t Y(\omega) (K_X^g)^{-1} Y(\omega)}$.

On obtient finalement l'expression explicite de $S(\omega)$, où la constante c n'intervient pas :

$$S(\omega) = \frac{{}^t Y(\omega) (K_X^g)^{-1} Y(\omega)}{c^2}.$$

Par ailleurs la matrice K_X^g étant symétrique réelle, il existe une matrice P orthogonale réelle (${}^t P = P^{-1}$) telle que ${}^t P K_X^g P = D$, matrice diagonale. Posons $Z \stackrel{\text{déf}}{=} {}^t P Y$; Z est comme $Y = X - E^g(X)$ une variable vectorielle g -centrée, et sa matrice de g -covariance est

$$K_Z^g = E^g(Z {}^t Z) = E^g({}^t P Y {}^t Y P) = {}^t P E^g(Y {}^t Y) P = {}^t P K_X^g P = D.$$

Ainsi les composantes Z_1, \dots, Z_d de Z sont non corrélées linéairement et les coefficients diagonaux de D sont égaux respectivement à $E^g(Z_1^2), \dots, E^g(Z_d^2)$; tous termes strictement positifs car D est inversible comme K_X^g .

Comme $D = {}^t P K_X^g P$, on a $D^{-1} = P^{-1} (K_X^g)^{-1} ({}^t P)^{-1} = {}^t P (K_X^g)^{-1} P$; et comme $Y = ({}^t P)^{-1} Z = P Z$, on a

$${}^t Y(\omega) (K_X^g)^{-1} Y(\omega) = {}^t Z(\omega) {}^t P (K_X^g)^{-1} P Z(\omega) = {}^t Z(\omega) D^{-1} Z(\omega);$$

soit $S(\omega) = \sum_{\ell=1}^d \frac{1}{E^g(Z_\ell^2)} Z_\ell(\omega)^2$; de sorte que $E^g(S) = \sum_{\ell=1}^d \frac{E^g(Z_\ell^2)}{E^g(Z_\ell^2)} = d$.

Remarque 8.1 Dans le cas où la matrice K_X^g n'est pas inversible, la matrice diagonale $K_Z^g = D$ définie ci-dessus a un nombre $d' < d$ de termes diagonaux non nuls, qu'on peut supposer occuper les d' premières rangées. Alors $\forall \ell \in \{d'+1, \dots, d\}$, on a $E^g(Z_\ell^2) = 0$ i.e. $Z_\ell \stackrel{P_X^g\text{-ps}}{=} 0$. On peut alors déterminer S via la variable g -centrée $Z' \stackrel{\text{déf}}{=} (Z_1, \dots, Z_{d'})$ à valeurs dans $\mathbb{R}^{d'}$ — dont la matrice $K_{Z'}^g$, sous-matrice de taille d'

de K_X^g est diagonale inversible — pour revenir ensuite à la variable $X = PZ + E^g(X)$. X est donc P_X^g -ps à valeurs dans le sous-espace affine de dimension d' de \mathbb{R}^d translaté par $E^g(X)$ de l'espace vectoriel engendré par les d' premiers vecteurs colonnes de P , qui sont les vecteurs propres de K_X^g associés aux valeurs propres $E^g(Z_\ell^2) > 0$.

Étant donné une fonction h vérifiant les mêmes propriétés (*) que g (cf. le paragraphe 4), on prend assez naturellement comme variable de confiance (ou de pondération) définitive

$$\mathcal{W}_X^{g,h} \stackrel{\text{déf}}{=} \frac{1}{E[h(\frac{S}{d})]} h\left(\frac{S}{d}\right) = \psi_X^{g,h}(X),$$

$$\text{où } \psi_X^{g,h}(x) \stackrel{\text{déf}}{=} \frac{1}{E[h(\frac{S}{d})]} h\left(\frac{1}{d} {}^t(x - E^g(X)) (K_X^g)^{-1} (x - E^g(X))\right) \geq 0.$$

