

## THESIS / THÈSE

### MASTER EN SCIENCES MATHÉMATIQUES À FINALITÉ APPROFONDIE

#### Robustesse

fonction d'influence de la profondeur de demi-espace scatter et du graphical lasso

LOUVET, Gaëtan

*Award date:*  
2022

*Awarding institution:*  
Universite de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



UNIVERSITE DE NAMUR

Faculté des Sciences

# Robustesse : fonction d'influence de la profondeur de demi-espace scatter et du graphical lasso

Promoteur : GERMAIN VAN BEVER

Mémoire présenté pour l'obtention du grade académique de master en sciences  
mathématiques à finalité approfondie

Gaëtan LOUVET

Juin 2022



# Remerciements

J'aimerais remercier toutes les personnes qui m'ont apporté leur soutien et qui ont contribué à l'aboutissement de ce mémoire.

Tout d'abord, je souhaite remercier mon promoteur, Germain Van Bever, pour son aide et ses conseils, sans lesquels ce mémoire n'aurait pu aboutir. Je tiens aussi à remercier les professeurs Ines Wilms et Jakob Raymaekers de l'Université de Maastricht, pour leur accueil et nos échanges durant mon stage de fin d'étude, qui ont mené à l'aboutissement du troisième chapitre de ce document.

Je remercie ensuite mes parents qui m'ont suivi tout au long de mon parcours et qui m'ont donné l'opportunité de poursuivre ma passion. Je remercie aussi particulièrement ma compagne Sarah qui me soutient au quotidien et qui m'a apporté le courage nécessaire durant mes études supérieures.

Enfin, en cette fin de parcours universitaire, je tiens à remercier l'ensemble des assistants et des professeurs du département de mathématiques de l'Université de Namur, qui se sont toujours montrés accueillants et disponibles.

Merci à toutes et à tous.



# Résumé

Lorsque nous recevons un jeu de données à analyser, il peut arriver que des valeurs aberrantes apparaissent, impactant alors les méthodes statistiques utilisées. Il existe différents outils permettant d'évaluer la robustesse de procédures, c'est-à-dire leur sensibilité lorsque les données sont contaminées. Ce récit s'intéresse à la robustesse de plusieurs méthodes statistiques. Dans le premier chapitre, nous commençons par définir différents outils afin d'évaluer la robustesse. Ensuite, nous introduisons les estimateurs robustes de position et de dispersion les plus connus et les plus utilisés. Dans le second chapitre, nous nous intéressons aux fonctions de profondeur, qui permettent de quantifier la centralité d'une valeur pour un certain paramètre. Nous discutons en particulier de la profondeur de demi-espace de position et de scatter. Après introduction de ces différents concepts, nous dérivons la fonction d'influence de la profondeur scatter. Nous commençons en considérant que l'estimateur de position dont cette dernière dépend est connu et fixé. Nous nous intéressons ensuite au cas où la distribution est discrète à support fini, et nous concluons enfin en donnant des bornes pour la fonction d'influence dans le cas où la distribution est absolument continue par rapport à la mesure de Lebesgue. Dans le dernier chapitre, rédigé en anglais, nous discutons du *graphical lasso*, une méthode permettant d'estimer les relations entre les noeuds d'un modèle graphique. Cette approche consiste en l'estimation d'une matrice, appelée matrice de précision. Après introduction des principaux concepts, nous dérivons la fonction d'influence de cette procédure.

**Mots clés :** statistiques, robustesse, fonction de profondeur, fonction d'influence, graphical lasso.

# Abstract

When we receive a dataset for analysis, outliers may appear, impacting the statistical methods used. There exist different tools to assess the robustness of procedures, i.e. their sensitivity when data are contaminated. This report focuses on the robustness of several statistical methods. In the first chapter, we start by defining different tools for evaluating the robustness. Then, we introduce the best known and most widely used robust position and dispersion estimators. In the second chapter, we focus on depth functions, which quantify the centrality of a value for a certain parameter of a distribution. In particular, we discuss the halfspace depth and scatter halfspace depth, the latter involving a position estimator in its definition. After introducing these different concepts, we derive the influence function of the scatter depth function. We start by considering that the position estimator is known and fixed. We then consider the case where the distribution is discrete with finite support, and finally conclude by giving bounds for the influence function in the case where the distribution is absolutely continuous with respect to the Lebesgue measure. In the last chapter, we discuss the *graphical lasso*, a method for estimating the relations between the nodes of a graphical model. This approach consists in the estimation of a matrix, called precision matrix. After introducing the main concepts, we derive the influence function of this procedure.

**Key words :** statistics, robustness, depth function, influence function, graphical lasso.



# Table des matières

<b>Mémoire</b>	<b>0</b>
<b>Introduction</b>	<b>1</b>
<b>1 La robustesse</b>	<b>3</b>
1.1 Motivation . . . . .	3
1.2 Des estimateurs comme fonctionnelles . . . . .	6
1.3 Des outils pour évaluer la robustesse . . . . .	8
1.3.1 Fonction d'influence . . . . .	8
1.3.2 Point de rupture . . . . .	16
1.3.3 La courbe de biais . . . . .	19
1.4 Estimateurs robustes de position . . . . .	21
1.4.1 M-estimateurs . . . . .	21
1.4.2 L-estimateurs . . . . .	26
1.5 Estimateurs robustes de dispersion . . . . .	29
1.6 Généralisation en dimension supérieure . . . . .	33
1.7 Conclusion . . . . .	41
<b>2 Notions de centralité et de profondeur</b>	<b>43</b>
2.1 Introduction aux fonctions de profondeur . . . . .	43
2.2 Fonction de profondeur de demi-espace . . . . .	45
2.2.1 Introduction . . . . .	45
2.2.2 Robustesse . . . . .	48
2.3 Fonction de profondeur scatter . . . . .	52
2.3.1 Fonction d'influence de la profondeur scatter lorsque $T_P = 0$ . . . . .	53
2.3.2 Fonction d'influence pour une distribution discrète . . . . .	56
2.3.3 Fonction d'influence pour une distribution continue . . . . .	58
2.3.4 Discussion sur la fonction d'influence . . . . .	73
2.3.5 Fonction d'influence empirique . . . . .	75
2.4 Conclusion . . . . .	77
<b>Stage</b>	<b>77</b>
<b>3 Robustness of graphical lasso</b>	<b>79</b>
3.1 Introduction . . . . .	79
3.2 Illustration of conditional dependence . . . . .	80
3.3 Estimation of the precision matrix . . . . .	83
3.4 Influence function of graphical lasso . . . . .	86
3.4.1 Influence function of MLE of $\Sigma$ . . . . .	86

3.4.2	Influence function of MLE of $\Omega$ . . . . .	87
3.4.3	Influence function of glasso for $\lambda > 0$ . . . . .	88
3.4.4	Discussion . . . . .	91
3.4.5	Gross error sensitivity . . . . .	92
3.5	Directions maximising the norm of the influence function . . . . .	93
3.6	Asymptotic variance . . . . .	99
3.7	Introduction to robust covariance matrix estimators . . . . .	101
3.7.1	Scale estimators . . . . .	102
3.7.2	Correlation estimators . . . . .	103
3.7.3	Influence functions and asymptotic variances . . . . .	104
3.7.4	Numerical illustrations . . . . .	107
3.8	Conclusion . . . . .	110
	<b>Conclusion</b>	<b>113</b>
	<b>Bibliographie</b>	<b>115</b>

# Introduction

Lorsqu'il s'agit d'analyser un jeu de données, il est courant que ce dernier contient des données aberrantes, pouvant alors fausser les méthodes statistiques utilisées. La robustesse vise à évaluer l'influence qu'a une perturbation dans les données sur une procédure statistique. Ainsi, un estimateur robuste aura tendance à être peu affecté par la présence de contaminations, tandis qu'un estimateur peu robuste risque de voir sa valeur fortement impactée. Il existe différents outils permettant d'évaluer la robustesse. Dans ce document, nous allons en introduire quelques-uns, ainsi que les utiliser afin d'observer le comportement des estimateurs les plus connus en présence de contamination.

Ce récit est divisé en trois chapitres. Les deux premiers constituent le mémoire de fin de cursus, tandis que le dernier, rédigé en anglais, contient le rapport de stage dans le cadre du master à finalité approfondie. Dans le premier chapitre, nous commençons par introduire différents outils permettant d'évaluer la robustesse d'une procédure statistique, en considérant dans un premier temps, et à finalité d'introduction, que la dimension des données est fixée à  $p = 1$ . Un exemple d'outil est la fonction d'influence, qui permet de quantifier l'impact qu'a une perturbation infinitésimale dans la distribution sur l'estimateur considéré, et sera un outil essentiel tout au long de ce récit. Nous introduisons aussi différents estimateurs robustes de position et de dispersion parmi les plus connus et les plus utilisés. Pour chacun d'eux, nous évaluons leur robustesse en utilisant les outils précédemment définis. Enfin, nous généralisons les différents concepts abordés aux distributions de dimension  $p > 1$ . Dans le second chapitre, nous introduisons le concept de fonctions de profondeur. Ces dernières ont été introduites dans le but de pouvoir estimer la centralité d'un élément par rapport à un jeu de données. Parmi celles-ci, nous discutons de la fonction de profondeur de demi-espace, ainsi que de la fonction de profondeur de scatter. Après une étude de la robustesse de la première, nous livrons pour la seconde la fonction d'influence dans différents cas. En effet, la profondeur de scatter dépend d'un paramètre de position, que nous considérons dans un premier temps connu et fixé. Ensuite, nous donnons la fonction d'influence pour une distribution discrète à support fini, avec un estimateur de position quelconque. Enfin, pour une distribution absolument continue par rapport à la mesure de Lebesgue, et sous certaines hypothèses sur la fonction de densité, nous bornons la fonction d'influence de la profondeur de scatter. Nous concluons ce chapitre par quelques remarques sur les résultats obtenus, ainsi que sur la variance asymptotique de la profondeur de scatter. Le troisième chapitre concerne une méthode d'estimation de liens entre différents noeuds d'un modèle graphique, que nous appelons *graphical lasso*. Cette procédure consiste en l'estimation d'une matrice via une méthode de pénalisation convexe. Nous commençons par introduire cet estimateur, en illustrant la manière dont il est défini. Nous livrons ensuite sa fonction d'influence ainsi que sa variance asymptotique. Nous clôturons par différentes analyses numériques, comme par exemple l'illustration de la fonction d'influence empirique, aussi appelée courbe de sensibilité.



# Chapitre 1

## La robustesse

Dans ce premier chapitre, nous allons introduire la notion de robustesse en statistique. Nous allons pour ce faire illustrer ce concept avec plusieurs exemples. Ensuite, nous allons définir la notion d'estimateur sous forme fonctionnelle. Nous poursuivrons en définissant plusieurs outils pour mesurer la robustesse d'un estimateur, comme par exemple la fonction d'influence, le point de rupture ainsi que la courbe de biais. Enfin, nous terminerons en définissant deux types d'estimateurs robustes, à savoir les M-estimateurs et les L-estimateurs. L'ensemble des définitions, propriétés et théorèmes abordés dans ce premier chapitre sont tirés des livres de référence [1] [2] et [3].

### 1.1 Motivation

Dans le cadre de la robustesse, les échantillons étudiés contiennent généralement ce que l'on appelle des données aberrantes ou *outliers*. Bien que nous pouvons distinguer différentes classes d'*outliers*, il reste difficile d'en donner une définition formelle. Ils sont décrits dans le document [4] comme des observations inconsistantes, à l'écart des autres données et leur détermination est souvent subjective, en fonction de l'auteur, des données étudiées et des objectifs de l'étude. Ils peuvent apparaître par erreur de calcul ou de recopiage, après un changement d'unité ou même lorsque plusieurs classes ou modèles sont présents dans le jeu de données.

Le modèle théorique d'où provient un échantillon contaminé est souvent décrit par une distribution donnée par

$$F = (1 - t)H + tG,$$

où  $H$  est la distribution principale dont nous cherchons à estimer les paramètres,  $G$  représente celle des *outliers* et  $t$  un réel positif. Une procédure ou un estimateur seront dits robustes s'ils sont peu perturbés par la présence (minoritaire) d'observations contaminées.

En guise d'exemple (inspiré de [5]), supposons vouloir estimer le paramètre de position de la loi de la variable aléatoire ayant engendré  $n$  observations  $x_1, \dots, x_n$ . Pour ce faire, il existe plusieurs estimateurs comme par exemple la moyenne empirique, donnée par

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \tag{1.1}$$

ou encore la médiane, définie par

$$\tilde{x} = \arg \min_m \frac{1}{n} \sum_{i=1}^n |x_i - m|. \tag{1.2}$$

Il est aussi possible de définir la moyenne avec une expression similaire à celle de la médiane. Sa définition est alors donnée par

$$\bar{x} = \arg \min_m \frac{1}{n} \sum_{i=1}^n (x_i - m)^2. \quad (1.3)$$

Les expressions (1.1) et (1.3) sont équivalentes. En effet, comme la fonction à minimiser est strictement convexe, son minimum vérifie

$$-\frac{2}{n} \sum_{i=1}^n (x_i - \bar{x}) = 0 \Leftrightarrow \sum_{i=1}^n (x_i - \bar{x}) = 0 \Leftrightarrow \frac{1}{n} \sum_{i=1}^n x_i = \bar{x},$$

et nous retrouvons alors bien la définition (1.1). Nous avons ici garantie de l'unicité de la solution, ce qui n'est pas le cas pour la médiane. Typiquement, lorsque la taille  $n$  d'un échantillon est paire, toute valeur entre les 2 données centrales (si les données sont ordonnées, celles en position  $n/2$  et  $n/2 + 1$ ) vérifie la condition (1.2). Nous considérons donc en général que la médiane est la moyenne des deux données centrales. Il existe une autre expression pour définir la médiane mais il faut pour cela introduire la notion de fonction de répartition empirique.

**Définition 1.1.** *La fonction de répartition empirique  $F_n$  basée sur un échantillon  $\{x_1, \dots, x_n\}$  est donnée par*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n F_{\delta_{x_i}}(x), \quad x \in \mathbb{R},$$

où  $F_{\delta_{x_i}}(x)$  est la distribution de Dirac définie par

$$F_{\delta_{x_i}}(x) = \begin{cases} 1 & \text{si } x_i \leq x \\ 0 & \text{sinon.} \end{cases}$$

En se rappelant que la médiane est en fait la donnée laissant 50% de l'échantillon de part et d'autre, nous pouvons aussi la définir comme

$$\tilde{x} = F_n^{-1} \left( \frac{1}{2} \right).^1$$

Considérons, en notant  $N(\mu, \sigma^2)$  la distribution d'une loi normale de moyenne  $\mu \in \mathbb{R}$  et de variance  $\sigma^2 \in \mathbb{R}$ , que notre échantillon est issu d'une distribution  $F = (1-t)N(0, 1) + tN(7, 0.5^2)$ . Si nous voulons estimer le paramètre de position de la loi cible  $N(0, 1)$ , nous voudrions que le résultat obtenu par l'estimateur de position soit proche de 0. Afin d'illustrer la différence entre les deux estimateurs définis précédemment, considérons 10 échantillons de taille 100 engendrés suivant la loi  $F$  et calculons la moyenne et la médiane de chacun d'eux. Celles-ci sont reportées sur les figures 1 à 3 pour différentes valeurs de  $t$ . Sur la FIGURE 1, nous observons que, pour  $t = 0$ , la moyenne et la médiane sont toutes les deux concentrées autour de 0, ce qui est cohérent étant donné que  $t = 0$  correspond à n'avoir aucune perturbation. Sur la FIGURE 2, nous constatons que les données commencent à être contaminées par la loi  $N(7, 0.5^2)$  et que les moyennes s'éloignent de 0, là où les médianes restent à proximité. Enfin, nous observons sur la FIGURE 3 que ces phénomènes s'amplifient lorsque  $t$  augmente. En effet, les médianes sont toujours proches de 0 tandis que les moyennes s'en éloignent davantage.

<sup>1</sup>Comme mentionné précédemment, la médiane n'est pas toujours unique et nous considérons dès lors un élément parmi ceux qui satisfont cette égalité.

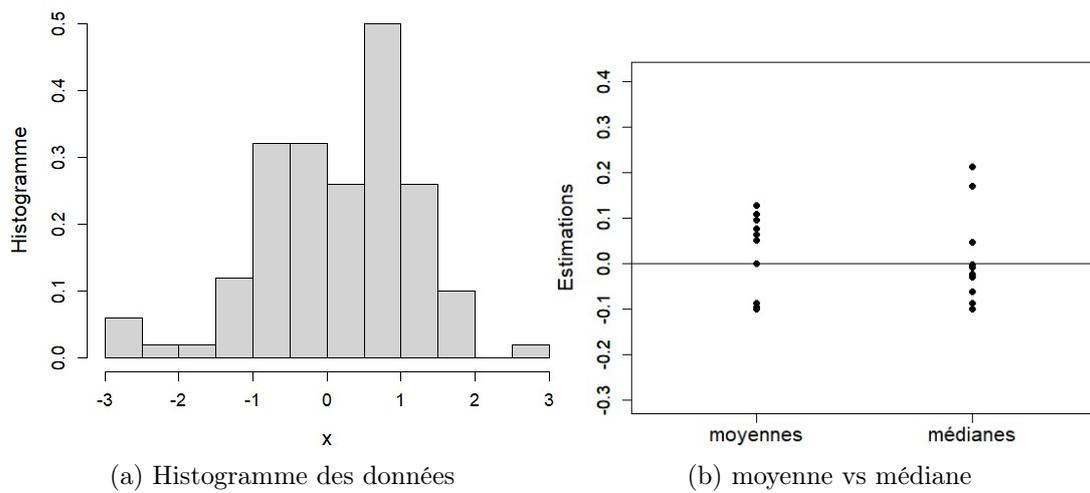


FIGURE 1 – Illustration de la robustesse de la moyenne et la médiane pour  $t = 0$ .

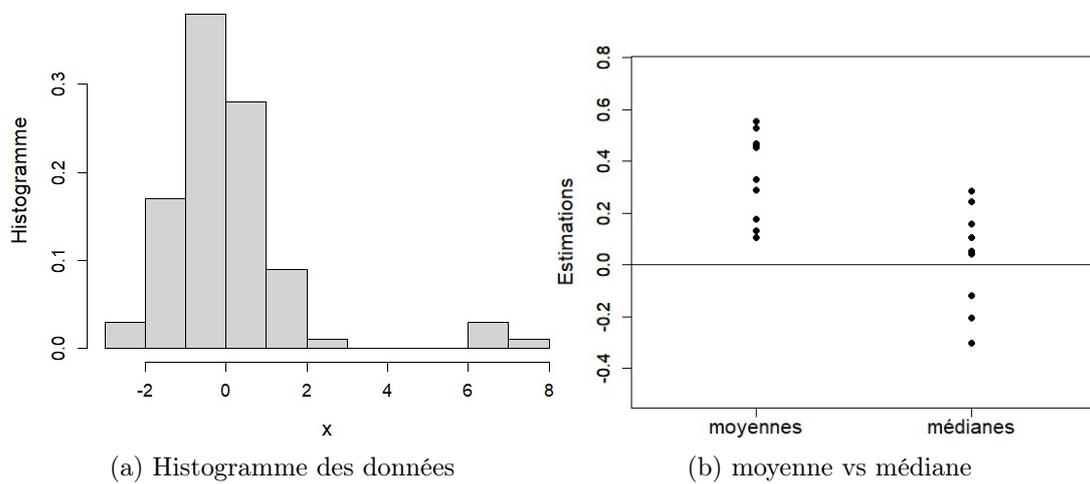


FIGURE 2 – Illustration de la robustesse de la moyenne et la médiane pour  $t = 0.05$ .

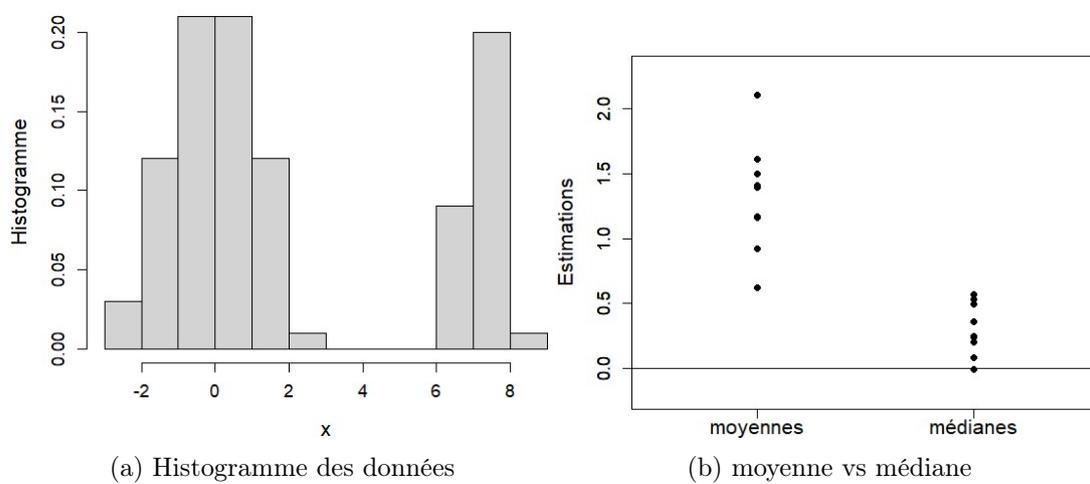


FIGURE 3 – Illustration de la robustesse de la moyenne et la médiane pour  $t = 0.2$ .

Une telle analyse peut aussi être effectuée pour le paramètre de dispersion, en considérant deux autres estimateurs. Le premier, l'écart type empirique, est défini par

$$\text{sd}(x_1, \dots, x_n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2},$$

tandis que le second, la déviation absolue centrée en la médiane (MAD) est donnée par la médiane de l'échantillon  $\{|x_1 - \tilde{x}|, \dots, |x_n - \tilde{x}|\}$ . Cet estimateur sera plus précisément introduit dans une autre section. Nous constatons à la FIGURE 4 que le MAD est bien plus robuste que l'écart type empirique, car beaucoup moins influencé par la présence de contaminations.

## 1.2 Des estimateurs comme fonctionnelles

Nous avons pu observer l'intérêt d'utiliser des estimateurs robustes lorsqu'il s'agit de retrouver les paramètres initiaux d'une loi à partir d'un échantillon. Il existe différents outils permettant d'évaluer la robustesse d'un estimateur mais avant de les définir, il nous faut introduire la notion d'estimateur sous forme fonctionnelle.

Un estimateur peut être défini comme une fonctionnelle  $T : \mathcal{F} \rightarrow \mathbb{R}$ , où  $\mathcal{F}$  représente un ensemble de distributions. Par exemple, pour la moyenne nous obtenons

$$T(F) = \int_{-\infty}^{+\infty} x dF(x),$$

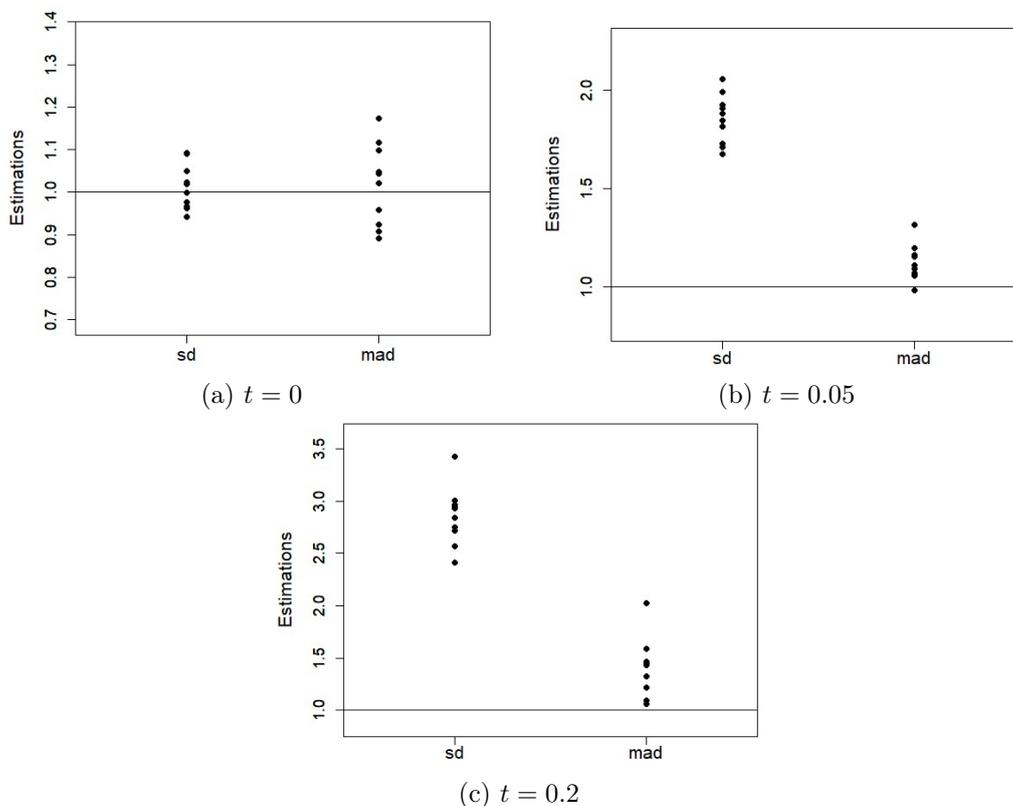


FIGURE 4 – Illustration de la robustesse de l'écart type et du MAD.

tandis que pour la médiane nous avons

$$T(F) = \arg \min_m \int_{-\infty}^{+\infty} |x - m| dF(x).$$

Notons que nous considérons  $\mathcal{F}$ , un ensemble de distributions pour lesquelles les expressions ci-dessus existent. Nous avons choisi pour la moyenne l'expression (1.1) pour ainsi observer que la fonctionnelle associée n'est rien d'autre que l'espérance mathématique. Nous aurions pu considérer évidemment l'autre expression pour obtenir une forme similaire à la médiane.

Nous pouvons évaluer chacune de ces deux expressions en la fonction de répartition empirique afin d'obtenir les estimateurs définis en (1.1) et (1.2). Ainsi, nous obtenons pour la moyenne

$$T(F_n) = \int_{-\infty}^{+\infty} x dF_n(x) = \frac{1}{n} \sum_{i=1}^n x_i,$$

et pour la médiane

$$T(F_n) = \arg \min_m \int_{-\infty}^{+\infty} |x - m| dF_n(x) = \arg \min_m \frac{1}{n} \sum_{i=1}^n |x_i - m|.$$

Nous comprenons alors bien que la fonctionnelle appliquée à une distribution  $F$  représente la version *population* de l'estimateur, là où, lorsque évaluée en la distribution empirique basée sur un échantillon, la fonctionnelle représente la version *empirique* de l'estimateur. De manière générale, si

$$T(F) = \int_{\mathbb{R}} h(x) dF(x),$$

où  $h$  est une fonction quelconque intégrable, alors l'homologue empirique de  $T(F)$  est donné par

$$T(F_n) = \frac{1}{n} \sum_{i=1}^n h(x_i).$$

Il s'agit là d'un résultat classique de théorie de la mesure, pouvant être démontré en utilisant par exemple le théorème de Radon-Nikodym avec la mesure de comptage.

Nous prendrons comme convention que  $T(F)$ , ou parfois même  $T$  lorsque le contexte rend la distribution  $F$  univoque, représente la version population de l'estimateur. Nous noterons  $T(F_n)$ , ou  $T_n$ , la version empirique de l'estimateur, évaluée alors en la fonction de répartition empirique obtenue via un échantillon de taille  $n$ .

Le fait de définir un estimateur comme une fonctionnelle pose la question de sa distribution asymptotique. Dans le cas de la moyenne, le Théorème Central Limite livre

$$\sqrt{n}(T(F_n) - T(F)) \xrightarrow{\mathcal{D}} N(0, \text{Var}_F(T(F))), \quad (1.4)$$

où  $\xrightarrow{\mathcal{D}}$  représente la convergence en distribution et  $\text{Var}_F(T(F))$  la variance asymptotique de l'estimateur, supposée strictement positive. Il existe un résultat général, pour n'importe quelle fonctionnelle  $T_P$ , mais qui nécessite alors d'autres hypothèses de différentiabilité (le résultat complet est disponible dans le document [1], aux pages 18-19), qui sont en pratique très souvent vérifiées.

## 1.3 Des outils pour évaluer la robustesse

### 1.3.1 Fonction d'influence

Nous pouvons à présent introduire les outils permettant de mesurer la robustesse d'un estimateur. Le premier, la fonction d'influence, a pour but de mesurer l'effet d'une *légère* perturbation dans la distribution (ou l'échantillon) sur un estimateur. La notion de *légère* perturbation est plus subjective que vague, et dépend du jeu de données ou de l'objectif recherché. Par exemple, si nous remplaçons une donnée  $x_i$  par  $x_i + c$  avec  $c$  un réel quelconque, la perturbation pourrait être considérée comme *légère* car seule une observation est modifiée. Cependant, la nouvelle donnée pourrait diverger à l'infini et donc radicalement s'éloigner de la valeur initiale.

Afin d'illustrer le concept de fonction d'influence, supposons avoir un échantillon ordonné  $x_1 \leq x_2 \leq \dots \leq x_n$  de taille  $n$  impaire. Nous noterons par la suite leur moyenne  $\hat{\mu}_n = n^{-1} \sum_{i=1}^n x_i$  et leur médiane  $\tilde{\mu}_n = x_{(n+1)/2}$ . Si une contamination intervient sur la dernière observation par exemple, c'est-à-dire que  $x_n$  devient  $x_n + c$ , nous nous demandons quelle influence elle aura sur la médiane et la moyenne. Notons  $\hat{\mu}_n^c$  et  $\tilde{\mu}_n^c$  les estimateurs de la moyenne et de la médiane basés sur l'échantillon contaminé. Nous obtenons alors pour la moyenne

$$\hat{\mu}_n^c = \frac{1}{n} \left( \sum_{i=1}^{n-1} x_i + (x_n + c) \right) = \frac{1}{n} \sum_{i=1}^n x_i + \frac{c}{n} = \hat{\mu}_n + \frac{c}{n},$$

et pour la médiane

$$\tilde{\mu}_n^c = \begin{cases} \tilde{\mu}_n & \text{si } x_n + c > x_{(n+1)/2} \\ x_n + c & \text{si } x_{(n-1)/2} \leq x_n + c \leq x_{(n+1)/2} \\ x_{(n-1)/2} & \text{sinon.} \end{cases} \quad (1.5)$$

Afin d'illustrer le résultat pour la médiane, nous avons à la FIGURE 5 considéré un échantillon de taille  $n = 5$  et les différents cas possibles. La première ligne représente l'échantillon de base, pour lequel aucune contamination n'est apportée. La médiane est représentée en vert et souvenons-nous qu'elle représente la donnée qui laisse de chaque côté 50% des observations. La seconde ligne illustre le cas où la dernière donnée contaminée reste supérieure à la médiane. Dans ce cas, cette dernière est inchangée. La troisième ligne représente le cas où la contamination  $c$ , alors négative, fait dévier  $x_n$  de sorte que ce soit cette dernière qui devient la médiane. Enfin, la quatrième ligne illustre l'ultime possibilité, où la donnée  $x_{(n-1)/2}$  devient alors la médiane.

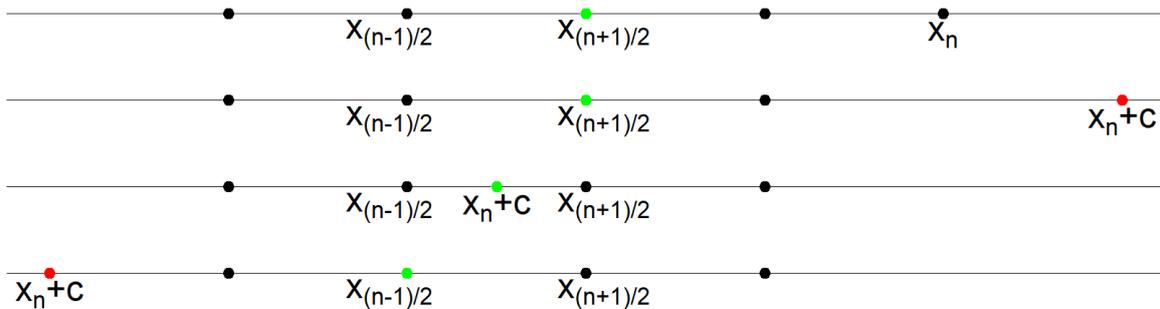


FIGURE 5 – Comportement de la médiane sous l'influence d'une contamination.

En considérant à présent la différence entre les estimateurs basés sur l'échantillon non contaminé et ceux basés sur l'échantillon contaminé, nous obtenons pour la moyenne que  $\hat{\mu}_n^c - \hat{\mu}_n = c/n$ . En réexprimant les conditions en fonction de la contamination  $c$ , nous obtenons pour la médiane

$$\tilde{\mu}_n^c - \tilde{\mu}_n = \begin{cases} 0 & \text{si } c > x_{(n+1)/2} - x_n \\ x_n + c - x_{(n+1)/2} & \text{si } x_{(n-1)/2} - x_n \leq c \leq x_{(n+1)/2} - x_n \\ x_{(n-1)/2} - x_{(n+1)/2} & \text{sinon.} \end{cases} \quad (1.6)$$

Nous observons à la FIGURE 6 que, dans le cas de la moyenne, la différence est une fonction non bornée en  $c$ . Dès lors, en choisissant une valeur appropriée pour la contamination, la moyenne peut atteindre n'importe quelle valeur. Autrement dit, la perturbation d'une seule observation peut avoir un impact arbitrairement grand sur la moyenne. Par contre, la différence pour la médiane est une fonction bornée en  $c$ . Dès lors, une seule contamination aura un impact limité sur la valeur de l'estimateur. En ce sens, nous pouvons dire que la médiane est plus robuste que la moyenne. La fonction d'influence essaye de fournir la même information que ci-dessus en regardant l'effet qu'a une infinitésimal perturbation dans la distribution sur l'estimateur, où ce dernier est à nouveau considéré comme une fonctionnelle. Cependant, avant de définir la fonction d'influence, il est intéressant de définir la notion de dérivée de *Gâteau*.

Considérons deux distributions  $F, H \in \mathcal{F}$  de sorte que, pour tout  $0 \leq t \leq 1$ ,  $(1-t)F + tH \in \mathcal{F}$ . Notons  $\varphi(t) = T((1-t)F + tH)$ , et supposons que la fonction  $\varphi(t)$  est dérivable  $n$  fois, est  $n-1$  fois continûment dérivable sur  $(0, 1)$  et est  $n-1$  fois continûment dérivable à droite en  $t = 0$ . Le

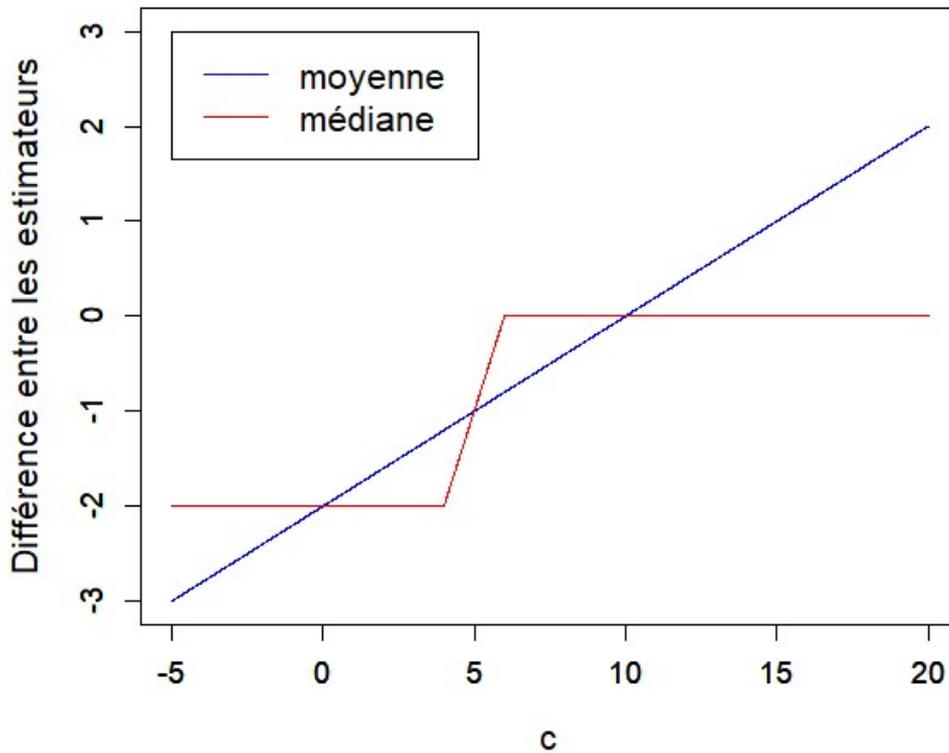


FIGURE 6 – Graphe de  $\hat{\mu}_n^c - \hat{\mu}_n$  et  $\tilde{\mu}_n^c - \tilde{\mu}_n$ .

développement de Taylor de  $\varphi(t)$  autour de  $u \in (0, 1)$  s'écrit

$$\varphi(t) = \varphi(u) + \sum_{k=1}^{n-1} \frac{\varphi^{(k)}(u)}{k!} (t-u)^k + \frac{\varphi^{(n)}(v)}{n!} (t-u)^n, \quad v \in [u, t]. \quad (1.7)$$

Nous nous intéressons ici particulièrement au développement à droite de  $u = 0$ , qui correspond à une petite contamination de  $F$ . Nous remplaçons alors la dérivée  $\varphi^{(k)}(0)$  par la dérivée à droite  $\varphi_+^{(k)}(0)$ . Cette dernière est appelée *dérivée de Gâteaux* de la fonctionnelle  $T$  en  $F$  dans la direction  $H$ .

**Définition 1.2.** Une fonctionnelle  $T$  est Gâteaux différentiable en  $F$  dans la direction  $H$  si, en notant  $F_{t,H} = F + t(H - F)$ , la limite

$$T'(F_{t,H}) = \left[ \frac{dT(F_{t,H})}{dt} \right]_{t=0} = \lim_{t \rightarrow 0^+} \frac{T(F + t(H - F)) - T(F)}{t}$$

existe.

Dans la définition de fonction d'influence, nous noterons plutôt  $F_{t,H} = (1-t)F + tH$  pour mettre en évidence la distribution liée aux contaminations. Lorsque  $H$  est la distribution de Dirac au point  $x \in \mathbb{R}$ , nous noterons  $F_{t,x}$  au lieu de  $F_{t,F_{\delta_x}}$ . Nous pouvons à présent définir la fonction d'influence.

**Définition 1.3.** La fonction d'influence d'un estimateur  $T$  en la distribution  $F$  est la dérivée de Gâteaux de  $T$  en  $F$  dans la direction  $F_{\delta_x}$ ,  $x \in \mathbb{R}$ . Son expression est donnée par

$$\text{IF}(x; T, F) = T'(F_{t,x}) = \lim_{t \rightarrow 0^+} \frac{T(F_{t,x}) - T(F)}{t}. \quad (1.8)$$

Remarquons que nous pouvons aussi écrire

$$\text{IF}(x; T, F) = \left[ \frac{dT(F_{t,x})}{dt} \right]_{t=0}. \quad (1.9)$$

Le théorème ci-dessous livre les deux principales propriétés de la fonction d'influence ([2]).

**Théorème 1.1.** Sous certaines hypothèses de régularité et différentiabilité, la fonction d'influence d'un estimateur  $T$  vérifie, pour une distribution  $F$ ,

- $E_F(\text{IF}(X; T, F)) = 0$ ,
  - $\sqrt{n}(T(F_n) - T(F)) \xrightarrow{D} N(0, \text{Var}_F(\text{IF}(X; T, F)))$ .
- (1.10)

La seconde propriété s'obtient grâce au développement de Von Mises, basé sur (1.7), donné par

$$T(G) = T(F) + \int_{-\infty}^{\infty} \text{IF}(x; T, F) d(G - F)(x) + \text{reste}.$$

Par la première propriété, nous avons que

$$\text{Var}_F(\text{IF}(X; T, F)) = E_F(\text{IF}(X; T, F)^2) - E_F(\text{IF}(X; T, F))^2 = E_F(\text{IF}(X; T, F)^2). \quad (1.11)$$

Grâce à la fonction d'influence, nous pouvons donc obtenir la variance asymptotique  $\text{Var}_F(T(F))$  de (1.4) de l'estimateur  $T(F)$ . Les conditions d'application de (1.10) sont cependant difficiles à

vérifier. Une fois la variance calculée, il est souvent plus simple de montrer la normalité asymptotique d'une autre manière plutôt que de vérifier toutes ces hypothèses. La fonction d'influence est d'ailleurs souvent introduite comme étant un outil *heuristique*, bien que des résultats comme (1.10) sont vérifiés dans la plupart des cas traités.

Afin de faire le lien entre la définition de la fonction d'influence et l'exemple introductif, considérons l'expression (1.8) avec  $t$  fixé. En réécrivant cette expression, nous obtenons

$$T(F_{t,x}) \approx T(F) + t \text{IF}(x; T, F). \quad (1.12)$$

Nous retrouvons alors le fait que la fonction d'influence quantifie la variation de l'estimateur lorsque la distribution est contaminée. Elle représente en effet le biais entre  $T(F_{t,x})$  et  $T(F)$ . Nous constatons de plus qu'il est préférable pour un estimateur d'avoir une fonction d'influence bornée. Dans ce cas là, une faible contamination de la distribution engendrera un effet borné sur l'estimateur.

Dans le but d'obtenir l'expression de la fonction d'influence de la moyenne et de la médiane, nous allons définir la fonction quantile ainsi que donner une propriété.

**Définition 1.4.** *La fonction quantile  $Q$  d'une distribution  $F$  est définie comme*

$$Q(u) = F^{-1}(u) := \inf\{x : F(x) \geq u\}, \quad 0 < u < 1. \quad (1.13)$$

La proposition ci-dessous livre l'expression de la fonction quantile pour la distribution contaminée  $F_{t,x}$ .

**Proposition 1.1.** *Considérons une distribution  $F$  absolument continue et strictement croissante là où elle prend des valeurs différentes de 0 et 1. En notant  $f$  la dérivée de  $F$ , la fonction quantile de  $F_{t,x}$  est donnée par*

$$F_{t,x}^{-1}(u) = \begin{cases} F^{-1}\left(\frac{u}{1-t}\right) & \text{si } u < (1-t)F(x) \\ x & \text{si } (1-t)F(x) \leq u \leq (1-t)F(x) + t \\ F^{-1}\left(\frac{u-t}{1-t}\right) & \text{si } u > (1-t)F(x) + t. \end{cases}$$

De plus, sa dérivée vaut

$$\frac{dF_{t,x}^{-1}(u)}{dt} = \begin{cases} \frac{u}{(1-t)^2} \frac{1}{f(F^{-1}(\frac{u}{1-t}))} & \text{si } u < (1-t)F(x) \\ 0 & \text{si } (1-t)F(x) < u < (1-t)F(x) + t \\ \frac{u-1}{(1-t)^2} \frac{1}{f(F^{-1}(\frac{u-t}{1-t}))} & \text{si } u > (1-t)F(x) + t. \end{cases}$$

*Démonstration.* Nous allons commencer par démontrer l'expression de la fonction quantile. Tout d'abord, remarquons que le dénominateur des premières et troisièmes lignes du résultat sont nuls lorsque  $t = 1$ . Ceci ne pose aucun problème étant donné que les conditions associées deviennent alors  $u < 0$  et  $u > 1$ . Or, la fonction quantile n'est définie que pour  $u$  dans l'intervalle  $(0, 1)$ . Nous avons d'ailleurs dans ce cas particulier que  $F_{1,x} = F_{\delta_x}$ . Cette fonction ne prend comme valeur que 0 et 1 et donc aucune valeur de  $u$  dans  $(0, 1)$  n'est atteinte. Nous allons alors utiliser la définition (1.13), avec laquelle nous trouvons, pour tout  $u$  dans  $(0, 1)$ ,

$$\begin{aligned} F_{\delta_x}^{-1}(u) &= \inf\{k : F_{\delta_x}(k) \geq u\} \\ &= \inf\{k : F_{\delta_x}(k) = 1\} \\ &= x. \end{aligned}$$

La dernière égalité vient du fait que le saut de  $F_{\delta_x}$  s'effectue en  $x$ , où nous avons,  $\forall y < x, F_{\delta_x}(y) = 0$  et  $F_{\delta_x}(x) = 1$ . Nous retrouvons bien l'expression attendue à la seconde ligne du résultat. Notons qu'il s'agit là d'un cas qui nous intéresse très peu étant donné que nous considérons pour la fonction d'influence uniquement d'infinitésimales perturbations, mais l'avoir traité nous permet de considérer à partir de maintenant  $t < 1$ . Nous allons à présent analyser le cas où  $u < (1-t)F(x)$ . Remarquons tout d'abord que, pour tout  $u$  dans  $(0, 1)$  vérifiant cette condition,

$$0 < u < (1-t)F(x) \Rightarrow 0 < \frac{u}{1-t} < F(x) \leq 1.$$

Nous avons pu observer grâce au cas où  $t = 1$  que la fonction quantile est constante lorsqu'un saut apparaît dans la distribution. Ici nous avons supposé  $F$  absolument continue (et donc continue), la fonction quantile est donc strictement croissante sur son domaine, de sorte que

$$\frac{u}{1-t} < F(x) \Leftrightarrow \left( F^{-1} \left( \frac{u}{1-t} \right) \right) < x.$$

Nous avons utilisé le fait que  $F^{-1}(F(x)) = x$  car  $F$  est strictement croissante là où elle prend une valeur différente de 0 et 1. Notons que si  $F(x) = 1$ ,  $F^{-1}(F(x))$  n'est pas défini mais l'inégalité reste vérifiée car nous sommes sûrs que  $\frac{u}{1-t}$  est quant à lui bien dans  $(0, 1)$ . Le cas où  $F(x) = 0$  n'est pas à envisager car  $u > 0$  et nous avons considéré précédemment que  $u < (1-t)F(x)$ . Nous pouvons donc conclure en écrivant

$$\begin{aligned} F_{t,x}(F_{t,x}^{-1}(u)) &= F_{t,x} \left( F^{-1} \left( \frac{u}{1-t} \right) \right) \\ &= (1-t)F \left( F^{-1} \left( \frac{u}{1-t} \right) \right) + tF_{\delta_x} \left( F^{-1} \left( \frac{u}{1-t} \right) \right) \\ &= u, \end{aligned}$$

où nous avons utilisé le fait que  $F(F^{-1}(y)) = y$  pour tout  $y$  dans  $(0, 1)$  étant donné que  $F$  est continue. Le résultat pour la troisième ligne se montre de manière similaire. Enfin, pour la seconde, il suffit d'appliquer un raisonnement semblable à celui utilisé pour le cas de  $F_{1,x}$ , en remarquant que cette fois-ci ce sont les valeurs comprises entre  $(1-t)F(x)$  et  $(1-t)F(x) + t$  qui ne sont pas atteintes (le saut de  $F_{t,x}$  ayant aussi lieu en  $x$ ). Pour la dérivée, si à nouveau nous considérons le cas où  $u < (1-t)F(x)$ , nous trouvons

$$\begin{aligned} \frac{dF_{t,x}^{-1}(u)}{dt} &= \frac{dF^{-1} \left( \frac{u}{1-t} \right)}{dt} \\ &= \left[ \frac{dF^{-1}(y)}{dy} \right]_{y=\left(\frac{u}{1-t}\right)} \frac{d}{dt} \left( \frac{u}{1-t} \right) \\ &= \frac{u}{(1-t)^2} \frac{1}{f \left( F^{-1} \left( \frac{u}{1-t} \right) \right)}, \end{aligned}$$

par la formule de dérivée de l'inverse d'une fonction. La troisième ligne de l'expression du résultat s'obtient de manière similaire, tandis que la seconde est triviale. Notons que les inégalités sont devenues strictes étant donné qu'en ces points, les limites à gauche et à droites diffèrent et donc la dérivée n'existe pas. □

Nous pouvons à présent calculer la fonction d'influence de la moyenne et de la médiane.

**Proposition 1.2.** Soit  $F$  une distribution et  $x \in \mathbb{R}$ . La fonction d'influence de la moyenne est donnée par

$$\text{IF}(x; \hat{\mu}, F) = x - \hat{\mu}(F). \quad (1.14)$$

De plus, si nous supposons  $F$  absolument continue et strictement croissante sur le support de la densité  $f$  (qui est la dérivée de  $F$ ), la fonction d'influence de la médiane vaut

$$\text{IF}(x; \tilde{\mu}, F) = \frac{\text{sign}(x - \tilde{\mu}(F))}{2f(\tilde{\mu}(F))}. \quad (1.15)$$

*Démonstration.* Considérons  $F$  une distribution et  $x \in \mathbb{R}$  quelconque. Nous allons tout d'abord démontrer le résultat pour la moyenne. Nous devons donc calculer

$$\text{IF}(x; \hat{\mu}, F) = \lim_{t \rightarrow 0_+} \frac{\hat{\mu}(F_{t,x}) - \hat{\mu}(F)}{t},$$

avec  $\hat{\mu}(F) = \int_{-\infty}^{+\infty} y dF(y)$ . Commençons par développer le premier terme du numérateur. Par linéarité de l'intégrale par rapport à la mesure, nous trouvons

$$\begin{aligned} \hat{\mu}(F_{t,x}) &= \int_{-\infty}^{+\infty} y dF_{t,x}(y) \\ &= \int_{-\infty}^{+\infty} y d((1-t)F + tF_{\delta_x})(y) \\ &= (1-t) \int_{-\infty}^{+\infty} y dF(y) + t \int_{-\infty}^{+\infty} y dF_{\delta_x}(y) \\ &= (1-t)\hat{\mu}(F) + tx. \end{aligned}$$

Nous trouvons directement

$$\begin{aligned} \text{IF}(x; \hat{\mu}, F) &= \lim_{t \rightarrow 0_+} \frac{(1-t)\hat{\mu}(F) + tx - \hat{\mu}(F)}{t} \\ &= x - \hat{\mu}(F). \end{aligned}$$

Afin de démontrer le résultat pour la médiane, nous allons utiliser la définition  $\tilde{\mu}(F) = F^{-1}(\frac{1}{2})$  ainsi que l'expression (1.9) pour la fonction d'influence. Nous devons donc calculer

$$\text{IF}(x; \tilde{\mu}, F) = \left[ \frac{dF_{t,x}^{-1}(\frac{1}{2})}{dt} \right]_{t=0}.$$

Par la **Proposition 1.1**, nous trouvons directement que

$$\left[ \frac{dF_{t,x}^{-1}(\frac{1}{2})}{dt} \right]_{t=0} = \begin{cases} \frac{1}{2} \frac{1}{f(F^{-1}(\frac{1}{2}))} & \text{si } \frac{1}{2} < F(x), \\ -\frac{1}{2} \frac{1}{f(F^{-1}(\frac{1}{2}))} & \text{si } \frac{1}{2} > F(x) \\ 0 & \text{sinon.} \end{cases}$$

Comme  $F$  est supposée strictement croissante,  $F^{-1}$  l'est aussi, de sorte que

$$\frac{1}{2} < F(x) \Leftrightarrow F^{-1}\left(\frac{1}{2}\right) < x \Leftrightarrow \tilde{\mu}(F) < x.$$

A nouveau, si  $F(x) = 1$ ,  $F^{-1}(F(x))$  n'est pas défini mais l'inégalité reste valide car  $\frac{1}{2}$  est dans  $(0, 1)$ . Nous obtenons enfin que

$$\left[ \frac{dF_{t,x}^{-1}(\frac{1}{2})}{dt} \right]_{t=0} = \begin{cases} \frac{1}{2} \frac{1}{f(\tilde{\mu}(F))} & \text{si } \tilde{\mu}(F) < x, \\ -\frac{1}{2} \frac{1}{f(\tilde{\mu}(F))} & \text{si } \tilde{\mu}(F) > x \\ 0 & \text{sinon} \end{cases}$$

$$= \frac{\text{sign}(x - \tilde{\mu}(F))}{2f(\tilde{\mu}(F))}.$$

□

Remarquons que si  $x = \tilde{\mu}(F)$ , la fonction d'influence vaut alors 0. Comme nous utilisons la **Proposition 1.1** pour  $u = \frac{1}{2}$ , il est suffisant que  $F$  soit strictement croissante uniquement dans un voisinage de  $\tilde{\mu}(F)$ . Nous avons sur la FIGURE 7 représenté les fonctions d'influence de la médiane et de la moyenne pour une loi normale centrée réduite  $N(0, 1)$ . Nous constatons alors que la fonction de la médiane est bornée (car nous avons bien dans notre cas que  $f(\tilde{\mu}(F)) \neq 0$ ), tandis que celle de la moyenne ne l'est pas.

Les définitions précédentes sont utilisées sur base d'estimateurs  $T$  en version population. Il est aussi possible de travailler avec une version empirique de la fonction d'influence.

**Définition 1.5.** Soit  $\{x_1, \dots, x_n\}$  un échantillon et  $F_n$  la fonction de répartition empirique associée. La fonction d'influence empirique pour un estimateur  $T$  est donnée par

$$\text{IF}(x; T_n, F_n) = (n+1)(T_{(n+1)}(x_1, \dots, x_n, x) - T_n(x_1, \dots, x_n)). \quad (1.16)$$

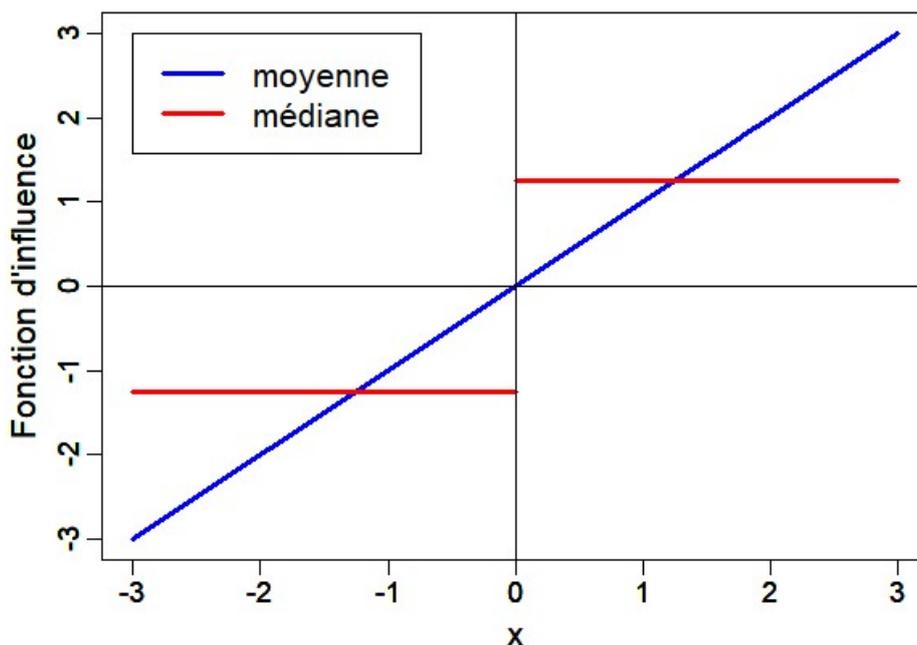


FIGURE 7 – Fonction d'influence pour la médiane et la moyenne avec  $F = N(0, 1)$ .

Nous avons dans l'introduction de ce paragraphe considéré le cas où une donnée était contaminée pour faciliter la compréhension du concept de fonction d'influence, pour ainsi ne pas faire intervenir de facteur  $n + 1$ . Ce dernier est d'ailleurs présent pour s'assurer que nous ayons dans la plupart des cas (bien que ce ne soit justement pas vérifié pour la médiane, comme démontré dans le document [6])

$$\lim_{n \rightarrow \infty} \text{IF}(x; T_n, F_n) = \text{IF}(x; T, F).$$

Il est cependant possible d'étudier une fonction d'influence empirique qui considère qu'une observation est remplacée par une autre au lieu d'en ajouter une, mais cette version est moins étudiée.

En utilisant l'expression (1.16), nous trouvons que la fonction d'influence empirique de la moyenne est donnée par

$$\begin{aligned} \text{IF}(x; \hat{\mu}_n, F_n) &= (n+1) \left( \frac{1}{n+1} \sum_{i=1}^n x_i + \frac{x}{n+1} - \frac{1}{n} \sum_{i=1}^n x_i \right) \\ &= (n+1) \left( \frac{-1}{n+1} \sum_{i=1}^n x_i + \frac{x}{n+1} \right) \\ &= x - \hat{\mu}_n. \end{aligned}$$

Pour la médiane, il faut à présent considérer l'ajout d'une contamination et non plus la contamination d'une donnée. Par un raisonnement analogue à celui illustré à la FIGURE 5, nous trouvons

$$\begin{aligned} \text{IF}(x; \tilde{\mu}_n, F_n) &= (n+1)(\tilde{\mu}_{n+1} - \tilde{\mu}_n) \\ &= \begin{cases} \frac{n+1}{2}(x_{(n+3)/2} - x_{(n+1)/2}) & \text{si } x \geq x_{(n+3)/2} \\ \frac{n+1}{2}(x - x_{(n+1)/2}) & \text{si } x_{(n+3)/2} \geq x \geq x_{(n-1)/2} \\ \frac{n+1}{2}(x_{(n-1)/2} - x_{(n+1)/2}) & \text{si } x \leq x_{(n-1)/2}. \end{cases} \end{aligned}$$

Remarquons que nous avons ici les termes divisés par 2 car nous avons considéré un échantillon de taille  $n$  impaire. Le fait d'ajouter une contamination fait que notre nouvel échantillon est de taille paire, disons  $m$ , et donc qu'il n'existe plus de donnée possédant 50% de l'échantillon de part et d'autre. Nous considérons dans ce cas comme médiane une donnée qui ne fait pas partie de l'échantillon et qui se trouve à mi-distance entre  $x_{m/2}$  et  $x_{m/2+1}$ , de sorte à avoir à nouveau  $m/2$  données de part et d'autre. La FIGURE 8 illustre les fonctions d'influence empiriques de la moyenne et la médiane sur le même échantillon utilisé pour la FIGURE 6. Remarquons que le graphe est fort semblable à celui de la FIGURE 6. Il est cependant différent car non seulement nous avons ici multiplié le résultat par  $(n+1)$  mais de plus, comme mentionné précédemment, nous considérons ici l'ajout d'une donnée et non la contamination d'une déjà existante.

A partir du concept de fonction d'influence, nous pouvons introduire un autre outil mesurant la robustesse.

**Définition 1.6.** *La gross error sensitivity de la fonctionnelle  $T$  en la distribution  $F$  est donnée par*

$$\gamma^*(T, F) = \sup_{x \in \mathbb{R}} |\text{IF}(x; T, F)|.$$

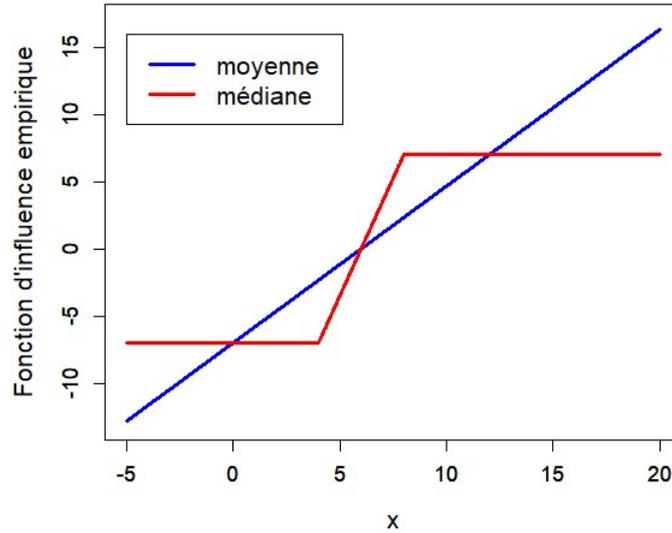


FIGURE 8 – Fonction d’influence empirique de la médiane et la moyenne pour un échantillon de taille  $n = 5$  similaire à celui utilisé pour la FIGURE 6.

Cet outil permet donc de mesurer la pire influence que peut avoir une perturbation sur une population. Il est alors préférable pour un estimateur d’avoir une *gross error sensitivity* finie, traduisant le fait que l’estimateur a une variation bornée si la population est contaminée. A partir des expressions (1.14) et (1.15), nous concluons que  $\gamma^*(\hat{\mu}, F) = \infty$  et  $\gamma^*(\tilde{\mu}, F) < \infty$  pour peu que  $f(\tilde{\mu}(F)) \neq 0$ .

Un second outil pour mesurer la robustesse dérivé de la fonction d’influence peut être vu comme l’impact sur l’estimateur  $T$  qu’a le changement d’une observation  $x_i$  en  $y$ .

**Définition 1.7.** La local sensitivity de la fonctionnelle  $T$  en la distribution  $F$  est donnée par

$$\lambda^*(T, F) = \sup_{x \neq y} \left| \frac{\text{IF}(y; T, F) - \text{IF}(x; T, F)}{y - x} \right|$$

Cet outil représente la pire variation relative que peut engendrer un changement de données sur  $T$ . Remarquons cependant que même une valeur infinie de  $\lambda^*(T, F)$  peut en fait représenter un changement limité à cause du dénominateur. A partir des expressions (1.14) et (1.15), nous obtenons pour nos deux estimateurs de position  $\lambda^*(\hat{\mu}(F), F) = 1$  et  $\lambda^*(\tilde{\mu}(F), F) = \infty$ . Le premier résultat s’obtient trivialement, tandis que pour le second, il suffit de remarquer que la fonction d’influence a un saut en  $\tilde{\mu}(F)$ . Nous remarquons alors que, contrairement à la médiane, la moyenne n’est pas sensible aux changements locaux de données. Cependant, cette caractéristique est beaucoup moins importante que celle d’avoir un  $\gamma^*(T, F)$  fini.

### 1.3.2 Point de rupture

Dans la section précédente, nous nous sommes intéressés à l’influence qu’a le changement d’une observation sur l’estimateur. Nous allons dans cette section nous concentrer sur le nombre maximum d’observations qui peuvent être modifiées avant que l’estimateur ne prenne une valeur arbitrairement grande. Ce dernier nous mènera à la définition de point de rupture.

Commençons tout d’abord par un exemple. Si nous considérons à nouveau l’exemple de la moyenne, nous avons déjà constaté que l’ajout d’une contamination  $c$  fait varier la moyenne

de  $\hat{\mu}$  à  $\hat{\mu}^c = \hat{\mu} + c/n$ . Par conséquent, si nous faisons diverger  $c$  à l'infini, il s'en suit que la moyenne diverge aussi à l'infini et donc prend une valeur arbitrairement grande avec seulement une observation aberrante. Nous en concluons que le nombre maximum de contamination que peut supporter la moyenne avant de diverger est 0, étant donné qu'une seule suffit à la faire diverger. Pour la médiane, nous avons remarqué par (1.5) que l'influence d'une contamination sur l'estimateur est bornée. Par facilité, si nous considérons  $n$  impair, nous pouvons en fait modifier  $(n-1)/2$  données sans que la valeur de la médiane ne tende arbitrairement vers l'infini. A partir de  $(n+1)/2$ , la médiane vaut  $x_k + c$  pour un certain  $1 \leq k \leq n$ , ainsi sa valeur peut devenir arbitrairement grande. Le nombre maximum de changements que peut supporter la médiane avant de diverger est donc de  $(n-1)/2$ . De la même manière, si  $n$  est pair, nous trouvons que ce nombre vaut  $n/2 - 1$ . La FIGURE 9 illustre ces propos. Nous observons que la valeur de la médiane devient arbitrairement grande à partir de 3 contaminations.

Nous pouvons à présent définir le point de rupture empirique.

**Définition 1.8.** *Considérons un échantillon  $x = \{x_1, \dots, x_n\}$  et notons  $x^{(m)}$  ce même échantillon dont  $m$  composantes quelconques ont été modifiées par des valeurs arbitraires. Le point de rupture empirique d'une fonctionnelle  $T_n$  est donné par*

$$\epsilon_n^*(T_n, x) = \frac{1}{n} \max\{m \in \mathbb{N} : \sup_{x^{(m)}} |T_n(x^{(m)})| < \infty\}.$$

Nous avons remarqué que la moyenne peut valoir l'infini après seulement une contamination. Nous obtenons donc que

$$\epsilon_n^*(\hat{\mu}_n, x) = 0.$$

Pour la médiane, nous pouvons modifier au maximum  $(n-1)/2$  observations avant de lui faire atteindre une valeur arbitrairement grande. Nous trouvons alors que

$$\epsilon_n^*(\tilde{\mu}_n, x) = \frac{(n-1)}{2n}.$$

Intéressons nous à présent à la version population du point de rupture. Nous avons précédemment introduit la fonction d'influence, qui mesure l'impact d'infiniment petites perturbations et est donc un concept local. En pratique, nous pouvons être intéressés par l'effet d'une fraction, petite mais fixée, de contamination, comme pour (1.12). Nous aimerions donc un outil plus global pouvant indiquer une région dans laquelle l'approximation (1.12) reste valable de sorte que l'estimateur continue à fournir des informations pertinentes sur le paramètre estimé. Pour

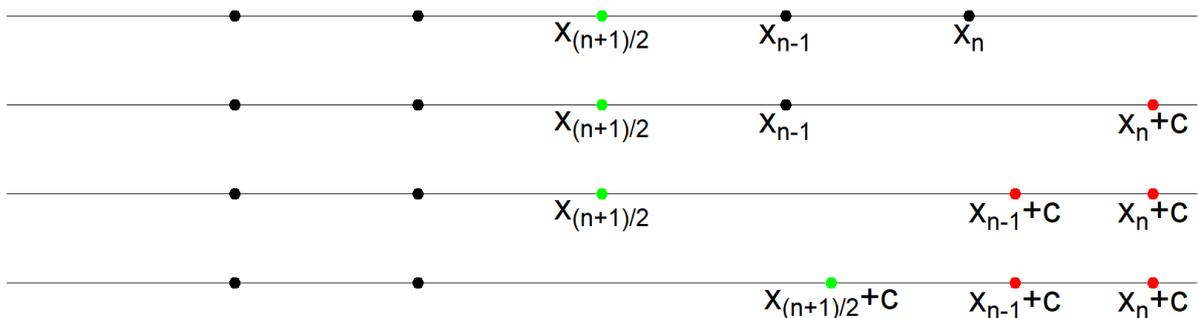


FIGURE 9 – Illustration du point de rupture de la médiane pour  $n = 5$ .

définir une telle région, il nous faut introduire une distance sur l'ensemble des distributions. Par exemple, la *distance de variation totale*.

**Définition 1.9.** Soit  $(\Omega, \mathcal{A}, \mathcal{P})$ , un espace probabilisé et  $F, G$ , deux distributions. La distance de variation totale entre ces deux distributions est donnée par

$$d(F, G) = \sup_{A \in \mathcal{A}} |F(A) - G(A)|.$$

Il faut ensuite se donner une idée de ce que veut dire *fournir des informations pertinentes*. Supposons pour cela que nous voulons estimer un paramètre de position d'une distribution  $F$  qui vaut 3. Considérons l'ensemble des distributions  $G$  se trouvant à une petite distance, disons 0.1, de  $F$  ou moins. Notre estimateur n'est plus du tout fiable si, pour les distributions se trouvant dans cet ensemble, il atteint des valeurs arbitrairement grandes. A partir de là, nous pouvons définir le point de rupture.

**Définition 1.10.** Soit  $F$  issue d'un ensemble de distributions paramétriques  $\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\}$ . Considérons aussi une distance  $d : \mathcal{P}_\theta \times \mathcal{P}_\theta \mapsto \mathbb{R}$ . Le point de rupture d'une fonctionnelle  $T$  est donné par

$$\epsilon^* = \sup_{0 \leq t \leq 1} \{t : \exists K_t \subset \Theta \text{ tel que } d(F, G) < t \text{ implique } P_G(T(F_n) \in K_t) \rightarrow 1 \text{ pour } n \rightarrow \infty\}, \quad (1.17)$$

où  $K_t$  représente un ensemble compact.

Nous retrouvons l'idée que, si une distribution  $G$  est à une distance maximale  $t$  de  $F$ , alors la probabilité que la valeur  $T(F_n)$  reste dans un certain compact, ne diverge pas donc, tend vers 1 pour un échantillon quelconque de taille allant à l'infini. Le point de rupture est la valeur supremum de  $t$  vérifiant cette propriété. Autrement dit, il s'agit de la perturbation maximum que peut subir  $F$  avant que l'estimateur basé sur la distribution perturbée ne diverge. Notons que dans de nombreux cas,

$$\lim_{n \rightarrow \infty} \epsilon_n^* = \epsilon^*.$$

Ainsi, pour la moyenne nous trouvons  $\epsilon^* = 0$ , tandis que pour la médiane nous avons  $\epsilon^* = \frac{1}{2}$ .

Un cas particulier de (1.17) souvent rencontré est le cas où  $d(F, G) < t$  est remplacé par  $G \in \{(1-t)F + tH, H \in \mathcal{F}\}$ . Nous parlons alors de *gross error breakdown point*. Il s'agit alors de la proportion de contamination maximum qui n'amènera jamais l'estimateur à prendre une valeur arbitrairement grande. Cet outil est plus simple à utiliser et correspond la plupart du temps au point de rupture classique. Nous comprenons alors que  $\frac{1}{2}$  est la valeur la plus grande que peut atteindre le point de rupture. En effet, si plus de la moitié des données sont contaminées, il n'est plus possible de distinguer la distribution de base à celle représentant les contaminations. Une telle valeur illustre donc que l'estimateur considéré est hautement robuste. Un point de rupture nul est par contre signe que l'estimateur peut prendre une valeur arbitraire à la moindre perturbation, illustrant alors que ce dernier est très peu robuste.

Notons que la notion de *rupture* dépend de l'espace des paramètres considérés. En effet, si nous prenons le cas où le paramètre de position vaut  $\tan(\tilde{\mu})$ , avec  $\tilde{\mu} \in (-\pi/2, \pi/2)$ , alors par construction, un bon estimateur de  $\tilde{\mu}$  ne divergera jamais pour n'importe quel nombre de contaminations étant donné que  $\tilde{\mu} \in (-\pi/2, \pi/2)$ . Un autre exemple est le cas où on estime des matrices de covariance, où la notion de rupture représente le cas où la plus petite valeur propre tend vers 0, synonyme d'implosion, ou la plus grande vers l'infini, et nous parlons alors d'explosion. Une autre définition est alors nécessaire pour le point de rupture. Une alternative parmi d'autres est de considérer le nombre maximum de contaminations avant que l'estimateur n'atteigne les bornes du domaine de l'espace des paramètres.

### 1.3.3 La courbe de biais

Le dernier outil que nous allons introduire et qui permet d'évaluer la robustesse d'un estimateur est la courbe de biais. Cette dernière étudie la variation maximale de l'estimateur lorsque l'on modifie une certaine proportion de l'échantillon ou de la population initiale.

En guise d'exemple, considérons à nouveau nos estimateurs de position que sont la moyenne et la médiane. Nous avons remarqué dans le paragraphe précédent qu'une seule contamination pouvait engendrer une variation infinie sur la moyenne. Dès lors, pour une fraction  $\epsilon_n = 0$  de contamination, la courbe de biais vaudra 0 tandis que pour tout autre fraction  $\epsilon_n = 1/n, 2/n, \dots, 1$ , la courbe de biais vaudra l'infini. Pour la médiane, supposons que les données sont triées par ordre croissant. A nouveau, pour  $\epsilon_n = 0$ , la variation maximale vaut 0. Lorsque nous considérons une seule donnée contaminée, nous obtenons que la variation maximale vaut

$$\max\{|x_{(n-1)/2} - x_{(n+1)/2}|, |x_{(n+3)/2} - x_{(n+1)/2}|\}.$$

Le premier terme vient de (1.6), dans le cas où  $x_n$  est déplacé en  $-\infty$  (ou au moins devient inférieur à  $x_{(n-1)/2}$ ), tandis que le second provient du cas où  $x_1$  est remplacé par  $+\infty$ . Notons que l'objectif est de capturer la variation la plus grande. Considérer dès lors le cas où  $x_n$  devient  $+\infty$  ou  $x_1$  devient  $-\infty$  est inintéressant étant donné que la médiane dans ces cas là est inchangée. De manière tout à fait similaire, nous trouvons dans le cas où deux données sont contaminées

$$\max\{|x_{(n-3)/2} - x_{(n+1)/2}|, |x_{(n+5)/2} - x_{(n+1)/2}|\},$$

où le premier terme apparaît si nous déplaçons les deux plus grandes données en  $-\infty$  et le second si ce sont les deux plus petites qui deviennent  $+\infty$ . De manière générale, la courbe de biais prend comme valeur

$$\max\{|x_{(n+1)/2-k} - x_{(n+1)/2}|, |x_{(n+1)/2+k} - x_{(n+1)/2}|\}, \quad (1.18)$$

pour  $\epsilon_n = k/n$ ,  $k = 0, \dots, (n-1)/2$ . Pour des fractions plus grandes que  $(n-1)/2$ , et donc supérieures strictement au point de rupture, nous avons, par définition du point de rupture, que la courbe de biais vaut l'infini. Nous avons représenté cette courbe de biais sur la FIGURE 10 pour un échantillon de taille  $n = 13$  issu d'une loi normale centrée réduite. Nous observons qu'au plus le nombre de contaminations augmente, au plus le biais augmente et nous n'avons pas représenté les résultats au-delà du point de rupture qui vaut 6 car la valeur est infinie.

Nous pouvons à présent définir la courbe de biais.

**Définition 1.11.** Soient une distribution initiale  $F_0$  et une fonctionnelle  $T$ . La courbe de biais est donnée par

$$b(t) = \sup_{F \in \mathcal{F}_{F_0,t}} |T(F_0) - T(F)|,$$

où

$$\mathcal{F}_{F_0,t} = \{F : F = (1-t)F_0 + tG, G \text{ une distribution quelconque}\}.$$

Le principal avantage de la courbe de biais par rapport au point de rupture est qu'elle fournit aussi le biais maximum pour chaque valeur de la perturbation inférieure au point de rupture.

Nous allons à présent déterminer l'expression de la courbe de biais pour la médiane. Lorsqu'il s'agit d'estimer le paramètre de position  $\theta$  de  $F(x - \theta)$ , où  $F$  est une distribution inconnue de  $\mathcal{F}_{F_0,t}$ , nous considérons généralement que  $F_0$  est symétrique autour de l'origine tandis que  $G$  peut être symétrique ou non. Nous parlons alors de contamination symétrique ou asymétrique. Beaucoup de fonctionnelles, comme la médiane par exemple, sont monotones par rapport à l'ordre sur les distributions défini de la manière suivante.

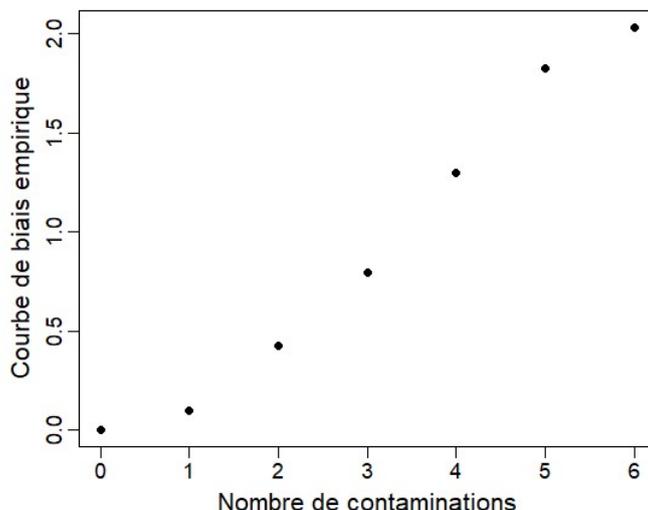


FIGURE 10 – Courbe de biais empirique pour un échantillon Normal de taille  $n = 13$ .

**Définition 1.12.** Une variable aléatoire  $X$  de distribution  $F$  est dite plus petite qu'une variable  $Y$  de distribution  $G$  si

$$F(x) \geq G(x), \quad \forall x \in \mathbb{R}.$$

Une fonctionnelle monotone atteint alors son biais maximum pour la distribution la plus grande notée  $F_\infty$  ou la plus petite notée  $F_{-\infty}$ . Notons que  $F_\infty$  et  $F_{-\infty}$  sont des distributions *limites* et ne vérifient pas les propriétés de base des fonctions de répartition. Nous pouvons donc écrire que

$$b(t) = \max \{ |T(F_\infty) - T(F_0)|, |T(F_{-\infty}) - T(F_0)| \}.$$

Remarquons que nous retrouvons une expression semblable à (1.18), avec  $T(F_{0,n}) = \tilde{\mu}(F_{0,n}) = x_{(n+1)/2}$ . Grâce aux remarques précédentes, nous pouvons déterminer la courbe de biais de la médiane.

**Proposition 1.3.** Soit  $F_0$  une distribution symétrique par rapport à l'origine. La courbe de biais de la médiane est donnée par

$$b(t) = \begin{cases} F_0^{-1} \left( \frac{1}{2(1-t)} \right) & \text{si } 0 \leq t \leq \epsilon^* \\ \infty & \text{si } t > \epsilon^*. \end{cases}$$

*Démonstration.* La seconde ligne du résultat est triviale étant donné que  $\epsilon^*$  représente le point de rupture, défini souvenons-nous comme la proportion maximum de contamination que peut encaisser l'estimateur avant de prendre une valeur arbitraire. Considérons  $t$  dans  $[0, \epsilon^*]$  fixé et notons que comme la distribution  $F_0$  est symétrique autour de l'origine,  $\tilde{\mu}(F_0) = 0$ . Remarquons ensuite que la plus grande distribution de  $\mathcal{F}_{F_0, t}$  est  $F_\infty = (1-t)F_0 + tF_{\delta_\infty}$ . En effet, comme  $F_0$  et  $t$  sont fixés, le seul paramètre qui varie est  $G$  et pour tout  $x \in \mathbb{R}$ , pour tout  $G$ ,  $G(x) \geq F_{\delta_\infty}(x)$ . Par définition de la médiane, cette dernière doit satisfaire

$$F_\infty(\tilde{\mu}(F_\infty)) = \frac{1}{2} \quad \Rightarrow \quad (1-t)F_0(\tilde{\mu}(F_\infty)) = \frac{1}{2} \quad \Rightarrow \quad F_0^{-1} \left( \frac{1}{2(1-t)} \right) = \tilde{\mu}(F_\infty).$$

Le point de rupture de la médiane valant 0.5, nous sommes sûrs que le dénominateur ne sera jamais nul. Par un raisonnement similaire,  $F_{-\infty} = (1-t)F_0 + tF_{\delta_{-\infty}}$  et sa médiane  $\tilde{\mu}(F_{-\infty})$  doit satisfaire

$$(1-t)F_0(\tilde{\mu}(F_{-\infty})) + t = \frac{1}{2} \quad \Rightarrow \quad \tilde{\mu}(F_{-\infty}) = F_0^{-1} \left( 1 - \frac{1}{2(1-t)} \right) = -\tilde{\mu}(F_\infty).$$

La dernière égalité venant du fait que  $F_0$  est symétrique par rapport à l'origine. Nous trouvons enfin que

$$\begin{aligned} b(t) &= \max \{ |\tilde{\mu}(F_\infty) - \tilde{\mu}(F_0)|, |\tilde{\mu}(F_{-\infty}) - \tilde{\mu}(F_0)| \} \\ &= \max \{ |\tilde{\mu}(F_\infty)|, |-\tilde{\mu}(F_\infty)| \} \\ &= \tilde{\mu}(F_\infty) \\ &= F_0^{-1} \left( \frac{1}{2(1-t)} \right). \end{aligned}$$

□

Nous pouvons montrer, de plus, que  $t\gamma^*$  approxime la courbe de biais. Cette approximation est évidemment très mauvaise pour  $t$  proche de  $\epsilon^*$  et même invalide pour  $t > \epsilon^*$ . Il est préférable de considérer cette approximation uniquement pour  $t < \epsilon^*/2$ . Tout cela illustre très bien le fait que la fonction d'influence, au travers de la gross error sensitivity, et le point de rupture fournissent des informations complémentaires. En effet, la fonction d'influence approxime le biais, là où le point de rupture détermine la région où cette approximation reste convenable. Nous avons à la FIGURE 11 représenté la courbe de biais pour la médiane avec  $F_0 = \Phi$ , la fonction de répartition de la loi normale centrée réduite. Nous avons aussi représenté l'approximation linéaire donnée par  $t\gamma^*$  ainsi que, verticalement, la zone de validité de cette approximation, donnée par  $\epsilon^*/2$ .

## 1.4 Estimateurs robustes de position

### 1.4.1 M-estimateurs

Nous avons jusqu'à maintenant considéré l'estimation de paramètres de position. Nous allons à présent introduire des formes visant à estimer les paramètres génériques dans un modèle paramétrique. Supposons à nouveau avoir un échantillon  $\{x_1, \dots, x_n\}$  et que nous voulons estimer un paramètre  $\theta$  de la distribution dont est issu l'échantillon. Un estimateur bien connu est l'estimateur de maximum de vraisemblance (MLE), défini comme la valeur  $T_n(x_1, \dots, x_n)$  qui maximise,

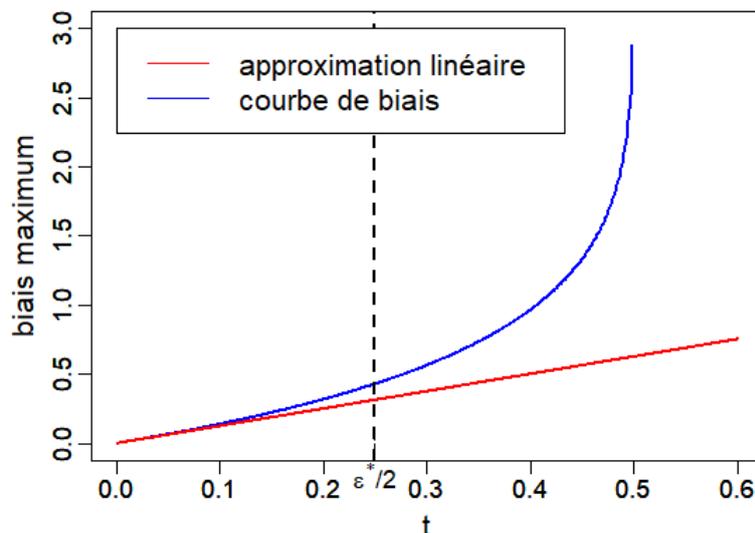


FIGURE 11 – Courbe de biais de la médiane ainsi que son approximation linéaire.

en  $\theta$ ,

$$\prod_{i=1}^n f_{\theta}(x_i) \quad \theta \in \Theta,$$

ou de manière équivalente, qui minimise

$$\sum_{i=1}^n [-\ln f_{\theta}(x_i)] \quad \theta \in \Theta. \quad (1.19)$$

La fonction  $f_{\theta}$  représente la fonction de densité de la distribution  $F_{\theta}$ , distribution que nous supposons alors continue, ou la fonction de masse de probabilité pour une distribution discrète. La distribution et la densité sont supposés connus mais leurs paramètres sont inconnus. Par exemple, nous pourrions obtenir des données issues d'une loi normale, où nous connaîtrions alors  $f_{\theta}$  comme fonction de  $\theta = (\mu, \sigma)$ . Ainsi, pour chaque estimateur  $T_n$  de  $\theta$ , nous pouvons calculer  $f_{T_n}$  de sorte à trouver celui ou ceux qui minimisent (1.19).

Les M-estimateurs sont une généralisation de l'idée ci-dessus.

**Définition 1.13.** Soit  $\{x_1, \dots, x_n\}$ , un échantillon. Un M-estimateur est défini comme étant la valeur minimisant

$$\sum_{i=1}^n \rho(x_i, \theta) \quad \theta \in \Theta, \quad (1.20)$$

où  $\rho$  est une fonction quelconque.

Si nous considérons un échantillon issu d'une loi normale  $N(\mu, 1)$  dont la densité est donnée par

$$f_{\mu}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2},$$

et que nous voulons estimer le paramètre de position  $\mu$ , nous pouvons considérer la fonction  $\rho(x, \theta) = -\ln f_{\theta}(x)$  avec  $\theta = \mu$ . Le M-estimateur associé minimise l'expression (1.19), qui vaut alors

$$\sum_{i=1}^n [-\ln f_{\theta}(x_i)] = \sum_{i=1}^n \left[ -\ln \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \right) \right].$$

Comme il s'agit d'un problème de minimisation, nous pouvons éliminer les constantes de sorte qu'il reste à minimiser

$$\sum_{i=1}^n \left[ -\ln \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} \right) \right] = \sum_{i=1}^n (x_i - \mu)^2.$$

L'expression obtenue correspond à la définition de la moyenne donnée en (1.3), ce qui confirme le fait que cette dernière est bien l'estimateur de maximum de vraisemblance dans le cas gaussien.

Si  $\rho$  est dérivable par rapport à  $\theta$ , alors, en notant

$$\psi(x, \theta) = \frac{\partial}{\partial \theta} \rho(x, \theta),$$

l'estimateur  $T_n$  minimisant (1.20) doit satisfaire aussi

$$\sum_{i=1}^n \psi(x_i, T_n) = 0. \quad (1.21)$$

Il est évident que cette dernière condition n'est pas équivalente à la première, mais elle s'avère cependant particulièrement utile dans la recherche d'estimateurs minimisant (1.20).

Supposons à présent que les fonctions  $\rho(\cdot, \theta)$  et  $\psi(\cdot, \theta)$  sont continûment dérivables par rapport à  $\theta$  et considérons  $F_n$ , la distribution empirique basée sur un échantillon. Nous pouvons alors écrire l'expression (1.20), à un facteur  $n$  près, comme

$$\int_{-\infty}^{\infty} \rho(x, T(F_n)) dF_n(x),$$

et la condition (1.21) devient

$$\int_{-\infty}^{\infty} \psi(x, T(F_n)) dF_n(x) = 0.$$

Remarquons que nous ne pouvons pas toujours définir la fonctionnelle  $T(F)$  correspondante à  $T(F_n)$  comme étant la donnée minimisant

$$\int_{-\infty}^{\infty} \rho(x, T(F)) dF(x).$$

En effet, dans le cas de la médiane, nous avons  $\rho(x, \theta) = |x - \theta|$  et

$$\int_{-\infty}^{\infty} |x - \theta| dF(x) = \infty,$$

indépendamment de la valeur de  $\theta$  et pour une distribution  $F$  n'ayant pas un premier moment absolu fini. Pour rappel, ce dernier est défini comme

$$E[|X|] = \int_{-\infty}^{\infty} |x| dF(x).$$

Une solution simple dans le cas de la médiane est de remplacer dans l'intégrale  $\rho(x, \theta)$  par  $\rho(x, \theta) - \rho(x, \theta_0)$  pour une valeur de  $\theta_0$  quelconque fixée, et chercher donc à minimiser

$$\int_{-\infty}^{\infty} (|x - \theta| - |x|) dF(x).$$

La condition pour  $T(F)$  équivalente à (1.21) devient

$$\int_{-\infty}^{\infty} \psi(x, T(F)) dF(x) = 0, \quad (1.22)$$

et cette dernière ne souffre pas du problème décrit précédemment. Nous pouvons à présent calculer la fonction d'influence d'un M-estimateur.

**Théorème 1.2.** *Soit  $F$  une distribution et supposons que les fonctions données  $\rho(\cdot, \theta)$  et  $\psi(\cdot, \theta)$  sont continûment dérivables par rapport à  $\theta$ . Considérons  $T(F)$ , un estimateur vérifiant (1.22) et  $x \in \mathbb{R}$ . La fonction d'influence de cet estimateur est donnée par*

$$\left[ \frac{dT(F_t)}{dt} \right]_{t=0} = \text{IF}(x; T, F) = \frac{\psi(x, T(F))}{-\int_{-\infty}^{\infty} \dot{\psi}(y, T(F)) dF(y)},$$

avec

$$\dot{\psi}(y, T(F)) = \left[ \frac{\partial}{\partial \theta} \psi(y, \theta) \right]_{\theta=T(F)}.$$

*Démonstration.* En remplaçant  $F$  par  $F_{t,x}$  dans (1.22) et en utilisant la linéarité de l'intégrale par rapport à la mesure, la condition devient

$$\int_{-\infty}^{\infty} \psi(y, T(F_{t,x})) d((1-t)F + tF_{\delta_x})(y) = 0$$

$$\Leftrightarrow (1-t) \int_{-\infty}^{\infty} \psi(y, T(F_{t,x})) dF(y) + t \psi(x, T(F_{t,x})) = 0.$$

En supposant qu'intégrale et dérivée permutent et en dérivant cette dernière expression par rapport à  $t$ , nous trouvons

$$- \int_{-\infty}^{\infty} \psi(y, T(F_{t,x})) dF(y) + (1-t) \frac{dT(F_{t,x})}{dt} \int_{-\infty}^{\infty} \left[ \frac{\partial}{\partial \theta} \psi(y, \theta) \right]_{\theta=T(F_{t,x})} dF(y)$$

$$+ \psi(x, T(F_{t,x})) + t \frac{dT(F_{t,x})}{dt} \left[ \frac{\partial}{\partial \theta} \psi(x, \theta) \right]_{\theta=T(F_{t,x})} = 0.$$

Si nous évaluons cette expression en  $t = 0$ , le dernier terme s'annule et, comme  $F_{0,x} = F$ , le premier terme est nul aussi par (1.22). Le reste de l'expression livre

$$\left[ \frac{dT(F_{t,x})}{dt} \right]_{t=0} = \text{IF}(x; T, F) = \frac{\psi(x, T(F))}{-\int_{-\infty}^{\infty} \dot{\psi}(y, T(F)) dF(y)},$$

où

$$\dot{\psi}(y, T(F)) = \left[ \frac{\partial}{\partial \theta} \psi(y, \theta) \right]_{\theta=T(F)}.$$

□

En supposant que le dénominateur de l'expression obtenue est non nul, nous obtenons directement l'expression de la fonction d'influence d'un M-estimateur, étant donné que tout M-estimateur doit vérifier (1.22). Remarquons que le dénominateur est constant, et donc la fonction d'influence a la même forme que la fonction  $\psi$  choisie. L'influence sera donc bornée si et seulement si  $\psi(\cdot, T(F))$  est bornée. Le choix de  $\psi$  est donc crucial pour la robustesse du M-estimateur formé.

Nous pouvons aussi déduire la valeur de la variance asymptotique de  $T(F)$  grâce à (1.11). Nous obtenons alors

$$\text{Var}_F T(F) = E_F[\text{IF}(x; T, F)^2] = \frac{\int_{-\infty}^{\infty} \psi^2(x, T(F)) dF(x)}{\left[ \int_{-\infty}^{\infty} \dot{\psi}(y, T(F)) dF(y) \right]^2}.$$

Un cas particulier souvent traité est le cas où l'échantillon  $\{x_1, \dots, x_n\}$  provient d'une distribution  $F(x - \theta)$ , où l'on veut alors estimer  $\theta$ , le paramètre (réel) de position. Dans ce cas, un M-estimateur  $T_n$  minimise

$$\sum_{i=1}^n \rho(x_i - \theta), \quad \theta \in \Theta. \quad (1.23)$$

A nouveau, si  $\psi$ , la dérivée de  $\rho$ , existe et est continue, alors  $T_n$  vérifie

$$\sum_{i=1}^n \psi(x_i - T_n) = 0. \quad (1.24)$$

La fonction d'influence de l'équivalent de  $T_n$  en version population pour une distribution  $G$  vaut

$$\text{IF}(x; T, G) = \frac{\psi(x - T(G))}{\int_{-\infty}^{\infty} \dot{\psi}(y - T(G)) dG(y)}. \quad (1.25)$$

Si l'on considère la distribution  $F$  de laquelle provient l'échantillon, nous avons  $T(F) = 0$  et

$$\text{IF}(x; T, F) = \frac{\psi(x)}{\int_{-\infty}^{\infty} \dot{\psi}(y) dF(y)},$$

en supposant toujours que le dénominateur est non-nul. La variance asymptotique est alors donnée par

$$\text{Var}_F T(F) = \frac{\int_{-\infty}^{\infty} \psi^2(x) dF(x)}{\left[ \int_{-\infty}^{\infty} \dot{\psi}(y) dF(y) \right]^2}.$$

Il est possible, à partir de la fonction  $\psi$ , de déterminer le point de rupture d'un M-estimateur de position. Si ce dernier estime le centre de symétrie de  $F(x - \theta)$ , que l'on suppose alors symétrique, nous pouvons dire que son point de rupture vaut 0 si  $\psi$  est non bornée et 1/2 si  $\psi$  est impaire et bornée. La classe des M-estimateurs contient donc des estimateurs très robustes et des estimateurs très peu robustes.

A partir de l'expression (1.23), nous observons que l'estimateur  $T_n$  est invariant par translation, c'est-à-dire qu'il vérifie

$$T_n(x_1 + c, \dots, x_n + c) = T_n(x_1, \dots, x_n) + c \quad \forall c \in \mathbb{R}.$$

En effet, si nous considérons  $\{x_1 + c, \dots, x_n + c\}$ , nous avons que

$$\arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho(x_i + c - \theta) = \left( \arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho(x_i - \theta) \right) + c.$$

Par contre, l'estimateur  $T_n$  n'est pas invariant par changement d'échelle, autrement dit il ne vérifie pas

$$T_n(cx_1, \dots, cx_n) = cT_n(x_1, \dots, x_n) \quad \forall c > 0.$$

Si nous considérons un échantillon  $\{cx_1, \dots, cx_n\}$ , de manière générale, l'égalité

$$\arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho(cx_i - \theta) = c \left( \arg \min_{\theta \in \Theta} \sum_{i=1}^n \rho(x_i - \theta) \right).$$

n'est pas vérifiée.

Si la distribution  $F$  est symétrique autour de 0, il est intéressant de choisir  $\rho$  symétrique autour de 0 et donc  $\psi$  est alors impaire. Si  $\rho(x)$  est une fonction strictement convexe, alors la fonction (1.23) est strictement convexe en  $\theta$ , et le M-estimateur est alors unique. Dans ce cas,  $\psi(x)$  est strictement croissante. Si enfin  $\rho(x)$  est linéaire sur un intervalle  $[a, b]$ , alors  $\psi$  est constante sur ce même intervalle, de sorte que l'expression (1.24) peut admettre plusieurs solutions.

Afin d'illustrer ces résultats, nous pouvons considérer le cas de la moyenne. La fonction  $\rho(x)$  à choisir est donnée par  $\rho(x) = x^2$ , et donc  $\psi(x) = 2x$  et  $\dot{\psi}(x) = 2$ . Comme  $\rho(x)$  est une fonction

strictement convexe, le problème de minimisation admet une unique solution qui doit vérifier la condition (1.22). Cette dernière est donnée dans notre cas par

$$2 \int_{-\infty}^{\infty} (x - T(F)) dF(x) = 0,$$

où nous trouvons alors directement  $T(F) = \int_{-\infty}^{\infty} x dF(x) = \hat{\mu}(F)$ . En utilisant l'expression (1.25), nous obtenons

$$\text{IF}(x; T, F) = \frac{2(x - \hat{\mu}(F))}{\int 2dF(x)} = x - \hat{\mu}(F),$$

ce qui est bien l'expression obtenue précédemment. De plus, comme la fonction  $\psi$  est non bornée, nous avons bien que le point de rupture vaut 0. Si nous reprenons la condition (1.24), nous obtenons

$$\begin{aligned} 2 \sum_{i=1}^n (x_i - T_n) &= 0 \\ \Leftrightarrow T_n &= \frac{1}{n} \sum_{i=1}^n x_i. \end{aligned}$$

Nous retrouvons alors l'expression de  $\bar{x}$ , la version empirique de l'espérance.