On a ainsi $\mathcal{W}_X^{g,h} \geq 0$, et $E(\mathcal{W}_X^{g,h}) = 1$: en effet sinon on aurait $E[h(\frac{S}{d})] = 0$, soit $h(\frac{S}{d}) \stackrel{\text{ps}}{=} 0$; on en déduirait $S \stackrel{\text{ps}}{>} d$ puisque h décroissante et $h(1) > 0$, et donc $E(\mathcal{V}_X^g S) > d E(\mathcal{V}_X^g) = d$: ce qui est incompatible avec $E(\mathcal{V}_X^g S) = E^g(S) = d$ (cf. ci-dessus).

On considère donc la nouvelle probabilité de référence qui se substitue à P_X^g

$$\underline{P_X^{g,h} = \mathcal{W}_X^{g,h} P}$$

et la nouvelle loi de probabilité $\mu^{g,h} = X(P_X^{g,h})$ de X associée à $P_X^{g,h}$, soit

$$\mu^{g,h} = \psi_X^{g,h} \mu.$$

$$\text{On a } \begin{cases} E^{g,h}(X) \stackrel{\text{déf}}{=} \int_{\mathbb{R}^d} x d\mu^{g,h}(x) = E(\mathcal{W}_X^{g,h} X) \\ K_X^{g,h} \stackrel{\text{déf}}{=} \int_{\mathbb{R}^d} (x - E^g(X)) {}^t(x - E^g(X)) d\mu^{g,h}(x) = E(\mathcal{W}_X^{g,h} Y {}^t Y) \end{cases}.$$

Une première idée pour définir la fonction h consiste à repartir du premier exemple du paragraphe 7, en cherchant un coefficient $\eta > 0$ raisonnable tel que $h(t) = g_c(\eta t)$ (ce qui impose $\eta \leq \frac{c}{2}$ pour que $h(1) = 1$). On sait (cf. le paragraphe 7) que pour $0 < p \leq \frac{1}{2}$, μ^g est une loi de Bernoulli de paramètre $E^g(X) \in [0, \frac{1}{2}[$, avec $E^g(X) = p \iff \varepsilon + \frac{1-2\varepsilon}{c} \leq p \leq \frac{1}{2}$. On a donc $K_X^g = \{E^g(X) (1 - E^g(X))\}$ et

$$h\left(\frac{1}{d} {}^t(x - E^g(X)) (K_X^g)^{-1} (x - E^g(X))\right) = \begin{cases} h\left(\frac{E^g(X)}{1 - E^g(X)}\right) = 1 \text{ pour } x = 0 \\ h\left(\frac{1 - E^g(X)}{E^g(X)}\right) \text{ pour } x = 1 \end{cases}.$$

$$\begin{aligned} \text{On en déduit } h\left(\frac{1}{d} {}^t(x - E^g(X)) (K_X^g)^{-1} (x - E^g(X))\right) \mu(dx) \\ = (1 - p) \delta_0(dx) + p h\left(\frac{1 - E^g(X)}{E^g(X)}\right) \delta_1(dx) \end{aligned}$$

et pour raison de normalisation,

$$\mu^{g,h} = \frac{1}{1 - p + p h\left(\frac{1 - E^g(X)}{E^g(X)}\right)} \left[(1 - p) \delta_0 + p h\left(\frac{1 - E^g(X)}{E^g(X)}\right) \delta_1 \right],$$

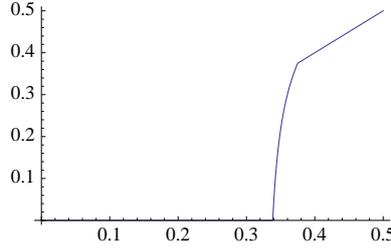
loi de Bernoulli de paramètre $E^{g,h}(X) = \frac{p h\left(\frac{1-E^g(X)}{E^g(X)}\right)}{1-p+p h\left(\frac{1-E^g(X)}{E^g(X)}\right)} \in \left[0, \frac{1}{2}\right]$.