### 1.4.2 L-estimateurs

Il existe d'autres classes d'estimateurs robustes, comme par exemple les L-estimateurs.

**Définition 1.14.** Soit  $\{x_1 \leq \dots \leq x_n\}$  un échantillon ordonné. La forme générale des L-estimateurs est donnée par

$$T_n = \sum_{i=1}^n c_i h(x_i) + \sum_{j=1}^k a_j h^*(x_{[np_j+1]}), \quad (1.26)$$

où  $c_1, \dots, c_n$  et  $a_1, \dots, a_k$  sont des coefficients donnés,  $0 < p_1 < \dots < p_k < 1$ ,  $h(\cdot)$  et  $h^*(\cdot)$  sont des fonctions.

Les coefficients  $c_i$ ,  $1 \leq i \leq n$ , sont générés par une fonction de poids bornée  $J : [0, 1] \rightarrow \mathbb{R}$  et sont déterminés soit par

$$c_i = \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(s) ds, \quad i = 1, \dots, n \quad (1.27)$$

ou alors par l'approximation

$$c_i = \frac{1}{n} J\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n.$$

La première partie de l'expression (1.26) prend en compte en général toutes les données, là où la seconde est une combinaison linéaire de certains quantiles. Beaucoup de L-estimateurs ne font intervenir que la première ou la seconde partie de l'expression (1.26). Nous parlons alors d'un estimateur de type I ou type II.

Comme exemple de L-estimateurs, nous connaissons déjà la médiane empirique. En effet, pour un échantillon de taille impaire, nous pouvons la retrouver en prenant par exemple  $a_1 = \dots = a_k = 0$ ,  $h(x) = x$  et

$$c_i = \begin{cases} 1 & \text{si } i = \frac{n+1}{2} \\ 0 & \text{sinon.} \end{cases}$$

Un autre estimateur, appelé *mid-range*, est défini comme

$$T_n = \frac{1}{2}(x_1 + x_n),$$

que l'on peut facilement obtenir en prenant les mêmes valeurs des paramètres que pour la médiane, sauf

$$c_i = \begin{cases} \frac{1}{2} & \text{si } i = 1, n \\ 0 & \text{sinon.} \end{cases}$$

Nous allons à présent nous concentrer plus particulièrement sur les L-estimateurs de type I. Supposons que notre L-estimateur a une fonction de poids  $J$  intégrable, vérifiant  $\int_0^1 J(u) du = 1$ . De la même manière que nous passons d'une somme à une intégrale par rapport à la distribution empirique  $F_n$ , dans la section (1.2) par exemple, nous allons dans notre cas plutôt utiliser la fonction quantile empirique étant donné que notre échantillon est ordonné. Cette dernière est définie comme

$$Q_n(t) = F_n^{-1}(t) = \inf\{x : F_n(x) \geq t\}, \quad 0 < t < 1.$$

Cette expression n'est rien d'autre que la version empirique de la fonction quantile définie en (1.13). L'expression de  $Q_n(t)$  est d'ailleurs donnée par

$$Q_n(t) = \begin{cases} x_i & \text{si } \frac{i-1}{n} < t \leq \frac{i}{n}, \quad \forall i = 1, \dots, n-1, \\ x_n & \text{si } \frac{n-1}{n} < t < 1. \end{cases} \quad (1.28)$$

Nous pouvons dès lors réécrire l'expression générale (1.26) pour un estimateur de type I comme

$$\begin{aligned} T_n &= \sum_{i=1}^n c_i h(x_i) = \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(s) h(x_i) ds \\ &= \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} J(s) h(Q_n(s)) ds \\ &= \int_0^1 J(s) h(Q_n(s)) ds, \end{aligned}$$

où nous avons appliqué successivement (1.27) et (1.28). La statistique exprimée comme fonctionnelle, c'est-à-dire la version population de  $T_n$ , peut alors être obtenue par

$$T(F) = \int_0^1 J(s) h(Q(s)) ds.$$

Nous allons à présent déterminer la fonction d'influence d'un L-estimateur.

**Théorème 1.3.** *Soient  $F$  une distribution absolument continue et strictement croissante là où elle est différente de 0 et 1 et  $T(F)$ , le L-estimateur défini ci-dessus, associé à la fonction de poids  $J$ . Sa fonction d'influence est donnée par*

$$\text{IF}(x; T, F) = \int_{-\infty}^{\infty} F(y) h'(y) J(F(y)) dy - \int_x^{\infty} h'(y) J(F(y)) dy.$$

*Démonstration.* Nous allons utiliser la **Proposition 1.1** étant donné que  $F$  vérifie ses hypothèses. Nous obtenons directement que

$$\begin{aligned} \frac{dT(F_{t,x})}{dt} &= \int_0^1 J(u) h'(F_{t,x}^{-1}(u)) \frac{dF_{t,x}^{-1}}{dt} du \\ &= \int_0^{(1-t)F(x)} \frac{u}{(1-t)^2} \frac{h'\left(F^{-1}\left(\frac{u}{1-t}\right)\right)}{f\left(F^{-1}\left(\frac{u}{1-t}\right)\right)} J(u) du \\ &\quad + \int_{(1-t)F(x)+t}^1 \frac{u-1}{(1-t)^2} \frac{h'\left(F^{-1}\left(\frac{u-t}{1-t}\right)\right)}{f\left(F^{-1}\left(\frac{u-t}{1-t}\right)\right)} J(u) du. \end{aligned}$$

Le fait de pouvoir échanger dérivée et intégrale doit évidemment être vérifié dans chaque cas particulier, par rapport aux fonctions  $h$  et  $J$  considérées. Comme  $F$  est strictement croissante, la fonction  $f$  ne peut s'annuler sur un intervalle (sinon  $F$  serait constante sur ce dernier) mais bien qu'en des ensembles de mesure de Lebesgue nulle. Le second dénominateur des intégrales obtenues peut donc être nul ponctuellement, mais ceci n'a pas de conséquence sur la valeur de l'intégrale. Le premier est quant à lui non nul car nous considérons des valeurs de  $t$  proches de 0. Évaluer l'expression obtenue en prenant  $t \rightarrow 0^+$  nous livre

$$\begin{aligned} \left. \frac{dT(F_{t,x})}{dt} \right|_{t=0} &= \int_0^{F(x)} u \frac{h'(F^{-1}(u))}{f(F^{-1}(u))} J(u) du + \int_{F(x)}^1 (u-1) \frac{h'(F^{-1}(u))}{f(F^{-1}(u))} J(u) du \\ &= \int_0^1 u \frac{h'(F^{-1}(u))}{f(F^{-1}(u))} J(u) du - \int_{F(x)}^1 \frac{h'(F^{-1}(u))}{f(F^{-1}(u))} J(u) du. \end{aligned}$$

En effectuant un changement de variable  $u = F(y)$ , nous trouvons que l'expression de la fonction d'influence d'un L-estimateur est donnée par

$$\text{IF}(x; T, F) = \int_{-\infty}^{\infty} F(y) h'(y) J(F(y)) dy - \int_x^{\infty} h'(y) J(F(y)) dy.$$

□

Remarquons que la dérivée de la fonction d'influence vérifie

$$\frac{d}{dx} \text{IF}(x; T, F) = h'(x) J(F(x)).$$

De plus, si nous considérons que  $h(x) = x$ ,  $J(u) = J(1-u)$ ,  $0 < u < 1$  et que  $F$  est symétrique autour de 0 (et donc  $F(x) = 1 - F(-x)$ ), la fonction d'influence devient

$$\begin{aligned} \text{IF}(x; T, F) &= \int_{-\infty}^{\infty} F(y) J(F(y)) dy - \int_x^{\infty} J(F(y)) dy \\ &= \int_0^{\infty} F(y) J(F(y)) dy + \int_{-\infty}^0 (1 - F(-y)) J(1 - F(-y)) dy - \int_x^{\infty} J(F(y)) dy \\ &= \int_0^{\infty} F(y) J(F(y)) dy + \int_0^{\infty} (1 - F(y)) J(F(y)) dy - \int_x^{\infty} J(F(y)) dy \\ &= \int_0^{\infty} J(F(y)) dy - \int_x^{\infty} J(F(y)) dy. \end{aligned}$$

Nous trouvons alors que

$$\text{IF}(x; T, F) = \int_0^x J(F(y)) dF(y), \quad \forall x \geq 0$$

et

$$\text{IF}(-x; T, F) = \text{IF}(x; T, F), \quad \forall x \in \mathbb{R}.$$

Comme pour les M-estimateurs, il existe un résultat livrant directement le point de rupture d'un L-estimateur sous certaines hypothèses. En effet, si la fonction  $J$  satisfait  $J(u) = 0$  pour  $0 < u \leq \alpha$  et  $1 - \alpha \leq u < 1$  avec  $0 < \alpha < \frac{1}{2}$ , alors  $\epsilon^* = \alpha$ .

## 1.5 Estimateurs robustes de dispersion

Nous allons à présent brièvement discuter de l'estimation du paramètre de dispersion. De tels problèmes sont en fait rares car le paramètre de dispersion intervient le plus souvent comme un paramètre de nuisance dans l'estimation robuste du paramètre de position. De manière générale, un estimateur de dispersion est une fonctionnelle positive  $S$  invariante par transformation d'échelle. Il existe différentes classes d'estimateurs de dispersion. Une première classe concerne celle des M-estimateurs. De manière similaire à (1.22), nous pouvons obtenir un M-estimateur de dispersion  $S(F)$  en résolvant

$$\int_{-\infty}^{\infty} \chi\left(\frac{x}{S(F)}\right) dF(x) = 0,$$

où  $\chi$  est une fonction adéquatement choisie. Par un raisonnement tout à fait similaire à celui proposé dans la démonstration du **Théorème 1.2**, nous trouvons que

$$\text{IF}(x; S, F) = \frac{\chi(x/S(F))S(F)}{\int_{-\infty}^{\infty} \dot{\chi}(y/S(F))(y/S(F))dF(y)}.$$

Si nous considérons par exemple  $\chi(x) = \text{sign}(|x| - 1)$ , nous obtenons

$$\begin{aligned} & 2 \int_{-\infty}^{\infty} \text{sign}\left(\left|\frac{x}{S(F)}\right| - 1\right) dF(x) = 0 \\ \Leftrightarrow & \int_{-\infty}^{\infty} \text{sign}(|x| - S(F)) dF(x) = 0 \\ \Leftrightarrow & \int_{-\infty}^{-S(F)} dF(x) + \int_{S(F)}^{\infty} dF(x) - \int_{-S(F)}^{S(F)} dF(x) = 0 \\ \Leftrightarrow & F(-S) + 1 - F(S) - F(S) + F(-S) = 0 \\ \Leftrightarrow & F(S) - F(-S) = \frac{1}{2} \end{aligned} \tag{1.29}$$

En utilisant la définition de la fonction de répartition, nous trouvons

$$\begin{aligned} P(X \leq S) - P(X \leq -S) = \frac{1}{2} & \Leftrightarrow P(-S \leq X \leq S) = \frac{1}{2} \Leftrightarrow P(|X| \leq S) = \frac{1}{2} \\ & \Leftrightarrow F_{|X|}(S) = \frac{1}{2} \Leftrightarrow S = F_{|X|}^{-1}\left(\frac{1}{2}\right) = \text{med}(|X|). \end{aligned}$$

Si  $F$  est symétrique autour de 0, nous avons que

$$F(S) - F(-S) = \frac{1}{2} \Leftrightarrow F(S) - (1 - F(S)) = \frac{1}{2} \Leftrightarrow F(S) = \frac{3}{4} \Leftrightarrow S = F^{-1}\left(\frac{3}{4}\right). \quad (1.30)$$

Afin de simplifier l'expression de la fonction d'influence pour un tel choix de fonction  $\chi$ , remarquons tout d'abord que (en se souvenant que la dérivée de  $\text{sign}(x)$  vaut  $2\delta(x)$ )

$$\left[ \frac{d}{dx} \text{sign}(|x| - 1) \right]_{x=\frac{y}{S}} = [2\delta(|x| - 1) \text{sign}(x)]_{x=\frac{y}{S}} = 2S\delta(|y| - S) \text{sign}(y).$$

Nous avons utilisé à la dernière égalité la propriété

$$\delta(\alpha x) = \frac{1}{|\alpha|} \delta(x),$$

pour tout  $x$  et  $\alpha \neq 0$ . La fonction d'influence devient alors

$$\begin{aligned} \text{IF}(x; S, F) &= \frac{\text{sign}(|x| - S)S}{\int_{-\infty}^{\infty} 2S\delta(|y| - S) \text{sign}(y)(y/S)dF(y)} \\ &= \frac{\text{sign}(|x| - S)}{\int_{-\infty}^{\infty} 2\delta(|y| - S)(|y|/S)f(y)dy} \\ &= \frac{\text{sign}(|x| - S)}{2(f(S) + f(-S))} \end{aligned}$$

Si  $F$  est symétrique, l'expression se simplifie encore en

$$\text{IF}(x; S, F) = \frac{\text{sign}(|x| - S)}{4f(S)}.$$

Un autre type de problème lié aux M-estimateurs, qui va d'ailleurs nous mener à l'un des estimateurs de dispersion les plus utilisés, est l'estimation simultanée de position et de dispersion. Il s'agit de résoudre pour  $S(F)$  et  $T(F)$

$$\begin{aligned} \int \psi\left(\frac{x - T(F)}{S(F)}\right) dF(x) &= 0, \\ \int \chi\left(\frac{x - T(F)}{S(F)}\right) dF(x) &= 0. \end{aligned}$$

En effectuant les mêmes opérations que dans la preuve du **Théorème 1.2** (à savoir, évaluer les expressions en  $F_{t,x}$ , les dériver par rapport à  $t$  puis considérer  $t = 0$ ), nous trouvons

$$\begin{aligned} \text{IF}(x; F, T) \int \psi'\left(\frac{y - T(F)}{S(F)}\right) dF(y) + \text{IF}(x; F, S) \int \psi'\left(\frac{y - T(F)}{S(F)}\right) \frac{y - T(F)}{S(F)} dF(y) \\ = \psi\left(\frac{x - T(F)}{S(F)}\right) S(F), \\ \text{IF}(x; F, T) \int \chi'\left(\frac{y - T(F)}{S(F)}\right) dF(y) + \text{IF}(x; F, S) \int \chi'\left(\frac{y - T(F)}{S(F)}\right) \frac{y - T(F)}{S(F)} dF(y) \\ = \chi\left(\frac{x - T(F)}{S(F)}\right) S(F). \end{aligned} \quad (1.31)$$

Si  $F$  est symétrique (autour de  $T(F)$ ),  $\chi$  est paire et  $\psi$  est impaire, certaines intégrales se simplifient et nous trouvons

$$\begin{aligned} \text{IF}(x; F, T) &= \frac{\psi\left(\frac{x}{S(F)}\right) S(F)}{\int \psi'\left(\frac{y}{S(F)}\right) dF(y)}, \\ \text{IF}(x; F, S) &= \frac{\chi\left(\frac{x}{S(F)}\right) S(F)}{\int \chi'\left(\frac{y}{S(F)}\right) \frac{y}{S(F)} dF(y)}. \end{aligned}$$

Considérons à présent  $\psi(x) = \text{sign}(x)$  et  $\chi(x) = \text{sign}(|x| - 1)$ . Notons que ces choix, pris seuls, sont ceux correspondant à  $T = \text{med}(X)$  et  $S = \text{med}|X|$ . Le système à résoudre devient

$$\begin{aligned} \int \text{sign}\left(\frac{x - T(F)}{S(F)}\right) dF(x) &= 0, \\ \int \text{sign}(|x - T(F)| - S(F)) dF(x) &= 0. \end{aligned}$$

La première équation livre à nouveau  $T(F) = \text{med}(X)$ . En effet, comme  $S$  est toujours positif,

$$\begin{aligned} \int \text{sign}\left(\frac{x - T(F)}{S(F)}\right) dF(x) = 0 &\Leftrightarrow \int \text{sign}(x - T(F)) dF(x) = 0 \\ \Leftrightarrow -\int_{-\infty}^T dF(x) + \int_T^{\infty} dF(x) = 0 &\Leftrightarrow -F(T) + 1 - F(T) = 0, \end{aligned}$$

et ainsi  $T = F^{-1}\left(\frac{1}{2}\right)$ . Maintenant que  $T(F)$  a été trouvé, la seconde égalité peut être résolue en fonction de  $S(F)$ . Par un calcul tout à fait similaire à celui proposé afin d'obtenir le résultat (1.29), nous trouvons

$$F(S + T) - F(-S + T) = \frac{1}{2}.$$

Enfin, il suffit d'effectuer un changement de variable  $Y = X - T$  et de remarquer que

$$F_X(S + T) = P[X \leq (S + T)] = P[X - T \leq S] = P[Y \leq S] = F_Y(S)$$

Ainsi, en utilisant (1.29), nous trouvons

$$F_Y(S) - F_Y(-S) = \frac{1}{2} \Leftrightarrow S = \text{med}(|Y|) \Leftrightarrow S = \text{med}|X - T|,$$

avec toujours  $T = \text{med}(X)$ . Cet estimateur porte le nom de *median absolute deviation* (MAD) et est parfois précédé de la constante  $(\Phi^{-1}\left(\frac{3}{4}\right))^{-1} \approx 1.48$  de sorte que, lorsque  $F = N(0, 1)$ , nous obtenons par (1.30) que  $S(F) = 1$ . Le théorème suivant livre la fonction d'influence de cet estimateur.

**Théorème 1.4.** *Soit  $F$  une distribution strictement croissante et  $x \in \mathbb{R}^p$ . La fonction d'influence de l'estimateur  $S = \text{MAD}$  est donnée par*

$$\text{IF}(x; F, S) = \frac{\text{sign}(|x - T| - S) - \frac{f(T+S) - f(T-S)}{f(T)} \text{sign}(x - T)}{2[f(T + S) + f(T - S)]},$$

où  $T = F^{-1}\left(\frac{1}{2}\right)$  et  $f$  est la fonction de densité. De plus, si  $F$  est symétrique, cette expression devient

$$\text{IF}(x; F, S) = \frac{\text{sign}(|x| - S)}{4f(S)}.$$

*Démonstration.* Si  $F$  est symétrique,  $T = F^{-1}\left(\frac{1}{2}\right) = 0$ , de sorte que la seconde partie du théorème découle directement de la première. Pour cette dernière, il faut considérer l'expression obtenue en (1.31) en utilisant  $\psi(x) = \text{sign}(x)$ ,  $\chi(x) = \text{sign}(|x| - 1)$  et  $\text{IF}(x; F, T) = \frac{\text{sign}(x - T)}{2f(T)}$  que nous avons obtenu en (1.15). Nous obtenons alors

$$\begin{aligned} & \frac{\text{sign}(x - T)}{2f(T)} \int 2S\delta(|y - T| - S)\text{sign}(y - T)dF(y) \\ & + \text{IF}(x; F, S) \int 2S\delta(|y - T| - S)\text{sign}(y - T)\frac{y - T}{S}dF(y) = \text{sign}(|x - T| - S)S. \end{aligned}$$

Nous allons nous concentrer sur la première intégrale. Nous trouvons

$$\begin{aligned} & \int \delta(|y - T| - S)\text{sign}(y - T)dF(y) \\ & = - \int_{-\infty}^T \delta(-y + T - S)dF(y) + \int_T^{\infty} \delta(y - T - S)dF(y) \\ & = f(T + S) - f(T - S). \end{aligned}$$

La seconde intégrale livre

$$\begin{aligned} & \int \delta(|y - T| - S)\text{sign}(y - T)\frac{y - T}{S}dF(y) \\ & = - \int_{-\infty}^T \delta(-y + T - S)\frac{y - T}{S}dF(y) + \int_T^{\infty} \delta(y - T - S)\frac{y - T}{S}dF(y) \\ & = f(T + S) + f(T - S). \end{aligned}$$

Nous obtenons donc, en remplaçant les valeurs des intégrales,

$$\begin{aligned} & \frac{\text{sign}(x - T)}{2f(T)} 2S(f(T + S) - f(T - S)) + \text{IF}(x; F, S) 2S(f(T + S) - f(T - S)) \\ & = \text{sign}(|x - T| - S)S \\ \Leftrightarrow \text{IF}(x; F, S) & = \frac{\text{sign}(|x - T| - S)S - \text{sign}(x - T)S \left( \frac{f(T + S) - f(T - S)}{f(T)} \right)}{2S(f(T + S) + f(T - S))} \\ \Leftrightarrow \text{IF}(x; F, S) & = \frac{\text{sign}(|x - T| - S) - \frac{f(T + S) - f(T - S)}{f(T)}\text{sign}(x - T)}{2[f(T + S) + f(T - S)]}. \end{aligned}$$

□

Remarquons que l'expression obtenue dans le cas où  $F$  est symétrique correspond à celle obtenue pour la fonction d'influence de  $S = \text{med}(|X|)$  qui concorde finalement avec le MAD dans ce cas-ci (car  $T = 0$ ). Pour autant que  $f(T)$  et  $2[f(T + S) + f(T - S)]$  sont non nuls, la fonction d'influence est bornée. Cet estimateur a de plus un point de rupture de 0.5 ce qui confirme qu'il s'agit d'un estimateur hautement robuste.

## 1.6 Généralisation en dimension supérieure

Nous allons à présent discuter des mêmes outils que ceux développés précédemment, mais cette fois-ci dans le cas où les données sont dans l'espace  $\mathbb{R}^p$ ,  $p \in \mathbb{N}_0$ .

Un estimateur peut à nouveau être défini comme une fonctionnelle  $T$  qui prend maintenant valeur dans un espace à plusieurs dimensions. Cet espace dépendra donc du paramètre estimé. Nous allons considérer pour l'instant des estimateurs de position, de sorte que  $T \in \mathbb{R}^p$ . La définition (1.3) de fonction d'influence peut tout à fait être transposée pour des données en dimension  $p$ , en considérant alors  $x \in \mathbb{R}^p$ . Le **Théorème 1.1** reste valable pour des données multivariées. Notons que, comme la matrice de covariance d'un vecteur aléatoire  $X$  est donnée par

$$\text{Var}(X) = E[(X - E[X])(X - E[X])'],$$

l'expression de la variance asymptotique obtenue en (1.11) devient

$$\text{Var}_F(\text{IF}(X; T, F)) = E[\text{IF}(X; T, F) \text{IF}(X; T, F)']. \quad (1.32)$$

Tous les outils définis dans la première section peuvent être adaptés au cas multivarié en considérant la norme euclidienne au lieu des valeurs absolues.

Nous allons à présent discuter brièvement des M-estimateurs. Ces derniers sont définis de manière similaire au cas univarié.

**Définition 1.15.** *Un M-estimateur est défini à partir d'une fonction  $\rho$  à valeur dans  $\mathbb{R}$  comme la donnée  $T(F) \in \mathbb{R}^p$  minimisant, en  $\theta$*

$$\int \rho(x, \theta) dF(x),$$

ou à travers une fonction  $\psi$  à valeur dans  $\mathbb{R}^p$  comme solution de l'équation vectorielle

$$\int \psi(x, \theta) dF(x) = 0. \quad (1.33)$$

A nouveau, les deux fonctions  $\rho$  et  $\psi$  sont liées par la relation

$$\psi(x, \theta) = \frac{\partial \rho(x, \theta)}{\partial \theta} = \left( \frac{\partial \rho(x, \theta)}{\partial \theta^{(1)}}, \dots, \frac{\partial \rho(x, \theta)}{\partial \theta^{(p)}} \right)'$$

Pour obtenir l'estimateur de maximum de vraisemblance, il suffit, comme pour le cas univarié de considérer  $\rho(x, \theta) = -\ln(f(x, \theta))$ , où  $f$  est la densité de la distribution paramétrique  $F_\theta$ .

Il est aussi possible dans le cas multivarié d'obtenir l'expression de la fonction d'influence d'un M-estimateur. Pour ce faire, il suffit de reproduire les étapes de la démonstration du **Théorème 1.2** en remarquant que

$$\frac{\partial}{\partial \theta} \psi(x, \theta)$$

forme une matrice de dimension  $p$ . Nous trouvons alors pour la fonction d'influence

$$\text{IF}(x; T, F) = M(\psi, F)^{-1} \psi(x, T(F)),$$

avec

$$M(\psi, F) \stackrel{\text{not.}}{=} - \int \left[ \frac{\partial}{\partial \theta} \psi(x, \theta) \right]_{T(F)} dF(x).$$

Pour obtenir la variance asymptotique de cet estimateur, il suffit d'utiliser la formule (1.32) avec l'expression obtenue ci-dessus. Nous trouvons directement

$$\text{Var}_F(T(F)) = M(\psi, F)^{-1}Q(\psi, F)M(\psi, F)^{-T},$$

avec

$$Q(\psi, F) \stackrel{\text{not.}}{=} \int \psi(x, T(F))\psi(x, T(F))^T dF(x).$$

Nous pouvons à présent nous intéresser à l'estimation du paramètre de dispersion, à savoir la matrice de covariance. Nous allons commencer par l'estimateur *minimum covariance determinant* (MCD) qui est l'un des premiers estimateurs robustes affins équivariants développés pour l'estimation de matrice de dispersion mais aussi de position. Avant de définir cet estimateur, notons  $[\cdot]$  la fonction qui arrondi un réel à l'entier inférieur.

**Définition 1.16.** *L'estimateur MCD de paramètre  $h$  vérifiant  $n/2 \leq h \leq n$  définit un estimateur de position et de dispersion de la manière suivante :*

- $\hat{\mu}_0$  est la moyenne des  $h$  observations pour lesquelles le déterminant de la matrice de covariance empirique est minimal.
- $\hat{\Sigma}_0$  est la matrice de covariance empirique correspondante multipliée par un facteur  $c_0$ .

Pour rappel, la matrice de covariance empirique pour un échantillon  $\{x_1, \dots, x_n\}$  est donnée par

$$S_n = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'. \quad (1.34)$$

Pour qu'une telle matrice soit non singulière, il faut considérer  $n > p$ . Ainsi, lors de l'évaluation de la matrice de covariance empirique pour  $h$  données, il est nécessaire que  $h$  soit strictement supérieur à  $p$ . Comme pour le MAD, la constante  $c_0$  est présente afin que l'estimateur corresponde à la matrice de covariance lorsque  $F$  est une distribution normale. Cette dernière est donnée par

$$c_0 = \frac{\alpha}{F_{\chi_{p+2}^2}(\chi_{p,\alpha}^2)}.$$

La notation  $\chi_{p,\alpha}^2$  représente le quantile  $\alpha$  de la distribution chi-carrée à  $p$  degrés de liberté, tandis que le paramètre  $\alpha$  vaut

$$\alpha = \lim_{n \rightarrow \infty} \frac{h(n)}{n}.$$

L'estimateur MCD a été conçu pour des distributions elliptiques symétriques et unimodales. Une distribution multivariée de paramètre  $\mu \in \mathbb{R}^p$  et  $\Sigma$ , une matrice définie positive de dimension  $p$ , est dite elliptique symétrique et unimodale s'il existe une fonction strictement décroissante  $g$  telle que la densité peut être écrite

$$f(x) = \frac{1}{\sqrt{|\Sigma|}} g((x - \mu)' \Sigma^{-1} (x - \mu)).$$

Le choix de  $h = \frac{n+p+1}{2}$  rend l'estimateur MCD le plus robuste (nous aurons l'occasion de voir pourquoi au paragraphe suivant), ce qui correspond à  $\alpha = 0.5$ . Afin de comparer différents estimateurs, il est commun d'utiliser l'efficacité relative asymptotique, définie comme suit.

**Définition 1.17.** *L'efficacité relative asymptotique (ARE) entre deux estimateurs  $T_1$  et  $T_2$  est donnée par*

$$e(T_1, T_2) = \frac{\text{Var}(T_2)}{\text{Var}(T_1)}.$$

Cette quantité n'a évidemment de sens que pour des estimateurs du même paramètre. Lorsque ce dernier est une matrice, la comparaison est faite élément par élément. Il est courant de considérer au numérateur l'estimateur ayant la plus faible variance, de sorte que l'ARE est comprise entre 0 et 1. De plus, lorsque nous parlons d'ARE d'un estimateur, cela sous-entend que nous le comparons avec le MLE du paramètre estimé, car ce dernier estimateur a la propriété d'avoir la plus petite variance asymptotique. Donc au plus l'efficacité relative asymptotique est proche de 1 ou 100%, au plus la variance asymptotique de l'estimateur considéré se rapproche de la valeur la plus petite possible et donc au plus il s'agit d'un bon estimateur. Le MCD a une efficacité relative asymptotique (ARE) très faible en comparant avec la matrice de variance empirique, qui est le MLE de  $\Sigma$ . Dans le cas du MCD avec  $\alpha = 0.5$ , l'ARE des éléments diagonaux n'est que de 6% pour  $p = 2$  et 20.5% pour  $p = 10$ . Cette valeur peut être augmentée en prenant  $\alpha = 0.75$ , où nous obtenons alors 26.2% lorsque  $p = 2$  et 45.9% lorsque  $p = 10$ . Un tel choix diminue cependant la robustesse de l'estimateur MCD, dans le sens où il est plus sensible aux données aberrantes. Dans le but d'augmenter son efficacité tout en conservant sa robustesse, nous pouvons considérer les estimateurs MCD pondérés donnés par

$$\hat{\mu}_{MCD} = \frac{\sum_{i=1}^n W(d_i^2)x_i}{\sum_{i=1}^n W(d_i^2)}$$

et

$$\hat{\Sigma}_{MCD} = c_1 \frac{1}{n} \sum_{i=1}^n W(d_i^2)(x_i - \hat{\mu}_{MCD})(x_i - \hat{\mu}_{MCD})',$$

avec  $d_i = \sqrt{(x_i - \hat{\mu}_0)' \hat{\Sigma}_0^{-1} (x_i - \hat{\mu}_0)}$  et  $W$  une fonction de poids. L'idée est de déterminer des outliers grâce aux estimateurs non pondérés au travers de  $d_i^2$ , et d'associer alors un poids plus faible aux données ayant une grande valeur de  $d_i^2$ . Un choix classique de fonction  $W$  est de considérer une fonction indicatrice qui associe donc 1 aux données dont la distance  $d_i^2$  est inférieure à une certaine valeur et 0 sinon. A nouveau, la constante  $c_1$  apparaît pour une question de consistance dans le cas gaussien. Cette pondération permet, avec  $\alpha = 0.5$ , d'obtenir une ARE de 45.5% pour  $p = 2$  et 82% pour  $p = 10$ .

Avant d'énoncer quelques propriétés sur l'estimateur MCD, nous définissons l'affin équivalence d'estimateurs de position et de dispersion.

**Définition 1.18.** *Des estimateurs de position  $T \in \mathbb{R}^p$  et de dispersion  $S \in \mathbb{R}^{p \times p}$  sont dits affins équivariants si, pour toute matrice non singulière  $A$  de dimension  $p \times p$  et pour tout vecteur  $b \in \mathbb{R}^p$*

$$\begin{aligned} T(XA' + 1_n b') &= T(X)A' + b, \\ S(XA' + 1_n b') &= AS(X)A', \end{aligned}$$

où  $1_n$  correspond au vecteur de dimension  $n$  ne contenant que des 1.

De telles propriétés assurent que l'estimateur se transforme bien lorsque les données sont modifiées par des transformations comme une rotation ou une translation. Nous pouvons à présent énoncer une première propriété de l'estimateur MCD.

**Proposition 1.4.** *L'estimateur MCD de position et dispersion est un estimateur afffin équivariant.*

*Démonstration.* Tout d'abord, notons  $X_h$  le sous-ensembles de  $h$  données et remarquons que  $\hat{\mu}_0$  est défini comme la moyenne de  $X_h$  et est donc afffin équivariant. De la même manière,  $\hat{\Sigma}_0$  est une matrice de covariance empirique basée sur  $X_h$  et est donc aussi afffin équivariant. Il est important de remarquer que le déterminant de la matrice de covariance basée sur  $X_h$  vérifie

$$|S(X_h A')| = |AS(X_h)A'| = |A|^2 |S(X_h)|.$$

Il s'en suit que le sous ensemble optimal, qui minimise donc le déterminant de  $S(X_h)$ , correspond à celui qui minimise  $S(X_h A' + 1_n b')$ . Enfin, remarquons que l'affin equivariance de  $\hat{\mu}_0$  et  $\hat{\Sigma}_0$  rendent la distance  $d_i$  afffin invariante, de sorte que  $\hat{\mu}_{MCD}$  et  $\hat{\Sigma}_{MCD}$  sont aussi afffin équivariants.  $\square$

Nous pouvons à présent discuter du point du rapture du MCD. Pour un estimateur de position  $T_n$ , nous allons ici considérer la définition

$$\epsilon_n^*(T_n, X_n) = \frac{1}{n} \min\{m \in \{1, \dots, n\} : \sup_m \|T_n(X_n) - T_n(X_{n,m})\| = +\infty\},$$

où  $X_{n,m}$  représente  $X_n$  où  $m$  données ont été remplacées par des valeurs arbitraires. Notons que la définition ici utilisée représente la proportion minimum suffisante pour que l'estimateur prenne une valeur arbitrairement grande. Cette dernière est donc légèrement différente de celle proposée précédemment, qui renvoyait la proportion maximum de contamination que pouvait supporter un estimateur avant de prendre une valeur arbitraire. De la même manière, pour un estimateur de dispersion  $C_n$ , nous considérons

$$\epsilon_n^*(C_n, X_n) = \frac{1}{n} \min\left\{m \in \{1, \dots, n\} : \sup_m \max_i |\log(\lambda_i(C_n(X_{n,m}))) - \log(\lambda_i(C_n(X_n)))| = +\infty\right\},$$

où  $0 < \lambda_p(C_n) \leq \dots \leq \lambda_1(C_n)$  sont les valeurs propres de  $C_n$ . Nous considérons donc l'estimateur comme obsolète lorsqu'une des valeurs propres devient arbitrairement grande ou proche de 0. Notons  $k(X_n)$  le nombre maximum d'observations dans le jeu de données se trouvant sur un hyperplan de  $\mathbb{R}^p$ . Sous l'hypothèse que  $k(X_n) < h$ , alors nous avons

$$\epsilon_n^*(\hat{\mu}_0, X_n) = \epsilon_n^*(\hat{\Sigma}_0, X_n) = \frac{\min(n - h + 1, h - k(X_n))}{n}.$$

Si les données sont issues d'une distribution continue, alors nous pouvons montrer que, presque sûrement,  $k(X_n) = p$ , de sorte que  $\epsilon_n^*(\hat{\mu}_0, X_n) = \epsilon_n^*(\hat{\Sigma}_0, X_n) = \min(n - h + 1, h - p)/n$ . Dès lors, si  $h = \lfloor \frac{n+p+1}{2} \rfloor$ , alors

$$\min\left(n - \left\lfloor \frac{n+p+1}{2} \right\rfloor + 1, \left\lfloor \frac{n+p+1}{2} \right\rfloor - p\right) = \min\left(\left\lfloor \frac{n-p+1}{2} \right\rfloor, \left\lfloor \frac{n-p+1}{2} \right\rfloor\right).$$

Ainsi,  $\epsilon_n^*(\hat{\mu}_0, X_n) = \epsilon_n^*(\hat{\Sigma}_0, X_n) = \left\lfloor \frac{n-p+1}{2} \right\rfloor$ . Il a été démontré que cette valeur est la plus grande que peut atteindre un estimateur de dispersion afffin equivariant pour des données en position générale (avec donc  $k(X_n) = p$ ). Remarquons enfin que

$$\lim_{n \rightarrow \infty} \epsilon_n^* = \min(\alpha, 1 - \alpha),$$

qui est maximum et vaut 0.5 lorsque  $\alpha = 0.5$ . Le point de rapture des estimateurs pondérés n'est pas inférieur à celui des estimateurs classiques, pour peu que la fonction  $W$  est décroissante,

bornée et nulle pour de grandes valeurs de  $d_i^2$ .

Nous allons à présent discuter de la fonction d'influence de l'estimateur MCD. Pour obtenir son expression, souvenons-nous qu'il faut considérer l'estimateur comme une fonctionnelle. Pour ce faire, notons

$$\mathcal{D}_F(\alpha) = \{A | A \subset \mathbb{R}^p \text{ mesurable et borné avec } P_F(A) = \alpha\},$$

où  $\alpha$  dénote comme précédemment la proportion de données considérées par l'estimateur MCD. Pour tout ensemble  $A \in \mathcal{D}_F(\alpha)$ , la moyenne et la matrice de covariance obtenues à partir de ce sous-ensemble sont données par

$$T_A(F) = \frac{\int_A y dF(y)}{\alpha}$$

et

$$\Sigma_A(F) = \frac{\int_A (y - T_A(F))(y - T_A(F))' dF(y)}{\alpha}.$$

L'élément  $A$  est appelé solution du problème du MCD si  $\det(\Sigma_A(F)) \leq \det(\Sigma_{\tilde{A}}(F))$ , pour tout autre  $\tilde{A} \in \mathcal{D}_F(\alpha)$ . Ainsi, la version population de l'estimateur MCD est donnée par

$$T(F) = T_A(F) \quad \text{et} \quad \Sigma(F) = c_\alpha \Sigma_A(F),$$

pour  $A$  une solution du problème du MCD.

La définition suggérée précédemment pose problème lorsque la distribution  $F$  n'est pas continue, auquel cas  $\mathcal{D}_F(\alpha)$  peut être vide. Nous définirons alors plutôt cet ensemble comme

$$\mathcal{D}_F(\alpha) = \{(A, x) | A \subset \mathbb{R}^p \text{ mesurable et borné, } x \in \mathbb{R}^p \setminus A \text{ et } \exists 0 \leq \delta \leq P_F(\{x\}) : P_F(A) + \delta = \alpha\}.$$

Les expressions des estimateurs deviennent alors

$$T_A(F) = \frac{\int_A y dF(y) + \delta x}{\alpha}$$

et

$$\Sigma_A(F) = \frac{\int_A (y - T_A(F))(y - T_A(F))' dF(y) + \delta(x - T_A(F))\delta(x - T_A(F))'}{\alpha}.$$

Si  $F$  est continue,  $P_F(\{x\}) = 0$ , de sorte que les définitions sont semblables à celles fournies précédemment. Nous considérons à présent  $(A, x) \in \mathcal{D}_F(\alpha)$  comme solution du problème du MCD si

$$\det(\Sigma_{(A,x)}(F)) \leq \det(\Sigma_{(\tilde{A},\tilde{x})}(F)),$$

pour tout autre  $(\tilde{A}, \tilde{x}) \in \mathcal{D}_F(\alpha)$ . Ainsi, la version population de l'estimateur MCD est donnée par

$$T(F) = T_{(A,x)}(F) \quad \text{et} \quad \Sigma(F) = c_\alpha \Sigma_{(A,x)}(F),$$

pour  $(A, x)$  une solution du problème du MCD.

La fonction d'influence obtenue ([7]) s'avère être bornée et, lorsque  $F$  est la distribution normale centrée réduite multivariée, la fonction d'influence pour l'estimateur de position vaut 0 pour tout  $x$  vérifiant  $\|x\| > \chi_{p,\alpha}^2$ . Ainsi, des outliers éloignés n'influencent pas la valeur de l'estimateur. Le même phénomène se produit pour les éléments hors-diagonale de l'estimateur de dispersion, tandis que pour les éléments diagonaux, la fonction d'influence reste constante pour des valeurs de  $\|x\|^2$  grandes. Les données aberrantes ont donc toujours une influence, bornée,

sur cet estimateur.

Nous allons à présent discuter d'un autre estimateur de dispersion multivarié très utilisé. Il s'agit du *minimum volume ellipsoid* (MVE). Ce dernier est souvent utilisé en pratique, en particulier pour détecter des observations aberrantes. Une approche classique pour réaliser cet objectif est de calculer la distance de Mahalanobis de chaque donnée en fonction de  $\bar{x}$  et  $S_n$ , c'est-à-dire

$$\text{MD}(x_i) = \sqrt{(x_i - \bar{x})' S_n^{-1} (x_i - \bar{x})} \quad i = 1, \dots, n.$$

Nous pouvons montrer que si les données suivent une loi normale en dimension  $p$ , alors la distance de Mahalanobis au carré suit approximativement une loi  $\chi_p^2$ , de sorte que nous comparons les distances à  $\chi_{p,0.975}^2$ . C'est ainsi que nous parlons d'ellipse, en considérant

$$\mathcal{E}(\bar{x}, S_n, 0.975) = \left\{ x : \text{MD}(x) \leq \sqrt{\chi_{p,0.975}^2} \right\}.$$

Comme les estimateurs utilisés sont peu robustes, il est tout à fait possible que des outliers ont des distances très petites et sont donc indétectables. Afin de pallier ce problème, et d'ainsi bien estimer le centre et la dispersion d'un jeu de données, il nous faut considérer des estimateurs robustes comme le sont ceux du MVE.

**Définition 1.19.** Soit  $h$  un paramètre compris entre  $\lfloor n/2 \rfloor + 1$  et  $n$ . L'estimateur de position  $T_n$  et de dispersion  $C_n$  appelés MVE minimisent le déterminant de  $C$  vérifiant la condition

$$\#\{i : (x_i - t)' C^{-1} (x_i - t) \leq c^2\} \geq h,$$

où la minimisation se fait sur  $t \in \mathbb{R}^p$  et  $C \in \text{PSD}$ , l'ensemble des matrices semi-définies positives de dimension  $p$ .

La constante  $c$  est un paramètre à fixer et pour avoir consistance dans le modèle gaussien, il faut considérer  $c = \chi_{p,\alpha}^2$ , avec à nouveau

$$\alpha = \lim_{n \rightarrow \infty} \frac{h(n)}{n}.$$

Le MVE va donc estimer le centre et la dispersion de (au moins)  $h$  données parmi les plus rassemblées.

Les estimateurs MVE sont, comme pour le MCD, des estimateurs affins équivariants. Il suffit de remarquer que la transformation non-singulière  $x \rightarrow A'x + b$  transforme un ellipsoïde de centre  $T$  et de matrice de dispersion  $S$  contenant au moins  $h$  données de  $X$  en un ellipsoïde de centre  $A'T + b$  de matrice de dispersion  $A'SA$  contenant au moins  $h$  données de  $XA + 1_n b'$ . Le volume du nouvel ellipsoïde vaut  $\det(A'SA)^{1/2} = |\det(A)| \det(S)^{1/2}$  de sorte que le MVE de  $XA + 1_n b'$  sont bien donnés par  $A'T_n + b$  et  $A'C_n A$ , où  $T_n$  et  $C_n$  sont les estimateurs MVE de  $X$ .

Lorsque les données sont en position générale, le MVE a le même point de rupture que celui du MCD. Les conclusions faites pour ce dernier s'appliquent donc aussi pour le MVE. Concernant l'efficacité asymptotique, il a été démontré que l'estimateur MVE converge vers une loi non-gaussienne avec un taux de  $n^{-1/3}$ , et non vers une loi gaussienne avec un taux de  $n^{-1/2}$  comme pour le MCD de position par exemple. Pour le MCD de dispersion, seul le taux de  $n^{-1/2}$  a été établi, mais pas la loi asymptotique. A nouveau, pour améliorer l'efficacité du MVE, nous pouvons considérer un estimateur pondéré comme pour le MCD. Notons que cette pondération améliorera l'efficacité pour un échantillon mais pas le taux de convergence. Notons enfin que le MVE est un cas particulier d'un ensemble d'estimateurs appelé les S-estimateurs.

**Définition 1.20.** Les  $S$ -estimateurs de position et dispersion  $(T_n, C_n)$  sont les solutions qui minimisent le déterminant de  $C$  sous la contrainte

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left( \sqrt{(x_i - t)' C^{-1} (x_i - t)} \right) \leq b,$$

avec  $t \in \mathbb{R}^p$  et  $C \in PSD$ .

Pour retrouver la définition du MVE, il faut considérer  $\rho_0 = 1 - \mathbb{I}_{[0,c]}$  et  $b = (n - h)/n$ . En effet, nous obtenons dans ce cas

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n 1 - \mathbb{I}_{[0,c]} \left( \sqrt{(x_i - t)' C^{-1} (x_i - t)} \right) \leq \frac{n - h}{n} \\ \Leftrightarrow & \quad n - \sum_{i=1}^n \mathbb{I}_{[0,c]} \left( \sqrt{(x_i - t)' C^{-1} (x_i - t)} \right) \leq n - h \\ \Leftrightarrow & \quad \#\{i : (x_i - t)' C^{-1} (x_i - t) \leq c^2\} \geq h. \end{aligned}$$

En général, une fonction  $\rho$  au moins une fois continument dérivable est utilisée, assurant que l'estimateur converge asymptotiquement vers une loi normale avec un taux de  $n^{-1/2}$ . Sous certaines conditions supplémentaires, nous pouvons obtenir l'expression générale de la fonction d'influence d'un  $S$ -estimateur, mais malheureusement le MVE ne remplit pas ces conditions.

Avant d'introduire le dernier estimateur, nous allons brièvement parler de  $M$ -estimateurs de dispersion. Pour cela, nous nous restreignons à la classe des distributions elliptiques. L'objectif est alors d'estimer les paramètres  $\mu$  et  $\Sigma$  de la distribution.

**Définition 1.21.** Les  $M$ -estimateurs de position et dispersion  $(\hat{\mu}, \hat{\Sigma})$  sont les solutions du système

$$\begin{aligned} & \sum_{i=1}^n u_1(d_i)(x_i - t) = 0, \quad t \in \mathbb{R}^p, \\ & \frac{1}{n} \sum_{i=1}^n u_2(d_i^2)(x_i - t)(x_i - t)' = C, \quad C \in PSD, \end{aligned}$$

avec  $d_i = d(x_i, t, C^{-1})$ ,  $u_1$  et  $u_2$  des fonctions définies sur  $\mathbb{R}^+$ .

Nous avons déjà parlé de  $M$ -estimateurs dans le cas multivarié pour estimer un paramètre de position. Il est possible de retomber sur une expression similaire à celle obtenue en (1.33) en considérant  $\theta = (\mu, \Sigma)$ ,  $\psi = (\psi_1, \psi_2)$  et

$$\begin{aligned} \psi_1(x, \theta) &= u_1(d)(x - \mu), \\ \psi_2(x, \theta) &= u_2(d^2)(x - \mu)(x - \mu)' - \Sigma. \end{aligned}$$

Ainsi, nous obtenons une seule équation donnée par

$$\frac{1}{n} \sum_{i=1}^n \psi(x, \theta) = 0,$$

dont la version population correspond bien à (1.33), avec cette fois-ci  $\theta = (\mu, \Sigma)$  au lieu de  $\theta = \mu$ . Remarquons que considérer  $u_1(d) = u_2(d^2) = 1$  renvoie la moyenne et la matrice de covariance empirique. Pour assurer existence et unicité de la solution, plusieurs conditions doivent être vérifiées. Parmi celles-ci, nous trouvons, en notant  $\phi_i(s) = u_i(s)s$  et  $\sup_{s \geq 0} \phi_2(s) = K_2$ ,

1.  $u_1$  et  $u_2$  sont décroissantes, positives et continues,
2.  $\phi_1$  et  $\phi_2$  sont bornées,
3.  $\phi_2$  est croissante et strictement croissante là où  $\phi_2 < K_2$ ,
4. Il existe  $s_0$  tel que  $\phi_2(s_0^2) > m$  et  $u_1(s) > 0$  pour tout  $s \leq s_0$  (ce qui implique  $K_2 > m$ ).