Or pour $p \leq \frac{1}{2}$, $E^g(X)$ vaut p et $h\left(\frac{1-p}{p}\right) = g_c\left(\eta \frac{1-p}{p}\right)$ vaut 1 (de sorte que $\mu^{g,h} = \mu$) si et seulement si

$$p \geq \varepsilon + \frac{1-2\varepsilon}{c} \text{ et } \eta \frac{1-p}{p} \leq \frac{c}{2} \iff p \geq \frac{1}{1 + \frac{c}{2\eta}}.$$

On a donc $E^g(X) = p$ et $E^{g,h}(X) = p$ pour les **mêmes** valeurs de p si et seulement si $\varepsilon + \frac{1-2\varepsilon}{c} = \frac{1}{1 + \frac{c}{2\eta}}$, c'est-à-dire pour la valeur $\eta = \frac{c}{2} \frac{1 + (c-2)\varepsilon}{c-1 - (c-2)\varepsilon}$; soit lorsque

$$\varepsilon = 0.25 \text{ et } c = 4, \text{ pour } \eta = \frac{6}{5} = 1.2.$$



$E^{g,h}(X)$ en fonction de $p \in]0, \frac{1}{2}[$ pour $\varepsilon = 0.25$, $c = 4$ et $\eta = 1.2$

Remarques 8.2

- Cette seconde phase $(X, \mathcal{V}_X^g) \xrightarrow{h} (X, \mathcal{W}_X^{g,h})$ est équivariante par similitude affine, et même par **toute** transformation affine bijective $x \in \mathbb{R}^d \xrightarrow{\Phi} Ax+b \in \mathbb{R}^d$ (A matrice inversible de taille d et $b \in \mathbb{R}^d$) : en effet on a, en notant $Y = X - E^g(X)$,

$$K_{AX+b}^g = K_{AX}^g = K_{AY}^g = E^g(AY {}^t(A Y)) = A K_X^g {}^t A;$$

donc pour tout $u \in \mathbb{R}^d$ non nul,

$$S_u^{A,b}(\omega) \stackrel{\text{déf}}{=} \frac{(AY(\omega) \cdot u)^2}{{}^t u K_{AX+b}^g u} = \frac{{}^t ({}^t(A Y(\omega)) u) {}^t(A Y(\omega)) u}{{}^t u A K_X^g {}^t A u} = \frac{{}^t({}^t A u) Y(\omega) {}^t Y(\omega) {}^t A u}{{}^t({}^t A u) K_X^g {}^t A u} = S_{{}^t A u}(\omega)$$

de sorte que $S^{A,b}(\omega) \stackrel{\text{déf}}{=} \sup_{u \neq 0} S_u^{A,b}(\omega) = \sup_{u \neq 0} S_{{}^t A u}(\omega) = \sup_{v \neq 0} S_v(\omega) = S(\omega)$. On en déduit que $\underline{\mathcal{W}_{AX+b}^{g,h}} = \underline{\mathcal{W}_X^{g,h}}$ pour Φ similitude affine (cf. remarque 4.2).

- On peut retrouver l'expression de $S(\omega)$ et la valeur de $E^g(S)$ sans se référer à la théorie des multiplicateurs de Lagrange, en utilisant la remarque précédente : en effet la matrice K_X^g étant symétrique définie positive, il existe des matrices A telles que ${}^t A A = (K_X^g)^{-1}$: car si P est une matrice orthogonale réelle telle que ${}^t P K_X^g P = D$, D matrice diagonale à coefficients positifs, et si P' est une matrice orthogonale réelle arbitraire, c'est le cas des matrices $A = P' \Delta {}^t P$ où Δ est une matrice diagonale telle que $\Delta^2 = D^{-1}$ (dans le cas particulier où on prend $P' = P$ et où les coefficients de Δ sont choisis positifs, A est la matrice *racine carrée* de $(K_X^g)^{-1}$, unique matrice symétrique positive telle que $A^2 = (K_X^g)^{-1}$).

Alors $K_{AX}^g = A K_X^g {}^t A = A ({}^t A A)^{-1} {}^t A = A A^{-1} ({}^t A)^{-1} {}^t A = I_d$, où I_d désigne la matrice unité de taille d ; de sorte qu'on a pour toute direction u de \mathbb{R}^d ,

$$S_u^A(\omega) \stackrel{\text{d\u00e9f}}{=} \frac{(AY(\omega) \cdot u)^2}{{}^t u K_{AX}^g u} = (AY(\omega) \cdot u)^2 \quad \text{car } {}^t u u = |u|_2^2 = 1.$$

On a donc d'apr\u00e8s le point pr\u00e9c\u00e9dent en question et les propri\u00e9t\u00e9s du produit scalaire,

$$S(\omega) = S^A(\omega) \stackrel{\text{d\u00e9f}}{=} \sup_{|u|_2=1} S_u^A(\omega) = |AY(\omega)|_2^2 = {}^t Y(\omega) {}^t A A Y(\omega) = {}^t Y(\omega) (K_X^g)^{-1} Y(\omega).$$

Enfin $E^g(S) = E^g({}^t(A Y) (A Y)) = \text{tr } E^g(A Y {}^t(A Y)) = \text{tr } K_{AX}^g = \text{tr } I_d = d$.

• Cette derni\u00e8re remarque fait appara\u00eetre une expression particuli\u00e8rement simple de S au moyen de la variable $Z \stackrel{\text{d\u00e9f}}{=} A Y$ d\u00e9duite de X par centrage et d\u00e9formation ($X = A^{-1} Z + E^g(X)$ o\u00f9 ${}^t A A = (K_X^g)^{-1}$) ; \u00e0 savoir $S = |Z|_2^2$.

Dans le cas de r\u00e9f\u00e9rence id\u00e9al o\u00f9 X est un vecteur gaussien (relativement \u00e0 P_X^g , ce qui suppose en particulier que g ne s'annule pas sur \mathbb{R}_+ !), Z est un vecteur gaussien centr\u00e9 tel que $K_Z^g = I_d$; de sorte que S est une variable du χ^2 de degr\u00e9 d . Ceci sugg\u00e8re une seconde id\u00e9e de d\u00e9finition de la fonction h , d\u00e9pendant de d :

$$h_d \text{ continue et affine par morceaux telle que } h_d|_{[0, s_d^1]} = 1 \text{ et } h_d|_{[s_d^2, +\infty[} = 0,$$

o\u00f9 $s_d^1 < s_d^2$ sont fix\u00e9s par exemple par $P(\chi_d^2 \geq s_d^1) = 0.25$ et $P(\chi_d^2 \geq s_d^2) = 0.05$.

Toutefois l'hypoth\u00e8se gaussienne sur X , qui assure la non corr\u00e9lation lin\u00e9aire des carr\u00e9s des composantes de Z , et donc une variance de $\frac{S}{d}$ proportionnelle \u00e0 $\frac{1}{d}$, est bien forte. Par exemple oppos\u00e9, soit V une variable r\u00e9elle telle que $E^g(V^2) = 1$ et soit $Z = ({}^t(B_1 V, \dots, B_d V))$, o\u00f9 B_1, \dots, B_d sont des variables de Rademacher (valeur 1 ou -1 avec la probabilit\u00e9 $\frac{1}{2}$) ind\u00e9pendantes entre elles et de V . Comme $\forall \ell \in \{1, \dots, d\}$, $E^g(B_\ell) = 0$ et $B_\ell^2 = 1$, on a encore $E^g(Z) = 0$ et $K_Z^g = I_d$; de sorte que pour toute matrice de taille d inversible A , Z est la variable d\u00e9duite de $X \stackrel{\text{d\u00e9f}}{=} A^{-1} Z$ par A -d\u00e9formation. On a donc $S = |Z|_2^2 = \sum_{\ell=1}^d V^2 = d V^2$ pour de telles X et $\text{var}^g \frac{S}{d} = \text{var}^g V^2 > 0$ (si V^2 n'est pas P_X^g -ps constant), variance ind\u00e9pendante de d .