La version population de ces estimateurs peut être obtenue en considérant

$$\begin{aligned} E_P [u_1(d)(X - t)] &= 0, \\ E_P [u_2(d^2)(X - t)(X - t)'] &= C. \end{aligned}$$

Concernant le point de rupture de ces estimateurs, ce dernier vérifie

$$\epsilon^* \leq \frac{1}{p+1}.$$

Si nous considérons  $\mu$  fixé et la fonction  $\phi_2$  croissante, le point de rupture de l'estimateur  $\hat{\Sigma}$  devient

$$\epsilon^* \leq \frac{1}{p}.$$

Malgré cette légère amélioration, nous remarquons que le point de rupture est très petit lorsque les données sont en grande dimension. Enfin, la fonction d'influence a aussi été obtenue et son expression s'avère être bornée si et seulement si les fonctions  $\phi_1$  et  $\phi_2$  le sont aussi.

Nous pouvons à présent discuter du M-estimateur de dispersion de Tyler. Ce dernier est obtenu en choisissant  $u_2(s) = ps^{-1}$ , de sorte que nous obtenons comme estimateur toute solution  $C$  de

$$\frac{p}{n} \sum_{i=1}^n \frac{(x_i - t)(x_i - t)'}{(x_i - \mu)'C^{-1}(x_i - \mu)} = C.$$

Notons qu'un tel choix de fonction livre  $K_2 = m$ , de sorte que la quatrième condition n'est pas vérifiée. L'existence de la solution à cependant quand même été démontrée en prenant en compte d'autres hypothèses. Notons que ces dernières sont toutes vérifiées avec probabilité 1 si l'échantillon provient d'une distribution continue et que  $n > p(p-1)$ , ce que nous allons considérer à partir de maintenant. La solution n'est évidemment pas unique, étant donné que si  $C$  vérifie l'expression, alors  $cV$  aussi pour tout constante  $c > 0$ . Cependant, à cette constante scalaire près, la solution est unique. Autrement dit si  $C_1$  et  $C_2$  sont solutions, alors il existe  $c > 0$  tel que  $C_1 = cC_2$ . Nous avons mentionné précédemment que la fonction d'influence d'un M-estimateur était bornée si et seulement si  $\phi_1$  et  $\phi_2$  sont bornées. Dans notre cas, seule la fonction d'influence du M-estimateur de dispersion nous intéresse, et cette dernière est bien bornée car  $\phi_2(s) = u_2(s)s = ms^{-1}s = m$  est bornée. Le document de référence [8] démontre aussi la normalité asymptotique de l'estimateur avec un taux de  $n^{-\frac{1}{2}}$ . La loi normale asymptotique est centrée, de sorte que l'estimateur est convergent. De plus, si la distribution sous-jacente à l'échantillon est elliptique de paramètre  $\mu = 0$ , alors la distribution asymptotique de l'estimateur ne dépend pas de la fonction  $h$  qui définit la distribution.

## 1.7 Conclusion

Nous avons dans ce premier chapitre introduit différents outils afin d'évaluer la robustesse d'une procédure statistique, dont les plus connus et utilisés sont la fonction d'influence ainsi que le point de rupture. Nous avons appliqué ces outils sur différents estimateurs de position, comme la moyenne et la médiane. Ainsi, nous avons pu observer que la médiane est théoriquement plus robuste que la moyenne. Nous avons aussi introduit différents estimateurs robustes, parmi lesquels nous trouvons les M-estimateurs ainsi que les L-estimateurs. Nous avons ensuite discuté des estimateurs de dispersion, un de plus connu portant le nom de *median absolute deviation*, dont nous avons démontré la fonction d'influence. Enfin, nous avons pu constater que les différents outils se généralisent facilement en dimension  $p > 1$ . Nous avons à nouveau introduit des estimateurs de position et de matrice de covariance, comme le MCD et le MVE.



## Chapitre 2

# Notions de centralité et de profondeur

Dans ce second chapitre, nous allons introduire la notion de fonction de profondeur. Nous définissons ensuite les deux fonctions d'intérêt dans ce mémoire, à savoir la fonction de profondeur de demi-espace ainsi que la fonction de profondeur scatter. Après avoir évalué la robustesse de la première, nous démontrons la fonction d'influence de la seconde dans différents cas. Nous commençons par considérer que le paramètre de position dont elle dépend est fixé. Ensuite, nous considérons une distribution discrète à support fini, et enfin nous bornons la fonction d'influence lorsque la distribution est absolument continue par rapport à la mesure de Lebesgue. Nous concluons en analysant les résultats obtenus ainsi qu'en fournissant quelques remarques sur la variance asymptotique. Ce chapitre est inspiré des références [9], [10], [11], [12] et [13].

### 2.1 Introduction aux fonctions de profondeur

La notion de centralité d'un ensemble d'objets a toujours été difficile à définir. Pour le cas particulier d'une variable  $X$  de distribution  $P$  en dimension 1, nous avons vu qu'il existe plusieurs définitions pour évaluer le centre, comme la moyenne et la médiane. Cette dernière est, comme nous l'avons développé dans le chapitre précédent, bien plus robuste que la moyenne. Définie comme le point  $\theta$  satisfaisant  $P[X \leq \theta] \geq 1/2$  et  $P[X \geq \theta] \leq 1/2$ , la médiane fait inévitablement intervenir une notion d'ordre. Le fait que  $\mathbb{R}^p$  n'ait pas d'ordre naturel rend dès lors difficile la généralisation de la définition de la médiane en dimension supérieure. Une première tentative fût de considérer la médiane de chaque composante. Pour un vecteur  $X = (X_1, \dots, X_p)$ , nous trouvons  $(\tilde{X}) = (\tilde{X}_1, \dots, \tilde{X}_p)$ . Malgré des propriétés de robustesse intéressantes, l'estimateur livre de très mauvaises performances, surtout lorsque les variables sont hautement corrélées. De plus, il n'est pas affiné équivariant, et donc la manière de mesurer les données aura un impact sur la valeur de l'estimateur. Afin d'améliorer cette définition, beaucoup d'auteurs ont considéré la *médiane spatiale*.

**Définition 2.1.** Soit  $X$  un vecteur aléatoire de distribution  $P$  sur  $\mathbb{R}^p$ . La médiane spatiale de  $X$  est la donnée  $\hat{\mu}_S(P) \in \mathbb{R}^p$  qui minimise  $E_P[\|X - \mu\|_2]$ .

Cette définition étend la définition de la médiane en une dimension, qui elle minimise  $E_P[\|X - \theta\|]$ . Si nous considérons une distribution empirique  $P^{(n)}$  de  $n$  données  $X_1, \dots, X_n$ , trouver la médiane spatiale revient à résoudre

$$\hat{\mu}_S^{(n)} = \arg \inf_{\mu \in \mathbb{R}^p} \sum_{i=1}^n \|X_i - \mu\|.$$

Il existe de nos jours de nombreux algorithmes permettant de résoudre un tel problème d'optimisation. Une autre manière de généraliser la médiane en plusieurs dimensions est de considérer par exemple la *médiane de projection*.

**Définition 2.2.** Soit  $X$  un vecteur aléatoire de distribution  $P$  sur  $\mathbb{R}^p$ . La médiane de projection de  $X$  est donnée par

$$\hat{\mu}_{Pr}(P) = \arg \inf_{\mu \in \mathbb{R}^p} \arg \sup_{u \in \mathcal{S}^{p-1}} \frac{|u'\mu - \tilde{\mu}(u'X)|}{\text{MAD}(u'X)},$$

où  $\mathcal{S}^{p-1} = \{x \in \mathbb{R}^p : \|x\| = 1\}$  représente la sphère unité de  $\mathbb{R}^p$  et  $\tilde{\mu}$  représente la médiane en une dimension.

Ces définitions ont en commun qu'elles livrent un estimateur en optimisant une fonction qui, d'une certaine manière, reflète la centralité d'un point pour une distribution. Ceci a motivé le développement d'outils plus généraux permettant de mesurer la centralité via des *fonctions de profondeur*. Plus précisément, une fonction de profondeur est une application  $D(\cdot, \cdot) : \mathbb{R}^p \times \mathcal{P} \rightarrow \mathbb{R} : (x, P) \mapsto D(x, P)$ , où  $\mathcal{P}$  représente un ensemble de distributions. Cette fonction va, pour chaque position  $x$ , associer une valeur qui reflète sa centralité par rapport à la distribution  $P$ . Il a été établi différentes propriétés qu'une fonction de profondeur devait idéalement satisfaire. Tout d'abord, la fonction doit être positive et bornée. Afin que la fonction de profondeur fournisse efficacement une notion de centralité ainsi que de distance sur  $\mathbb{R}^p$ , elle doit satisfaire en plus les quatre propriétés suivantes :

1. *Affine invariance.* La profondeur d'un point  $x \in \mathbb{R}^p$  doit être indépendante de la manière dont nous avons mesuré les données (l'échelle, le système de coordonnées, ...).
2. *Maximum au centre.* Si la distribution possède un centre naturel, comme le point de symétrie pour une distribution symétrique par exemple, alors la valeur de la fonction de profondeur est maximale en ce point.
3. *Monotone par rapport au point le plus profond.* Si un point  $x \in \mathbb{R}^p$  s'éloigne de la donnée ayant la valeur maximum de profondeur, la profondeur de  $x$  doit décroître de manière monotone.
4. *Converger vers 0 à l'infini.* La valeur de la profondeur pour un point  $x \in \mathbb{R}^p$  qui va à l'infini doit converger vers 0.

Sur bases de ces quatre propriétés intuitives, nous pouvons définir rigoureusement une fonction de profondeur.

**Définition 2.3.** Soit la fonction  $D : \mathbb{R}^p \times \mathcal{P} \rightarrow \mathbb{R}$  bornée et positive. Supposons qu'elle satisfait

1.  $D(Ax + b; F_{AX+b}) = D(x; F_X)$  pour tout vecteur aléatoire  $X \in \mathbb{R}^p$ , pour toute matrice  $A$  de dimension  $p \times p$  et non singulière et pour tout vecteur de dimension  $p$ .
2.  $D(\theta; F) = \sup_{x \in \mathbb{R}^p} D(x; F)$  pour n'importe quelle distribution  $F$  de centre de symétrie  $\theta$ .
3.  $D(x; F) \leq D(\theta + \alpha(x - \theta); F)$  est vérifié pour  $\alpha \in [0, 1]$ , pour toute distribution  $F$  dont le point le plus profond est  $\theta$ .
4.  $D(x; F) \rightarrow 0$  lorsque  $\|x\| \rightarrow \infty$  pour toute distribution  $F$ .

Alors,  $D(\cdot; F)$  est appelée fonction de profondeur statistique.

Afin d'obtenir une version empirique de la fonction  $D(x; F)$ , que nous notons  $D(x; F_n)$ , il suffit de remplacer la distribution  $F$  par la distribution empirique  $F_n$ .

## 2.2 Fonction de profondeur de demi-espace

### 2.2.1 Introduction

Nous avons jusqu'ici régulièrement discuté de centre en référence à un point de symétrie. Il existe cependant de nombreuses définitions de symétrie en statistiques multivariées. La plus connue affirme qu'un vecteur aléatoire  $X \in \mathbb{R}^p$  est symétrique autour de  $\theta$  si la distribution de  $X - \theta$  est la même que celle de  $\theta - X$ . La symétrie de demi-espace affirme que  $X$  est symétrique autour de  $\theta$  si  $P[X \in HS(\theta)] \geq 0.5$  pour tout demi-espace  $HS(\theta)$  contenant  $\theta$ . Lorsque la distribution vérifie cette propriété, nous dirons qu'elle est H-symétrique. Cette dernière notion est centrale dans la fonction de profondeur de demi-espace. Il s'agit là de la fonction de profondeur la plus connue, et elle est définie comme suit.

**Définition 2.4.** Soient  $x \in \mathbb{R}^p$  et  $X$  un vecteur aléatoire dans  $\mathbb{R}^p$  de distribution  $P$ . La profondeur de demi-espace de  $x$  par rapport à  $P$  est donnée par

$$D_H(x; P) = \inf_{u \in \mathcal{S}^{p-1}} P[u'(X - x) \geq 0]. \quad (2.1)$$

Nous pouvons adapter cette définition en utilisant la notion de demi-espace. La profondeur se réécrit alors

$$D_H(x; P) = \inf_{\mathcal{H}_x} P[HS(x)],$$

où  $\mathcal{H}_x$  représente l'ensemble des sous-espace  $HS(x)$ . Le lien entre cette définition et la définition (2.1) peut paraître peu évident. Remarquons tout d'abord que

$$\bigcup_{u \in \mathcal{S}^{p-1}} \{x \in \mathbb{R}^p : u'(X - x) \geq 0\} \subset \mathcal{H}_x,$$

de sorte que

$$\inf_{u \in \mathcal{S}^{p-1}} P[u'(X - x) \geq 0] \geq \inf_{\mathcal{H}_x} P[HS(x)].$$

Pour obtenir l'égalité dans l'autre sens, il suffit de se convaincre que, pour tout sous-espace  $HS(x)$ , il existe une direction  $u \in \mathcal{S}^{p-1}$  telle que  $P[HS(x)] \geq P[u'(X - x) \geq 0]$  ( $u$  est en quelque sorte la direction définissant le sous-espace que nous avons parallèlement ramené à  $x$ ).

Nous pouvons à présent montrer que la fonction de profondeur de demi-espace ainsi définie vérifie bien les conditions pour être une fonction de profondeur.

**Théorème 2.1.** La fonction de profondeur  $D_H(x; P)$  est une fonction de profondeur au sens de la définition (2.3).

*Démonstration.* Il est clair que la fonction de profondeur de demi-espace est positive et bornée. Il reste donc à vérifier les quatre propriétés.

1. *Affine invariance.* Trivial
2. *Maximum au centre.* Supposons que  $P$  est H-symétrique autour d'un unique point  $\theta \in \mathbb{R}^p$ . Par définition, nous avons que  $P[HS(\theta)] \geq 0.5$ . Il s'en suit que  $D_H(\theta; P) \geq 0.5$ . Supposons par l'absurde qu'il existe un point  $x_0 \in \mathbb{R}^p$ ,  $x_0 \neq \theta$ , tel que  $D_H(x_0; P) > D_H(\theta; P)$ . Ceci entraîne que  $\inf_{\mathcal{H}_{x_0}} P[HS(x_0)] > 0.5$ , de sorte que  $P[HS(x_0)] > 0.5$  pour tout demi-espace contenant  $x_0$ . Cela implique que  $P$  est H-symétrique par rapport à  $x_0$ , ce qui contredit le fait que la distribution est symétrique autour de l'unique point  $\theta$ .

3. *Monotone par rapport au point le plus profond.* Nous allons pour ce faire montrer que, pour tout  $\theta_1$  et  $\theta_2 \neq \theta_1$  dans  $\mathbb{R}^p$ , et pour tout  $0 < \alpha < 1$ ,  $D_H(\alpha\theta_1 + (1 - \alpha)\theta_2; P) \geq \min(D_H(\theta_1; P), D_H(\theta_2; P))$ . Notons  $v = \alpha\theta_1 + (1 - \alpha)\theta_2$ , et remarquons que tout demi-espace  $HS(v)$  contenant  $v$  contient soit  $\theta_1$ , soit  $\theta_2$ , ou les deux. Si  $\theta_1 \in HS(v)$ , alors il existe un sous espace  $HS(\theta_1) \subseteq HS(v)$  tel que  $P[HS(\theta_1)] \leq P[HS(v)]$ , et il en est de même pour  $\theta_2$ . Nous trouvons dès lors

$$\begin{aligned} D_H(z; P) &= \inf_{\mathcal{H}_z} P[HS(z)] = \min \left( \inf_{\mathcal{H}_z \cap \mathcal{H}_{\theta_1}} P[HS(z)], \inf_{\mathcal{H}_z \cap \mathcal{H}_{\theta_2}} P[HS(z)] \right) \\ &\geq \min \left( \inf_{\mathcal{H}_{\theta_1}} P[HS(\theta_1)], \inf_{\mathcal{H}_{\theta_2}} P[HS(\theta_2)] \right) \\ &= \min(D_H(\theta_1; P), D_H(\theta_2; P)). \end{aligned}$$

Ainsi, si  $\theta_1 = x$  quelconque et  $\theta_2 = \theta$ , le point le plus profond, nous avons

$$D_H(\alpha\theta_1 + (1 - \alpha)\theta_2; P) = D_H(\theta + \alpha(x - \theta); P) \geq \min\{D_H(\theta; P), D_H(x; P)\} = D_H(x; P).$$

4. *Converge vers 0 à l'infini.* Comme nous avons que  $P[\|X\| > \|x\|] \rightarrow 0$  pour  $\|x\| \rightarrow \infty$ , et que, pour chaque  $x$  et  $X$ , il existe un demi-espace  $HS(x)$  tel que  $HS(x) \subset \{\|X\| > \|x\|\}$ , alors  $D_H(x; P) \rightarrow 0$  lorsque  $\|x\| \rightarrow \infty$ . □

Pour une distribution empirique  $P^{(n)}$ , la profondeur de demi-espace d'un point  $x \in \mathbb{R}^p$  est donnée par

$$D_H(x; P^{(n)}) = \frac{1}{n} \min_{u \in \mathcal{S}^{p-1}} \#\{i : u'(X_i - x) \geq 0\}.$$

Il existe des définitions équivalentes à celles proposées justement, comme par exemple

$$D_H(x; P) = \inf_{u \in \mathcal{S}^{p-1}} \min(P[u'(X - x) \geq 0], P[u'(X - x) \leq 0]),$$

dont la version empirique vaut

$$D_H(x; P^{(n)}) = \frac{1}{n} \min_{u \in \mathcal{S}^{p-1}} \min(\#\{i : u'(X_i - x) \geq 0\}, \#\{i : u'(X_i - x) \leq 0\}).$$

Afin d'illustrer cette dernière définition, nous allons considérer un ensemble de 10 données en dimension 2. Sur la FIGURE 1, nous avons représenté les 10 données en bleu et le point rouge représente le point  $x$  dont nous voulons calculer la profondeur. Nous avons considéré sur le panel de gauche une certaine direction, en rouge, où nous trouvons que le nombre minimum de valeurs projetées de part et d'autre de  $x$  vaut 2. Le panel de droite est similaire à celui de gauche, où nous avons trouvé une direction telle que  $\#\{i : u'(X_i - x) \geq 0\} = 0$ . Comme  $D_H(x; P^{(n)}) \geq 0$ , nous trouvons que la profondeur du point rouge vaut 0. A la FIGURE 2, nous avons discrétisé l'espace  $\mathbb{R}^2$  et représenté chaque point par la valeur de sa profondeur. Notons que, par soucis de clarté, nous avons sur chacune des figures 1 et 2 représenté la profondeur empirique sans le facteur  $1/n$ . Nous constatons que, au plus nous allons vers le milieu du jeu de données, au plus la valeur de la profondeur augmente. Ainsi, nous pouvons définir un estimateur de position comme la donnée qui maximise la profondeur de demi-espace.

**Définition 2.5.** Soit  $X$  un vecteur aléatoire dans  $\mathbb{R}^p$  de distribution  $P$ . La médiane de demi-espace  $T_P$  est donnée par

$$T_P = \arg \max_{x \in \mathbb{R}^p} D_H(x; P).$$

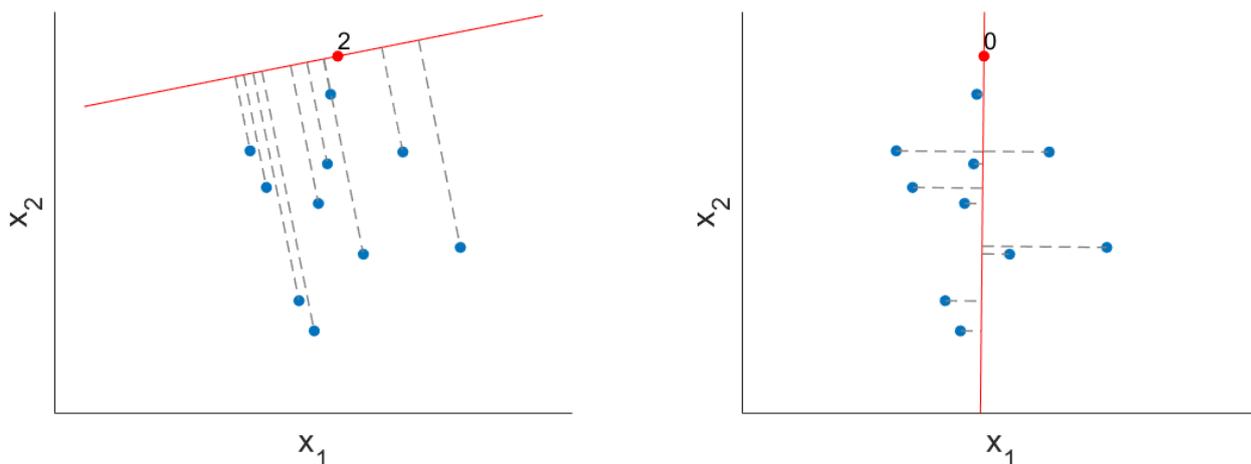


FIGURE 1 – Illustration de la définition de la profondeur de demi-espace.

Notons que la médiane ainsi définie n'est pas toujours unique, et il convient dès lors de considérer le centre de masse de l'ensemble des éléments satisfaisant l'égalité. De plus, pour  $p = 1$ , la définition ci-dessus correspond à la définition de la médiane usuelle.

A la FIGURE 2, les régions de profondeur définies par

$$D_\alpha(P_n) = \{x \in \mathbb{R}^p : D_H(x; P_n) \geq \alpha\}$$

semblent former des ensembles convexes. La proposition suivante montre que cette affirmation est en effet vérifiée.

**Proposition 2.1.** *Soient  $\{x_1, \dots, x_n\}$  un échantillon dans  $\mathbb{R}^p$  et  $P_n$  la mesure de probabilité empirique associée. Pour n'importe quelle valeur de  $\alpha$  comprise entre 0 et 1, nous avons que l'ensemble  $D_\alpha(P_n)$  est convexe.*

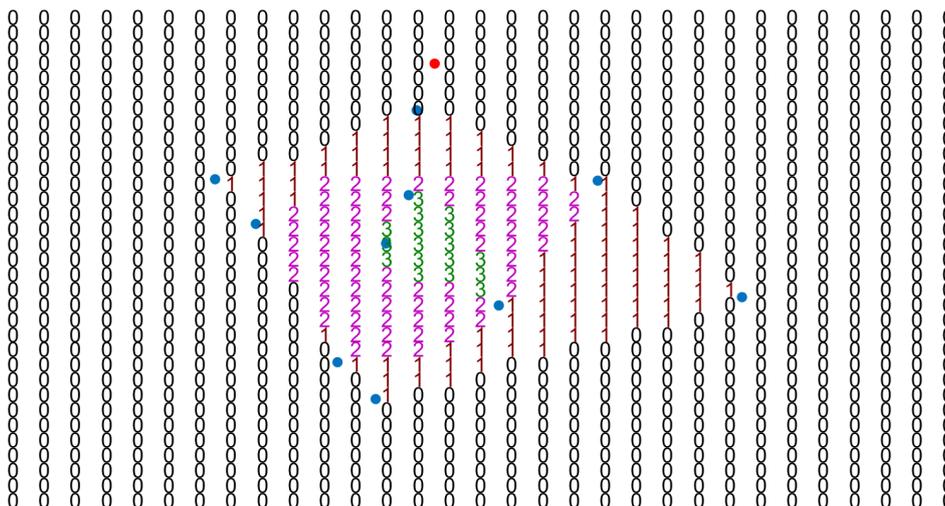


FIGURE 2 – Illustration de la médiane de demi-espace.

*Démonstration.* Soient  $\theta_1$  et  $\theta_2 \neq \theta_1$  dans  $D_\alpha(P_n)$ , de sorte que  $D_H(\theta_1; P_n) \geq \alpha$  et  $D_H(\theta_2; P_n) \geq \alpha$ . Il faut montrer que, pour tout  $0 < \gamma < 1$ ,  $\gamma\theta_1 + (1 - \gamma)\theta_2$  est dans  $D_\alpha(P_n)$ . Ceci découle directement du fait que  $D_H(\gamma\theta_1 + (1 - \gamma)\theta_2; P_n) \geq \min\{D_H(\theta_1; P_n), D_H(\theta_2; P_n)\} \geq \alpha$ .  $\square$

### 2.2.2 Robustesse

Concernant la robustesse de la fonction de profondeur de demi-espace, le document [10] en donne la fonction d'influence. Pour une valeur de  $z \in \mathbb{R}^p$ , nous définissons les sous-ensembles

$$\begin{aligned}\mathcal{H}_{x,z} &= \{HS(x) \in \mathcal{H}_x : z \in HS(x)\} \\ \mathcal{H}_{x,\bar{z}} &= \{HS(x) \in \mathcal{H}_x : z \notin HS(x)\}.\end{aligned}$$

Le premier sous-ensemble correspond aux demi-espaces contenant à la fois  $x$  et  $z$ , là où le second contient uniquement  $x$  et pas  $z$ . Nous avons  $\mathcal{H}_{x,z} \cap \mathcal{H}_{x,\bar{z}} = \emptyset$  et  $\mathcal{H}_{x,z} \cup \mathcal{H}_{x,\bar{z}} = \mathcal{H}_x$ . En écrivant

$$\begin{aligned}D_H^{(z)}(x; P) &= \inf_{\mathcal{H}_{x,z}} P[HS(x)], \\ D_H^{(\bar{z})}(x; P) &= \inf_{\mathcal{H}_{x,\bar{z}}} P[HS(x)],\end{aligned}$$

nous avons  $D_H(x; P) = \min(D_H^{(z)}(x; P), D_H^{(\bar{z})}(x; P))$ . Nous pouvons à présent donner le résultat ainsi que la preuve pour la fonction d'influence de la profondeur de demi-espace.

**Théorème 2.2.** *Soient  $X$  un vecteur aléatoire dans  $\mathbb{R}^p$  de distribution  $F$  et  $z \in \mathbb{R}^p$ . La fonction d'influence de la fonction de profondeur de demi-espace  $D_H(x; P_F)$  pour  $x \in \mathbb{R}^p$  fixé arbitrairement est donnée par*

$$\text{IF}(z; D_H(x; P), F) = \begin{cases} -D_H(x; P) & \text{si } D_H^{(\bar{z})}(x; P) < D_H^{(z)}(x; P), \\ 1 - D_H(x; P) & \text{sinon.} \end{cases}$$

*Démonstration.* Sous la distribution perturbée  $F_{t,x} \stackrel{\text{not}}{=} \tilde{F}$ , nous avons

$$P_{\tilde{F}}[HS(x)] \stackrel{\text{not}}{=} \tilde{P}[HS(x)] = \begin{cases} (1-t)P_F[HS(x)] + t & \text{si } HS(x) \in \mathcal{H}_{x,z}, \\ (1-t)P_F[HS(x)] & \text{si } HS(x) \in \mathcal{H}_{x,\bar{z}}. \end{cases}$$

Ceci livre que

$$\begin{aligned}D_H(x; \tilde{P}) &= \min\left(D_H^{(z)}(x; \tilde{P}), D_H^{(\bar{z})}(x; \tilde{P})\right) \\ &= \min\left((1-t)D_H^{(z)}(x; P) + t, (1-t)D_H^{(\bar{z})}(x; P)\right).\end{aligned}$$

Afin d'obtenir la fonction d'influence, nous allons discuter deux cas.

1. Si  $D_H^{(z)}(x; P) \geq D_H^{(\bar{z})}(x; P)$ , alors  $D_H(x; P) = D_H^{(\bar{z})}(x; P)$  et  $D_H(x; \tilde{P}) = (1-t)D_H^{(\bar{z})}(x; P) = (1-t)D_H(x; P)$ . En utilisant la définition de la fonction d'influence, nous trouvons

$$\begin{aligned}\text{IF}(z; D_H(x; P), F) &= \lim_{t \rightarrow 0^+} \frac{(1-t)D_H(x; P) - D_H(x; P)}{t} \\ &= -D_H(x; P).\end{aligned}$$

2. Considérons à présent le cas où  $D_H^{(z)}(x; P) < D_H^{(\bar{z})}(x; P)$ . Nous avons alors que  $D_H(x; P) = D_H^{(z)}(x; P)$ . Étant donné que

$$\lim_{t \rightarrow 0_+} D_H^{(z)}(x; \tilde{P}) = D_H^{(z)}(x; P)$$

et

$$\lim_{t \rightarrow 0_+} D_H^{(\bar{z})}(x; \tilde{P}) = D_H^{(\bar{z})}(x; P)$$

il existe  $t_0$  suffisamment petit tel quel, pour tout  $t < t_0$ ,  $D_H^{(z)}(x; \tilde{P}) < D_H^{(\bar{z})}(x; \tilde{P})$ . Ainsi,  $D_H(x; \tilde{P}) = (1 - t)D_H^{(z)}(x; P) + t = (1 - t)D_H(x; P) + t$  et la fonction d'influence vaut

$$\begin{aligned} \text{IF}(z; D_H(x; P), F) &= \lim_{t \rightarrow 0_+} \frac{(1 - t)D_H(x; P) + t - D_H(x; P)}{t} \\ &= 1 - D_H(x; P). \end{aligned}$$

Notons que dans le cas particulier où  $z = x$ ,  $D_H^{(z)}(x; P) = D_H(x; P)$  et  $P_{\tilde{F}}[HS(x)] = (1 - t)P_F[HS(x)] + t$ , étant donné que  $\mathcal{H}_x = \mathcal{H}_{x,z}$ . Ainsi,  $D_H(x; \tilde{P}) = (1 - t)D_H(x; P) + t$  et  $\text{IF}(z; D_H(x; P), F) = 1 - D_H(x; P)$ . □

Comme la profondeur est une probabilité, comprise entre 0 et 1, la fonction d'influence est bornée par  $-1$  et  $1$ . Souvenons-nous qu'une fonction d'influence bornée reflète la robustesse de la fonctionnelle. Nous pouvons, grâce à la fonction d'influence, évaluer la variance asymptotique. Pour ce faire, nous allons désigner par *demi espace optimal*  $HO$  un élément de  $\mathcal{H}_x$  satisfaisant  $D_H(x; P) = P[HO]$ . La fonction d'influence est une variable qui prend uniquement deux valeurs. La probabilité  $p_x$  qu'elle prenne la valeur  $1 - D_H(x; P)$  correspond à la probabilité que  $z$  appartienne à tous les sous-espaces optimaux, et donc à l'intersection de ceux-ci. En effet, supposons qu'il existe deux sous-espaces optimaux  $HO_1$  et  $HO_2$  et que  $z$  appartient au premier mais pas au second. Dans ce cas,  $P_{\tilde{F}}[HO_1] = (1 - t)P[HO_1] + t$  et  $P_{\tilde{F}}[HO_2] = (1 - t)P[HO_2]$ . Comme  $P[HO_1] = P[HO_2]$ , nous avons que  $D_H(x; \tilde{P}) = P_{\tilde{F}}[HO_2] < P_{\tilde{F}}[HO_1]$ . Ainsi,  $D_H(x; \tilde{P}) = (1 - t)D_H(x; P)$  et donc  $\text{IF}(z; D_H(x; P), F) = -D_H(x; P)$ . Nous trouvons pour l'espérance et la variance

$$\begin{aligned} E_F[\text{IF}(Z; D_H(x; P), F)] &= (1 - D_H(x; P))p_x - D_H(x; P)(1 - p_x) \\ &= p_x - D_H(x; P) \end{aligned}$$

et

$$\begin{aligned} \text{Var}_F[\text{IF}(Z; D_H(x; P), F)] &= (1 - D_H(x; P))^2 p_x + D_H(x; P)^2 (1 - p_x) - (p_x - D_H(x; P))^2 \\ &= p_x(1 - p_x). \end{aligned}$$

S'il existe un unique sous-espace optimal  $HO$ , alors  $p_x = P[HO] = D_H(x; P)$ , et donc

$$E_F[\text{IF}(Z; D_H(x; P), F)] = 0$$

et

$$\text{Var}_F[\text{IF}(Z; D_H(x; P), F)] = D_H(x; P)(1 - D_H(x; P)).$$

Lorsque, au lieu de considérer un vecteur aléatoire  $X$  de distribution  $F$ , nous avons un échantillon de  $n$  observations  $\{X_1, \dots, X_n\}$  issues de cette distribution, nous pouvons utiliser des estimateurs afin d'approximer la fonction d'influence. Nous considérons ici l'estimation donnée par

$$SI(X_i; T(F)) = -(n-1)(T(\hat{F}_{n-1}^{(i)}) - T(\hat{F}_n)),$$

où  $\hat{F}_{n-1}^{(i)}$  représente la distribution empirique basée sur l'échantillon où nous avons retiré la  $i^{\text{ème}}$  observation. Avant de donner la valeur de la fonction d'influence empirique pour la profondeur de demi-espace, considérons  $\mathcal{H}_{x, X_i}$  et  $\mathcal{H}_{x, \bar{X}_i}$ , les sous-espaces contenant respectivement  $x$  et  $X_i$  et uniquement  $x$ . Le résultat ainsi que la preuve sont similaires à celles pour la fonction d'influence basée sur une fonctionnelle.

**Théorème 2.3.** *Soit  $\{X_1, \dots, X_n\}$  un échantillon issu d'une distribution  $F$  et soit  $x \in \mathbb{R}^p$ . La fonction d'influence empirique de la profondeur de demi-espace est donnée par*

$$SI(X_i; D_H(x; \hat{F}_n)) = \begin{cases} -D_H(x, \hat{F}_n) & \text{si } D_H^{(\bar{X}_i)}(x, \hat{F}_n) < D_H^{(X_i)}(x, \hat{F}_n) \\ 1 - D_H(x, \hat{F}_n) & \text{sinon,} \end{cases}$$

où

$$D_H^{(X_i)}(x, \hat{F}_n) = \min_{\mathcal{H}_{x, X_i}} P[HS(x)] \quad \text{et} \quad D_H^{(\bar{X}_i)}(x, \hat{F}_n) = \min_{\mathcal{H}_{x, \bar{X}_i}} P[HS(x)].$$

*Démonstration.* Notons  $D_H^{(\bar{X}_i)}(x, \hat{F}_n) = m(x)/n$  et  $D_H^{(X_i)}(x, \hat{F}_n) = l(x)/n$ . Nous avons dès lors que  $D_H^{(\bar{X}_i)}(x, \hat{F}_{n-1}^{(i)}) = m(x)/(n-1)$  et  $D_H^{(X_i)}(x, \hat{F}_{n-1}^{(i)}) = (l(x) - 1)/(n-1)$ . Nous allons à nouveau considérer deux cas.

1. Si  $m(x) < l(x)$ , alors  $D_H(x, \hat{F}_n) = m(x)/n$  et  $D_H^{(X_i)}(x, \hat{F}_{n-1}^{(i)}) = m(x)/(n-1)$ . En effet, nous avons que  $D_H^{(\bar{X}_i)}(x, \hat{F}_{n-1}^{(i)}) = m(x)/(n-1) \leq (l(x) - 1)/(n-1) = D_H^{(X_i)}(x, \hat{F}_{n-1}^{(i)})$ . Ainsi, la fonction d'influence empirique vaut

$$SI(X_i; D_H(x; \hat{F}_n)) = -(n-1) \left( \frac{m(x)}{n-1} - \frac{m(x)}{n} \right) = -\frac{m(x)}{n} = -D_H(x; \hat{F}_n).$$

2. Si nous considérons que  $m(x) \geq l(x)$ , alors  $D_H(x, \hat{F}_n) = l(x)/n$  et, par un raisonnement similaire à celui proposé au cas précédent,  $D_H^{(X_i)}(x, \hat{F}_{n-1}^{(i)}) = (l(x) - 1)/(n-1)$ . Nous trouvons pour la fonction d'influence empirique

$$SI(X_i; D_H(x; \hat{F}_n)) = -(n-1) \left( \frac{(l(x) - 1)}{n-1} - \frac{l(x)}{n} \right) = 1 - \frac{l(x)}{n} = 1 - D_H(x; \hat{F}_n).$$

□

Le fait que la fonction d'influence empirique ait la même forme que sa version population n'est pas anodin. En effet, la fonction d'influence empirique converge vers sa version population, bien que ce ne soit pas toujours vérifié pour n'importe quelle fonctionnelle.

Concernant la médiane de demi-espace, sa fonction d'influence a aussi été obtenue dans le document [11] pour les distributions H-symétriques. Dans ce cas particulier, nous pouvons considérer sans perdre de généralité que la distribution est symétrique autour de l'origine. Avant de donner le théorème permettant d'établir l'expression de la fonction d'influence, nous introduisons quelques concepts et notations. Tout d'abord, l'ensemble

$$L(\alpha; P) := D_{\Pi(P) - \alpha}(P) = \{x \in \mathbb{R}^p : D_H(x; P) \geq \Pi(P) - \alpha\},$$

avec  $\Pi(P) = \max_{x \in \mathbb{R}^p} D_H(x; P)$ . Dans le cas de distributions H-symétrique,  $\Pi(P) = 0.5$  et cette valeur est atteinte uniquement au centre de symétrie. Nous notons aussi  $r(\alpha, u)$  le rayon de cet ensemble le long d'une direction  $u$ . Plus précisément,  $r(\alpha, u)$  correspond à la longueur du segment  $L(\alpha; P) \cap \{tu : t \geq 0\}$ . Nous pouvons à présent énoncer le théorème qui nous permettra d'obtenir la fonction d'influence.

**Théorème 2.4.** *Soit  $P$  une distribution absolument continue sur  $\mathbb{R}^p$  et H-symétrique autour de l'origine. Pour tout  $x \in \mathbb{R}^p$  et  $0 \leq t < 1/3$ ,*

$$T_{\tilde{P}} = \begin{cases} z, & \text{si } z \text{ est à l'intérieur de } L\left(\frac{t}{1-t}; P\right) \\ \frac{1}{2}r\left(\frac{t}{1-t}, \frac{z}{\|z\|}\right) \frac{z}{\|z\|} & \text{si } z \notin L\left(\frac{t}{1-t}; P\right) \text{ ou sur la frontière.} \end{cases}$$

Nous observons que, pour une valeur de  $t$  fixée,  $T_{\tilde{P}}$  ne croit pas de manière monotone. En effet, d'abord la valeur croit tant que  $z$  est à l'intérieur de  $L\left(\frac{t}{1-t}; P\right)$ . Puis lorsque  $z$  arrive au bord de cet ensemble, la valeur de la médiane ne reste pas sur le bord mais redescend à  $\frac{1}{2}r\left(\frac{t}{1-t}, \frac{z}{\|z\|}\right) \frac{z}{\|z\|}$ .

Nous pouvons à présent énoncer le théorème livrant l'expression de la fonction d'influence de la médiane.

**Théorème 2.5.** *Sous les hypothèses du Théorème 2.4, la fonction d'influence de la médiane de demi-espace est donnée par*

$$\text{IF}(z; T, P) = \frac{1}{2} \lambda\left(\frac{z}{\|z\|}\right) \frac{z}{\|z\|},$$

où  $\lambda\left(\frac{z}{\|z\|}\right) = \lim_{\alpha \rightarrow 0} r\left(\alpha, \frac{z}{\|z\|}\right) / \alpha$ , pour peu que la limite existe. Pour  $z = 0$ ,  $\text{IF}(z; T, P) = 0$ .

*Démonstration.* Il suffit d'utiliser le résultat du Théorème 2.4 et de remarquer que, pour  $z$  fixé, il existe  $t$  suffisamment petit tel quel  $z \notin L\left(\frac{t}{1-t}; P\right)$ . Ainsi, comme nous avons supposé que la distribution est symétrique autour de  $T_P = 0$ , nous trouvons

$$\text{IF}(z; T, P) = \lim_{t \rightarrow 0} \frac{T_{\tilde{P}}}{t} = \frac{1}{2} \lambda\left(\frac{z}{\|z\|}\right) \frac{z}{\|z\|}.$$

Par contre, lorsque  $z = 0$ , il est clair que, pour tout  $\epsilon > 0$ ,  $z \in L\left(\frac{t}{1-t}; P\right)$ , de sorte que  $T_{\tilde{P}} = 0$ .  $\square$

Pour une distribution elliptique, lorsque  $\Sigma = I_p$ , la densité de  $u'X$ , notée  $f_0$ , est indépendante de  $u$ . Si la matrice de covariance est différente de l'identité, alors la densité de  $u'X$  est donnée par

$$f_u(x) = \frac{f_0(x/\sqrt{u'\Sigma u})}{\sqrt{u'\Sigma u}}. \quad (2.2)$$

Nous pouvons montrer que pour de telles distributions,  $\lambda(u) = \left\{f_0(0)\sqrt{u'\Sigma u}\right\}^{-1}$ . Ainsi, pour des distributions de matrice de covariance  $\Sigma = I_p$ ,  $\lambda(u) = 1/f_0(0)$ . Pour la loi normale standard par exemple, nous trouvons  $\lambda(u) = \sqrt{2\pi}$ .

## 2.3 Fonction de profondeur scatter

Nous allons à présent introduire la fonction de profondeur qui va nous occuper jusqu'à la fin de ce chapitre. Souvenons d'abord que l'idée principale de la médiane de demi-espace est de projeter les données sur une direction et d'en évaluer la profondeur sur ces espaces à une dimension. En prenant le point le plus profond, nous obtenons un estimateur de centralité. Une telle recette peut être utilisée pour estimer les matrices de covariance. Plus formellement, si nous considérons  $X \sim N(0, \Sigma)$ , la médiane de  $|u'X|^2$  est donnée par  $\beta u'\Sigma u$ , et pour  $\beta$  une certaine constante. En estimant la variance dans chaque direction, nous pouvons obtenir un estimateur de la matrice de covariance. La fonction de profondeur est donnée par l'expression suivante.

**Définition 2.6.** Soient  $P$  une distribution et  $\Sigma \in \mathbb{R}^{p \times p}$  une matrice semi-définie positive. La fonction de profondeur de scatter est donnée par

$$HD_{P,T}^{\text{sc}}(\Sigma) = \inf_{u \in \mathcal{S}^{p-1}} \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u'\Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u'\Sigma u} \right] \right),$$

où  $T_P$  est un estimateur de position choisi.

Afin d'illustrer cette définition, nous avons à la FIGURE 3 représenté en rouge tout d'abord l'estimateur de position  $T_P$ . L'ellipse noire représente l'ensemble des valeurs de  $\left\{ \sqrt{u'\Sigma u}, u \in \mathcal{S}^{p-1} \right\}$ . Il s'agit donc de parcourir toutes les directions  $u$ , et de prendre la valeur minimum entre la probabilité de la bande intérieure, définie par les traits en pointillés, et la bande extérieure.

En considérant

$$\hat{\Sigma} = \operatorname{argmax}_{\Sigma > 0} HD_{P,T}^{\text{sc}}(\Sigma),$$

nous obtenons un estimateur de matrice de covariance. Souvenons-nous que dans le cas de la profondeur de demi-espace, pour une distribution H-symétrique, la donnée qui maximisait la

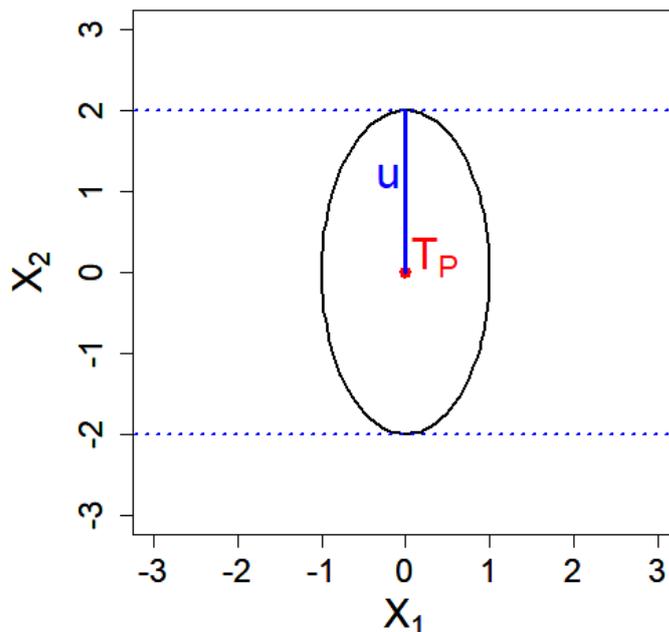


FIGURE 3 – Illustration de la définition de la fonction de profondeur de demi-espace.

profondeur correspondait exactement au centre de symétrie. Pour la profondeur scatter avec  $P \sim N(0, \Sigma)$ , la matrice qui maximise la profondeur est donnée par  $\beta\Sigma$ , avec  $\beta$  défini au travers de l'égalité

$$\Phi\left(\sqrt{\beta}\right) = \frac{3}{4},$$

où  $\Phi$  est la distribution d'une loi normale  $N(0, 1)$ . En d'autres termes, nous avons que

$$HD_{P,T}^{\text{sc}}(\beta\Sigma) = \frac{1}{2}.$$

Un bon estimateur de la matrice de covariance est dès lors donné par  $\hat{\Sigma}/\beta$ . Il existe de nombreux résultats sur cet estimateur (convergence, borne min max,...), mais nous allons nous concentrer dans ce document sur la fonction de profondeur elle-même.

Sauf mentionné autrement, les résultats obtenus dans les prochaines sections constituent l'apport personnel de ce document.

### 2.3.1 Fonction d'influence de la profondeur scatter lorsque $T_P = 0$

Dans cette sous-section, nous livrons la fonction d'influence lorsque le paramètre de position  $T_P$  est connu et vaut 0. Nous commençons par réécrire la profondeur de scatter comme

$$\begin{aligned} HD_{P,T}^{\text{sc}}(\Sigma) &= \inf_{u \in \mathcal{S}^{p-1}} \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u'\Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u'\Sigma u} \right] \right) \\ &\stackrel{\text{not.}}{=} \inf_{u \in \mathcal{S}^{p-1}} \min (P[A_{\text{in}}(u)], P[A_{\text{out}}(u)]) \\ &= \min \left( \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)] \right). \end{aligned}$$

Pour une distribution perturbée  $F_{t,z}$  et sa mesure de probabilité associée  $\tilde{P}$ , nous trouvons, pour une direction  $u \in \mathcal{S}^{p-1}$  fixée,

$$\tilde{P}[A_{\text{in}}(u)] = \begin{cases} (1-t)P[A_{\text{in}}(u)] + t & \text{si } z \in A_{\text{in}}(u) \\ (1-t)P[A_{\text{in}}(u)] & \text{sinon.} \end{cases}$$

Nous définissons à présent différents ensembles  $U_{\text{in}}^z$ ,  $U_{\text{in}}^{\bar{z}}$ ,  $U_{\text{out}}^z$  et  $U_{\text{out}}^{\bar{z}}$  comme

$$\begin{aligned} U_{\text{in}}^z &= \{u \in \mathcal{S}^{p-1} : z \in A_{\text{in}}(u)\}, \\ U_{\text{in}}^{\bar{z}} &= \{u \in \mathcal{S}^{p-1} : z \notin A_{\text{in}}(u)\}, \\ U_{\text{out}}^z &= \{u \in \mathcal{S}^{p-1} : z \in A_{\text{out}}(u)\}, \\ U_{\text{out}}^{\bar{z}} &= \{u \in \mathcal{S}^{p-1} : z \notin A_{\text{out}}(u)\}. \end{aligned}$$

La profondeur de demi-espace, sous la distribution perturbée, peut dès lors se réécrire

$$\begin{aligned}
HD_{\tilde{P},T}^{\text{sc}}(\Sigma) &= \min \left( \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[A_{\text{out}}(u)] \right) \\
&= \min \left[ \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{in}}^{\bar{z}}} \tilde{P}[A_{\text{in}}(u)] \right), \right. \\
&\quad \left. \min \left( \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)], \inf_{u \in U_{\text{out}}^{\bar{z}}} \tilde{P}[A_{\text{out}}(u)] \right) \right] \\
&= (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)] + \frac{t}{1-t}, \inf_{u \in U_{\text{in}}^{\bar{z}}} P[A_{\text{in}}(u)], \right. \\
&\quad \left. \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] + \frac{t}{1-t}, \inf_{u \in U_{\text{out}}^{\bar{z}}} P[A_{\text{out}}(u)] \right).
\end{aligned}$$

Le théorème suivant livre la fonction d'influence lorsque le paramètre de position est connu et fixé à l'origine. Notons que ce résultat est aussi valide si  $\tilde{P}$  n'a pas d'influence sur  $T_P$ , de sorte que la perturbation ne change pas la position.