9 Application et exemples

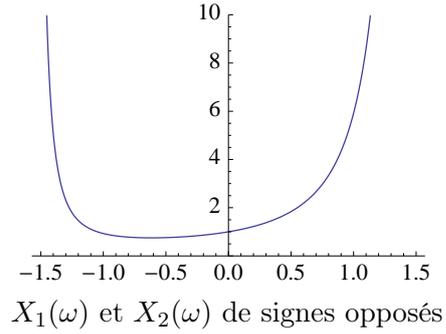
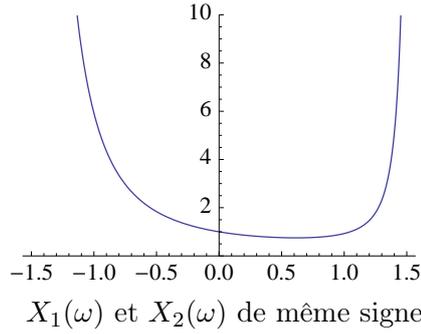
Illustration avec $d = 2$: on consid\u00e8re un couple al\u00e9atoire $X = (X_1, X_2)$ centr\u00e9 ($E^g(X) = (0, 0)$) et de matrice de covariance $K_X^g = \begin{pmatrix} 1 & \sin \alpha \\ \sin \alpha & 1 \end{pmatrix}$, o\u00f9 $-\frac{\pi}{2} < \alpha < \frac{\pi}{2}$.

On a $(K_X^g)^{-1} = \frac{1}{\cos^2 \alpha} \begin{pmatrix} 1 & -\sin \alpha \\ -\sin \alpha & 1 \end{pmatrix}$, donc

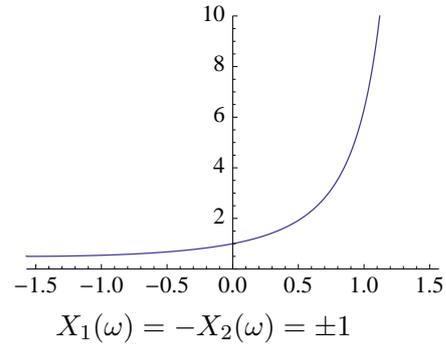
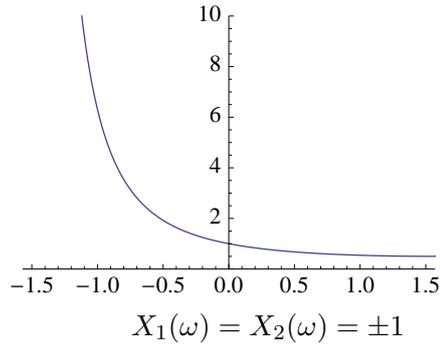
$$S(\omega) = {}^t X(\omega) (K_X^g)^{-1} X(\omega) = \frac{X_1(\omega)^2 + X_2(\omega)^2 - 2 \sin \alpha X_1(\omega) X_2(\omega)}{\cos^2 \alpha}.$$

Pour $X_1(\omega)^2 + X_2(\omega)^2 = E(X_1^2 + X_2^2) = 2$, on obtient les graphes suivants de $\frac{S(\omega)}{d}$ en fonction de α :

- Cas général où $|X_1(\omega)| \neq |X_2(\omega)|$ (en l'occurrence $|X_1(\omega)| = \sqrt{\frac{3}{2}}$ et $|X_2(\omega)| = \frac{1}{\sqrt{2}}$)



- Cas particulier où $|X_1(\omega)| = |X_2(\omega)|$



Dans tous les cas $\frac{1}{d} S(\omega)$ vaut 1 lorsque $\alpha = 0$ (c'est-à-dire lorsque les composantes X_1 et X_2 de X sont non corrélées linéairement).