**Théorème 2.6.** *Soient  $P$  une mesure de probabilité sur  $\mathbb{R}^p$ ,  $\Sigma$  une matrice définie positive, et supposons que le paramètre de position  $T_P$  est connu et fixé à 0. Notons*

$$A = \min \left( \inf_{u \in U_{\text{in}}^{\bar{z}}} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right)$$

et

$$B = \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^{\bar{z}}} P[A_{\text{out}}(u)] \right).$$

La fonction d'influence de la profondeur de scatter est donnée par

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) = \begin{cases} 1 - HD_{P,T}^{\text{sc}}(\Sigma) & \text{si } A > B \\ -HD_{P,T}^{\text{sc}}(\Sigma) & \text{si } A \leq B. \end{cases}$$

*Démonstration.* Commençons par remarquer que, pour  $U_{\text{in}}^z$  par exemple,

$$(1-t) \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)] \leq \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)] \leq (1-t) \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)] + t,$$

et donc

$$\lim_{t \rightarrow 0} \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)] = \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)].$$

Ce résultat est évidemment valide pour les trois autres ensembles. Ainsi, dans le cas où

$$HD_{P,T}^{\text{sc}}(\Sigma) = \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) < \min \left( \inf_{u \in U_{\text{in}}^{\bar{z}}} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^{\bar{z}}} P[A_{\text{out}}(u)] \right),$$

il existe  $t$  suffisamment petit tel quel

$$\begin{aligned} \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right) &= (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) + t \\ &< (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) \\ &= \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right). \end{aligned}$$

Nous trouvons donc

$$\begin{aligned} HD_{\tilde{P},T}^{\text{sc}}(\Sigma) &= \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right) \\ &= (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) + t \\ &= (1-t) HD_{P,T}^{\text{sc}}(\Sigma) + t, \end{aligned}$$

de sorte que la fonction d'influence est donnée par

$$\begin{aligned} \text{IF}(z; HD_{\tilde{P},T}^{\text{sc}}(\Sigma), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(\Sigma) - HD_{P,T}^{\text{sc}}(\Sigma)}{t} \\ &= \lim_{t \rightarrow 0} \frac{(1-t) HD_{P,T}^{\text{sc}}(\Sigma) + t - HD_{P,T}^{\text{sc}}(\Sigma)}{t} \\ &= 1 - HD_{P,T}^{\text{sc}}(\Sigma). \end{aligned}$$

Dans l'autre cas, si

$$HD_{P,T}^{\text{sc}}(\Sigma) = \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) \leq \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right),$$

nous avons que

$$\begin{aligned} \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right) &= (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) \\ &\leq (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) + t \\ &= \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right). \end{aligned}$$

Nous obtenons donc pour la profondeur sous la distribution perturbée

$$\begin{aligned} HD_{\tilde{P},T}^{\text{sc}}(\Sigma) &= \min \left( \inf_{u \in U_{\text{in}}^z} \tilde{P}[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} \tilde{P}[A_{\text{out}}(u)] \right) \\ &= (1-t) \min \left( \inf_{u \in U_{\text{in}}^z} P[A_{\text{in}}(u)], \inf_{u \in U_{\text{out}}^z} P[A_{\text{out}}(u)] \right) \\ &= (1-t) HD_{P,T}^{\text{sc}}(\Sigma), \end{aligned}$$

de sorte que la fonction d'influence est donnée par

$$\begin{aligned} \text{IF}(z; HD_{\tilde{P},T}^{\text{sc}}(\Sigma), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(\Sigma) - HD_{P,T}^{\text{sc}}(\Sigma)}{t} \\ &= \lim_{t \rightarrow 0} \frac{(1-t)HD_{\tilde{P},T}^{\text{sc}}(\Sigma) - HD_{P,T}^{\text{sc}}(\Sigma)}{t} \\ &= -HD_{\tilde{P},T}^{\text{sc}}(\Sigma). \end{aligned}$$

□

Nous constatons que l'expression obtenue est similaire à celle de la fonction d'influence pour la profondeur de demi-espace. Ainsi, les conclusions obtenues pour cette dernière, en terme de robustesse mais aussi pour la variance asymptotique par exemple, s'appliquent au cas de la fonction de profondeur scatter, lorsque le paramètre de position est fixé. Dans les sections suivantes, nous livrons la fonction d'influence pour un estimateur de position quelconque, en commençant lorsque la distribution  $P$  est discrète à support fini.

### 2.3.2 Fonction d'influence pour une distribution discrète

Nous nous intéressons à présent aux distributions discrètes. Tout d'abord, nous introduisons un nouvel estimateur de position. Ce dernier correspond au centre de masse de l'ensemble des points maximisant la profondeur. Il est donc défini comme

$$T_P = \frac{\int_{\text{Med}(P)} y \, dy}{\int_{\text{Med}(P)} dy}, \quad \text{où } \text{Med}(P) = \{x \in \mathbb{R}^p : D_H(x, P) = \Pi(P)\}. \quad (2.3)$$

Nous utilisons cette définition afin d'avoir un estimateur de position unique, et qui correspond donc à la médiane de demi-espace si la donnée maximisant la profondeur est unique. La proposition suivante livre le comportement de cet estimateur sous la distribution perturbée.

**Proposition 2.2.** *Soit  $F$  une distribution discrète à support fini sur  $\mathbb{R}^p$ . Soient  $z \in \mathbb{R}^p$  et  $\tilde{P}$  la mesure de probabilité associée à la distribution contaminée  $F_{t,z}$ . Il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $T_{\tilde{P}}$  est indépendant de  $t$ .*

*Démonstration.* Comme la distribution est discrète à support fini, l'ensemble des valeurs que prend la fonction de profondeur de demi-espace est fini. L'objectif est de montrer qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $\text{Med}(\tilde{P})$  est indépendant de  $t$  et donc son centre de masse l'est aussi. Commençons par montrer qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $\text{Med}(\tilde{P}) \subseteq \text{Med}(P)$ . Nous allons pour ce faire montrer qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $\text{Med}(P)^C \subseteq \text{Med}(\tilde{P})^C$ . Nous avons que

$$\max_{x \in \text{Med}(P)^C} D_H(x; P) < \Pi(P),$$

et donc il existe  $t_0 > 0$ , tel que pour tout  $t < t_0$

$$\max_{x \in \text{Med}(P)^C} D_H(x; \tilde{P}) \leq (1-t) \max_{x \in \text{Med}(P)^C} D_H(x; P) + t < (1-t)\Pi(P) \leq \Pi(\tilde{P}).$$

Ainsi,  $\exists t_0 > 0 : \forall t < t_0, \forall x \in \text{Med}(P)^C, D_H(x; \tilde{P}) < \Pi(\tilde{P})$  et donc  $\exists t_0 > 0 : \forall t < t_0, \text{Med}(P)^C \subseteq \text{Med}(\tilde{P})^C$ . Le fait d'avoir montré  $\text{Med}(\tilde{P}) \subseteq \text{Med}(P)$  suggère que les seuls candidats  $x$  tels que  $D_H(x; \tilde{P}) = \Pi(\tilde{P})$  se trouvent dans  $\text{Med}(P)$ . Dans ce dernier ensemble, la profondeur  $D_H(x, \tilde{P})$  ne prend que deux valeurs, soit  $(1-t)\Pi(P)$ , soit  $(1-t)\Pi(P) + t$ . Si, pour tout  $x$  dans  $\text{Med}(P)$ , nous avons  $D_H(x, \tilde{P}) = (1-t)D_H(x, P)$  ou  $D_H(x, \tilde{P}) = (1-t)D_H(x, P) + t$ , alors  $\text{Med}(\tilde{P}) = \text{Med}(P)$ ,

qui est indépendant de  $t$ , et donc  $T_{\tilde{P}} = T_P$ , toujours pour tout  $t < t_0$ . Cependant, si, toujours dans  $\text{Med}(P)$ , il existe  $x$  tel que  $D_H(x, \tilde{P}) = (1-t)D_H(x, P)$  et  $y$  tel que  $D_H(y, \tilde{P}) = (1-t)D_H(y, P) + t$ , nous trouvons

$$\text{Med}(\tilde{P}) = \{x \in \text{Med}(P) : D_H(x, \tilde{P}) = (1-t)D_H(x, P) + t\}.$$

qui est aussi indépendant de  $t$ , pour tout  $t < t_0$ . □

Nous allons illustrer ce dernier résultat en abordant les deux différents cas exposés à la fin de la démonstration. Pour ce faire, nous considérons une distribution discrète  $P$  de  $\mathbb{R}$  sur cinq éléments  $a < b < c < d < e$ . Ainsi, nous avons  $\Pi(P) = 3/5$  et cette valeur est atteinte uniquement en  $c$ . Ainsi,  $\text{Med}(P) = \{c\}$ . Notons que, pour une contamination  $z > e$ , la profondeur de  $c$  sous distribution perturbée vaut  $D_H(c, \tilde{P}) = (1-t)3/5$ . Les autres valeurs atteintes par la fonction de profondeur  $D_H(x, P)$  sont  $2/5, 1/5$  et  $0$ , de sorte qu'il existe  $t$  suffisamment petit tel que, pour tout  $x \neq c$ ,  $D_H(x, \tilde{P}) \leq (1-t)2/5 + t < (1-t)3/5$ . Ainsi,  $c$  est aussi l'unique point qui maximise la profondeur sous la distribution perturbée. Cet exemple soulève une conclusion très importante de la proposition précédente. En effet, lorsque la donnée qui maximise la profondeur  $D_H(x, P)$  est unique, et donc  $\text{Med}(P)$  est un singleton, nous avons qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $\text{Med}(\tilde{P}) = \text{Med}(P)$  et donc  $T_{\tilde{P}} = T_P$ . Pour illustrer le second cas, considérons une distribution discrète  $P$  de  $\mathbb{R}$  sur quatre éléments  $a < b < c < d$ , où chaque valeur a la même probabilité. Nous trouvons  $\Pi(P) = 0.5$  et cette valeur est atteinte pour l'ensemble de points entre  $b$  et  $c$ , de sorte que  $\text{Med}(P) = \{x \in \mathbb{R} : b \leq x \leq c\}$ . Ainsi, nous avons que  $T_P = \frac{b+c}{2}$ . Si nous ajoutons une contamination en  $z > d$ , le seul point qui aura une profondeur de  $(1-t)0.5 + t$  est  $c$ , de sorte que  $\text{Med}(\tilde{P}) = \{c\} \subset \text{Med}(P)$ . Ainsi, il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $T_{\tilde{P}} \neq T_P$ .

Notons enfin que, de manière générale, un estimateur pour lequel nous connaissons la fonction d'influence vérifiera toujours  $\lim_{t \rightarrow 0} T_{\tilde{P}} = T_P$ , étant donné que pour que la dérivée de cet estimateur existe en  $t = 0$ , il doit être continu en  $t = 0$ .

Nous avons démontré qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ , soit  $T_{\tilde{P}} = T_P$ , soit  $T_{\tilde{P}} \neq T_P$ . Dans le premier cas, nous pouvons directement conclure à l'expression de la fonction d'influence en utilisant le **Théorème 2.6** pour  $Y = X - T_P$ . Notons qu'une condition suffisante pour être dans ce cas est que  $T_P$  est unique. Concernant le second cas, il est impossible de conclure directement et il faut étudier chaque cas séparément. Afin d'illustrer cela, reprenons notre exemple d'une distribution discrète sur quatre éléments de  $\mathbb{R}$ , avec  $a = -5$ ,  $b = -2$ ,  $c = 2$  et  $d = 5$ , de sorte que  $T_P = 0$ . Nous avons obtenu que pour une contamination en  $z > d$ , il existe  $t_0 > 0$  tel que pour tout  $t < t_0$ ,  $T_{\tilde{P}} = c$ . Considérons initialement  $\Sigma = 1$ , pour lequel nous trouvons  $HD_{P,T}^{\text{sc}}(\Sigma) = 0$  étant donné que  $P[|X| \leq 1] = 0$ . Cependant, nous avons que  $\tilde{P}[|(X-2)| \leq 1] = (1-t)1/4$ , de sorte que

$$\begin{aligned} \text{IF}(z; HD_{P,T}^{\text{sc}}(1), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(1) - HD_{P,T}^{\text{sc}}(1)}{t} \\ &= \lim_{t \rightarrow 0} \frac{(1-t)1/4}{t} \\ &= +\infty. \end{aligned}$$

Par contre, si nous prenons  $\Sigma = 10$ , nous avons  $HD_{P,T}^{\text{sc}}(10) = P[|X| \geq 10] = 0$  et  $HD_{\tilde{P},T}^{\text{sc}}(\Sigma) =$

$\tilde{P}[|(X - 2)| \geq 10] = 0$ , de sorte que

$$\begin{aligned} \text{IF}(z; HD_{\tilde{P},T}^{\text{sc}}(10), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(10) - HD_{P,T}^{\text{sc}}(10)}{t} \\ &= 0. \end{aligned}$$

En conclusion, si  $T_P$  n'est pas unique, il n'existe pas d'expression universelle pour la fonction d'influence.

Nous avons principalement discuté de l'estimateur lié à la profondeur de demi-espace, mais les conclusions restent similaires pour d'autres estimateurs de position. Par exemple, si nous considérons la moyenne, nous avons, en considérant le dernier exemple, que  $T_P = 0$  et  $T_{\tilde{P}} = tz$ . Ainsi, nous avons que  $HD_{P,T}^{\text{sc}}(10) = P[|X| \geq 10] = 0$  et  $HD_{\tilde{P},T}^{\text{sc}}(\Sigma) = \tilde{P}[|(X - tz)| \geq 10] = 0$  pour  $t$  suffisamment petit, de sorte que

$$\begin{aligned} \text{IF}(z; HD_{P,T}^{\text{sc}}(10), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(10) - HD_{P,T}^{\text{sc}}(10)}{t} \\ &= 0. \end{aligned}$$

Si nous considérons par contre  $\Sigma = 2$ , nous avons que  $HD_{P,T}^{\text{sc}}(2) = P[|X| \geq 2] = 0.5$ . En revanche, sous la distribution perturbée, nous trouvons  $HD_{\tilde{P},T}^{\text{sc}}(\Sigma) = \tilde{P}[|(X - tz)| \geq 2] = (1 - t)0.25$ , de sorte que

$$\begin{aligned} \text{IF}(z; HD_{P,T}^{\text{sc}}(2), F) &= \lim_{t \rightarrow 0} \frac{HD_{\tilde{P},T}^{\text{sc}}(2) - HD_{P,T}^{\text{sc}}(2)}{t} \\ &= \infty. \end{aligned}$$

Nous remarquons donc que, pour la moyenne aussi, il est compliqué de dégager une expression générale pour la fonction d'influence.

### 2.3.3 Fonction d'influence pour une distribution continue

Dans cette section, nous livrons la fonction d'influence de la profondeur scatter pour une fonctionnelle de position  $T_P$  quelconque. Nous voulons pour cela montrer que l'infimum qui apparaît dans la fonction de profondeur est atteint. Si la distribution est discrète à support fini, il est évident qu'il suffit uniquement d'un nombre fini de direction sur la sphère pour évaluer le minimum qui est donc atteint. Avant de démontrer que l'infimum est aussi atteint pour une distribution absolument continue, nous avons besoin d'une propriété sur la convergence de la suite  $(u'_n X)_{n \in \mathbb{N}}$ , où  $(u_n)$  est une suite de  $\mathbb{R}^p$  que l'on suppose converger vers une direction  $u_0$  et  $X$  un vecteur aléatoire. Les définitions de la convergence presque sûre et en distribution sont reprises ci-dessous.

**Définition 2.7.** *La suite de variable aléatoire  $(X_n)$  converge presque sûrement vers la variable aléatoire  $X$  si*

$$P \left[ \lim_{n \rightarrow \infty} X_n = X \right] = 1.$$

**Définition 2.8.** *La suite de variable aléatoire  $(X_n)$  converge en distribution vers la variable aléatoire  $X$  si*

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

en tout point où la fonction  $F_X$  est continue.

Notons que la convergence presque sûre implique la convergence en distribution. Nous pouvons à présent montrer que la suite  $(u'_n X)$  converge en distribution vers  $u'_0 X$ .

**Proposition 2.3.** *Soient  $X$  un vecteur aléatoire et  $(u_n)$  une suite de  $\mathbb{R}^p$  convergente vers  $u_0$ . Alors, la suite  $(u'_n X)$  converge en distribution vers  $u'_0 X$ .*

*Démonstration.* Il suffit de remarquer que

$$P \left[ \lim_{n \rightarrow \infty} u'_n X = u'_0 X \right] = 1,$$

de sorte que la suite  $(u'_n X)$  converge presque sûrement vers  $u'_0 X$  et donc en distribution.  $\square$

Nous pouvons à présent énoncer et démontrer que l'infimum apparaissant dans la fonction de profondeur est atteint lorsque la distribution considérée est continue.

**Proposition 2.4.** *Soit  $P$  une mesure de probabilité sur  $\mathbb{R}^p$  absolument continue par rapport à la mesure de Lebesgue, et soit  $\Sigma$  une matrice définie positive. La fonction de profondeur de scatter peut dès lors se réécrire*

$$\begin{aligned} HD_{P,T}^{\text{sc}}(\Sigma) &= \inf_{u \in \mathcal{S}^{p-1}} \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u' \Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u' \Sigma u} \right] \right) \\ &= \min_{u \in \mathcal{S}^{p-1}} \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u' \Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u' \Sigma u} \right] \right). \end{aligned}$$

*Démonstration.* L'objectif est de montrer que la fonction

$$\mathcal{S}^{p-1} \rightarrow [0, 1] : u \mapsto \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u' \Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u' \Sigma u} \right] \right)$$

est continue sur le compact  $\mathcal{S}^{p-1}$ , et donc atteint ses bornes. Pour ce faire, nous montrons que la fonction

$$\mathcal{S}^{p-1} \rightarrow [0, 1] : u \mapsto f(u) := P \left[ |u'(X - T_P)| \leq \sqrt{u' \Sigma u} \right]$$

est continue. Soit  $u_0$  et  $\eta > 0$  fixés. Nous devons montrer que, pour  $\delta > 0$ , nous avons

$$\|u - u_0\| < \delta \quad \implies \quad |f(u) - f(u_0)| < \eta.$$

Considérons l'ensemble

$$A(u, u_0) = \left\{ x \in \mathbb{R}^p : |u'(x - T_P)| \leq \sqrt{u'_0 \Sigma u_0} \right\}.$$

Comme la forme quadratique  $u \mapsto u' \Sigma u$  est continue (et donc sa racine carrée aussi) et comme la variable aléatoire  $u'(X - T_P)$  est absolument continue par rapport à la mesure de Lebesgue (et donc sa fonction de répartition est continue), il existe  $\delta_1 > 0$  tel que

$$\begin{aligned} |f(u) - P[A(u, u_0)]| &= \left| F_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) - F_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) \right. \\ &\quad \left. - F_{u'(X-T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) + F_{u'(X-T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) \right| \\ &\leq \frac{\eta}{2}. \end{aligned}$$

Il reste à montrer que  $|P[A(u, u_0)] - f(u_0)|$  est arbitrairement petit lorsque  $\|u - u_0\|$  est petit. Comme nous avons montré que, pour une suite  $(u_n)$  qui converge vers  $u_0$ , la suite  $u'_n(X - T_P)$

converge en distribution vers  $u'_0(X - T_P)$ , de sorte que pour tout  $x \in \mathbb{R}^p$ ,  $\lim_{n \rightarrow \infty} F_{u'_n(X - T_P)}(x) = F_{u'_0(X - T_P)}(x)$ , il existe  $\delta_2 > 0$  tel que

$$\begin{aligned} |P[A(u, u_0)] - f(u_0)| &= \left| F_{u'(X - T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) - F_{u'(X - T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) \right. \\ &\quad \left. - F_{u'_0(X - T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) + F_{u'_0(X - T_P)} \left( \sqrt{u'_0 \Sigma u_0} \right) \right| \\ &\leq \frac{\eta}{2}. \end{aligned}$$

En conclusion, nous trouvons que, pour  $\|u - u_0\| < \min(\delta_1, \delta_2)$ ,

$$|f(u) - f(u_0)| < |f(u) - P[A(u, u_0)]| + |P[A(u, u_0)] - f(u_0)| < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

En utilisant le même argument, nous pouvons montrer que la deuxième partie du minimum dans la fonction de profondeur est continue, de sorte que le minimum entre deux fonctions continues est aussi continu, ce qui conclut la preuve.  $\square$

Nous allons à présent discuter du comportement de l'estimateur de position sous la distribution contaminée, à savoir  $T_{\tilde{P}}$ . Lorsque la fonctionnelle de position est de la forme  $\int gdF$ , comme c'est le cas pour la moyenne avec  $g(x) = x$ , nous trouvons directement, par linéarité de l'intégrale,  $T_{\tilde{P}} = (1 - t)T_P + tz$ . Ainsi, nous avons que

$$\lim_{t \rightarrow 0} T_{\tilde{P}} = T_P.$$

Concernant l'estimateur défini en 2.3, la proposition suivante montre qu'il s'agit d'un estimateur convergent sous la distribution perturbée.

**Proposition 2.5.** *Soit  $P$  une mesure de probabilité absolument continue par rapport à la mesure de Lebesgue sur  $\mathbb{R}^p$  et supposons que  $\inf_{u \in S^{p-1}} f_u(u'T_P) > 0$ . Soient  $z \in \mathbb{R}^p$  et  $\tilde{P}$  la mesure de probabilité associée à la distribution contaminée  $F_{t,z}$ . Alors*

$$\lim_{t \rightarrow 0} T_{\tilde{P}} = T_P.$$

*Démonstration.* Il suffit d'utiliser le théorème 4.3 du document [11]. Ce dernier montre en effet que

$$\limsup_{t \rightarrow 0} \frac{\sup_{z \in \mathbb{R}^p} \|T_{\tilde{P}} - T_P\|}{t} < \infty,$$

de sorte que le numérateur tend vers 0.  $\square$

Notons que, pour une distribution elliptique, il est suffisant, au vue de l'expression (2.2), que  $f_0(u'T_P / \sqrt{u'\Sigma u}) > 0$ .

Nous livrons à présent les bornes sur la fonction d'influence lorsque la distribution est continue. Nous commençons tout d'abord par donner un lemme qui sera utile pour la preuve de la fonction d'influence. Nous allons dans celui-ci utiliser la notion d'équidifférentiabilité.

**Définition 2.9.** Une famille de fonction  $\{f(u, \cdot)\}_{u \in \mathcal{S}^{p-1}}$  est équidifférentiable en  $t \in [0, 1]$  si l'expression

$$\frac{f(u, t') - f(u, t)}{t' - t}$$

converge uniformément pour  $t' \rightarrow t$ .

Notons que l'équidifférentiabilité d'une fonction  $f$  à deux arguments permet d'écrire

$$\frac{f(x(t), t) - f(x(t), t_0)}{t - t_0} = f_t(x(t), t_0) + \frac{o(t - t_0)}{t - t_0}. \quad (2.4)$$

Nous allons dorénavant, sans perdre de généralité, considérer dans les différentes propositions que  $u'(T_{\tilde{P}} - T_P) \geq 0$ , de sorte que si  $u$  vérifie l'inégalité opposée, il suffit alors de considérer  $-u$ , pour lequel les discussions sont similaires. Définissons tout d'abord les ensembles  $\tilde{A}_{\text{in}}(u)$ ,  $\tilde{A}_{\text{out}}(u)$ ,  $B_t(u)$  et  $C_t(u)$  comme

$$\begin{aligned} \tilde{A}_{\text{in}}(u) &= \left\{ x \in \mathbb{R}^p : |u'(x - T_{\tilde{P}})| \leq \sqrt{u' \Sigma u} \right\}, \\ \tilde{A}_{\text{out}}(u) &= \left\{ x \in \mathbb{R}^p : |u'(x - T_{\tilde{P}})| \geq \sqrt{u' \Sigma u} \right\}, \\ B_t(u) &= \left\{ x \in \mathbb{R}^p : \sqrt{u' \Sigma u} \leq u'(x - T_P) \leq \sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right\}, \\ C_t(u) &= \left\{ x \in \mathbb{R}^p : -\sqrt{u' \Sigma u} \leq u'(x - T_P) \leq -\sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right\}. \end{aligned}$$

**Lemme 2.1.** Soit  $P$  une distribution absolument continue sur  $\mathbb{R}^p$  et  $z \in \mathbb{R}^p$ . Supposons que  $T_{\tilde{P}}$  est continument différentiable par rapport à  $t$  et que la densité  $f_{u'(X - T_P)}(x)$  est continue par rapport à  $u$  et  $x$ . Alors, les familles de fonctions  $\{P[B_t(u)]\}_{u \in \mathcal{S}^{p-1}}$  et  $\{P[C_t(u)]\}_{u \in \mathcal{S}^{p-1}}$  définies comme

$$\begin{aligned} P[B_t(u)] &= P \left[ \left\{ x \in \mathbb{R}^p : \sqrt{u' \Sigma u} \leq u'(x - T_P) \leq \sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right\} \right], \\ P[C_t(u)] &= P \left[ \left\{ x \in \mathbb{R}^p : -\sqrt{u' \Sigma u} \leq u'(x - T_P) \leq -\sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right\} \right] \end{aligned}$$

sont équidifférentiables.

*Démonstration.* Nous donnons la preuve pour la première famille de fonctions, que nous écrivons

$$\{P[B_t(u)]\}_{u \in \mathcal{S}^{p-1}} = \{g(u, t)\}_{u \in \mathcal{S}^{p-1}}.$$

Une condition suffisante pour l'équidifférentiabilité est que la famille des dérivées partielles par rapport au second argument, notée  $\{g_t(u, t)\}_{u \in \mathcal{S}^{p-1}}$ , est équicontinue. De plus, si la fonction est définie sur un compact, ce qui est le cas ici, la continuité de  $g_t(u, t)$  sur  $\mathcal{S}^{p-1} \times [0, 1]$  implique que  $\{g_t(u, t)\}_{u \in \mathcal{S}^{p-1}}$  est équicontinu. Il suffit donc de prouver que  $g_t(u, t)$  est continu. Nous avons

$$P[B_t(u)] = F_{u'(X - T_P)} \left( \sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right) - F_{u'(X - T_P)} \left( \sqrt{u' \Sigma u} \right)$$

ce qui donne

$$\frac{d}{dt} P[B_t(u)] = f_{u'(X - T_P)} \left( \sqrt{u' \Sigma u} + u'(T_{\tilde{P}} - T_P) \right) u' \frac{d}{dt} T_{\tilde{P}}.$$

Sous l'hypothèse de différentiabilité continue de  $T_{\tilde{P}}$ , il est clair que la fonction est continue par rapport  $t$ . Sous les hypothèses sur  $f_X(x)$ , la fonction est continue sur  $\mathcal{S}^{p-1} \times [0, 1]$  sachant que l'unique terme dépendant de à la fois  $t$  et  $u$  est un produit scalaire qui lui même est continu sur  $\mathcal{S}^{p-1} \times [0, 1]$ .

□

Nous pouvons à présent discuter des différentes hypothèses de ce lemme. Notons tout d'abord qu'il faudrait pour être totalement rigoureux demander que  $f_{u'(X-T_P)}(g(u,t))$  soit continue conjointement par rapport à  $u$  et  $t$ , où  $g(u,t) = \sqrt{u'\Sigma u} + u'(T_{\tilde{P}} - T_P)$  est elle-même une fonction continue sur  $\mathcal{S}^{p-1} \times [0,1]$ . Bien que souvent équivalente, cette dernière condition est dans certains cas plus forte que celle supposée dans la proposition. Concernant la différentiabilité continue de  $T_{\tilde{P}}$ , il est évident que c'est le cas pour des estimateurs de la forme  $\int g dF$ , étant donné que nous avons alors  $T_{\tilde{P}} = (1-t)T_P + tz$ , qui est continument différentiable par rapport à  $t$ . Concernant l'estimateur basé sur la profondeur de demi-espace, en utilisant le **Théorème 2.4**, nous avons que, pour  $t$  suffisamment petit,

$$T_{\tilde{P}} = \frac{1}{2} r \left( \frac{t}{1-t}, \frac{z}{\|z\|} \right) \frac{z}{\|z\|}.$$

Il s'agit donc de voir si

$$\frac{d}{dt} r \left( \frac{t}{1-t}, \frac{z}{\|z\|} \right)$$

est continue par rapport à  $t$ . Pour une distribution elliptique, [11] livre, pour une direction  $u$  et  $\alpha \in \mathbb{R}$ ,

$$r(\alpha, u) = F_0^{-1} \left( \frac{1}{2} + \alpha \right) \left( \sqrt{u'\Sigma^{-1}u} \right)^{-1}$$

où  $F_0$  est la distribution commune de  $u'X$  si la matrice de covariance  $\Sigma$  de la distribution de  $X$  est l'identité. Ainsi, nous trouvons

$$\frac{d}{dt} r \left( \frac{t}{1-t}, u \right) = \left( \sqrt{u'\Sigma^{-1}u} \right)^{-1} \frac{1}{(1-t)^2} \frac{1}{f_0 \left( F_0^{-1} \left( \frac{1}{2} + \frac{t}{(1-t)} \right) \right)},$$

qui est continu en  $t$  sur l'intervalle  $[0,0.5]$  par exemple. La seconde hypothèse moins triviale est que  $f_{u'(X-T_P)}$  est continue par rapport à  $u$ . A nouveau, pour des distributions elliptiques, l'égalité (2.2) montre que nous avons toujours cette continuité. Pour une distribution quelconque par contre, le théorème de Scheffé affirme que la convergence des densités de variables aléatoires implique la convergence en distribution. Pour une suite de variables aléatoires de la forme  $u'_n X$ , où  $u_n$  est une suite convergente vers une direction  $u_0$  par exemple, nous avons la convergence en distribution, mais, bien que ce soit souvent le cas, cela n'implique pas que les fonctions de densité convergent. Il existe cependant des résultats permettant de conclure à la convergence de la suite de densités, toujours sous l'hypothèse de convergence en loi de la suite de variables aléatoires. Par exemple, dans le document [14], ils prouvent que  $\|f_{u'_n X}(x) - f_{u'_0 X}(x)\|_\infty \rightarrow 0$ , sous les hypothèses que

1. il existe une valeur  $M(x)$  telle que  $\sup_n |f_{u'_n X}(x)| < M(x)$ ,
2.  $\{f_{u'_n X}\}$  est uniformément continue.

Ou encore le document [15] montre qu'il est possible d'obtenir la même thèse sans supposer que les fonctions  $f_{u'_n X}$  soient continues.

Nous pouvons à présent aborder le théorème principal de cette section, qui livre les bornes sur la fonction d'influence de la profondeur de scatter. Pour une distribution discrète, le résultat a déjà été obtenu et nous considérons donc une distribution absolument continue. Il nous faudra de plus ajouter quelques hypothèses sur la densité, par rapport au lemme précédent. La première est que la densité  $f_X(x)$  est bornée, et la seconde que  $\sup_{u \in \mathcal{S}^{p-1}} f_{u'(X-T_P)}(x) < \infty$ . Notons que, pour

une distribution elliptique, l'expression 2.2 suffit à montrer que seule la condition  $f_X(x)$  bornée est nécessaire, car implique que  $f_0(x)$  est bornée et donc la seconde condition est vérifiée. Pour une distribution quelconque, il existe différents résultats permettant de conclure que si la densité est bornée, alors  $\sup_{u \in \mathcal{S}^{p-1}} f_{u'(X-T_P)}(x) < \infty$ . Par exemple, le document [16] montre que si la densité peut s'écrire comme  $f_X = \prod_{i=1}^n f_i(x_i)$ , alors la densité projetée sur n'importe quel sous-espace de  $\mathbb{R}^p$  est bornée. Le document [17] montre quant à lui que, sans hypothèse supplémentaire au fait que  $f_X$  est bornée, que la densité de la variable aléatoire projetée est bornée partout sauf possiblement sur des ensembles de petites mesures. Il livre aussi d'autres conclusions sur certaines propriétés à vérifier, dans des cas particuliers, pour s'assurer que toutes les densités marginales sont partout bornées.

Nous donnons un autre lemme utile pour la démonstration de la fonction d'influence

**Lemme 2.2.** *Soient  $P$  une distribution absolument continue sur  $\mathbb{R}^p$ ,  $\Sigma$  une matrice définie positive et  $z \in \mathbb{R}^p$ . Supposons que  $T_{\tilde{P}}$  converge vers  $T_P$ . Alors, sous les hypothèses du **Lemme 2.4**, ainsi qu'en supposant  $\sup_{u \in \mathcal{S}^{p-1}} f_{u'(X-T_P)}(x) < \infty$ , la fonction de profondeur de demi-espace scatter est continue par rapport à  $t$ , dans le sens où*

$$\lim_{t \rightarrow 0} HD_{\tilde{P}, T}^{\text{sc}}(\Sigma) = HD_{P, T}^{\text{sc}}(\Sigma).$$

*Démonstration.* Il nous faut montrer que

$$\lim_{t \rightarrow 0} \min \left( \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{out}}(u)] \right) = \min \left( \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)] \right).$$

Nous montrons que

$$\lim_{t \rightarrow 0} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)],$$

de sorte que l'argument fonctionne aussi pour l'autre partie du minimum, et donc nous avons bien la thèse. Par définition de l'infimum, nous avons que pour tout  $t > 0$ , il existe  $\tilde{u}_{\text{in}}$ , dépendant de  $t$ , tel que

$$\begin{aligned} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] &\geq (1-t)(P[A_{\text{in}}(\tilde{u}_{\text{in}})] + P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]) - t \\ &\geq (1-t) \left( \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] + P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})] \right) - t. \end{aligned}$$

De plus, nous avons que, pour un  $u_0 \in U_0$ ,

$$\begin{aligned} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] &\leq \tilde{P}[\tilde{A}_{\text{in}}(u_0)] \\ &\leq (1-t) \left( \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] + P[B_t(u_0)] - P[C_t(u_0)] \right) + t. \end{aligned}$$

Il nous suffit donc de montrer que, pour n'importe quelle direction dépendante de  $t$ ,

$$\lim_{t \rightarrow 0} P[B_t(\tilde{u}_{\text{in}})] = \lim_{t \rightarrow 0} P[C_t(\tilde{u}_{\text{in}})] = 0.$$

Nous effectuons la preuve pour l'ensemble  $B_t(\tilde{u}_{\text{in}})$ . En utilisant le **Lemme 2.1** ainsi que la remarque (2.4) avec  $g(u(t), t) = P[B_t(\tilde{u}_{\text{in}})]$  et  $t_0 = 0$ , nous avons, en remarquant que  $P[B_0(\tilde{u}_{\text{in}})] = 0$ ,

$$\frac{P[B_t(\tilde{u}_{\text{in}})]}{t} = f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}_{\text{in}}} \right) \tilde{u}'_{\text{in}} \text{IF}(z; T, P) + \frac{o(t)}{t}.$$

Ainsi, nous obtenons

$$\begin{aligned} \lim_{t \rightarrow 0} P[B_t(\tilde{u}_{\text{in}})] &= \lim_{t \rightarrow 0} t f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}_{\text{in}}} \right) \tilde{u}'_{\text{in}} \text{IF}(z; T, P) + o(t) \\ &= 0, \end{aligned}$$

par hypothèse que  $\sup_{u \in \mathcal{S}^{p-1}} f_{u'(X-T_P)}(x) < \infty$ , ce qui conclut la preuve.  $\square$

Comme c'est le cas pour  $T_{\tilde{P}}$ , nous supposons généralement que la fonctionnelle est continue par rapport à  $t$  sinon la fonction d'influence n'existerait pas. Dans le cas de la profondeur de demi-espace scatter, nous avons montré que c'était bien le cas. Notons de plus que, comme

$$\lim_{t \rightarrow 0} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)],$$

alors, si  $\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)]$ , il existe  $t_0 > 0$  tel que pour tout  $t < t_0$ ,  $\inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{out}}(u)]$ . Le précédent lemme implique aussi que

$$\begin{aligned} \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] &= \lim_{t \rightarrow 0} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] \\ &= \lim_{t \rightarrow 0} (1-t)(P[A_{\text{in}}(\tilde{u}_{\text{in}})] + P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]) + tI[z \in \tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})] \\ &= \lim_{t \rightarrow 0} P[A_{\text{in}}(\tilde{u}_{\text{in}})]. \end{aligned} \tag{2.5}$$

Nous définissons enfin au préalable différents ensembles :

$$\begin{aligned} U_0 &= \left\{ u \in \mathcal{S}^{p-1} : HD_{\tilde{P}, T}^{\text{sc}}(\Sigma) = \min \left( P \left[ |u'(X - T_P)| \leq \sqrt{u' \Sigma u} \right], P \left[ |u'(X - T_P)| \geq \sqrt{u' \Sigma u} \right] \right) \right\}, \\ U_{\text{in}} &= \left\{ u^* \in \mathcal{S}^{p-1} : \min_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = P[A_{\text{in}}(u^*)] \right\}, \\ U_{\text{out}} &= \left\{ u^* \in \mathcal{S}^{p-1} : \min_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)] = P[A_{\text{out}}(u^*)] \right\}, \\ U_{\text{in}}(\delta) &= \{ u \in \mathcal{S}^{p-1} : \exists u^* \in U_{\text{in}} : \|u - u^*\| \leq \delta \}, \\ U_{\text{out}}(\delta) &= \{ u \in \mathcal{S}^{p-1} : \exists u^* \in U_{\text{out}} : \|u - u^*\| \leq \delta \} \end{aligned}$$

ainsi que la fonction  $g : \mathcal{S}^{p-1} \rightarrow \mathbb{R}$  comme

$$g(u) = u' \text{IF}(z; T, P) \left( f_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) - f_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) \right).$$

Le théorème livrant la fonction d'influence de la profondeur de demi-espace scatter est repris ci-dessous.

**Théorème 2.7.** *Soient  $P$  une distribution absolument continue sur  $\mathbb{R}^p$ ,  $\Sigma$  une matrice définie positive et  $z \in \mathbb{R}^p$ . Supposons que  $T_{\tilde{P}}$  est continument différentiable par rapport à  $t$  et que la densité  $f_{u'(X-T_P)}(x)$  est continue par rapport à  $x$  et  $u$ . Considérons de plus que  $f_X(x)$  est bornée et que  $\sup_{u \in \mathcal{S}^{p-1}} f_{u'(X-T_P)}(x) < \infty$ . Soient  $u_{\text{in}}^* \in U_{\text{in}}$  et  $u_{\text{out}}^* \in U_{\text{out}}$  définis comme vérifiant*

$$g(u_{\text{in}}^*) = \min_{u \in U_{\text{in}}} g(u) \quad \text{et} \quad -g(u_{\text{out}}^*) = \min_{u \in U_{\text{out}}} -g(u).$$

Alors, nous avons que

- *si*

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$-HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*) \leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*),$$

- *si*

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$-HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*) \leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*),$$

- *si*

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$\begin{aligned} -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)) &\leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \\ &\leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)). \end{aligned}$$

*Démonstration.* Nous commençons par réécrire la fonction de profondeur comme

$$\begin{aligned} HD_{P,T}^{\text{sc}}(\Sigma) &= \inf_{u \in \mathcal{S}^{p-1}} \min \left( \tilde{P}[\tilde{A}_{\text{in}}(u)], \tilde{P}[\tilde{A}_{\text{out}}(u)] \right) \\ &= \min \left( \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{out}}(u)] \right). \end{aligned} \quad (2.6)$$

Nous pouvons supposer que chaque infimum est atteint pour des directions  $\tilde{u}_{\text{in}}$  et  $\tilde{u}_{\text{out}}$  dépendante de  $t$ , étant donné que

$$\begin{aligned} \min \left( \tilde{P}[\tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \tilde{P}[\tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right) - t^2 &\leq \min \left( \inf \tilde{P}[\tilde{A}_{\text{in}}(u)], \inf \tilde{P}[\tilde{A}_{\text{out}}(u)] \right) \\ &\leq \min \left( \tilde{P}[\tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \tilde{P}[\tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right), \end{aligned}$$

et le terme  $t^2$  disparaît lorsque nous considérons la limite dans la fonction d'influence.

Nous commençons par la borne inférieure. Nous avons que

$$\begin{aligned} &\min \left( \tilde{P}[\tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \tilde{P}[\tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right) \\ &= \min \left( (1-t)(P[A_{\text{in}}(\tilde{u}_{\text{in}})] + P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]) + t I[z \in \tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \right. \\ &\quad \left. (1-t)(P[A_{\text{out}}(\tilde{u}_{\text{out}})] + P[C_t(\tilde{u}_{\text{out}})] - P[B_t(\tilde{u}_{\text{out}})]) + t I[z \in \tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right) \\ &\geq \min \left( (1-t)(P[A_{\text{in}}(\tilde{u}_{\text{in}})] + P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]), \right. \\ &\quad \left. (1-t)(P[A_{\text{out}}(\tilde{u}_{\text{out}})] + P[C_t(\tilde{u}_{\text{out}})] - P[B_t(\tilde{u}_{\text{out}})]) \right) \\ &\geq (1-t)HD_{P,T}^{\text{sc}}(\Sigma) + \min \left( (1-t)(P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]), \right. \\ &\quad \left. (1-t)(P[C_t(\tilde{u}_{\text{out}})] - P[B_t(\tilde{u}_{\text{out}})]) \right) \end{aligned} \quad (2.7)$$

En utilisant la remarque (2.4), nous trouvons

$$\begin{aligned}
& \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \\
&= \lim_{t \rightarrow 0} \frac{HD_{P,T}^{\text{sc}}(\Sigma) - HD_{P,T}^{\text{sc}}(\Sigma)}{t} \\
&\geq -HD_{P,T}^{\text{sc}}(\Sigma) + \lim_{t \rightarrow 0} \min \left( \tilde{u}'_{\text{in}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}'_{\text{in}}} \right) - f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}'_{\text{in}}} \right) \right), \right. \\
&\quad \left. \tilde{u}'_{\text{out}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}'_{\text{out}}} \right) - f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}'_{\text{out}}} \right) \right) \right).
\end{aligned}$$

A présent, nous montrons que les ensembles  $U_{\text{in}}$ ,  $U_{\text{out}}$ ,  $U_{\text{in}}(\delta)$  et  $U_{\text{out}}(\delta)$  sont compacts. Notons qu'il suffit de montrer qu'ils sont fermés car sont inclus dans le compact  $\mathcal{S}^{p-1}$ . L'ensemble  $U_{\text{in}}$  est fermé, car équivalent à l'ensemble des zéros de la fonction continue  $P[A_{\text{in}}(u)] - \inf_{v \in \mathcal{S}^{p-1}} P[A_{\text{in}}(v)]$ . De la même manière,  $U_{\text{out}}$  est fermé. Concernant l'ensemble  $U_{\text{in}}(\delta)$ , montrons que si une suite  $u_n$  appartient à cet ensemble, alors sa limite aussi. Le fait que  $(u_n)_n \subset U_{\text{in}}(\delta)$  implique que

$$\forall n \in \mathbb{N}, \exists u_n^* \in U_{\text{in}} : \|u_n - u_n^*\| \leq \delta.$$

Notons que  $u_n^*$  est une suite dans le compact  $U_{\text{in}}$ , et donc possède une sous-suite  $u_m^*$  convergente vers  $u^* \in U_{\text{in}}$ , car ce dernier est fermé, et vérifiant l'inégalité ci-dessus. La suite  $u_n$  étant convergente, sa limite correspond à la limite de n'importe quelle sous-suite, et donc de  $u_m$ . Nous trouvons donc

$$\lim_{m \rightarrow \infty} \|u_m - u_m^*\| = \left\| \lim_{m \rightarrow \infty} u_m - u^* \right\| \leq \delta,$$

de sorte que  $\lim_{m \rightarrow \infty} u_m = \lim_{n \rightarrow \infty} u_n$  appartient à  $U_{\text{in}}(\delta)$ . De la même manière, l'ensemble  $U_{\text{out}}(\delta)$  est fermé.

Nous voulons montrer que la direction  $\tilde{u}_{\text{in}}$  finit par être arbitrairement proche des directions minimisant  $P[A_{\text{in}}(u)]$ , c'est-à-dire que

$$\forall \delta > 0, \exists t_0 > 0 : t < t_0 \Rightarrow \tilde{u}_{\text{in}} \in U_{\text{in}}(\delta).$$

Supposons par l'absurde que

$$\begin{aligned}
& \exists \delta > 0, \forall t_0 > 0, \exists t < t_0 : \tilde{u}_{\text{in}} \notin U_{\text{in}}(\delta) \\
& \Rightarrow \exists \delta > 0, \forall t_0 > 0, \exists t < t_0 : \forall u^* \in U_{\text{in}}, \|\tilde{u}_{\text{in}} - u^*\| > \delta.
\end{aligned}$$

Ainsi, pour chaque valeur de  $t_0 > 0$ , il existe une direction parmi l'ensemble des directions

$$\{\tilde{u}_{\text{in}} ; 0 \leq t \leq 1\}$$

dépendantes de  $t_0$  qui vérifient l'inégalité. Définissons dès lors la suite de directions  $u_n$ , qui à chaque naturel  $n$  associe la direction qui vérifie l'inégalité pour  $t_0 = 1/n$ . Cette suite définie sur le compact  $\mathcal{S}^{p-1}$  possède une sous-suite  $u_m$  qui converge, et qui satisfait aussi l'inégalité. Ceci implique que la sous-suite  $u_m$  converge vers une direction n'appartenant pas à  $U_{\text{in}}$ . La fonction  $P[A_{\text{in}}(u)]$  étant continue en  $u$ , nous trouvons

$$\lim_{m \rightarrow \infty} P[A_{\text{in}}(u_m)] = P \left[ A_{\text{in}} \left( \lim_{m \rightarrow \infty} u_m \right) \right] \neq \min_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)].$$

Or, nous avons déjà prouvé que

$$\begin{aligned} \lim_{m \rightarrow \infty} P[A_{\text{in}}(u_m)] &= \lim_{t \rightarrow 0} P[A_{\text{in}}(\tilde{u}_{\text{in}})] \\ &\stackrel{(2.5)}{=} \min_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)]. \end{aligned}$$

Nous pouvons utiliser le même argument pour montrer que

$$\forall \delta > 0, \exists t_0 > 0 : t < t_0 \Rightarrow \tilde{u}_{\text{out}} \in U_{\text{out}}(\delta).$$

Ainsi, nous avons  $\forall \delta > 0$

$$\begin{aligned} &\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \\ &\geq -HD_{P,T}^{\text{sc}}(\Sigma) + \lim_{t \rightarrow 0} \min \left( \tilde{u}'_{\text{in}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}_{\text{in}}} \right) - f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}_{\text{in}}} \right) \right), \right. \\ &\quad \left. \tilde{u}'_{\text{out}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}_{\text{out}}} \right) - f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}_{\text{out}}} \right) \right) \right) \\ &\geq -HD_{P,T}^{\text{sc}}(\Sigma) + \min \left( \inf_{u \in U_{\text{in}}(\delta)} u' \text{IF}(z; T, P) \left( f_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) - f_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) \right), \right. \\ &\quad \left. \inf_{u \in U_{\text{out}}(\delta)} u' \text{IF}(z; T, P) \left( f_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) - f_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) \right) \right) \\ &\stackrel{\text{not.}}{=} -HD_{P,T}^{\text{sc}}(\Sigma) + \min \left( \inf_{u \in U_{\text{in}}(\delta)} g(u), \inf_{u \in U_{\text{out}}(\delta)} -g(u) \right) \\ &= -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(\tilde{u}_{\text{in}}(\delta)), -g(\tilde{u}_{\text{out}}(\delta))) \end{aligned}$$

Remarquons que la fonction  $g(u)$  est continue par hypothèse que la densité  $f_{u'(X-T_P)}(x)$  est continue par rapport à  $u$  et  $x$ . En conséquence, il existe une direction  $\tilde{u}_{\text{in}}(\delta)$  qui atteint l'infimum sur ce compact.