En pratique, étant donné un échantillon $\{x^1, \dots, x^n\}$ de taille n de \mathbb{R}^d , on a déjà introduit les probabilités portées par l'échantillon :

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x^i} \quad \text{et} \quad \mu_n^g = \frac{1}{n} \sum_{i=1}^n v_n^i \delta_{x^i},$$

où m_n est la médiane empirique de $\{x^1, \dots, x^n\}$ et où pour tout $i \in \{1, \dots, n\}$,

$$r_n^i \stackrel{\text{déf}}{=} |x^i - m_n|_2, \quad \bar{r}_n \stackrel{\text{déf}}{=} \sum_{j=1}^n \left[F^\tau \left(\frac{j}{n} \right) - F^\tau \left(\frac{j-1}{n} \right) \right] r_n^{(j)} \quad \text{et} \quad v_n^i \stackrel{\text{déf}}{=} \frac{g \left(\frac{r_n^i}{\bar{r}_n} \right)}{\frac{1}{n} \sum_{j=1}^n g \left(\frac{r_n^j}{\bar{r}_n} \right)};$$

de sorte que $\bar{x}_n^g = \int_{\mathbb{R}^d} x d\mu_n^g(x) = \frac{1}{n} \sum_{i=1}^n v_n^i x^i$.

On considère à présent la matrice de covariance empirique $k_n^g = (\sigma_{\ell,m}^g)_{1 \leq \ell, m \leq d}$ où

$$\begin{aligned} \forall \ell, m, \quad \sigma_{\ell,m}^g &= \int x_\ell x_m d\mu_n^g(x) - \left(\int x_\ell d\mu_n^g(x) \right) \left(\int x_m d\mu_n^g(x) \right) \\ &= \frac{1}{n} \sum_{i=1}^n v_n^i x_\ell^i x_m^i - \left(\frac{1}{n} \sum_{i=1}^n v_n^i x_\ell^i \right) \left(\frac{1}{n} \sum_{j=1}^n v_n^j x_m^j \right) \end{aligned}$$

(on ne se préoccupe pas du biais dans l'estimation de K_X^g). Autrement dit

$$k_n^g = \frac{1}{n} \sum_{i=1}^n v_n^i x^i t(x^i) - \left(\frac{1}{n} \sum_{i=1}^n v_n^i x^i \right) t \left(\frac{1}{n} \sum_{j=1}^n v_n^j x^j \right)$$

est la *matrice d'inertie* du nuage des points x^1, \dots, x^n affectés des poids $\frac{v_n^1}{n}, \dots, \frac{v_n^n}{n}$.

Sous réserve que k_n^g soit inversible, on calcule son inverse et on forme

$$s_n^i \stackrel{\text{déf}}{=} t(x^i - \bar{x}_n^g) (k_n^g)^{-1} (x^i - \bar{x}_n^g).$$

On pose enfin $w_n^i \stackrel{\text{déf}}{=} \frac{h(\frac{s_n^i}{d})}{\frac{1}{n} \sum_{j=1}^n h(\frac{s_n^j}{d})}$ et on obtient $\mu_n^{g,h} = \frac{1}{n} \sum_{i=1}^n w_n^i \delta_{x^i}$, c'est-à-dire $\frac{w_n^i}{n}$

comme coefficient de pondération de chaque point x^i de l'échantillon.

Les indicateurs associés, de point d'effondrement inchangé ε du fait de l'utilisation de \bar{x}_n^g et k_n^g dans la définition de s_n^i , sont de manière naturelle :

- pour l'espérance, l'estimation de $E^{g,h}(X) : \bar{x}_n^{g,h} = \int_{\mathbb{R}^d} x d\mu_n^{g,h}(x) = \frac{1}{n} \sum_{i=1}^n w_n^i x^i ;$
- pour la covariance, l'estimation de $K_X^{g,h} : k_n^{g,h} = (\sigma_{\ell,m}^{g,h})_{1 \leq \ell, m \leq d}$ avec

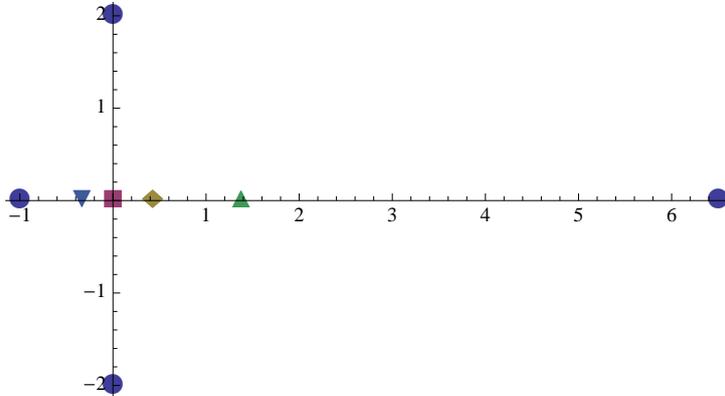
$$\begin{aligned} \forall \ell, m, \sigma_{\ell,m}^{g,h} &= \int x_\ell x_m d\mu_n^{g,h}(x) - \left(\int x_\ell d\mu_n^{g,h}(x) \right) \left(\int x_m d\mu_n^{g,h}(x) \right) \\ &= \frac{1}{n} \sum_{i=1}^n w_n^i x_\ell^i x_m^i - \left(\frac{1}{n} \sum_{i=1}^n w_n^i x_\ell^i \right) \left(\frac{1}{n} \sum_{j=1}^n w_n^j x_m^j \right) \end{aligned}$$

($k_n^{g,h}$ matrice d'inertie du nuage des points x^1, \dots, x^n affectés des poids $\frac{w_n^1}{n}, \dots, \frac{w_n^n}{n}$).

Exemples où $d = 2$, $\varepsilon = 0.25$, $c = 4$ et (première définition, cf. le paragraphe 8), $h(t) = g_c(\eta t)$ avec $\eta = 1.2$; $\bar{x}_n^{g,h}$ apparaît ci-dessous comme un triangle bleu pointant vers le bas, les autre codes couleurs étant inchangés :

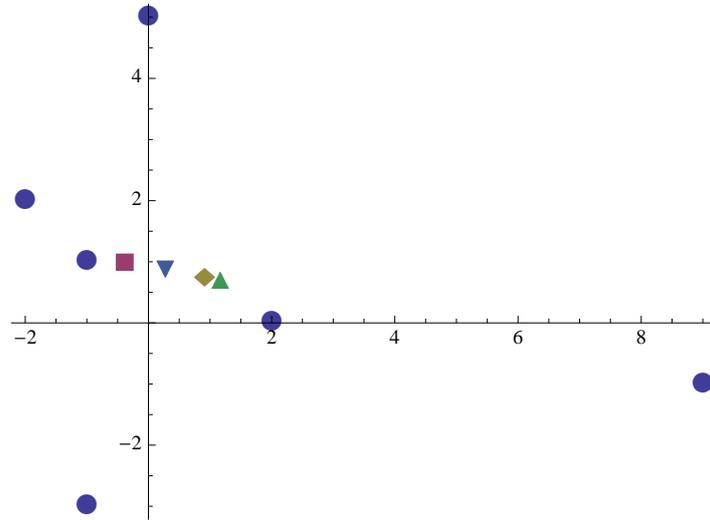
- Échantillon de $n = 4$ points :