De plus, nous avons que  $\bigcap_{\delta > 0} U_{\text{in}}(\delta) = U_{\text{in}}(0) = U_{\text{in}}$ . L'inclusion  $\supset$  est directe. Pour l'inclusion dans l'autre sens, supposons qu'il existe  $x \in \bigcap_{\delta > 0} U_{\text{in}}(\delta)$ , ce qui implique que

$$\forall \delta > 0, \exists u(\delta) \in U_{\text{in}} : \|x - u(\delta)\| \leq \delta.$$

En prenant la limite de chaque coté, nous trouvons

$$\|x - \lim_{\delta \rightarrow 0} u(\delta)\| = 0 \Rightarrow x = \lim_{\delta \rightarrow 0} u(\delta).$$

Nous concluons que  $\lim_{\delta \rightarrow 0} u(\delta) = x \in U_{\text{in}}$  car ce dernier ensemble est fermé.

La fonction  $g(\tilde{u}_{\text{in}}(\delta))$  est bornée et croît lorsque  $\delta$  tend vers 0, étant donné que les ensembles  $U_{\text{in}}(\delta)$  sont emboîtés. Donc, la limite en  $\delta$  existe, ce qui veut dire qu'il existe une constante  $L$  telle que

$$\lim_{\delta \rightarrow 0} g(\tilde{u}_{\text{in}}(\delta)) = L.$$

De plus, la suite  $\tilde{u}_{\text{in}}(\delta)$  est une suite dans le compact  $U_{\text{in}}(\delta)$ , et donc possède une sous-suite  $\tilde{u}_{\text{in}}(\delta_n)$  qui converge. Cette limite appartient à  $U_{\text{in}}$  étant donné que nous avons prouvé que

$\bigcap_{\delta>0} U_{\text{in}}(\delta) = U_{\text{in}}$ . Enfin, comme la limite en  $\delta$  de  $g(\tilde{u}_{\text{in}}(\delta))$  existe, nous pouvons considérer la limite correspondante à la sous-suite  $\tilde{u}_{\text{in}}(\delta_n)$ , ce qui livre

$$\begin{aligned} \lim_{\delta \rightarrow 0} g(\tilde{u}_{\text{in}}(\delta)) &= \lim_{n \rightarrow \infty} g(\tilde{u}_{\text{in}}(\delta_n)) \\ &= g\left(\lim_{n \rightarrow \infty} \tilde{u}_{\text{in}}(\delta_n)\right) \\ &= g(u_{\text{in}}^*), \end{aligned}$$

où  $u_{\text{in}}^* \in U_{\text{in}}$ . Nous trouvons finalement comme borne inférieure

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \geq -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)),$$

où  $u_{\text{out}}^* \in U_{\text{out}}$ . Souvenons-nous que dans le cas où  $\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)]$ , nous pouvons considérer uniquement le premier terme dans la profondeur de demi-espace de scatter sous la distribution perturbée. La borne devient alors

$$\begin{aligned} \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) &= \lim_{t \rightarrow 0} \frac{\inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] - \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)]}{t} \\ &\geq \dots \geq -HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*). \end{aligned}$$

En conclusion pour la borne inférieure, nous avons que, si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \geq -HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*),$$

tandis que si  $\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)]$ , alors

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \geq -HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*).$$

Enfin, dans le cas particulier où  $\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)]$ , nous trouvons

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \geq -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)).$$

Nous n'avons pas beaucoup d'informations sur les directions  $u_{\text{in}}^*$  et  $u_{\text{out}}^*$ . Il est clair que si  $U_{\text{in}}$  et  $U_{\text{out}}$  possèdent un unique élément (au signe près), ces directions sont connues. Dans le cas où ces ensembles possèdent plusieurs éléments, nous pouvons quand même affirmer que la direction  $u_{\text{in}}^*$  par exemple satisfait

$$g(u_{\text{in}}^*) = \min_{u \in U_{\text{in}}} g(u).$$

Ceci découle du fait que, pour tout  $\delta > 0$ ,

$$\inf_{u \in U_{\text{in}}(\delta)} g(u) \leq \min_{u \in U_{\text{in}}} g(u),$$

de sorte que la limite du terme de gauche en  $\delta$  conserve cette inégalité.

Concernant la borne supérieure, nous commençons dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

et nous considérons dès lors uniquement  $\inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)]$ . Tout d'abord, remarquons que, en utilisant l'équidifférentiabilité, nous trouvons

$$\begin{aligned} \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] &\leq \tilde{P}[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] \\ &= (1-t)(P[A_{\text{in}}(u_{\text{in}}^*)] + P[B_t(u_{\text{in}}^*)] - P[C_t(u_{\text{in}}^*)]) + tI[z \in \tilde{A}_{\text{in}}(u_{\text{in}}^*)] \\ &\leq (1-t)(P[A_{\text{in}}(u_{\text{in}}^*)] + tg(u_{\text{in}}^*)) + t. \end{aligned}$$

Comme  $P[A_{\text{in}}(u_{\text{in}}^*)] = HD_{P,T}^{\text{sc}}(\Sigma)$ , la borne supérieure est donnée par

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*).$$

Dans l'autre cas, où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

nous trouvons de la même manière  $1 - HD_{P,T}^{\text{sc}}(\Sigma) + h(\tilde{u}_{\text{in}}^*)$  comme borne supérieure. Enfin, si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

nous avons que  $P[A_{\text{in}}(u_{\text{in}}^*)] = P[A_{\text{out}}(u_{\text{out}}^*)]$ , de sorte que

$$\begin{aligned} &\min \left( \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)], \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{out}}(u)] \right) \\ &\leq \min \left( \tilde{P}[\tilde{A}_{\text{in}}(u_{\text{in}}^*)], \tilde{P}[\tilde{A}_{\text{out}}(u_{\text{out}}^*)] \right) \\ &= \min \left( (1-t)(P[A_{\text{in}}(u_{\text{in}}^*)] + P[B_t(u_{\text{in}}^*)] - P[C_t(u_{\text{in}}^*)]) + tI[z \in \tilde{A}_{\text{in}}(u_{\text{in}}^*)], \right. \\ &\quad \left. (1-t)(P[A_{\text{out}}(u_{\text{out}}^*)] + P[C_t(u_{\text{out}}^*)] - P[B_t(u_{\text{out}}^*)]) + tI[z \in \tilde{A}_{\text{out}}(u_{\text{out}}^*)] \right) \\ &\leq (1-t)HD_{P,T}^{\text{sc}}(\Sigma) + (1-t)t \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)) + t. \end{aligned}$$

La borne devient

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)).$$

En conclusion, nous trouvons comme bornes pour la fonction d'influence

$$-HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*) \leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*)$$

si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)].$$

Dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

nous trouvons

$$-HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*) \leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*).$$

Enfin, si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

les bornes valent

$$-HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)) \leq \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq 1 - HD_{P,T}^{\text{sc}}(\Sigma) + \min(g(u_{\text{in}}^*), -g(u_{\text{out}}^*)).$$

Remarquons que, dans la borne supérieure, nous nous sommes débarrassés rapidement de l'indicatrice. Notons que nous aurions pu la laisser pour obtenir comme borne supérieure, dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)]$$

par exemple,

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq \lim_{t \rightarrow 0} I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] - HD_{P,T}^{\text{sc}}(\Sigma) + g(\tilde{u}_{\text{in}}^*) \quad (2.8)$$

Et donc, si  $\lim_{t \rightarrow 0} I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] = 0$ , nous avons l'égalité entre les deux bornes.  $\square$

Le corollaire suivant montre que, pour certaines valeurs de  $z \in \mathbb{R}^p$ , nous avons l'égalité entre les bornes de la fonction d'influence.

**Corollaire 2.1.** *Soient  $P$  une distribution absolument continue sur  $\mathbb{R}^p$ ,  $\Sigma$  une matrice définie positive et  $z \in \mathbb{R}^p$  tel que, pour tout  $u$  dans  $U_0$ ,  $|u'(z - T_P)| \neq \sqrt{u'\Sigma u}$ . Sous les hypothèses du Théorème 2.7, nous avons que*

- si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) = I[A_{\text{in}}(u_{\text{in}}^*)] - HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*),$$

- si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) = I[A_{\text{out}}(u_{\text{out}}^*)] - HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*),$$

- si

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

alors

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) = -HD_{P,T}^{\text{sc}}(\Sigma) + \min(I[A_{\text{in}}(u_{\text{in}}^*)] + g(u_{\text{in}}^*), I[A_{\text{out}}(u_{\text{out}}^*)] - g(u_{\text{out}}^*)).$$

*Démonstration.* Nous commençons par la borne inférieure. Tout d'abord, nous montrons qu'il existe  $\delta_0 > 0$  tel que, pour tout  $\delta < \delta_0$ , la fonction  $I[A_{\text{in}}(u)]$  est continue sur  $U_{\text{in}}(\delta)$ . Nous définissons pour cela l'ensemble

$$U_{\text{eq}} = \left\{ u \in \mathcal{S}^{p-1} : |u'(z - T_P)| = \sqrt{u'\Sigma u} \right\}.$$

Remarquons que l'hypothèse que  $z$  est tel que, pour tout  $u \in U_{\text{in}}$ ,  $|u'(z - T_P)| \neq \sqrt{u'\Sigma u}$  s'exprime aussi comme  $U_{\text{in}} \cap U_{\text{eq}} = \emptyset$ . De plus, l'ensemble  $U_{\text{eq}}$  contient l'ensemble de discontinuité de la fonction  $I[A_{\text{in}}(u)]$  (pour  $p \geq 2$ ; pour  $p = 1$ , la propriété que nous prouvons est évidemment vérifiée car  $U_{\text{in}}(\delta)$  ne contient qu'un élément). Nous devons donc montrer qu'il existe  $\delta_0 > 0$  tel que, pour tout  $\delta < \delta_0$ ,  $U_{\text{in}}(\delta) \cap U_{\text{eq}} = \emptyset$ . Nous avons déjà montré que  $U_{\text{in}}(\delta)$  est

fermé, et il est évident que  $U_{\text{eq}}$  l'est aussi. Supposons par l'absurde que pour tout  $\delta_0 > 0$ , il existe  $\delta < \delta_0$  tel que  $U_{\text{in}}(\delta) \cap U_{\text{eq}} \neq \emptyset$ . A chaque  $\delta_0$ , on peut donc associer une direction  $u(\delta_0) \in U_{\text{in}}(\delta) \cap U_{\text{eq}} \subset U_{\text{in}}(\delta_0) \cap U_{\text{eq}}$ . Nous définissons dès lors une suite  $u_n$  telle que, pour tout  $n$ ,  $u_n$  est la direction  $u(\delta_0)$  correspondante au  $\delta_0 = 1/n$ . Comme la suite  $u_n$  est définie sur un compact, elle admet une sous-suite  $u_m$  qui converge et qui satisfait aussi  $u_m \in U_{\text{in}}(\delta_0) \cap U_{\text{eq}}$ . Ceci implique que la limite de  $u_m$  pour  $\delta_0$  tendant vers 0 (et donc  $m$  et  $n$  tendant vers l'infini) appartient à  $U_{\text{in}} \cap U_{\text{eq}}$ , ce qui est en contradiction avec l'hypothèse que  $U_{\text{in}} \cap U_{\text{eq}} = \emptyset$ .

La seconde propriété que nous montrons est que, pour tout  $\delta < \delta_0$ , il existe  $t_0 > 0$  tel que pour tout  $t < t_0$ , pour tout  $u \in U_{\text{in}}(\delta)$ ,  $I[B_t(u)] = I[C_t(u)] = 0$ . Nous faisons la preuve pour  $I[B_t(u)]$ . Soit  $\delta < \delta_0$  and supposons par l'absurde que pour tout  $t_0 > 0$ , il existe  $t < t_0$  et  $u(t_0) \in U_{\text{in}}(\delta)$  tel que  $I[B_t(u(t_0))] = 1$ . Cela veut dire que

$$\sqrt{u(t_0)' \Sigma u(t_0)} \leq u(t_0)'(z - T_P) \leq \sqrt{u(t_0)' \Sigma u(t_0)} + u(t_0)'(T_{\tilde{P}} - T_P).$$

En utilisant le même argument que précédemment, nous pouvons définir une suite  $u_n$  qui, à chaque  $n$ , associe la direction  $u(t_0)$  pour  $t_0 = 1/n$ . Cette suite admet une sous-suite  $u_m$  qui satisfait l'inégalité ci-dessus et converge dans  $U_{\text{in}}(\delta)$  vers  $u_*$ . Comme ceci est vrai pour tout  $t_0 > 0$ , et donc pour tout  $n$  et tout  $m$ , nous trouvons, en prenant la limite

$$\sqrt{u_*' \Sigma u_*} \leq u_*'(z - T_P) \leq \sqrt{u_*' \Sigma u_*},$$

ce qui est en contradiction avec le fait que pour tout  $\delta < \delta_0$ ,  $U_{\text{in}}(\delta) \cap U_{\text{eq}} = \emptyset$ .

Nous pouvons à présent obtenir la borne inférieure. Nous trouvons, en utilisant l'expression en (2.7) mais en conservant cette fois-ci les indicatrices,

$$\begin{aligned} HD_{P,T}^{\text{sc}}(\Sigma) &\geq \min \left( \tilde{P}[\tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \tilde{P}[\tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right) \\ &\geq (1-t)HD_{P,T}^{\text{sc}}(\Sigma) \\ &\quad + \min \left( (1-t)(P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]) + t(I[A_{\text{in}}(\tilde{u}_{\text{in}})] + I[B_t(\tilde{u}_{\text{in}})] - I[C_t(\tilde{u}_{\text{in}})]), \right. \\ &\quad \left. (1-t)(P[C_t(\tilde{u}_{\text{out}})] - P[B_t(\tilde{u}_{\text{out}})]) + t(I[A_{\text{in}}(\tilde{u}_{\text{in}})] + I[C_t(\tilde{u}_{\text{in}})] - I[B_t(\tilde{u}_{\text{in}})]) \right). \end{aligned}$$

Soit  $\delta < \delta_0$ . Nous avons montré dans la première partie de la démonstration qu'il existe  $t_0$  suffisamment petit tel que, pour tout  $t < t_0$ ,  $\tilde{u}_{\text{in}} \in U_{\text{in}}(\delta)$ . De plus, nous avons qu'il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ , pour tout  $u \in U_{\text{in}}(\delta)$ ,  $I[B_t(u)] = I[C_t(u)] = 0$ . Donc la borne inférieure devient

$$\begin{aligned} \min \left( \tilde{P}[\tilde{A}_{\text{in}}(\tilde{u}_{\text{in}})], \tilde{P}[\tilde{A}_{\text{out}}(\tilde{u}_{\text{out}})] \right) &\geq (1-t)HD_{P,T}^{\text{sc}}(\Sigma) \\ &\quad + \min \left( (1-t)(P[B_t(\tilde{u}_{\text{in}})] - P[C_t(\tilde{u}_{\text{in}})]) + t I[A_{\text{in}}(\tilde{u}_{\text{in}})], \right. \\ &\quad \left. (1-t)(P[C_t(\tilde{u}_{\text{out}})] - P[B_t(\tilde{u}_{\text{out}})]) + t I[A_{\text{in}}(\tilde{u}_{\text{in}})] \right). \end{aligned}$$

Ainsi, la fonction d'influence est bornée inférieurement par

$$\begin{aligned}
& -HD_{P,T}^{\text{sc}}(\Sigma) \\
& + \lim_{t \rightarrow 0} \min \left( \tilde{u}'_{\text{in}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}'_{\text{in}}} \right) - f_{\tilde{u}'_{\text{in}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{in}} \Sigma \tilde{u}'_{\text{in}}} \right) \right) + I[A_{\text{in}}(\tilde{u}_{\text{in}})] \right. \\
& \quad \left. , \tilde{u}'_{\text{out}} \text{IF}(z; T, P) \left( f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( -\sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}'_{\text{out}}} \right) - f_{\tilde{u}'_{\text{out}}(X-T_P)} \left( \sqrt{\tilde{u}'_{\text{out}} \Sigma \tilde{u}'_{\text{out}}} \right) \right) + I[A_{\text{out}}(\tilde{u}_{\text{out}})] \right) \\
& \geq -HD_{P,T}^{\text{sc}}(\Sigma) \\
& + \min \left( \inf_{u \in U_{\text{in}}(\delta)} u' \text{IF}(z; T, P) \left( f_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) - f_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) \right) + I[A_{\text{in}}(\tilde{u}_{\text{in}})], \right. \\
& \quad \left. \inf_{u \in U_{\text{out}}(\delta)} u' \text{IF}(z; T, P) \left( f_{u'(X-T_P)} \left( -\sqrt{u' \Sigma u} \right) - f_{u'(X-T_P)} \left( \sqrt{u' \Sigma u} \right) \right) + I[A_{\text{out}}(\tilde{u}_{\text{out}})] \right) \\
& \stackrel{\text{not.}}{=} -HD_{P,T}^{\text{sc}}(\Sigma) + \min \left( \inf_{u \in U_{\text{in}}(\delta)} g_2(u), \inf_{u \in U_{\text{out}}(\delta)} h_2(u) \right) \\
& = -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g_2(\tilde{u}_{\text{in}}(\delta)), h_2(\tilde{u}_{\text{out}}(\delta))),
\end{aligned}$$

où comme pour la fonctions  $g$  dans la première partie de la preuve, les fonctions  $g_2$  et  $h_2$  sont continues sur  $U_{\text{in}}(\delta)$ , ce qui était la seule hypothèse sur ces fonctions nécessaire à la suite du raisonnement. Nous trouvons donc comme borne inférieure

$$\begin{aligned}
\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) & \geq -HD_{P,T}^{\text{sc}}(\Sigma) + \min(g_2(u_{\text{in}}^*), h_2(u_{\text{out}}^*)) \\
& \geq -HD_{P,T}^{\text{sc}}(\Sigma) + \min(I[A_{\text{in}}(u_{\text{in}}^*)] + g(u_{\text{in}}^*), I[A_{\text{out}}(u_{\text{out}}^*)] - g(u_{\text{out}}^*)),
\end{aligned}$$

Notons que, comme précédemment, cette borne est nécessaire que dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)].$$

Dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

ou

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

nous pouvons à nouveau considérer uniquement le terme correspondant dans la profondeur sous la distribution perturbée, de sorte à trouver respectivement comme borne

$$-HD_{P,T}^{\text{sc}}(\Sigma) + I[A_{\text{in}}(u_{\text{in}}^*)] + g(u_{\text{in}}^*)$$

et

$$-HD_{P,T}^{\text{sc}}(\Sigma) + I[A_{\text{out}}(u_{\text{out}}^*)] - g(u_{\text{out}}^*).$$

Concernant la borne supérieure, nous commençons par traiter le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] < \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)].$$

Nous avons, en considérant le  $u_{\text{in}}^* \in U_{\text{in}}$  apparaissant dans la borne inférieure, que

$$\begin{aligned}
HD_{P,T}^{\text{sc}}(\Sigma) & = \inf_{u \in \mathcal{S}^{p-1}} \tilde{P}[\tilde{A}_{\text{in}}(u)] \\
& \leq \tilde{P}[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] \\
& \leq (1-t)(P[A_{\text{in}}(u_{\text{in}}^*)] + P[B_t(u_{\text{in}}^*)] - P[C_t(u_{\text{in}}^*)]) + t I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)].
\end{aligned}$$

Nous montrons à présent que

$$\lim_{t \rightarrow 0} I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] = I[A_{\text{in}}(u_{\text{in}}^*)].$$

Si  $|u_{\text{in}}^*(z - T_P)| > \sqrt{u_{\text{in}}^* \Sigma u_{\text{in}}^*}$ , alors  $I[A_{\text{in}}(u_{\text{in}}^*)] = 0$ . Comme  $T_{\tilde{P}} \rightarrow T_P$ , nous trouvons

$$\lim_{t \rightarrow 0} |u_{\text{in}}'(z - T_{\tilde{P}})| = |u_{\text{in}}'(z - T_P)| > \sqrt{u_{\text{in}}^* \Sigma u_{\text{in}}^*}.$$

Par conséquent, il existe  $t_0 > 0$  tel que, pour tout  $t < t_0$ ,  $|u_{\text{in}}'(z - T_{\tilde{P}})| > \sqrt{u_{\text{in}}^* \Sigma u_{\text{in}}^*}$  et donc  $I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] = 0$ . Si  $|u_{\text{in}}'(z - T_P)| < \sqrt{u_{\text{in}}^* \Sigma u_{\text{in}}^*}$ , nous trouvons, en utilisant le même argument, que  $|u_{\text{in}}'(z - T_{\tilde{P}})| < \sqrt{u_{\text{in}}^* \Sigma u_{\text{in}}^*}$  et donc  $I[A_{\text{in}}(u_{\text{in}}^*)] = I[\tilde{A}_{\text{in}}(u_{\text{in}}^*)] = 1$ . Ainsi, la borne supérieure obtenue en (2.8) devient

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq I[A_{\text{in}}(u_{\text{in}}^*)] - HD_{P,T}^{\text{sc}}(\Sigma) + g(u_{\text{in}}^*),$$

Dans le cas

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] > \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

la borne supérieure vaut

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq I[A_{\text{out}}(u_{\text{out}}^*)] - HD_{P,T}^{\text{sc}}(\Sigma) - g(u_{\text{out}}^*),$$

et dans le cas où

$$\inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{in}}(u)] = \inf_{u \in \mathcal{S}^{p-1}} P[A_{\text{out}}(u)],$$

nous trouvons

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) \leq -HD_{P,T}^{\text{sc}}(\Sigma) + \min(I[A_{\text{in}}(u_{\text{in}}^*)] + g(u_{\text{in}}^*), I[A_{\text{out}}(u_{\text{out}}^*)] - g(u_{\text{out}}^*)).$$

□

### 2.3.4 Discussion sur la fonction d'influence

Remarquons tout d'abord que, pour toute valeur de  $z$  à l'intérieur de l'ellipse formée par  $\Sigma$ , nous avons qu'il existe aucune direction  $u$  satisfaisant  $|u'(z - T_P)| = \sqrt{u' \Sigma u}$ . Ainsi, pour ces valeurs nous avons toujours l'égalité pour la fonction d'influence, avec en plus, comme  $z \in A_{\text{in}}(u)$  pour tout  $u$ ,  $I[A_{\text{out}}(u_{\text{out}}^*)] = 0$  et  $I[A_{\text{in}}(u_{\text{in}}^*)] = 1$ . De plus, si  $U_0$  ne possède qu'un nombre fini d'éléments, nous avons l'égalité pour la fonction d'influence, sauf pour

$$z \in \bigcup_{u \in U_0} \left\{ x \in \mathbb{R}^p : |u'(x - T_p)| = \sqrt{u' \Sigma u} \right\},$$

qui est de mesure de Lebesgue nulle.

Dans le cas où il existe  $u_0 \in U_0$  tel que

$$|u_0'(z - T_P)| = \sqrt{u_0' \Sigma u_0},$$

chaque situation doit être analysée séparément, et il est même possible de quand même conclure à l'égalité pour la fonction d'influence. Par exemple, pour  $p = 1$ , nous avons toujours que les indicatrices dans les bornes valent 1. Si nous considérons une distribution uniforme sur  $\mathcal{B}(0, 1)$ ,

$\Sigma = 10I_2$  et comme estimateur de position  $T_P$  la moyenne, qui vaut 0 dans notre cas, nous avons  $U_0 = \mathcal{S}^1$ , car pour tout  $u$ ,

$$HD_{P,T}^{\text{sc}}(\Sigma) = P[A_{\text{out}}(u)] = P[|u'(X - T_P)| > \sqrt{u'\Sigma u}] = 0.$$

Pour  $z$  quelconque, comme  $T_{\tilde{P}}$  converge vers  $T_P$ , nous avons qu'il existe  $t_0 > 0$  tel que pour tout  $t < t_0$ , pour  $u$  orthogonal à  $z - T_P$ ,

$$HD_{\tilde{P},T}^{\text{sc}}(\Sigma) = P[|u'(X - T_{\tilde{P}})| > \sqrt{u'\Sigma u}] = 0,$$

et la fonction indicatrice vaut donc 0 (étant donné que pour ce  $u$ ,  $z \in A_{\text{in}}(u)$ ). Nous obtenons donc que la fonction d'influence est nulle pour tout  $z \in \mathbb{R}^p$ , même s'il existe  $u_0 \in U_0$  tel que  $|u'_0(z - T_P)| = \sqrt{u'_0 \Sigma u_0}$ . C'est pourquoi la condition est suffisante mais pas nécessaire. Le cas où la fonction n'est pas atteinte nécessite plus de discussion pour déterminer la valeur de la fonction d'influence, qui dépendra alors fortement de la distribution considérée. Cependant, même dans ce cas, les bornes obtenues restent valides.

L'expression obtenue dans la section précédente montre d'abord que la fonction d'influence de la profondeur de demi-espace de scatter dépend de la fonction d'influence de l'estimateur de position considéré. Ainsi, si cette dernière est une fonction non bornée, comme c'est le cas pour la moyenne par exemple, la fonction d'influence de la fonction de profondeur est également non bornée. Le cas le plus intéressant est celui où la distribution  $f_X$  est symétrique autour de  $T_P$ , de sorte que chaque distribution  $f_{w'(X - T_P)}$  est aussi symétrique, et le terme faisant intervenir la fonction d'influence de l'estimateur de position disparaît. Autrement dit, si la distribution est symétrique, la fonction d'influence de la profondeur ne dépend pas de l'estimateur de position utilisé. Nous obtenons donc une expression proche de la fonction d'influence de la profondeur de demi-espace de position, qui est bornée entre  $-1$  et  $1$ . Cette dernière valeur correspond donc à la gross error sensitivity de la profondeur de demi-espace scatter, pour une distribution symétrique. Pour une distribution continue quelconque, nous pouvons borner la gross error sensitivity par le maximum, en valeur absolue, des deux valeurs prises par la fonction d'influence.

Comme nous avons obtenu la fonction d'influence, il est possible de déduire la variance asymptotique de la fonction de profondeur de scatter. Le fait que nous ne sachions pas déterminer la valeur exacte de la fonction d'influence pour chaque valeur de  $z \in \mathbb{R}^p$  rend l'évaluation de la variance asymptotique (ASV) plus complexe. Cependant, si nous supposons que le cardinal de  $U_0$  est fini, nous trouvons

$$\begin{aligned} \text{ASV}(HD_{P,T}^{\text{sc}}(\Sigma), F) &= \int_{\mathbb{R}} \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F)^2 dF(z) \\ &= \int_A \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F)^2 + \int_{A^C} \text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F)^2, \end{aligned}$$

où  $A$  est un certain sous espace de  $\mathbb{R}^p$  pour lequel l'indicatrice dans la fonction d'influence vaut 1, de sorte que cette dernière vaut 0 pour les valeurs de  $z$  dans  $A^C$ . L'intérêt de supposer que  $U_0$  est un ensemble fini est que le sous-espace pour lequel nous n'avons pas l'égalité pour la fonction d'influence est de mesure de Lebesgue nulle, de sorte qu'il n'apparaît pas dans l'évaluation de l'ASV. Nous ne développons pas plus l'expression car celle-ci dépend de nombreux cas (quel infimum correspond à la valeur de la profondeur de demi-espace,...). Notons simplement que la fonction  $g$  dans la fonction d'influence de la profondeur dépend de  $z$  au travers de  $\text{IF}(z; T, P)$ . Ainsi, la variance asymptotique de la fonction de profondeur de scatter dépendra de la variance asymptotique de l'estimateur de position choisi. Notons enfin que, même si  $U_0$  est infini, nous pouvons toujours borner l'ASV en prenant la valeur maximum parmi les bornes de la fonction d'influence.

### 2.3.5 Fonction d'influence empirique

Concernant la fonction d'influence empirique, il est clair que, comme pour une distribution discrète, le résultat va dépendre fortement du comportement de  $T_P^{(i)}$ . Si ce dernier est différent de  $T_P$ , nous ne pouvons de manière générale quantifier le changement de  $HD_{P,T}^{\text{sc}}(\Sigma)$  à  $HD_{P,T}^{\text{sc},(i)}(\Sigma)$ . Par exemple, si nous supposons que toutes les données  $X_i$  se trouvent autour ou à l'intérieur de  $\mathcal{S}^{p-1}$ , avec  $\Sigma = \alpha I_p$  et  $\alpha$  suffisamment grand, nous avons, de manière analogue au dernier exemple de la section précédente,  $HD_{P,T}^{\text{sc}}(\Sigma) = HD_{P,T}^{\text{sc},(i)}(\Sigma) = 0$ . Nous trouvons donc, pour chaque  $X_i$ ,

$$SI(X_i; HD_{P,T}^{\text{sc}}(\Sigma)) = 0.$$

Si nous considérons quatre données formant un triangle (voir FIGURE 4, panel de gauche), nous avons que la profondeur vaut 0.5. Cependant, si nous retirons une donnée, nous observons sur le panel de droite que  $T_P^{(i)}$  a beaucoup bougé, de sorte que nous pouvons trouver une direction telle que

$$HD_{P,T}^{\text{sc},(i)}(\Sigma) = P[|u'(X - T_P^{(i)})| \geq 1] = 0.$$

Ainsi, nous voyons que la profondeur est passée de 0.5, soit la plus grande valeur qu'elle peut prendre, à 0, la plus petite. Si par contre nous supposons que  $T_P$  est connu et fixé à 0, alors nous pouvons obtenir la valeur de la fonction d'influence empirique. La démonstration va être similaire à celle proposée pour la fonction d'influence de la profondeur scatter avec l'estimateur de position fixé, nous introduisons dès lors les ensembles  $U_{\text{in}}^{X_i}$ ,  $U_{\text{in}}^{\overline{X_i}}$ ,  $U_{\text{out}}^{X_i}$  et  $U_{\text{out}}^{\overline{X_i}}$  comme

$$\begin{aligned} U_{\text{in}}^{X_i} &= \{u \in \mathcal{S}^{p-1} : X_i \in A_{\text{in}}(u)\}, \\ U_{\text{in}}^{\overline{X_i}} &= \{u \in \mathcal{S}^{p-1} : X_i \notin A_{\text{in}}(u)\}, \\ U_{\text{out}}^{X_i} &= \{u \in \mathcal{S}^{p-1} : X_i \in A_{\text{out}}(u)\}, \\ U_{\text{out}}^{\overline{X_i}} &= \{u \in \mathcal{S}^{p-1} : X_i \notin A_{\text{out}}(u)\}. \end{aligned}$$

**Théorème 2.8.** Soit  $\{X_1, \dots, X_n\}$  un échantillon issu d'une distribution  $F$  et soit  $\Sigma$  une matrice définie positive. Notons

$$A = \min \left( \min_{U_{\text{in}}^{X_i}} P[A_{\text{in}}(u)], \min_{U_{\text{out}}^{\overline{X_i}}} P[A_{\text{out}}(u)] \right)$$

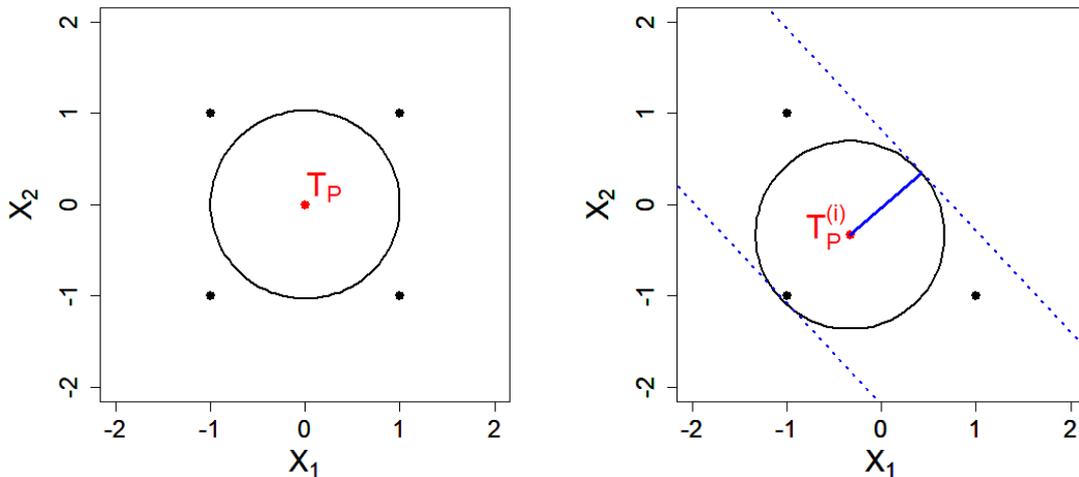


FIGURE 4 – Illustration de la variation de profondeur lorsqu'une donnée est retirée.

et

$$B = \min \left( \min_{U_{\text{in}}^{X_i}} P[A_{\text{in}}(u)], \min_{U_{\text{out}}^{X_i}} P[A_{\text{out}}(u)] \right).$$

La fonction d'influence empirique de la profondeur de scatter est donnée par

$$\text{IF}(z; HD_{P,T}^{\text{sc}}(\Sigma), F) = \begin{cases} 1 - HD_{P,T}^{\text{sc}}(\Sigma) & \text{si } A \geq B \\ -HD_{P,T}^{\text{sc}}(\Sigma) & \text{si } A < B. \end{cases}$$

*Démonstration.* Nous commençons par écrire la profondeur comme

$$HD_{P,T}^{\text{sc}}(\Sigma) = \min \left( \min_{U_{\text{in}}^{X_i}} P[A_{\text{in}}(u)], \min_{\bar{U}_{\text{in}}^{X_i}} P[A_{\text{in}}(u)], \min_{U_{\text{out}}^{X_i}} P[A_{\text{out}}(u)], \min_{\bar{U}_{\text{out}}^{X_i}} P[A_{\text{out}}(u)] \right).$$

Supposons que

$$\min_{U_{\text{in}}^{X_i}} P[A_{\text{in}}(u)] = \frac{j}{n}, \quad \min_{\bar{U}_{\text{in}}^{X_i}} P[A_{\text{in}}(u)] = \frac{k}{n}, \quad \min_{U_{\text{out}}^{X_i}} P[A_{\text{out}}(u)] = \frac{l}{n}, \quad \min_{\bar{U}_{\text{out}}^{X_i}} P[A_{\text{out}}(u)] = \frac{m}{n},$$

de sorte que

$$HD_{P,T}^{\text{sc}}(\Sigma) = \frac{1}{n} \min(j, k, l, m).$$

Lorsque nous retirons  $X_i$ , nous trouvons

$$HD_{P,T}^{\text{sc},(i)}(\Sigma) = \frac{1}{n-1} \min(j-1, k, l-1, m).$$

Si  $j$  ou  $l$  correspond à la valeur de la profondeur, c'est-à-dire si

$$\min(j, l) \leq \min(k, m),$$

alors

$$HD_{P,T}^{\text{sc},(i)}(\Sigma) = \frac{\min(j, l) - 1}{n-1},$$

et donc

$$\begin{aligned} SI(X_i; HD_{P,T}^{\text{sc}}(\Sigma)) &= -(n-1) \left( \frac{\min(j, l) - 1}{n-1} - \frac{\min(j, l)}{n} \right) \\ &= 1 - HD_{P,T}^{\text{sc}}(\Sigma). \end{aligned}$$

Cependant, si

$$\min(k, m) < \min(j, l),$$

alors

$$HD_{P,T}^{\text{sc},(i)}(\Sigma) = \frac{\min(k, m)}{n-1},$$

et donc

$$\begin{aligned} SI(X_i; HD_{P,T}^{\text{sc}}(\Sigma)) &= -(n-1) \left( \frac{\min(k, m)}{n-1} - \frac{\min(k, m)}{n} \right) \\ &= -HD_{P,T}^{\text{sc}}(\Sigma). \end{aligned}$$

□

Nous constatons que l'expression obtenue est semblable à celle que nous avons dérivée pour la fonction d'influence empirique de la profondeur de demi-espace. Notons enfin que, dans le cas où le paramètre de position n'est pas fixé, nous aurions une fonction d'influence empirique qui prendrait valeur dans

$$\{\dots, -1 - HD_{P,T}^{\text{sc}}(\Sigma), -HD_{P,T}^{\text{sc}}(\Sigma), 1 - HD_{P,T}^{\text{sc}}(\Sigma), \dots\}. \quad (2.9)$$

Ainsi, même si nous arrivions à montrer que, pour un nombre de données  $n$  qui tend vers l'infini, les deux valeurs possibles sont  $-HD_{P,T}^{\text{sc}}(\Sigma)$  et  $1 - HD_{P,T}^{\text{sc}}(\Sigma)$ , le fait que les termes faisant intervenir les fonctions de densité n'apparaissent pas montre que la fonction d'influence empirique ne converge pas vers la fonction d'influence (sauf si la distribution considérée est symétrique).

## 2.4 Conclusion

Dans ce second chapitre, nous avons introduit la notion de fonction de profondeur, en définissant les propriétés qu'une telle fonction doit satisfaire. Nous avons défini les fonctions de profondeur de demi-espace ainsi que de scatter. Pour la première, nous avons constaté qu'elle est robuste car sa fonction d'influence est bornée. Pour la seconde, nous avons démontré que, pour une distribution discrète, la fonction d'influence pouvait ne pas exister, et donc qu'il était difficile d'obtenir une expression générale. Pour une distribution absolument continue par rapport à la mesure de Lebesgue, nous avons obtenu des bornes de la fonction d'influence. Ainsi, nous avons constaté que ces dernières dépendent de la fonction d'influence de l'estimateur de position utilisé, mais surtout que si la distribution est symétrique, le choix de l'estimateur de position n'influence pas le résultat. Si la distribution n'est pas symétrique et que la fonction d'influence de l'estimateur de position est non bornée, alors la fonction d'influence de la profondeur scatter ne l'est pas non plus. Nous avons aussi constaté que le résultat dépend de la fonction de densité, qui peut dès lors prendre une valeur significative et ainsi accroître la borne.



## Chapter 3

# Robustness of graphical lasso

In this last chapter, we talk about the graphical lasso (glasso), a method for estimating sparse graphical models. We start with a brief introduction of this method and the estimator employed. Then, we derive its influence function, its asymptotic variance and we give an upper bound of its gross error sensitivity. Afterwards, we discuss the directions maximising the norm of the influence function. We conclude this chapter by introducing robust covariance matrix estimators based on pairwise correlations and we illustrate numerically the results obtained in the different sections.

### 3.1 Introduction

Suppose we have an undirected graph where each vertex represents a variable, and the absence of an edge between two vertices means that the corresponding variables are conditionally independent, given the other variables. In order to understand the relationships between these variables, we need to estimate the precision matrix, which correspond to the inverse of the covariance matrix (the link between this matrix and the conditional dependence is illustrated in the next section). Estimating graphical models with many more parameters than observations is challenging and led to an active area of research on the estimation of high-dimensional inverse covariance matrices. Indeed, the problem which appears when estimating the precision matrix is that, when the dimension  $p$  is bigger than the number of data points  $n$ , the sample covariance matrix is not invertible. Several methods attempt to curb the curse of dimensionality through regularised estimation procedures. The objective is to incorporate sparsity in the estimation, which corresponds to networks with few edges. A common method for estimating sparse graphical models is the graphical lasso ([18], [19], [20], [21]). The estimator is obtained via a convex penalisation method, which involves an  $l_1$ -penalty and the negative log-likelihood of a sample of multivariate normal random variables. It is given by

$$T_{\Omega} = \underset{\Omega}{\operatorname{argmin}} \left\{ -\log |\Omega| + \operatorname{tr}(S\Omega) + \lambda \mathcal{P}(\Omega) \text{ s.t. } \Omega = \Omega^T, \Omega \succ 0 \right\}, \quad (3.1)$$

where  $S$  denotes the empirical covariance matrix.

Various studies have been conducted on the properties of this estimator. For example, for any positive semi-definite estimator of the covariance matrix and  $\lambda > 0$ , [21] proved that the solution of the expression above is unique and positive definite, even if  $p > n$ . Looking at the robustness properties of the glasso, [22] and [23] showed that the estimator is not robust in presence of cellwise outliers. Therefore, in order to obtain a robust estimate of the precision matrix, they proposed a plug-in approach, using a robust covariance matrix estimator in equation

(3.3). There are several robust estimators of the covariance matrix, but their computation is very time-consuming. To overcome this problem, they used pairwise robust procedures, which will be introduced in a next section. In the paper [23] but also in [24], they provide results concerning the breakdown point of the estimator.

The main objective of this chapter is to evaluate the influence function of the glasso. Remember that the influence function is defined, for any estimator  $T(F)$ , as

$$\text{IF}(z; T, F) = \lim_{t \rightarrow 0^+} \frac{T(F_{t,z}) - T(F)}{t},$$

where  $F_{t,z} = (1-t)F + t\delta_z$ . Note that we will sometimes use the equivalent definition

$$\text{IF}(z; T, F) = \left[ \frac{dT(F_{t,z})}{dt} \right]_{t=0}.$$

This tool is often used in robust statistics to evaluate the influence on an estimator of an infinitesimal contamination in the distribution. We will also evaluate the gross error sensitivity, given by

$$\gamma^*(T, F) = \sup_{z \in \mathbb{R}^p} \|\text{IF}(z; T, F)\|.$$

This robustness tool evaluates the worst possible contamination. At last, we will also use the formula

$$\text{ASV}(T, F) = \int \text{IF}(z; T, F) \text{IF}(z; T, F)^T dF(z).$$

to obtain the asymptotic variance.

## 3.2 Illustration of conditional dependence

The reconstruction of a network requires the estimation of a precision matrix we denote  $\Omega$ . We introduce in this first section the notion of conditional independence. Let  $S$  denote the empirical covariance matrix, defined by

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T,$$

where  $n$  is the number of observations  $X_i$  in  $\mathbb{R}^p$ , and  $\bar{X}$  denotes the mean, defined by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

We consider only random vectors of a normal distribution, which depends on a parameter of location denoted  $\mu$  and a covariance matrix which we denote by  $\Sigma$ . We will denote this distribution by  $N(\mu, \Sigma)$ . As a reminder, the covariance matrix is defined, for a random vector  $X = (X_1, X_2, X_3)$  for example, as

$$\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \text{Cov}(X_2, X_3) \\ \text{Cov}(X_1, X_3) & \text{Cov}(X_2, X_3) & \text{Var}(X_3) \end{pmatrix}, \quad (3.2)$$

with  $\text{Var}(X_i)$  representing the variance of the random variable  $X_i$  and  $\text{Cov}(X_i, X_j)$  the covariance between the variables  $X_i$  and  $X_j$ . Note that if  $X_i$  and  $X_j$  are independent, the covariance

between them is zero. Our objective is to estimate  $\Omega = \Sigma^{-1}$  because, for a Gaussian model (in other words when the random variables follow a normal distribution), we have  $\Omega_{i,j} = 0$  if and only if the variables  $X_i$  and  $X_j$  are independent conditionally to the other variables.

To illustrate the concept of conditional independence, let us consider a Markov chain. The latter represents in a way a random walk, where for each following step, we advance by a value obtained via a  $N(0, 1)$  distribution. Formally, we write the chain  $X_1, X_2, \dots$  such that the distribution of  $X_{i+1}|X_i \sim N(X_i, 1)$ . The notation  $X_{i+1}|X_i$  represents the conditional distribution, i.e. the value  $X_{i+1}$  can take supposing we know the value of  $X_i$ . In order to obtain such a chain, we initialise  $x_1$  by a draw obtained from a  $N(0, 1)$  distribution and then  $x_i$  will be obtained by taking a draw from a  $N(x_{i-1}, 1)$  distribution (see FIGURE 1).

In the following, we rewrite this chain by considering  $n$  random variables  $Z_1, \dots, Z_n \sim N(0, 1)$  so that

$$\begin{aligned} X_1 &= Z_1 \\ X_2 &= X_1 + Z_2 = Z_1 + Z_2 \\ X_3 &= X_2 + Z_3 = Z_1 + Z_2 + Z_3 \\ &\dots \\ X_n &= \sum_{i=1}^n Z_i. \end{aligned}$$

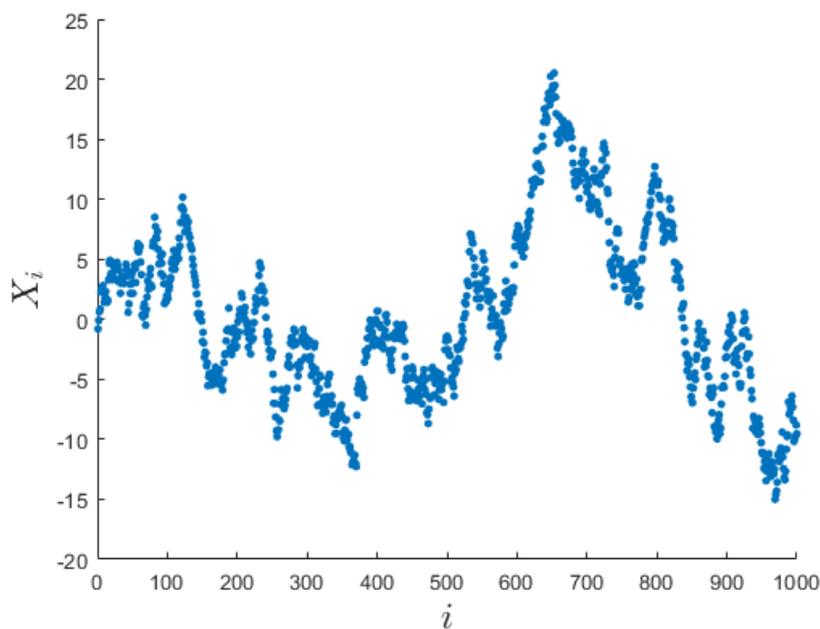


Figure 1 – Example of a Markov chain.