$$\{x^1 = (-1, 0), x^2 = (0, 2), x^3 = (0, -2), x^4 = (6.5, 0)\}.$$

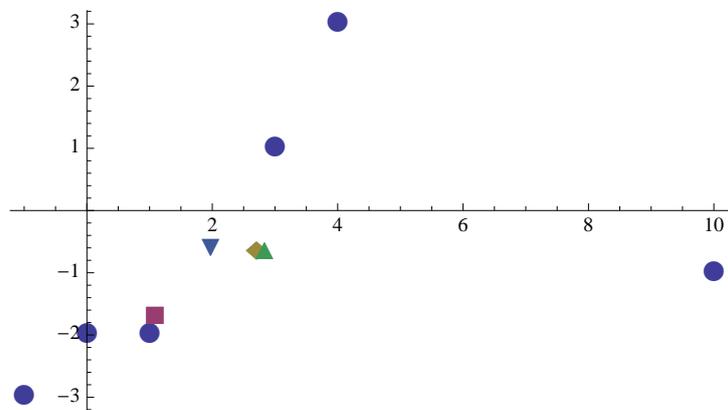


$$\{v_n^i\}_{1 \leq i \leq n} \approx \{1.185, 1.185, 1.185, 0.445\}; \{w_n^i\}_{1 \leq i \leq n} = \left\{ \frac{4}{3}, \frac{4}{3}, \frac{4}{3}, 0 \right\}.$$

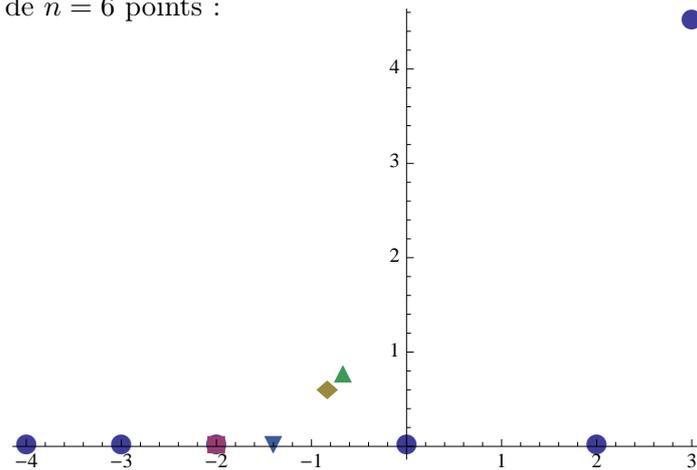
- Échantillon de $n = 6$ points :



- Échantillon de $n = 6$ points :



- Échantillon de $n = 6$ points :



$$\{v_n^i\}_{1 \leq i \leq n} \approx \{1.04553, 1.04553, 1.04553, 1.04553, 1.04553, 0.772356\};$$

$$\{w_n^i\}_{1 \leq i \leq n} = \{1.2, 1.2, 1.2, 1.2, 1.2, 0\};$$

$$k_n^g \approx \begin{pmatrix} 6.21404 & 2.22068 \\ 2.22068 & 2.27115 \end{pmatrix}; k_n^{g,h} = \begin{pmatrix} 4.64 & 0 \\ 0 & 0 \end{pmatrix}.$$

- Échantillons de $n = 9$ points :

On considère la suite d'échantillons de 9 points de \mathbb{R}^2 dépendant de $k \in \{0, 60\}$ déjà vus au paragraphe 7 :

$$\{x^1 = x^2 = (1, 1), x^3 = x^4 = (-1, 1), x^5 = x^6 = (-1, -1),$$

$$x^7 = x^8 = (1, -1), x^9 = (\frac{k}{5} - 5, \frac{k}{10})\}.$$

Les codes couleurs sont les mêmes que précédemment.