We can therefore write  $X = AZ$ , where  $X = (X_1, \dots, X_n)^T$ ,  $Z = (Z_1, \dots, Z_n)^T$  and

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 1 & 0 & \dots & 0 \\ & & & \vdots & & & \\ 1 & 1 & 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

We have that  $X$  is a multivariate normally distributed random variable. To obtain its covariance matrix  $\Sigma$ , we use the formula

$$\text{Cov}(AZ) = A \text{Cov}(Z) A^T.$$

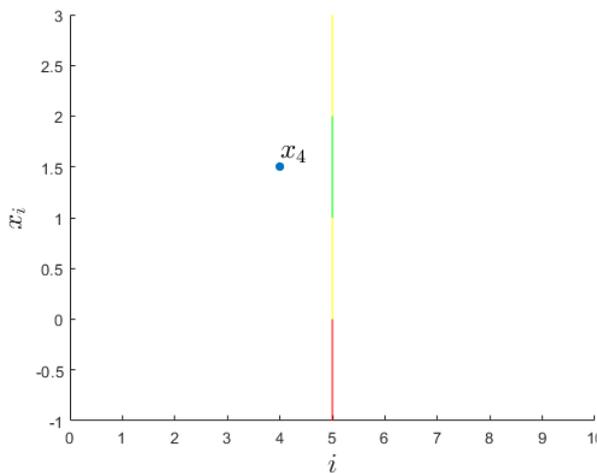
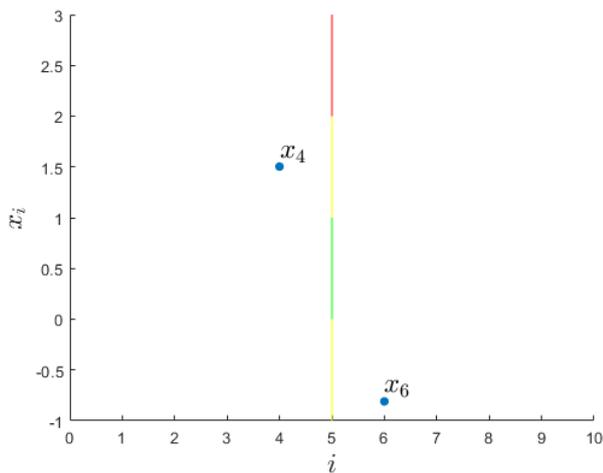
In our case, as  $Z$  is a random vector consisting of independent random variables of variance equal to 1, by the definition (3.2), we have that  $\text{Cov}(Z) = I_p$ , the identity matrix of dimension  $p \times p$ . Thus, if we restrict ourselves to  $n = p = 10$ , we find

$$\Sigma = \text{Cov}(X) = AA^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 & 4 & 4 & 4 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 & 5 & 5 & 5 & 5 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 & 6 & 6 & 6 & 6 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 7 & 7 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 8 & 8 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{pmatrix}.$$

In a graphical model analysis, the matrix of interest is the precision matrix, which is the inverse of  $\Sigma$ . In our case, we find

$$\Omega = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}.$$

Note that the matrix is sparse, which means that there are many entries equal to 0. To illustrate the relationship between this matrix and the conditional dependence, suppose we want to estimate the value  $x_5$  of a Markov chain, and we want to find what other values might help us achieve this goal. It is obvious that knowing the value of  $x_4$  is very important since  $x_5$  is obtained by considering  $x_4 + z_4$ , with  $z_4$  the result of  $Z_4 \sim N(0, 1)$ . The vertical line  $x = 5$  at the FIGURE 2 represents the set of values that  $x_5$  can take from the moment we know  $x_4$ . We have represented different intervals in different colours, from green, which represents the most probable values, given that  $x_5$  comes from a normal distribution centred in  $x_4$ , to red, which represents the least probable values. Due to the normal distribution being symmetrical around

Figure 2 – Estimation of  $x_5$  based on  $x_4$ .Figure 3 – Estimation  $x_5$  based on  $x_4$  on  $x_6$ .

its mean, we have the same colours on each side of  $x_4$ . There is therefore at this moment as much chance that  $x_5$  is between 2 and 3 as there is that it is between 0 and 1 (the two intervals in yellow). Now suppose we also know the value of  $x_6$ . Intuitively, we can adapt the colours of our vertical line (see FIGURE 3). Indeed, in view of the position of  $x_6$ , which is nothing other than  $x_5 + z_5$ , with  $z_5$  the result of  $Z_5 \sim N(0, 1)$ , it seems much more likely that  $x_5$  is between 0 and 1 (in green) rather than between 2 and 3 (in red). Knowing  $x_6$  therefore brings even more precision to the estimation of  $x_5$ . Finally, we can easily convince ourselves that knowing the values of  $x_1$  to  $x_3$  or  $x_7$  to  $x_{10}$  will not improve the estimate of  $x_5$ . We can actually prove that the conditional distribution of  $X_i$ , knowing  $X_j$ ,  $j \neq i$ , depends only on  $X_{i-1}$  and  $X_{i+1}$ . Indeed, we have for  $i = 5$  for example,

$$X_5 | X_1, \dots, X_4, X_6, \dots, X_{10} = X_5 | X_4, X_6 \sim N((X_4 + X_6)/2, 1/2).$$

We will thus say that the variable  $X_i$  is, conditionally to  $X_{i-1}$  and  $X_{i+1}$ , independent of the variables  $X_s$ ,  $s = 1, 2, \dots, i-2, i+2, \dots, p$ . The  $\Omega$  matrix provides this information directly. We observe that in line 5, only the values 4 5 and 6 are non-zero.

### 3.3 Estimation of the precision matrix

To sparsely estimate the matrix  $\Omega$ , it is common to use a convex penalisation method. The estimator is given by

$$T_\Omega = \underset{\Omega}{\operatorname{argmin}} \left\{ -\log |\Omega| + \operatorname{tr}(S\Omega) + \lambda \mathcal{P}(\Omega) \text{ s.t. } \Omega = \Omega^T, \Omega \succ 0 \right\}, \quad (3.3)$$

where  $|\cdot|$  and  $\operatorname{tr}(\cdot)$  denote respectively the determinant and the trace,  $\mathcal{P}(\cdot)$  is a penalty function and  $\lambda$  controls the intensity of this penalty. The problem is solved under the constraints that the matrix is symmetric and positive definite. To better understand where this expression comes from, let us recall what a maximum likelihood estimator is.

Suppose we want to estimate a parameter (we will consider for the illustration that we want to estimate  $\mu$ ) from a sample  $\{x_1, \dots, x_n\}$  of dimension  $p = 1$ . To do this, we will maximise the likelihood, given by

$$L(x_1, \dots, x_n; \mu) = \prod_{i=1}^n f(x_i; \mu),$$

with  $f$  the density of a normal distribution. We often consider the logarithm of the above expression, which does not change the maximisation problem. Why is it interesting to maximise this expression? It actually represents the match between the model considered, depending on the parameter to be estimated, and our data. The idea is to find the parameter that will ensure that most of the data are associated with a high density. For example, let us consider the FIGURE 4 where a sample from a  $N(0, 1)$  distribution is shown in red. We thus seek to estimate  $\mu$ , whose true value, used to generate the data, is in our case 0. We observe that if we try  $\tilde{\mu} = 2$ , the represented density  $f(x; 2)$  seems to disagree with the data, so that when we calculate the logarithm of  $L$ , we find  $-27.0039$ . If we now try  $\tilde{\mu} = 1$  in the figure on the right, we observe that a number of the data start to have a significant density, which delivers  $\log(L) = -14.0628$ . Finally, if we take  $\tilde{\mu} = 0$ , which is the true value of the parameter, we find  $\log(L) = -11.1216$ .

In order to obtain the maximum likelihood estimator for  $\Omega$ , we will consider without loss of generality that  $\mu = 0$ . The expression to be maximised is given by

$$\begin{aligned} L(X; \Omega) &= \prod_{i=1}^n f(X_i; \Omega) \\ &= \prod_{i=1}^n \left[ \left( \frac{1}{2\pi} \right)^{\frac{p}{2}} \frac{1}{|\Omega^{-1}|^{\frac{1}{2}}} \exp \left( -\frac{1}{2} X_i^T \Omega X_i \right) \right]. \end{aligned}$$

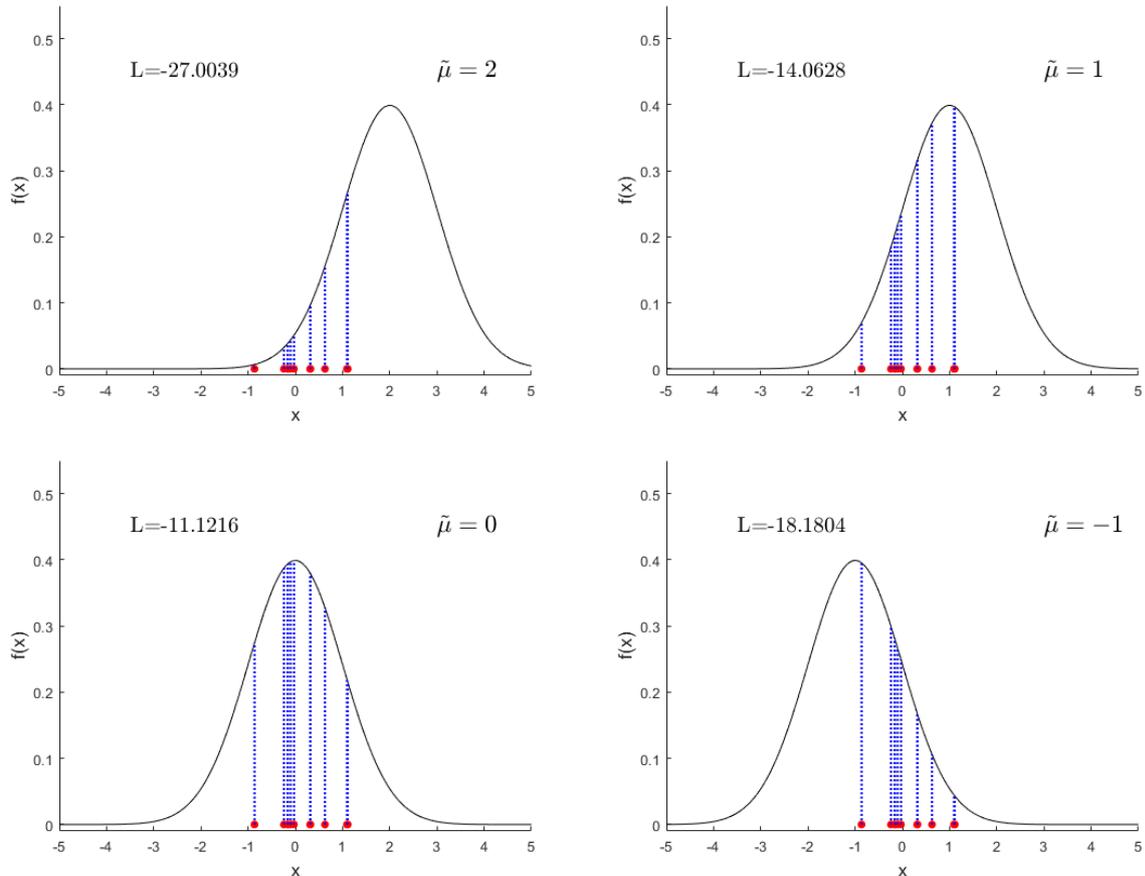


Figure 4 – Toy illustration of the maximum likelihood estimator for the mean  $\mu$ .

The notation  $|\Omega^{-1}|$  represents the determinant of the matrix  $\Omega^{-1}$ . Taking the logarithm, we find

$$\log L(X; \Omega) = C - \frac{n}{2} \log |\Omega^{-1}| - \frac{1}{2} \sum_{i=1}^n (X_i^T \Omega X_i),$$

where  $C$  is a constant independent of  $\Omega$ . Remember that, for any matrix  $A, B$  and  $C$  of suitable dimension, we have  $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$ . Thus, we find

$$\sum_{i=1}^n (X_i^T \Omega X_i) = \sum_{i=1}^n \text{tr} (X_i^T \Omega X_i) \quad (3.4)$$

$$\begin{aligned} &= \sum_{i=1}^n \text{tr} (\Omega X_i X_i^T) \\ &= \text{tr} \left[ \Omega \sum_{i=1}^n (X_i X_i^T) \right] \\ &= n \text{tr} (\Omega S). \end{aligned} \quad (3.5)$$

We have used several properties of the trace, such as the fact that the trace of a number is the number itself (3), or that the trace is linear (4). We therefore find

$$\log L(X; \Omega) = C - \frac{n}{2} \log |\Omega^{-1}| - \frac{n}{2} \text{tr} (\Omega S).$$

Finally, it only remains to use the properties of the determinant and the logarithm to obtain

$$-\frac{n}{2} \log |\Omega^{-1}| = -\frac{n}{2} \log |\Omega|^{-1} = \frac{n}{2} \log |\Omega|.$$

The expression to be maximised becomes

$$\log L(X; \Omega) = C + \frac{n}{2} (\log |\Omega| - \text{tr} (\Omega S)).$$

In order to find the first part of the expression (3.3), it is enough to remove the constants  $C$  and  $\frac{n}{2}$  which do not vary the optimisation problem, and to change the maximum into a minimum of the opposite expression.

The term  $\lambda \mathcal{P}(\Omega)$  represents a penalty term which allows in a certain way to add constraints on the estimator. For example, if we take the function

$$\mathcal{P}(\Omega) = \|\Omega^{-\text{diag}}\|_1,$$

where  $\Omega^{-\text{diag}}$  represents a vector containing the off-diagonal elements of the matrix, we obtain the graphical lasso. We can also write

$$\|\Omega^{-\text{diag}}\|_1 = \sum_{i \neq j} |\Omega_{i,j}|.$$

In order for  $T_\Omega$  to minimise this penalty, it must be sparse, which corresponds to a network with few edges. The tuning parameter  $\lambda$  judges the importance of the penalty. If it is equal to 0, the classical MLE is obtained. For small values of  $\lambda$ , few entries in  $\Omega$  will be estimated as 0, and if  $\lambda$  is large, many entries will be estimated as 0.

### 3.4 Influence function of graphical lasso

In this section, we talk about the main objective of our report, which is evaluating the influence function of the graphical lasso. We first need to consider our estimator as a functional. We then take

$$T_{\Omega}(F) = \operatorname{argmin}_{\Omega} \{-\log|\Omega| + \operatorname{tr}(S(F)\Omega) + \lambda\|\Omega^{-\text{diag}}\|_1 \text{ s.t. } \Omega = \Omega^T, \Omega \succ 0\}, \quad (3.6)$$

where  $F$  is the distribution and  $S(F)$  is the classical covariance matrix estimator defined by

$$S(F) = E_F[XX^T].$$

Note that we will always consider that our functionals are continuous w.r.t.  $t$ . If we do not have this condition, we would not be able to compute the influence function, since the derivative of the functional would not exist.

In this section, we start by evaluating the influence function of the MLE of the covariance matrix  $\Sigma$ . We then use a similar procedure to obtain the MLE of the precision matrix  $\Omega$ , which corresponds to the graphical lasso when considering  $\lambda = 0$ . Afterwards, we give the proof of the influence function of the graphical lasso for  $\lambda > 0$ . We conclude this section with some remarks on the asymptotic variance.

#### 3.4.1 Influence function of MLE of $\Sigma$

Before evaluating the influence function of the graphical lasso, we start with the MLE of  $\Sigma$ . We first check that the method we use with matrix differentiation is consistent with the results already known. In other words, we prove the influence function of the empirical covariance matrix in two different ways and show that the result is the same. The influence function of  $S(F)$  is equal to

$$\text{IF}(z; S, F) = zz^T - S(F).$$

A first method to get this expression is to use the fact that the expected value is linear w.r.t. the distribution, which means

$$E_{F_{t,z}}[XX^T] = (1-t)E_F[XX^T] + tzz^T,$$

and therefore

$$\begin{aligned} \text{IF}(z; S, F) &= \lim_{t \rightarrow 0} \frac{(1-t)E_F[XX^T] + tzz^T - E_F[XX^T]}{t} \\ &= zz^T - S(F). \end{aligned}$$

For the second method, we slightly change the expression (3.6) to get the MLE of the covariance matrix. We find

$$\begin{aligned} T_{\Sigma}(F) &= \operatorname{argmin}_{\Sigma} \{-\log|\Sigma^{-1}| + \operatorname{tr}(S(F)\Sigma^{-1}) \text{ s.t. } \Sigma = \Sigma^T, \Sigma \succ 0\} \\ &= \operatorname{argmin}_{\Sigma} \{\log|\Sigma| + \operatorname{tr}(S(F)\Sigma^{-1}) \text{ s.t. } \Sigma = \Sigma^T, \Sigma \succ 0\}. \end{aligned}$$

It is well known that the MLE is given by  $S(F)$ , so the solution of the above minimisation problem must return  $S(F)$ . We get this result using matrix differentiation. First, we write

$$\begin{aligned} T_{\Sigma}(F) &= \operatorname{argmin}_{\Sigma} \{\log|\Sigma| + \operatorname{tr}(S(F)\Sigma^{-1}) \text{ s.t. } \Sigma = \Sigma^T, \Sigma \succ 0\} \\ &= \operatorname{argmin}_{\Sigma} g(\Sigma). \end{aligned}$$

If we look at  $dg(\Sigma)/d\Sigma$ , we find

$$\begin{aligned}\frac{d}{d\Sigma}g(\Sigma) &= \frac{d}{d\Sigma}\log|\Sigma| + \frac{d}{d\Sigma}\text{tr}(S(F)\Sigma^{-1}) \\ &= \Sigma^{-1} - \Sigma^{-1}S(F)\Sigma^{-1}.\end{aligned}$$

The first order condition (F.O.C.) is given by

$$T_{\Sigma}(F)^{-1} - T_{\Sigma}(F)^{-1}S(F)T_{\Sigma}(F)^{-1} = 0,$$

which gives

$$\begin{aligned}I_p - T_{\Sigma}(F)^{-1}S(F) &= 0 \\ \Leftrightarrow T_{\Sigma}(F) &= S(F).\end{aligned}$$

In order to derive the influence function, we need  $T_{\Sigma}(F_{t,z})$ , which is given by

$$\begin{aligned}T_{\Sigma}(F_{t,z}) &= \underset{\Sigma}{\text{argmin}} \{ \log|\Sigma| + \text{tr}(S(F_{t,z})\Sigma^{-1}) \text{ s.t. } \Sigma = \Sigma^T, \Sigma \succ 0 \} \\ &= \underset{\Sigma}{\text{argmin}} g(t, \Sigma).\end{aligned}$$

The first order condition becomes

$$T_{\Sigma}(F_{t,z}) = S(F_{t,z}),$$

and so

$$\text{IF}(z; T_{\Sigma}, F) = \text{IF}(z; S, F) = zz^T - S(F).$$

Note that, in this case, it is clear that the functional is continuous w.r.t  $t$  since

$$S(F_{t,z}) = (1-t)S(F) + tzz^T.$$

### 3.4.2 Influence function of MLE of $\Omega$

We now consider the influence function of the graphical lasso when  $\lambda = 0$ . The estimator under the contaminated distribution is given by

$$\begin{aligned}T_{\Omega}(F_{t,z}) &= \underset{\Omega}{\text{argmin}} \{ -\log|\Omega| + \text{tr}(S(F_{t,z})\Omega) \text{ s.t. } \Omega = \Omega^T, \Omega \succ 0 \} \\ &= \underset{\Omega}{\text{argmin}} g(t, \Omega).\end{aligned}$$

To get the F.O.C., we derive the function to be minimised w.r.t.  $\Omega$ . We find

$$\frac{d}{d\Omega} g(t, \Omega) = -\Omega^{-1}(t) + S(F_{t,z}) = 0.$$

The F.O.C. becomes

$$T_{\Omega}(F_{t,z}) = S(F_{t,z})^{-1}.$$

The influence function of the MLE equals

$$\begin{aligned}\text{IF}(z; T_{\Omega}, F) &= \left[ \frac{d}{dt} T_{\Omega}(F_{t,z}) \right]_{t=0} \\ &= \left[ \frac{d}{dt} S(F_{t,z})^{-1} \right]_{t=0} \\ &= -S(F)^{-1} \text{IF}(z; S, F) S(F)^{-1} \\ &= -T_{\Omega}(F) \text{IF}(z; S, F) T_{\Omega}(F).\end{aligned}$$

It is not surprising that the influence function of the covariance matrix estimator put in the definition of the graphical lasso appears in the influence function of the glasso itself. Note that  $S(F_{t,z})^{-1}$  does not exist if  $p > n$ , which is one of the reasons why the graphical lasso has been introduced.

### 3.4.3 Influence function of glasso for $\lambda > 0$

We derive the influence function of the graphical lasso with  $L_1$  penalisation. We write

$$\begin{aligned} T_\Omega(F_{t,z}) &= \underset{\Omega}{\operatorname{argmin}} \left\{ -\log|\Omega| + \operatorname{tr}(S(F_{t,z})\Omega) + \lambda \|\Omega^{-\operatorname{diag}}\|_1 \text{ s.t. } \Omega = \Omega^T, \Omega \succ 0 \right\} \\ &= \underset{\Omega}{\operatorname{argmin}} g(t, \Omega). \end{aligned}$$

Using the same techniques as before leads to a problem when evaluation the first order condition. Indeed, we have

$$\left. \frac{d}{d\Omega} g(t, \Omega) \right|_{\Omega=T_\Omega(F_{t,z})} = 0 \Leftrightarrow -T_\Omega(F_{t,z})^{-1} + S(F_{t,z}) + \lambda\Gamma(t) = 0, \quad (3.7)$$

with

$$\Gamma(t)_{i,j} \begin{cases} = \operatorname{sign}(T_\Omega(F_{t,z})_{i,j}) & \text{if } T_\Omega(F_{t,z})_{i,j} \neq 0 \\ \in [-1, 1] & \text{if } T_\Omega(F_{t,z})_{i,j} = 0. \end{cases}$$

Since we can not properly derive the third term w.r.t.  $t$ , we need an other technique to obtain the influence function. Note that if we could prove that the null elements of the precision matrix remain null under the contaminated distribution (which is the case), we have

$$\Gamma(t)_{i,j} \begin{cases} = \operatorname{sign}(T_\Omega(F)_{i,j}) & \text{if } T_\Omega(F)_{i,j} \neq 0 \\ \in [-1, 1] & \text{if } T_\Omega(F)_{i,j} = 0 \end{cases}$$

which does not depend on  $t$  any more. So deriving (3.7) w.r.t.  $t$  and taking the limit in  $t \rightarrow 0$  gives

$$T_\Omega(F)^{-1} \operatorname{IF}(z; T_\Omega, F) T_\Omega(F)^{-1} + \operatorname{IF}(z; S, F) = 0.$$

We could as in the previous section, isolate  $\operatorname{IF}(z; T_\Omega, F)$  to get its expression. However, in this case, we do not take into account that there exists  $T_\Omega(F)_{i,j} = 0$  and so  $\operatorname{IF}(z; (T_\Omega)_{i,j}, F) = 0$ . In other words, if  $\forall i, j, (T_\Omega(F_{t,z}))_{i,j} \neq 0$ , then we have, as in the previous section,

$$\operatorname{IF}(z; T_\Omega, F) = -T_\Omega(F) \operatorname{IF}(z; S, F) T_\Omega(F).$$

Moreover, note that when  $\lambda = 0$ , we found, if it exists,  $T_\Omega = S(F)^{-1}$ . This latter estimator is Fisher consistent for  $\Sigma^{-1}$ , since  $S(F)$  is itself Fisher consistent for  $\Sigma$ . However, when  $\lambda > 0$ , there is no reason why the estimator  $T_\Omega(F)$  should be Fisher consistent for  $\Sigma^{-1}$ . This is why we do not replace  $T_\Omega(F)$  by  $\Omega$  in the previous expressions.

To derive the influence function of the graphical lasso, we use the *vec* operator, which transforms a matrix into a column vector by vertically stacking the columns of the matrix. For example,

$$\operatorname{vec} \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 2 \\ 3 \end{pmatrix}.$$

We write  $\text{mat}$  the inverse operator of  $\text{vec}$ , defined such that  $\text{mat}(\text{vec}(X)) = X$ . We define

$$\omega := \text{vec}(\Omega) \in \mathbb{R}^{p^2}.$$

Note that  $\text{tr}(S(F)\Omega) = \text{vec}(S(F)^T)\omega$ , so the glasso estimator becomes

$$T_\omega(F) = \arg \min_{\omega \in \mathbb{R}^{p^2}} \left\{ -\log |\text{mat}(\omega)| + \text{vec}(S(F)^T)\omega + \lambda \|\omega\|_1 \right\}.$$

In the proof of the influence function of the graphical lasso, we will use an approximation  $p_m$  of the norm to obtain a differentiable expression. We use the notation  $C^\infty(\mathbb{R}^{p^2})$  for the set of continuous and infinitely differentiable functions on  $\mathbb{R}^{p^2}$  and  $W^{2,2}(\mathbb{R}^{p^2})$  for the Sobolev space given by the subset of functions in  $L^2(\mathbb{R}^{p^2})$  whose second weak derivative is integrable in  $L^2$  norm. The theorem below gives the influence function of the graphical lasso estimator.

**Theorem 3.1.** *Let  $F$  a distribution over  $\mathbb{R}^p$  and  $S$  the classical covariance matrix estimator. The influence function of the graphical lasso estimator is given by*

$$\text{IF}(z; DT_\omega, F) = - \left[ \begin{array}{c} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} \text{IF}(z; D \text{vec}(S), F)_{1:s} \\ 0 \end{array} \right],$$

where  $\otimes$  denotes the kronecker product and  $D \in \mathbb{R}^{p^2 \times p^2}$  is a symmetric permutation matrix such that  $(DT_\omega(F))_{1:s} \neq 0$  and  $(DT_\omega(F))_{(s+1):p^2} = 0$ .

*Proof.* In order to prove the proposition, we want to use the implicit function theorem, but the function to be minimised in the estimator is not differentiable because of the norm. We therefore use the same technique as in [25], which uses an approximation  $p_m$  in  $C^\infty(\mathbb{R}^{p^2})$  which converge to the norm in the Sobolev space  $W^{2,2}(\mathbb{R}^{p^2})$  when  $m \rightarrow \infty$ . It leads to another estimator defined as

$$T_{\omega, p_m}(F) = \arg \min_{\omega \in \mathbb{R}^{p^2}} \left\{ -\log |\text{mat}(\omega)| + \text{vec}(\widehat{S}(F))^T \omega + \lambda p_m(\omega) \right\}.$$

We can now use the implicit function theorem to get the influence function of  $T_{\omega, p_m}(F)$ . First, note that

$$\frac{d}{d\omega_{i,j}} (-\log |\text{mat}(\omega)|) = -\frac{1}{|\text{mat}(\omega)|} \frac{d}{d\omega_{i,j}} |\text{mat}(\omega)| = -(\text{mat}(\omega)^{-1})_{j,i}$$

(see [26] page 134 for example) so

$$\frac{d}{d\omega} (-\log |\text{mat}(\omega)|) = -\text{vec}(\text{mat}(\omega)^{-T}).$$

The first order condition (F.O.C.) is then given by

$$0 = -\text{vec}(\text{mat}(T_{\omega, p_m}(F))^{-T}) + \text{vec}(S(F)) + \lambda \nabla p_m(T_{\omega, p_m}(F)). \quad (3.8)$$

Consider  $z \in \mathbb{R}^p$  and define

$$f(t, \omega) := -\text{vec}(\text{mat}(\omega)^{-1}) + \text{vec}(S(F_{t,z})) + \lambda \nabla p_m(\omega)$$

such that

$$f(0, T_{\omega, p_m}(F)) = 0.$$

The partial derivatives w.r.t.  $\omega$  and  $t$  in  $(0, T_{\omega, p_m}(F))$  are given by

$$\left. \frac{d f(t, \omega)}{dt} \right|_{t=0} = \text{vec}(\text{IF}(S, F, z))$$

and

$$\left. \frac{d f(t, \omega)}{d \omega} \right|_{\omega=T_{\omega, p_m}(F)} = \lambda \nabla^2 p_m(T_{\omega, p_m}(F)) + U(T_{\omega, p_m}(F)),$$

where

$$\begin{aligned} U(\omega) &:= \frac{d(-\text{vec}(\Omega^{-1}))}{d \text{vec}(\Omega)} \\ &= \frac{-d(\Omega^{-1})}{d\Omega} && \text{(see [26] page 127 for example)} \\ &= \Omega^{-1} \otimes \Omega^{-1} && \text{(see [26] page 131 for example)} \\ &= \text{mat}(\omega)^{-1} \otimes \text{mat}(\omega)^{-1}. \end{aligned}$$

In order to use the classical implicit function theorem, one should show that the derivatives are continuous. However, in [27], a proof of the implicit function theorem is given only supposing that the derivative w.r.t. the second argument is nowhere vanishing. If we consider  $p_m$  convex, which is reasonable since it approximates the convex  $L_1$  norm, we have that  $\nabla^2 p_m(\omega)$  is p.s.d.. Moreover, we have by definition that  $U(\omega)$  is p.d., such that the derivative of  $f(t, \omega)$  w.r.t. the second argument is p.d. and therefore nowhere vanishing. The implicit function theorem gives

$$\text{IF}(z; T_{\omega, p_m}, F) = -(\lambda \nabla^2 p_m(T_{\omega, p_m}(F)) + U(T_{\omega, p_m}(F)))^{-1} \text{vec}(\text{IF}(z; S, F)).$$

Note that the theorem also gives that the function  $T_{\omega, p_m}(F)$  is continuous w.r.t.  $t$ . Before giving the influence function of the glasso estimator, we consider a symmetric permutation matrix  $D \in \mathbb{R}^{p^2 \times p^2}$  such that  $(D_{\omega}(F))_{1:s} \neq 0$  and  $(DT_{\omega}(F))_{(s+1):p^2} = 0$ . We find

$$\begin{aligned} \lim_{m \rightarrow \infty} \text{IF}(z; T_{\omega, p_m}, F) &= \lim_{m \rightarrow \infty} -(\lambda \nabla^2 p_m(T_{\omega, p_m}(F)) + U(T_{\omega, p_m}(F)))^{-1} \text{vec}(\text{IF}(z; S, F)) \\ &= \lim_{m \rightarrow \infty} -DD^{-1}(\lambda \nabla^2 p_m(T_{\omega, p_m}(F)) + U(T_{\omega, p_m}(F)))^{-1} D^{-T} D^T \text{vec}(\text{IF}(z; S, F)) \\ &= \lim_{m \rightarrow \infty} -D(D^T \lambda \nabla^2 p_m(T_{\omega, p_m}(F))D + D^T U(T_{\omega, p_m}(F))D)^{-1} D^T \text{vec}(\text{IF}(z; S, F)) \\ &= \lim_{m \rightarrow \infty} -D(\lambda \nabla^2 p_m(DT_{\omega, p_m}(F)) + D^T U(T_{\omega, p_m}(F))D)^{-1} D^T \text{vec}(\text{IF}(z; S, F)). \end{aligned}$$

If a matrix  $A$  is partitioned and all necessary inverses exist, we have

$$\begin{aligned} A^{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & A^{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \\ A^{12} &= -A^{11}A_{12}A_{22}^{-1} & A^{21} &= -A_{22}^{-1}A_{21}A^{11}. \end{aligned}$$

The hessian matrix  $\nabla^2 p_m(T_{\omega, p_m}(F))$  is diagonal and  $(\nabla^2 p_m(DT_{\omega}(F)))_{jj} \rightarrow 0$  for  $j = 1, \dots, s$  and  $(\nabla^2 p_m(DT_{\omega}(F)))_{jj} \rightarrow \infty$  for  $j = s+1, \dots, p^2$ . It gives

$$\lim_{m \rightarrow \infty} (\lambda \nabla^2 p_m(DT_{\omega, p_m}(F)) + D^T U(T_{\omega, p_m}(F))D)^{-1} = \begin{bmatrix} (D^T(\text{mat}(T_{\omega}(F))^{-1} \otimes \text{mat}(T_{\omega}(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Since the permutation matrix  $D$  is symmetric, we have  $D^2 = I$ , and the influence function is given by

$$\begin{aligned} \lim_{m \rightarrow \infty} \text{IF}(z; DT_{\omega, p_m}, F) &= D \lim_{m \rightarrow \infty} \text{IF}(z; T_{\omega, p_m}, F) \\ &= - \begin{bmatrix} (D(\text{mat}(T_{\omega}(F))^{-1} \otimes \text{mat}(T_{\omega}(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \text{IF}(z; D \text{vec}(S), F) \\ &= - \begin{bmatrix} (D(\text{mat}(T_{\omega}(F))^{-1} \otimes \text{mat}(T_{\omega}(F))^{-1})D)_{11}^{-1} \text{IF}(z; D \text{vec}(S), F)_{1:s} \\ 0 \end{bmatrix}, \end{aligned} \tag{3.9}$$

which ends the proof. □

In the following lemma, we prove that the influence function obtained using an approximation  $p_m$  is independent of the choice of  $p_m$ .

**Lemma 3.1.** *Let  $F$  a distribution over  $\mathbb{R}^p$  and  $S$  the classical covariance matrix estimator. The influence function of the graphical lasso estimator based on an approximation  $p_m$  of the norm is independent of the chosen approximation.*

*Proof.* In order to show that  $\text{IF}(z; T_\omega, F) := \lim_{m \rightarrow \infty} \text{IF}(z; T_{\omega, p_m}, F)$  is independent of the approximation  $p_m$ , we need to prove first that  $\lim_{m \rightarrow \infty} T_{\omega, p_m}(F)$  is unique. Using Lemma 2 in [25] with  $\Lambda_\lambda(\omega; F, p) = -\log |\text{mat}(\omega)| + \text{vec}(S)^T \omega + \lambda p_m(\omega)$  gives

$$\lim_{m \rightarrow \infty} T_{\omega, p_m}(F) = T_\omega(F). \quad (3.10)$$

We can now prove the independence of  $\lim_{m \rightarrow \infty} \text{IF}(z; T_{\omega, p_m}, F)$  of  $p_m$ . Let  $\{p_m\}_{m \geq 1}$  and  $\{p'_m\}_{m \geq 1}$  be two sequences in  $C^\infty(\mathbb{R}^{p^2})$  converging to  $p$  in  $W^{2,2}(\mathbb{R}^{p^2})$ . We have

$$\begin{aligned} \text{IF}(z; T_{\omega, p_m}, F) - \text{IF}(z; T_{\omega, p'_m}, F) &= (-\lambda \nabla^2 p_m(T_{\omega, p_m}(F)) + U(T_{\omega, p_m}(F)))^{-1} \\ &\quad + (\lambda \nabla^2 p'_m(T_{\omega, p'_m}(F)) + U(T_{\omega, p'_m}(F)))^{-1} \text{vec}(\text{IF}(\widehat{\Sigma}, F, z)). \end{aligned}$$

Since the inverse of a matrix is continuous and  $\nabla^2 p_m$  and  $\nabla^2 p'_m$  are sequences in  $C^\infty(\mathbb{R}^{p^2})$ , we find, using (3.10),

$$\lim_{m \rightarrow \infty} (\text{IF}(z; T_{\omega, p_m}, F) - \text{IF}(z; T_{\omega, p'_m}, F)) = 0.$$

□

In the next subsection, we discuss and give some remarks on the results obtained.

### 3.4.4 Discussion

Note that in the previous calculus, we should have considered the fact that  $\Omega$  is symmetric when derivating w.r.t.  $x$ . However, [28] proved that the gradient of a function w.r.t. a symmetric matrix is the symmetric of the gradient. Moreover, all the derivatives we obtained are symmetric expressions. For example, if we take the equation (3.8) and use the *mat* operator on both sides, we get

$$0 = -(T_\Omega(F))^{-T} + S(F) + \lambda \nabla p_m(T_\Omega(F)).$$

We note that the result is already symmetric, so considering that  $\Omega$  is symmetric does not change anything in the derivatives. Moreover, the logdet function in the precision matrix estimator acts as a barrier to the positive definite cone and ensures that the solution of the F.O.C. is p.d.

We continue the discussion with the influence function in this form as it is easier to read without using the *mat* operator. This result shows that the derivative of  $(T_\Omega)_{i,j} = 0$  is zero, and so under infinitesimal contamination, the value in the precision matrix remains zero. A somewhat surprising conclusion at first sight is that the elements of the covariance matrix estimator corresponding to the null elements of the precision matrix do not appear in the influence function of the latter. Intuitively, since we estimate the precision matrix which is the inverse of the covariance matrix, we would have expected that putting contamination in an element of the covariance matrix would have an impact on any element of the precision matrix, independently

of the value of the corresponding element of the contaminated value in the precision matrix. However, this phenomenon can be explained mathematically. Indeed, if we look at the definition of the graphical lasso, the contamination only appears through the covariance matrix estimator in the trace. Moreover, we have that

$$\text{tr}(S(F_{t,z})\Omega) = \text{vec}(S(F_{t,z}))^T \text{vec}(\Omega).$$

So if  $(T_\Omega(F))_{i,j} = (T_\Omega(F_{t,z}))_{i,j} = 0$ , it is clear that putting contamination in the corresponding element of the covariance matrix estimator will not change the scalar product, and so will not influence the estimation of the precision matrix.

Remember we conclude previously that, in the case of  $(T_\Omega)_{i,j} \neq 0$  for all  $1 \leq i, j \leq p$ , the influence function is given by

$$\text{IF}(z; T_\Omega, F) = -T_\Omega \text{IF}(z; S, F) T_\Omega.$$

This is actually consistent with the previous result, since in the case  $\Omega_{i,j} \neq 0$  for all  $1 \leq i, j \leq p$ , we can take  $D = I_p$  and so

$$\begin{aligned} \text{IF}(z; T_\omega, F) &= -((\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1}))^{-1} \text{IF}(z; \text{vec}(S), F) \\ &= -(T_\Omega \otimes T_\Omega) \text{IF}(z; \text{vec}(S), F), \end{aligned}$$

since for every matrix  $A, B$ , we have  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ . Using the property that  $\text{vec}(AXB) = (B^T \otimes A) \text{vec}(X)$ , we find

$$\begin{aligned} \text{IF}(z; T_\omega, F) &= \text{vec}(\text{IF}(z; T_\Omega, F)) \\ &= \text{vec}(-T_\Omega \text{IF}(z; S, F) T_\Omega) \\ &= -(T_\Omega \otimes T_\Omega) \text{vec}(\text{IF}(z; S, F)). \end{aligned}$$

### 3.4.5 Gross error sensitivity

Using the expression of the influence function, we can find an upper bound for the gross error sensitivity. Remember that it is defined as

$$\gamma^*(T_\Omega, F) = \sup_{z \in \mathbb{R}^p} \|\text{IF}(z; T_\Omega, F)\|.$$

Using expression(3.9), we find

$$\begin{aligned} \gamma^*(T_\omega, F) &= \sup_{z \in \mathbb{R}^p} \left\| \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \text{IF}(z; D \text{vec}(S), F) \right\| \\ &\leq \sup_{z \in \mathbb{R}^p} \left\| \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right\| \|\text{IF}(z; D \text{vec}(S), F)\| \\ &= \left\| \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right\| \sup_{z \in \mathbb{R}^p} \|\text{IF}(z; D \text{vec}(S), F)\| \\ &= \left\| \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \right\| \gamma^*(\text{vec}(S), F) \\ &= \left\| (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} \right\| \gamma^*(\text{vec}(S), F). \end{aligned}$$

Note that if we considered the expression of the influence function as

$$\text{IF}(z; T_\omega, F) = \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} \text{IF}(z; D \text{vec}(S), F)_{1:s} \\ 0 \end{bmatrix},$$

the bound is given by

$$\gamma^*(T_\omega, F) \leq \|(D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1}\| \gamma^*(\text{vec}(S)_{1:s}, F).$$

At last, if we suppose that, for all  $1 \leq i, j \leq p$ ,  $(T_\Omega)_{i,j} \neq 0$ , we can simplify the expression to find

$$\begin{aligned} \gamma^*(T_\omega, F) &\leq \|(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})^{-1}\| \gamma^*(\text{vec}(S), F) \\ &= \|\Omega \otimes \Omega\| \gamma^*(\text{vec}(S), F) \\ &= \|\Omega\|^2 \gamma^*(\text{vec}(S), F). \end{aligned}$$

### 3.5 Directions maximising the norm of the influence function

For numerical purposes, we were also interested in the direction which maximises the influence function. In other words, we looked for

$$\operatorname{argmax}_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; T_\Omega, F)\|_F^2.$$

Since  $\text{IF}(z; T_\Omega, F)$  depends on  $\text{IF}(z; S, F)$ , we first evaluated

$$\operatorname{argmax}_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; S, F)\|_F^2 = \operatorname{argmax}_{z \in \mathcal{S}^{p-1}} \|zz^T - \Sigma\|_F^2.$$

The next proposition gives the solution of the above maximisation expression.

**Proposition 3.1.** *Let  $F$  a c.d.f. and  $S$  the classical covariance matrix estimator. The vector which maximises the Frobenius norm of the influence function of  $S$  is the eigenvector corresponding to the smallest eigenvalue of the true covariance matrix  $\Sigma$ . Moreover, we have*

$$\max_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; S, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 - 2\lambda_p + 1,$$

where  $\lambda_1, \dots, \lambda_p$  are the eigenvalues of  $\Sigma$ .

*Proof.* Using the spectral decomposition of  $\Sigma = O\Lambda O^T$ , we find

$$\|(zz^T - \Sigma)\|_F^2 = \|(O^T zz^T O - \Lambda)\|_F^2.$$

The matrix  $O$  contains the eigenvectors  $v_i$ ,  $i = 1, \dots, p$ , which form an orthonormal basis. So we can decompose  $z$  in this basis to get  $z = \sum_{i=1}^p \alpha_i v_i$ . We have

$$\begin{aligned} \|zz^T - \Sigma\|_F^2 &= \|O^T zz^T O - \Lambda\|_F^2 \\ &= \left\| \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j O^T v_i v_j^T O - \Lambda \right\|_F^2. \end{aligned}$$

Moreover, we have that  $O^T v_i = e_i$ , so each matrix  $O^T v_i v_j^T O$  has only one non-zero component in  $(i, j)$ . We find

$$\begin{aligned} \|zz^T - \Sigma\|_F^2 &= \left\| \sum_{i=1}^p \sum_{j=1}^p \alpha_i \alpha_j O^T v_i v_j^T O - \Lambda \right\|_F^2 \\ &= \sum_{i=1}^p (\alpha_i^2 - \lambda_i)^2 + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \alpha_i^2 \alpha_j^2. \end{aligned}$$

The first term corresponds to the diagonal elements, when  $i = j$ , and the second term corresponds to the off-diagonal elements. Note that  $\|z\|_2 = 1$ , so

$$\begin{aligned}\|z\|_2^2 &= \left\| \sum_{i=1}^p \alpha_i v_i \right\|_2^2 \\ &= \sum_{i=1}^p \alpha_i^2 \|v_i\|_2^2 \\ &= \sum_{i=1}^p \alpha_i^2 = 1.\end{aligned}$$

We can therefore write that

$$\sum_{\substack{j=1 \\ j \neq i}}^p \alpha_j^2 = (1 - \alpha_i^2).$$

Putting this in the previous equation, we have

$$\begin{aligned}\|zz^T - \Sigma\|_F^2 &= \sum_{i=1}^p (\alpha_i^2 - \lambda_i)^2 + \sum_{i=1}^p \alpha_i^2 (1 - \alpha_i^2) \\ &= \sum_{i=1}^p (\alpha_i^4 - 2\alpha_i^2 \lambda_i + \lambda_i^2) + \sum_{i=1}^p (\alpha_i^2 - \alpha_i^4) \\ &= \sum_{i=1}^p \lambda_i^2 - 2 \sum_{i=1}^p \alpha_i^2 \lambda_i + 1\end{aligned}$$

We want to maximise this expression, which means to minimise

$$\begin{aligned}\sum_{i=1}^p \alpha_i^2 \lambda_i &\geq \sum_{i=1}^p \alpha_i^2 \lambda_p \\ &= \lambda_p,\end{aligned}$$

which is reached when  $\alpha_p = 1$  and  $\alpha_i = 0$ ,  $i = 1, \dots, p-1$ . In conclusion, we find

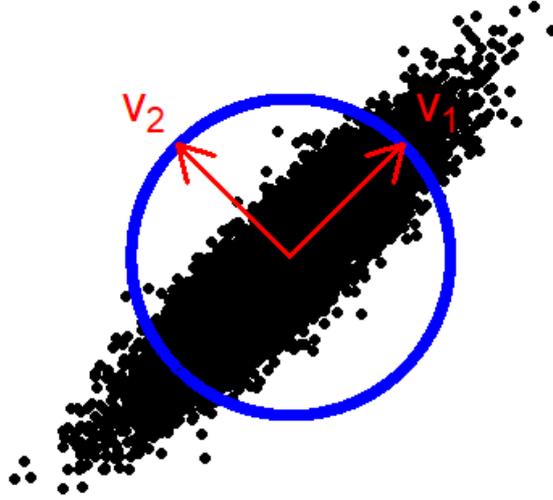
$$v_p = \operatorname{argmax}_{z \in \mathcal{S}^{p-1}} \|\operatorname{IF}(z; S, F)\|_F^2,$$

where  $v_p$  is the eigenvector associated to the smallest eigenvalue of  $\Sigma$ , and

$$\max_{z \in \mathcal{S}^{p-1}} \|\operatorname{IF}(z; S, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 - 2\lambda_p + 1.$$

□

After all, this conclusion is not very surprising. Indeed, FIGURE 5 displays in black data from a normal distribution in dimension 2. The blue circle represents the unit disk and each arrow represents an eigenvector of  $\Sigma$  used to generate the data. It is clear that putting contamination in the direction of  $v_2$ , which is the eigenvector corresponding to the smallest eigenvalue, is more likely to break the structure and the estimation of the covariance matrix.

Figure 5 – Illustration of the **Proposition 2**.

We can now focus on

$$\operatorname{argmax}_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; T_\Omega, F)\|_F^2.$$

Again, we first consider the MLE, for which the estimation is Fisher consistent, which means that  $T_\Omega = \Omega$ . The proposition below gives the solution of this maximisation expression.

**Proposition 3.2.** *Let  $F$  a c.d.f. and  $S$  the classical covariance matrix estimator. Suppose  $\lambda = 0$  in the expression of the graphical lasso. Let  $g$  the function defined as  $g(x) = x^4 - 2x^3$  and  $\lambda_1, \dots, \lambda_p$  the eigenvalues of  $\Omega$ . The vector which maximises the Frobenius norm of the influence function of  $T_\Omega$  is either the eigenvector corresponding to the smallest eigenvalue, if  $g(\lambda_1) < g(\lambda_p)$ , or the eigenvector corresponding to the largest eigenvalue, if  $g(\lambda_1) > g(\lambda_p)$ . Moreover, we have*

$$\max_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; T_\Omega, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 + \max(g(\lambda_1), g(\lambda_p)).$$

*Proof.* Using the expressions of the influence function of  $T_\Omega$  and  $S$ , we find

$$\begin{aligned} \|\text{IF}(z; T_\Omega, F)\|_F^2 &= \|\Omega \text{IF}(z; S, F) \Omega\|_F^2 \\ &= \|\Omega (zz^T - \Sigma) \Omega\|_F^2. \end{aligned}$$

The square of the Frobenius norm is actually the trace of the square of the matrix, since in our case, the matrix is symmetric. Using the fact that  $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$  for any matrices  $A, B, C$  of suitable dimensions, we find

$$\begin{aligned} \|\Omega (zz^T - \Sigma) \Omega\|_F^2 &= \|\Omega \Omega (zz^T - \Sigma)\|_F^2 \\ &= \|\Omega \Omega zz^T - \Omega\|_F^2 \end{aligned}$$

Using the spectral decomposition of  $\Omega$ , we find

$$\begin{aligned} \|\text{IF}(z; T_\Omega, F)\|_F^2 &= \|O \Lambda O^T O \Lambda O^T zz^T - O \Lambda O^T\|_F^2 \\ &= \|\Lambda^2 O^T zz^T O - \Lambda\|_F^2 \\ &= \|\Lambda^2 (O^T z)(O^T z)^T - \Lambda\|_F^2 \\ &= \|\Lambda^2 xx^T - \Lambda\|_F^2, \end{aligned}$$

with  $x = O^T z$ . The diagonal elements  $(i, i)$  are of the form  $\lambda_i^2 x_i^2 - \lambda_i$  and the off-diagonal elements  $(i, j)$  look like  $\lambda_i^2 x_i x_j$ . Therefore, we have

$$\|\text{IF}(z; T_\Omega, F)\|_F^2 = \sum_{i=1}^p (\lambda_i^2 x_i^2 - \lambda_i)^2 + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \lambda_i^4 x_i^2 x_j^2.$$

Note that  $\|x\| = \|O^T z\| = \|z\| = 1$ , so  $\sum_{\substack{j=1 \\ j \neq i}}^p x_j^2 = (1 - x_i^2)$ . Using this, we find

$$\begin{aligned} \|\text{IF}(z; T_\Omega, F)\|_F^2 &= \sum_{i=1}^p (\lambda_i^4 x_i^4 - 2\lambda_i^3 x_i^2 + \lambda_i^2) + \sum_{i=1}^p \lambda_i^4 x_i^2 (1 - x_i^2) \\ &= \sum_{i=1}^p \lambda_i^2 - 2 \sum_{i=1}^p \lambda_i^3 x_i^2 + \sum_{i=1}^p \lambda_i^4 x_i^2 \\ &= \sum_{i=1}^p \lambda_i^2 + \sum_{i=1}^p x_i^2 (\lambda_i^4 - 2\lambda_i^3) \end{aligned}$$

Moreover, we have  $x_i = (O^T z)_i = O_{i,\cdot}^T z = v_i^T z$ . Using  $z = \sum_{j=1}^p \alpha_j v_j$ , we find  $x_i = \sum_{j=1}^p \alpha_j v_i^T v_j = \alpha_i$ . At last, we find

$$\begin{aligned} \|\text{IF}(z; T_\Omega, F)\|_F^2 &= \sum_{i=1}^p \lambda_i^2 + \sum_{i=1}^p \alpha_i^2 (\lambda_i^4 - 2\lambda_i^3) \\ &\leq \sum_{i=1}^p \lambda_i^2 + \max_{i \in \{1, \dots, p\}} (\lambda_i^4 - 2\lambda_i^3) \end{aligned}$$

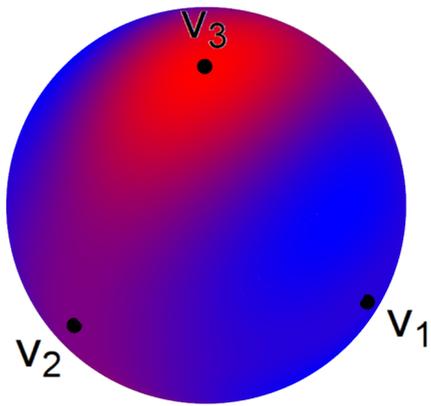
which is reached for the eigenvector corresponding to the eigenvalue which maximises  $(\lambda^4 - 2\lambda^3)$ . Note that all the eigenvalues are positive since  $T_\Omega$  is pd and the function decreases strictly between 0 and 1.5 and then increases strictly. It means that the only eigenvalue which can maximise the function is the first or the last one. Therefore, we have

$$\max_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; T_\Omega, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 + \max(\lambda_1^4 - 2\lambda_1^3, \lambda_p^4 - 2\lambda_p^3),$$

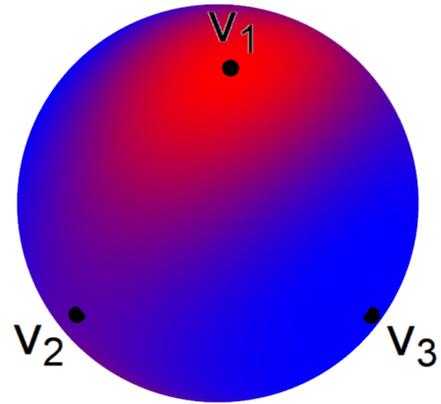
which completes the proof. □

If all the eigenvalues (and equivalently, the biggest eigenvalue) are smaller than 1.5, then  $\lambda_p$  reaches the maximum and so it is  $v_p$  which maximises the influence function. On the other hand, if all the eigenvalues (and equivalently, the smallest eigenvalue) are bigger than 1.5 then  $\lambda_1$  reaches the maximum and so it is  $v_1$  which maximises the influence function.

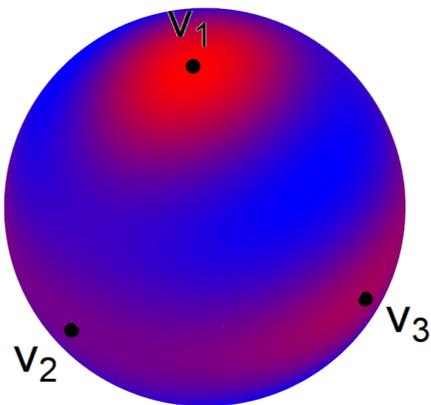
To illustrate those results, we use several matrices in dimension  $p = 3$  and plot each direction in a colour corresponding to the square of the Frobenius norm of the influence function (see FIGURE 6). The points in red have a high influence, where the points in blue have a small one. We start on the top left panel with a first true precision matrix which has as eigenvalues 1.4, 1 and 0.1. Since all the eigenvalues are smaller than 1.5, it is the smallest eigenvalues which maximises the function  $g$  in the norm of the influence function, and so the latter is maximised for  $z = v_3$ . On the top right panel, we took  $\Omega$  whose eigenvalues equal 1.7, 2.5 and 3. In this case,



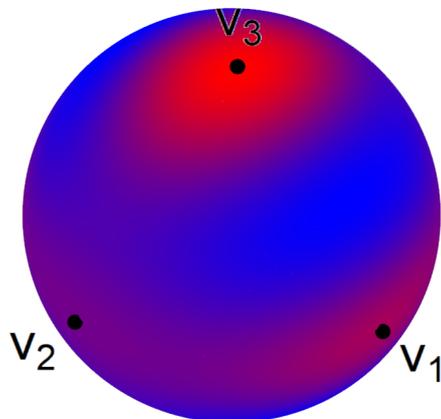
Eigenvalues : 1.4, 1 and 0.1



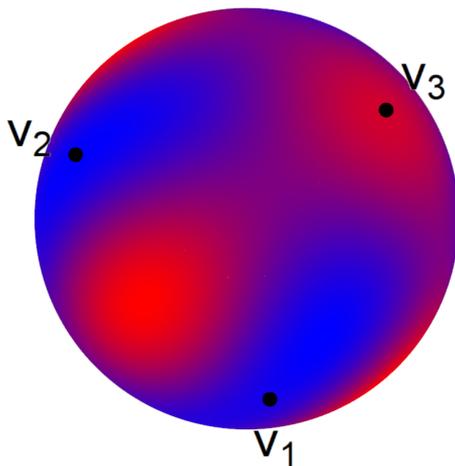
Eigenvalues : 1.7, 2.5 and 3



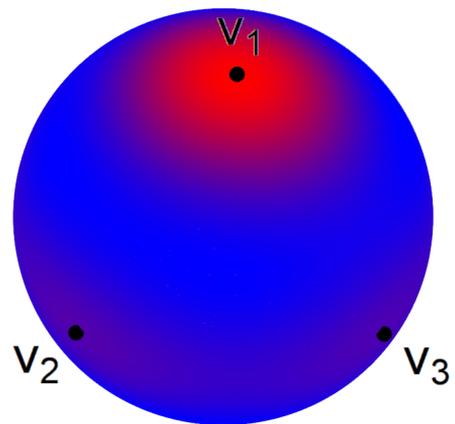
Eigenvalues : 0.5, 1.2 and 2.2



Eigenvalues : 0.5, 1.5 and 1.8



sparse matrix with  $\lambda = 0.01$



Diagonal matrix

Figure 6 – Illustration of the maximiser of the norm of the influence function.

since they all are bigger than 1.5 it is the biggest eigenvalue which maximises the expression in the norm of the influence function and so  $v_1$  is the direction corresponding to the biggest change. The bottom panels display the results when the eigenvalues are 0.5, 1.2 and 2.2 and 0.5, 1.5 and 1.8. In both cases, the smallest eigenvalues is smaller than 1.5 and the biggest one is bigger than 1.5. So we have to look at the expression  $\max(\lambda_1^4 - 2\lambda_1^3, \lambda_p^4 - 2\lambda_p^3)$ . In the first case, since  $2.2^4 - 2 * 2.2^3 = 2.1296 > -0.1875 = 0.5^4 - 2 * 0.5^3$ , the eigenvector which maximises the norm of the influence is the vector associated to the largest eigenvalues. On the second case, since  $-0.1875 > -1.1664 = 1.8^4 - 2 * 1.8^3$ , it is the eigenvector associated to the smallest eigenvalue which maximises the norm of the influence function.

When considering  $\lambda > 0$ , we could not get a nice result for the vector which maximises the norm of the influence function. The main problem is that we can not say much about  $T_\Omega(F)$ . Even if for all  $1 \leq i, j \leq p$ ,  $(T_\Omega)_{i,j} \neq 0$  and so we have the same form for the influence function that the one obtained when  $\lambda = 0$ , the estimation may not be Fisher consistent for  $\Sigma^{-1}$  and so we can not write  $\widehat{\Omega}\Sigma = I_p$  for example. Numerical experiments tend to confirm that, depending on the matrix, they may be no relationship between this vector and the eigenvectors of  $T_\Omega$ . Looking at the bottom left panel, we see that no eigenvector maximises the norm of the influence function. However we could notice numerically that when lambda is taken sufficiently large, such that the estimation is diagonal, again the only vectors which maximise the norm of the influence function are the first or last eigenvector. The rigorous proof is given below.

**Proposition 3.3.** *Let  $F$  a c.d.f. and  $S$  the classical covariance matrix estimator. Suppose  $\lambda > 0$  is large enough such that the precision matrix estimator is diagonal. Let  $g$  the function defined as  $g(x) = x^4 - 2x^3$  and  $\lambda_1, \dots, \lambda_p$  the eigenvalues of  $\Omega$ . The vector which maximises the Frobenius norm of the influence function of  $T_\Omega$  is either the eigenvector corresponding to the smallest eigenvalue, if  $g(\lambda_1) < g(\lambda_p)$ , or the eigenvector corresponding to the largest eigenvalue, if  $g(\lambda_1) > g(\lambda_p)$ . Moreover, we have*

$$\max_{z \in S^{p-1}} \|\text{IF}(z; T_\Omega, F)\|_F^2 = \sum_{i=1}^p \lambda_i^2 + \max(g(\lambda_1), g(\lambda_p)).$$

*Proof.* To prove the proposition, we have to consider the expression of the influence function obtained when  $\lambda > 0$ , that is

$$\text{IF}(z; D \text{vec}(T_\Omega), F) = - \begin{bmatrix} (D(T_\Omega^{-1} \otimes T_\Omega^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \text{IF}(z; D \text{vec}(S), F).$$

Now  $T_\Omega$  is diagonal, so we write it  $T_\Omega = \text{diag}(\lambda_1, \dots, \lambda_p)$ . Therefore,  $(T_\Omega)^{-1} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_p^{-1})$  and  $T_\Omega^{-1} \otimes T_\Omega^{-1} = \text{diag}(\lambda_1^{-2}, \lambda_1^{-1}\lambda_2^{-1}, \dots, \lambda_1^{-1}\lambda_p^{-1}, \lambda_1^{-1}\lambda_2^{-1}, \lambda_2^{-2}, \dots, \lambda_p^{-1}\lambda_2^{-1}, \dots, \lambda_p^{-2})$ , which is a  $p^2 \times p^2$  matrix. Remember that  $D$  is defined such that all the non-zero values of  $\text{vec}(T_\Omega)$  are at the top of the vector. In other words,  $D$  will put the  $p$  diagonal elements of the matrix at the start, and then all the 0. The permutation matrix will therefore exchange the  $i^{\text{th}}$  element with the element  $i + (i - 1)p$ , for  $2 \leq i \leq p$ . Applying this permutation matrix on both sides of a diagonal matrix will exchange the same diagonal elements. Therefore, we find  $(D(T_\Omega^{-1} \otimes T_\Omega^{-1})D)_{11} = \text{diag}(\lambda_1^{-2}, \dots, \lambda_p^{-2})$  and the inverse is given by  $(D(T_\Omega^{-1} \otimes T_\Omega^{-1})D)_{11}^{-1} = \text{diag}(\lambda_1^2, \dots, \lambda_p^2)$ . Note that applying  $D$  to  $\text{vec}(\text{IF}(z; S, F)) = \text{vec}(zz^T - \Sigma)$  will, as for  $\text{vec}(T_\Omega)$  put all the diagonal elements at the beginning of the vector. It gives  $(z_1^2 - S(1, 1), \dots, z_p^2 - S(p, p), \dots)^T$ , where the elements after  $z_p^2 - S(p, p)$  are not of interest there since

$$\begin{aligned} \|\text{IF}(z; \text{vec}(T_\Omega), F)\|_F^2 &= \|(D(T_\Omega^{-1} \otimes T_\Omega^{-1})D)_{11}^{-1}(z_1^2 - S(1, 1), \dots, z_p^2 - S(p, p), \dots)^T\|_F^2 \\ &= \|\text{diag}(\lambda_1^2, \dots, \lambda_p^2)(z_1^2 - S(1, 1), \dots, z_p^2 - S(p, p))^T\|_F^2. \end{aligned}$$

Note that, since  $T_\Omega$  is diagonal,  $\lambda\|\Omega^{-\text{diag}}\|_1 = 0$  and so

$$\begin{aligned} \min_{\Omega > 0} \{ -\log|\Omega| + \text{tr}(S(F)\Omega) + \lambda\|\Omega^{-\text{diag}}\|_1 \} &= -\log|T_\Omega| + \text{tr}(S(F)T_\Omega) \\ &= -\log|T_\Omega| + \text{tr}(\text{diag}(S(1,1), \dots, S(p,p))T_\Omega) \\ &\geq \min_{\Omega > 0} \{ -\log|\Omega| + \text{tr}(\text{diag}(S(1,1), \dots, S(p,p))\Omega) \} \\ &= -\log|\text{diag}(S(1,1), \dots, S(p,p))^{-1}| + \text{tr}(I_p), \end{aligned}$$

where the last equality comes from the definition of the MLE. Moreover, we have

$$\begin{aligned} \min_{\Omega > 0} \{ -\log|\Omega| + \text{tr}(S(F)\Omega) + \lambda\|\Omega^{-\text{diag}}\|_1 \} &\leq \min_{\substack{\Omega > 0 \\ \Omega \text{ diagonal}}} \{ -\log|\Omega| + \text{tr}(S(F)\Omega) + \lambda\|\Omega^{-\text{diag}}\|_1 \} \\ &= \min_{\substack{\Omega > 0 \\ \Omega \text{ diagonal}}} \{ -\log|\Omega| + \text{tr}(\text{diag}(S(1,1), \dots, S(p,p))\Omega) \} \\ &\leq -\log|\text{diag}(S(1,1), \dots, S(p,p))^{-1}| + \text{tr}(I_p). \end{aligned}$$

Since the solutions are unique, we find  $T_\Omega = \text{diag}(S(1,1)^{-1}, \dots, S(p,p)^{-1})$ . Therefore, the norm of the influence function becomes

$$\begin{aligned} \|\text{IF}(z; \text{vec}(T_\Omega), F)\|_F^2 &= \|\text{diag}(\lambda_1^2, \dots, \lambda_p^2)(z_1^2 - \lambda_1^{-1}, \dots, z_p^2 - \lambda_p^{-1})^T\|_F^2 \\ &= \sum_{i=1}^p (\lambda_i^2 z_i^2 - \lambda_i)^2 \\ &\leq \sum_{i=1}^p (\lambda_i^2 z_i^2 - \lambda_i)^2 + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \lambda_i^4 z_i^2 z_j^2. \end{aligned}$$

We proved previously that this upper bound is reached when  $z$  is equal to the eigenvector  $v$  corresponding to the eigenvalue which maximises the function  $g$  ( we had  $\alpha_i$  instead of  $z_i$  but since  $\Omega$  is diagonal,  $O$  contains the vectors of the canonical basis and so  $z_i = \alpha_i$ ). However, in our case the eigenvectors are the vectors of the canonical basis since  $T_\Omega$  is diagonal, and so

$$\sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \lambda_i^4 v_i^2 v_j^2 = 0.$$

In conclusion, we have

$$\begin{aligned} \max_{z \in \mathcal{S}^{p-1}} \|\text{IF}(z; \text{vec}(T_\Omega), F)\|_F^2 &= \max \left( (\lambda_1^2 - \lambda_1)^2 + \sum_{i=2}^p \lambda_i^2, \sum_{i=1}^{p-1} \lambda_i^2 + (\lambda_p^2 - \lambda_p)^2 \right) \\ &= \sum_{i=1}^p \lambda_i^2 + \max(g(\lambda_1), g(\lambda_p)). \end{aligned}$$

□

### 3.6 Asymptotic variance

As in the previous sections, we first start with the asymptotic variance of the covariance matrix estimators. For the empirical covariance matrix, the asymptotic variance of  $\text{vec}(S)$  is given in [29] and equals

$$\text{ASV}(\text{vec}(S), F) = E[(X - \mu)(X - \mu)' \otimes (X - \mu)(X - \mu)'] - \text{vec}(\Sigma) \text{vec}(\Sigma)',$$

where  $X \sim F$  and  $\mu = E[X]$ . We consider in this section  $F = N_3(0, \Sigma)$ , where  $\Sigma$  is the same matrix that the one we used in the previous section. We therefore have  $\mu = 0$  and we can obtain this ASV matrix theoretically. For example,

$$\begin{aligned} [\text{ASV}(\text{vec}(S), F)]_{1,1} &= E[X_1^4] - 1 \\ &= 3\text{Var}(X_1)^4 - 1 \\ &= 2. \end{aligned}$$

Since the vector  $X$  is Gaussian, we have  $X_2 = \rho X_1 + \sqrt{1 - \rho^2}Z$ , where  $Z \sim N(0, 1)$  is independent of  $X_1$ . The second diagonal element is given by

$$\begin{aligned} [\text{ASV}(\text{vec}(S), F)]_{2,2} &= E[X_1^2 X_2^2] - 0.25 \\ &= E[X_1^2(\rho^2 X_1^2 + (1 - \rho^2)Z^2 + 2\rho\sqrt{1 - \rho^2}X_1 Z)] - 0.25 \\ &= \rho^2 E[X_1^4] + (1 - \rho^2)E[X_1^2 Z^2] + 2\rho\sqrt{1 - \rho^2}E[X_1^3 Z^2] - 0.25 \\ &= 2\rho^2 + 1 - 0.25, \end{aligned}$$

since  $E[X_1^2 Z^2] = E[X_1^2]E[Z]^2 = \text{Var}(X_1)\text{Var}(Z) = 1$  and  $E[X_1^3] = 0$  since it is the integral of an odd function. In this case,  $\rho = 0.5$  so we have  $[\text{ASV}(\text{vec}(S), F)]_{2,2} = 1.25$ . Using the same kind of argument, we have  $E[X_1^3 X_2] = 3\rho\text{Var}(X)^4 = 1.5$  and so  $[\text{ASV}(\text{vec}(S), F)]_{2,1} = 1$ , and in our case, the whole matrix is given by

$$\text{ASV}(\text{vec}(S), F) = \begin{bmatrix} 2 & 1 & 0.5 & 1 & 0.5 & 0.25 & 0.5 & 0.25 & 0.125 \\ 1 & 1.25 & 0.625 & 1.25 & 1 & 0.5 & 0.625 & 0.5 & 0.25 \\ 0.5 & 0.625 & 1.0625 & 0.625 & 0.5 & 0.625 & 1.0625 & 0.625 & 0.5 \\ 1 & 1.25 & 0.625 & 1.25 & 1 & 0.5 & 0.625 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 & 1 & 2 & 1 & 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.625 & 0.5 & 1 & 1.25 & 0.625 & 1.25 & 1 \\ 0.5 & 0.625 & 1.0625 & 0.625 & 0.5 & 0.625 & 1.0625 & 0.625 & 0.5 \\ 0.25 & 0.5 & 0.625 & 0.5 & 1 & 1.25 & 0.625 & 1.25 & 1 \\ 0.125 & 0.25 & 0.5 & 0.25 & 0.5 & 1 & 0.5 & 1 & 2 \end{bmatrix}.$$

Note that the values of high interest there are the diagonal elements, which correspond to the asymptotic variance of each component of the matrix  $S$ . Therefore, we observe that the diagonal elements of this latter have a bigger asymptotic variance, where the component (1, 3) and (3, 1) have an asymptotic variance nearly twice as small.

We can now focus on the ASV of the graphical lasso estimator. The asymptotic variance can be obtained via the formula

$$\text{ASV}(DT_\omega, F) = \int \text{IF}(z; DT_\omega, F) \text{IF}(z; DT_\omega, F)^T dF(z).$$

Using equation (3.9), we find for the asymptotic variance of the glasso

$$\text{ASV}(DT_\omega, F) = \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \text{ASV}(D \text{vec}(S), F) \\ \begin{bmatrix} (D(\text{mat}(T_\omega(F))^{-1} \otimes \text{mat}(T_\omega(F))^{-1})D)_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Note that the asymptotic variance of the null elements of the estimation of the precision matrix is, as expected, 0 since under contamination they stay at 0. Again, the ASV of the graphical lasso estimator is a simple function of the ASV of the covariance matrix estimator used.

### 3.7 Introduction to robust covariance matrix estimators

In this section, we introduce several robust covariance matrix estimators. Note that the expression obtained for the influence function depends, as expected, explicitly of the influence function of the covariance matrix estimator. In order to get a better influence function than the unbounded one obtained when using the empirical covariance matrix, we can use several robust estimators, whose influence function is bounded. We therefore introduce estimators based on pairwise correlations. We consider several correlation estimators and we use their influence function to obtain the influence function of the graphical lasso. Then, we plot the results and the sensitivity curves, both in 2 dimensions, with a fixed direction, and in 3 dimensions.

The first estimator we use is the MCD, that we have already largely introduced in the first chapter. Remember that it is defined as the empirical covariance matrix based on the  $h$  data which lead to the empirical covariance matrix with the smallest determinant, where  $h$  is a constant between  $n/2$  and  $n$ . As mentioned earlier, that kind of estimator is time consuming so we also use estimators based on pairwise correlations. For the data  $X = (x^1, \dots, x^p) \in \mathbb{R}^{n \times p}$ , these estimators are defined as

$$\widehat{\Sigma}_{i,j} = \text{scale}(x^i)\text{scale}(x^j)r(x^i, x^j), \quad i, j = 1, \dots, p, \quad (3.11)$$

where  $\text{scale}(\cdot)$  is a robust scale estimate and  $r(\cdot, \cdot)$  is a robust correlation estimator. To obtain the influence function of this covariance matrix estimator, remember that the influence function is defined as the derivative of the estimator, evaluated in 0. Using the product rule, we find, for a distribution  $F$  in dimension  $p$ ,

$$\begin{aligned} \text{IF}(z; \widehat{\Sigma}_{i,j}, F) &= \text{IF}(z; \text{scale}, F_i)\text{scale}(X^j)r(X^i, X^j) + \text{scale}(X^i)\text{IF}(z; \text{scale}, F_j)r(X^i, X^j) \\ &\quad + \text{scale}(X^i)\text{scale}(X^j)\text{IF}(z; r, H_{i,j}). \end{aligned}$$

The distribution  $H$  is the bivariate distribution of the vector  $(X^i, X^j)$ , and the distributions  $F_i$  and  $F_j$  represent the univariate distribution of  $X^i$  and  $X^j$ . When the distribution is  $N_p(0, \Sigma)$ , we can easily obtain the marginal distributions. Indeed, we have

$$\begin{aligned} X^i &\sim N_1(0, \Sigma_{i,i}), \\ X^j &\sim N_1(0, \Sigma_{j,j}), \\ (X^i, X^j) &\sim N_2(0, \Sigma^{i,j}), \end{aligned}$$

where  $\Sigma^{i,j}$  is defined as

$$\Sigma^{i,j} = \begin{bmatrix} \Sigma_{i,i} & \Sigma_{i,j} \\ \Sigma_{j,i} & \Sigma_{j,j} \end{bmatrix}.$$

Note that we write  $\text{IF}(z; \text{scale}, F_i)$  even if it is clear that only the  $i$ th component of  $z$  influences the scale of  $F_i$ . We have

$$\begin{aligned} \text{IF}(z; \widehat{\Sigma}_{i,j}, P) &= \text{IF}(z_i; \text{scale}, F_i)\text{scale}(X^j)r(X^i, X^j) + \text{scale}(X^i)\text{IF}(z_j; \text{scale}, F_j)r(X^i, X^j) \\ &\quad + \text{scale}(X^i)\text{scale}(X^j)\text{IF}((z_i, z_j); r, H_{i,j}). \end{aligned} \quad (3.12)$$

### 3.7.1 Scale estimators

As scale estimator, we could use the median absolute deviation. However, as this needs a location estimate, we prefer to use the  $Q_n$  estimator [30] defined below.

**Definition 3.1.** Let  $\{x_1, \dots, x_n\}$  a sample of data in dimension  $p = 1$ . The  $Q_n$  estimator is given by

$$Q_n = d \{ |x_i - x_j|; 1 \leq i < j \leq n \}_{(k)},$$

where  $d$  is a constant factor and  $k = \binom{[n/2]+1}{2}$ .

The estimator corresponds to the  $k$ th order statistic of the  $\binom{n}{2}$  interpoint distances. It has a 50% breakdown point and is fast to compute. For symmetric distributions, the estimator is Fisher consistent for  $\sigma$  if we take  $d = 1/((F^{*2})^{-1}(5/8))$ , where  $F^{*2}$  denotes the convolution  $F * F$ . For example, if  $F = N(0, \sigma^2)$ , we find  $F * F = N(0, 2\sigma^2)$ . For the standard normal distribution, we have

$$d = \frac{1}{\sqrt{2}\Phi^{-1}(5/8)} \approx 2.219144.$$

Note that the  $Q_n$  estimator is also more efficient than the MAD at the Gaussian model. The following proposition gives the influence function of this estimator.

**Proposition 3.4.** Suppose  $F$  is a continuous c.d.f. with density function  $f$ . The influence function of the  $Q_n$  estimator is given by

$$\text{IF}(z; Q, F) = d \frac{\frac{1}{4} - F(x + d^{-1}) + F(x - d^{-1})}{\int f(y + d^{-1})f(y)dy}.$$

In FIGURE 7, we plotted the influence function. We note that the influence of an inlier at the centre of the distribution is smaller (in absolute value) than the influence of an outlier at infinity.

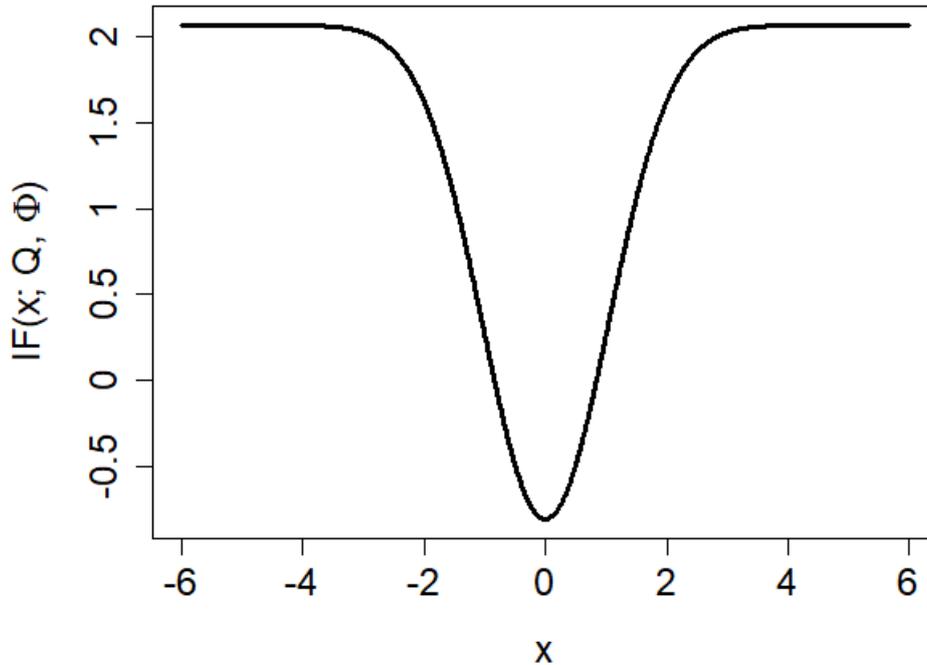


Figure 7 – Influence function of  $Q$  at the standard normal distribution.

We also see that the maximum value is slightly bigger than 2. Since  $F$  is an increasing function, we have  $F(x - d^{-1}) - F(x + d^{-1}) \leq 0$ , which means that

$$\text{IF}(z; Q, F) \leq \frac{d}{4 \int f(y + d^{-1})f(y)dy} \approx 2.069.$$

This value is approached when  $x$  goes to  $\pm\infty$ , since

$$\lim_{x \rightarrow +\infty} F(x - d^{-1}) - F(x + d^{-1}) = 1 - 1 = 0$$

and

$$\lim_{x \rightarrow -\infty} F(x - d^{-1}) - F(x + d^{-1}) = 0 - 0 = 0.$$

### 3.7.2 Correlation estimators

There are different choices of robust correlation estimators. Remember first that the correlation  $\rho$  of two random variables is defined as

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

For a bivariate sample  $\{(x_1, y_1), \dots, (x_n, y_n)\}$ , the classical Pearson correlation estimator is given by

$$\hat{\rho}_P = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

where  $\bar{x}$  and  $\bar{y}$  denote the sample mean. It is well known that this estimator is highly affected by the presence of a small amount of outliers. A popular alternative to the Pearson estimator is the Spearman's rank correlation  $r_S$ , defined as the Pearson correlation between the ranks of the observations. Note that, in order to compare the different correlation estimators, we take the consistent version for the population correlation  $\rho$  at the bivariate normal distribution. For the Spearman rank correlation, we obtain

$$\hat{\rho}_S = 2 \sin \left( \frac{1}{6} \pi r_S \right).$$

An other alternative to those estimators is Kendall's correlation, defined as

$$r_K = \frac{2}{n(n-1)} \sum_{i < j} \text{sign}((x_i - x_j)(y_i - y_j)).$$

Its consistent version is given by

$$\hat{\rho}_K = \sin \left( \frac{1}{2} \pi r_K \right).$$

An other correlation estimator using the sign function is given by

$$r_Q = \frac{1}{n} \sum_{i=1}^n \text{sign} \{ (x_i - \text{med}_i(x_i))(y_i - \text{med}_i(y_i)) \}.$$

It's consistent version is given by

$$\hat{\rho}_Q = \sin \left( \frac{1}{2} \pi r_Q \right).$$

We introduce a last estimator, which is the Gaussian Rank correlation. We first compute the ranks of each observation, and we write them  $R(x_i)$  and  $R(y_i)$ ,  $1 \leq i \leq n$ . Then we rescale the ranks between 0 and 1 and plug the values in the quantile function  $\Phi^{-1}$ . We get

$$\hat{\rho}_G = c_n \sum_{i=1}^n \Phi^{-1} \left( \frac{R(x_i)}{n+1} \right) \Phi^{-1} \left( \frac{R(y_i)}{n+1} \right),$$

where the constant  $c_n$  satisfies

$$\frac{1}{c_n} = \sum_{i=1}^n \Phi^{-1} \left( \frac{i}{n+1} \right)^2.$$

Note that this estimator is already Fisher consistent. Thanks to the use of the ranks and sign function, those estimators are more resistant to small amounts of outliers than the Pearson correlation estimator.

### 3.7.3 Influence functions and asymptotic variances

In order to compute the influence function of the covariance matrix estimator, we first need the influence functions of these correlation estimators. They have all been derived in [31]. We start with the classical estimator, whose influence function is given by

$$\text{IF}((x, y); \hat{\rho}_P, H) = xy - \rho \left( \frac{x^2 + y^2}{2} \right),$$

where  $H$  is the distribution of  $(X, Y)$ . We keep  $\rho$  instead of  $\hat{\rho}_P$  in the expression since the estimator is Fisher consistent, so it is easier to compare with the other estimators. The Gaussian Rank estimator has actually the same influence function. Note that this result is somehow misleading, since it is based on an asymptotic version of the estimator. For a finite sample, the Gaussian Rank estimator is actually much more robust. The influence function of the Spearman correlation estimator is given by

$$\begin{aligned} \text{IF}((x, y); \hat{\rho}_S, H) = \frac{\pi}{3} \text{sign}(\rho) \sqrt{1 - \frac{\rho^2}{4}} \left[ -3\rho - 9 + 12 \{F(x)G(y) \right. \\ \left. + E_H[F(X)I(Y \geq y)] + E_H[G(Y)I(X \geq x)] \} \right], \end{aligned}$$

where  $F$  and  $G$  are the marginal cumulative distribution functions, defined by

$$F(t) = P_H[X \leq t] \quad G(t) = P_H[Y \leq t].$$

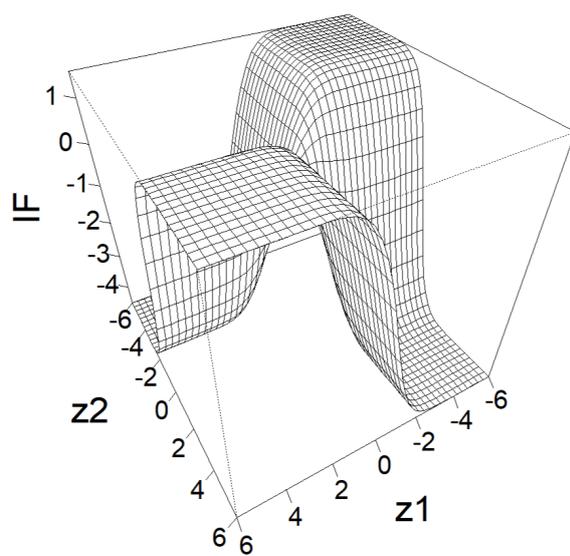
For the Kendall correlation estimator, we have

$$\text{IF}((x, y); \hat{\rho}_K, H) = \frac{\pi}{2} \text{sign}(\rho) \sqrt{1 - \rho^2} \left[ 2 \{2P_H[(X - x)(Y - y) > 0] - 1 - \rho\} \right]$$

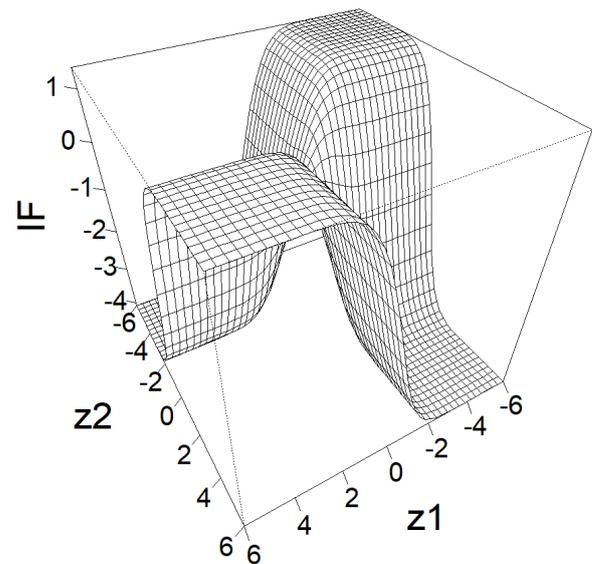
Finally, we find for the Quadrant correlation estimator

$$\text{IF}((x, y); \hat{\rho}_Q, H) = \frac{\pi}{2} \text{sign}(\rho) \sqrt{1 - \rho^2} \left[ \text{sign}(xy) - \rho \right].$$

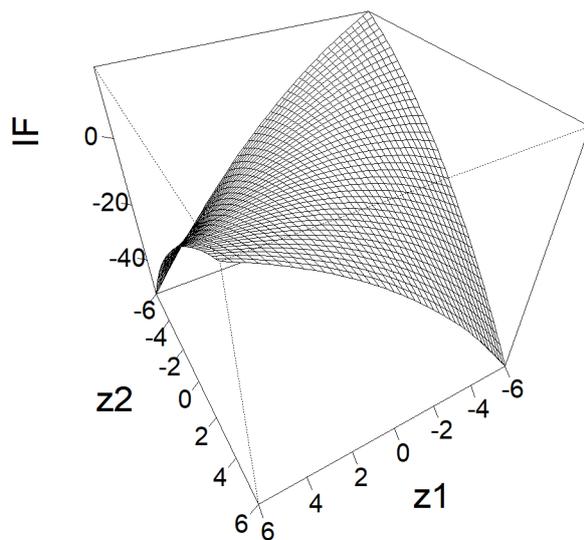
Note that the factors  $\pi/2 \text{sign}(\rho) \sqrt{1 - \rho^2}$  and  $\pi/3 \text{sign}(\rho) \sqrt{1 - \frac{\rho^2}{4}}$  are added because we use the Fisher consistent version of each estimator, in order to compare them properly. In FIGURE 8, we plot the different influence functions for  $H$  the bivariate normal distribution and  $\rho = 0.5$ . We



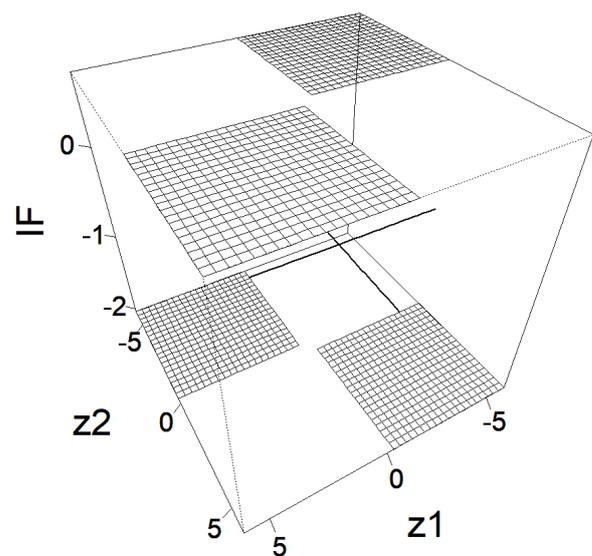
Spearman correlation estimator



Kendall correlation estimator



Gaussian Rank estimator



Quadrant correlation estimator

Figure 8 – Influence function of correlation estimators at the bivariate normal distribution.

note that the influence function of the Gaussian Rank estimator is unbounded as expected when looking at its expression. The influence function of the Quadrant estimator is bounded but has jumps at the coordinate axes, where it is equal to 0. It means that, close to these axes, small changes lead to relatively large changes in the estimator. The influence functions of the Kendall and Spearman correlation estimator are bounded, smooth and are actually very similar. One can check that they are actually equal to  $4\pi(\Phi(x) - 0.5)(\Phi(y) - 0.5)$  when  $\rho = 0$  and slightly differ when  $\rho \neq 0$ .

We can now discuss the asymptotic variance of the covariance matrix estimator (3.11). As mentioned before, in general we are mainly interested in

$$\text{ASV}(\widehat{\Sigma}_{i,j}, F) = \int \text{IF}(z; \widehat{\Sigma}_{i,j}, F)^2 dF(z).$$

Using equation (3.12), we find

$$\begin{aligned} \text{IF}(z; \widehat{\Sigma}_{i,j}, F)^2 &= (\text{IF}(z; Q, F_i)Q(X^j)r(X^i, X^j) + Q(X^i)\text{IF}(z; Q, F_j)r(X^i, X^j) \\ &\quad + Q(X^i)Q(X^j)\text{IF}(z; r, F_{i,j}))^2 \\ &= \text{IF}(z; Q, F_i)^2Q(X^j)^2r(X^i, X^j)^2 + Q^2(X^i)\text{IF}(z; Q, F_j)^2r(X^i, X^j)^2 \\ &\quad + 2r(X^i, X^j)^2Q(X^i)Q(X^j)\text{IF}(z; Q, F_i)\text{IF}(z; Q, F_j) \\ &\quad + 2Q(X^i)Q(X^j)^2r(X^i, X^j)\text{IF}(z; Q, F_i)\text{IF}(z; r, F_{i,j}) \\ &\quad + 2Q(X^j)Q(X^i)^2r(X^i, X^j)\text{IF}(z; Q, F_j)\text{IF}(z; r, F_{i,j}) \\ &\quad + Q(X^i)^2Q(X^j)^2\text{IF}(z; r, F_{i,j})^2. \end{aligned}$$

Integrating this expression, we have

$$\begin{aligned} \text{ASV}(\widehat{\Sigma}_{i,j}, F) &= \text{ASV}(Q, F_i)Q(X^j)^2r(X^i, X^j)^2 + \text{ASV}(Q, F_j)Q(X^i)^2r(X^i, X^j)^2 \\ &\quad + 2r(X^i, X^j)^2Q(X^i)Q(X^j) \int \text{IF}(z; Q, F_i)\text{IF}(z; Q, F_j) dF_{i,j}(z) \\ &\quad + 2Q(X^i)Q(X^j)^2r(X^i, X^j) \int \text{IF}(z; Q, F_i)\text{IF}(z; r, F_{i,j}) dF_{i,j}(z) \\ &\quad + 2Q(X^j)Q(X^i)^2r(X^i, X^j) \int \text{IF}(z; Q, F_j)\text{IF}(z; r, F_{i,j}) dF_{i,j}(z) \\ &\quad + Q(X^i)^2Q(X^j)^2\text{ASV}(r, F_{i,j}). \end{aligned}$$

Note that, if  $i = j$ , we have

$$\int \text{IF}(z; Q, F_i)\text{IF}(z; Q, F_i) dF_i(z) = \text{ASV}(Q, F_i)$$

Moreover, if  $r(X^i, X^j) = 0$ , the expression becomes

$$\text{ASV}(\widehat{\Sigma}_{i,j}, F) = Q(X^i)^2Q(X^j)^2\text{ASV}(r, F_{i,j}).$$

As for the influence function, we notice that the ASV of a component of the covariance matrix is a function of the ASV of the scale and correlation estimators.

The asymptotic variances of the different correlation estimators have been obtained in [31], and are given by

$$\begin{aligned} \text{ASV}(\widehat{\rho}_P, \Phi_\rho) &= (1 - \rho^2)^2 \\ \text{ASV}(\widehat{\rho}_Q, \Phi_\rho) &= (1 - \rho^2) \left( \frac{\pi^2}{4} - \arcsin^2(\rho) \right) \\ \text{ASV}(\widehat{\rho}_K, \Phi_\rho) &= \pi^2(1 - \rho^2) \left( \frac{1}{9} - \frac{4}{\pi^2} \arcsin^2\left(\frac{\rho}{2}\right) \right) \\ \text{ASV}(\widehat{\rho}_S, \Phi_\rho) &= \frac{\pi^2}{9} \left( 1 - \frac{\rho^2}{4} \right) 144 \left\{ \frac{1}{144} - \frac{9}{4\pi^2} \arcsin^2\left(\frac{\rho}{2}\right) \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\pi^2} \int_0^{\arcsin(\frac{\rho}{2})} \arcsin\left(\frac{\sin(x)}{1+2\cos(2x)}\right) dx \\
& + \frac{2}{\pi^2} \int_0^{\arcsin(\frac{\rho}{2})} \arcsin\left(\frac{\sin(2x)}{\sqrt{1+2\cos(2x)}}\right) dx \\
& + \frac{1}{\pi^2} \int_0^{\arcsin(\frac{\rho}{2})} \arcsin\left(\frac{\sin(2x)}{2\sqrt{\cos(2x)}}\right) dx \\
& + \frac{1}{2\pi^2} \int_0^{\arcsin(\frac{\rho}{2})} \arcsin\left(\frac{3\sin(x) - \sin(3x)}{4\cos(2x)}\right) dx \Bigg\}.
\end{aligned}$$

Note that, since the Gaussian Rank estimator and the Pearson correlation estimator have the same influence function, they also have the same asymptotic variance. The asymptotic variance of the  $Q_n$  estimator does not seem to be evaluated in the literature, so, for numerical experiment, we would need to approximate it using

$$\text{ASV}(Q_n, F) = \int \text{IF}(Q_n, F, z)^2 dF(z).$$

### 3.7.4 Numerical illustrations

We can now plug in this expression in the influence function of the graphical lasso (3.9). Note that we will use for illustrations based on the graphical lasso the matrix

$$\Sigma = \begin{bmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{bmatrix}$$

whose inverse is given by

$$\Omega = \begin{bmatrix} 4/3 & -2/3 & 0 \\ -2/3 & 5/3 & -2/3 \\ 0 & -2/3 & 4/3 \end{bmatrix}.$$

We consider this since the precision matrix has a sparse structure. Moreover, the eigenvalues of  $\Omega$  are equal to  $\lambda_1 = 2.457$ ,  $\lambda_2 = 1.333$  and  $\lambda_3 = 0.542$ . Since  $g(\lambda_1) = 6.788 > -0.232 = g(\lambda_3)$ , it is the eigenvector corresponding to the largest eigenvalue which would maximise the norm of the influence function, if we considered  $\lambda = 0$ . Since  $\Omega = \Sigma^{-1}$ , this eigenvector corresponds to the eigenvector related to the smallest eigenvalue of  $\Sigma$ , which maximises the norm of the influence function of the classical covariance matrix. This vector is therefore a good candidate to introduce contamination. We also the penalty parameter at  $\lambda = 3.10^{-3}$ .

In the FIGURE 9, we illustrate the influence function for the graphical lasso estimator based on several covariance matrix estimators. We plot the norm, but also the influence of the first diagonal element and the influence of a non-zero off-diagonal element. We increased the norm of the outlier in the direction. Looking at the previous section, we chose as the direction the eigenvector corresponding to the smallest eigenvalue of  $\Sigma$ . For both the classical and Gaussian Rank estimator, the influence is unbounded, which is consistent with the fact that their own influence function is unbounded. The estimator based on the Kendall and Pearson correlation estimators display similar behaviour, whereas the quadrant looks slightly more robust. However, similar to the latter's influence function, the influence of an outlier of small norm is bigger for

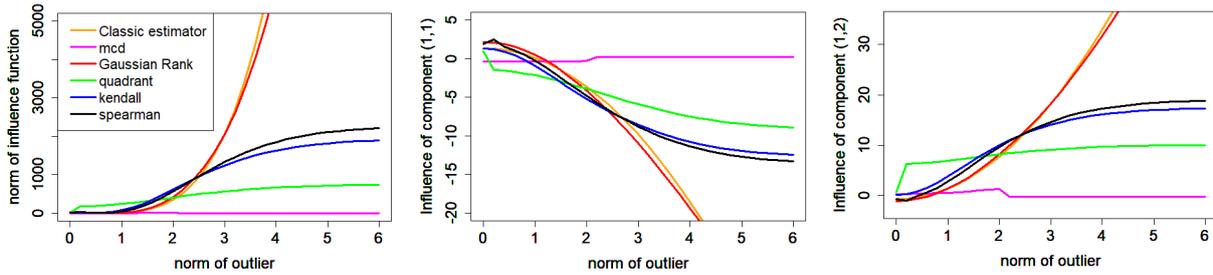


Figure 9 – Influence functions of the graphical lasso based on several covariance matrix estimators.

the quadrant than for the other estimators. At last, the graphical lasso estimator based on the MCD delivers the best influence function, since the variation is very small compare to the others, and is even constant when the norm is bigger than 2.

We also computed the sensitivity curve in FIGURE 10. We have computed it for different estimators and the results are reported in the FIGURE 9. The MCD is less affected by outliers than the other estimators, but it is however more time-consuming so we didn't produce the sensitivity curve for this latter. The results for the sensitivity curves of all estimators apart from the Quadrant are similar to the results on the influence functions displayed in FIGURE 9. However, the result is quite different for the covariance matrix estimator based on the Quadrant correlation estimator. Indeed, the sensitivity curve of the Quadrant estimator changes for fixed parameter  $(n, p, \lambda, \dots)$ , and so depends on the sample.

In the FIGURE 11, we compute the norm of the influence function of the graphical lasso based on several covariance matrix estimators. We considered an outlier  $z = (z_1, z_2, 0)$ . We note that, when using the empirical covariance matrix, the influence is the highest compare to the other estimators. As for the two-dimensional plot, the norm of the influence function when using the MCD estimator is very small. Again, the plots based on the Kendall and Spearman correlation estimators are very similar. At last, we note that the influence function of the graphical lasso using the Quadrant correlation estimator is discontinuous on the axis  $z_1 = 0$  and  $z_2 = 0$  because of the sign function in the definition of the correlation estimator. We also illustrate in FIGURE 12 the sensitivity curves. We observe that the shapes are again very similar to the one obtained in the previous pictures. For the same reasons as mentioned before, we didn't plot the sensitivity curve for the Quadrant and MCD estimators.

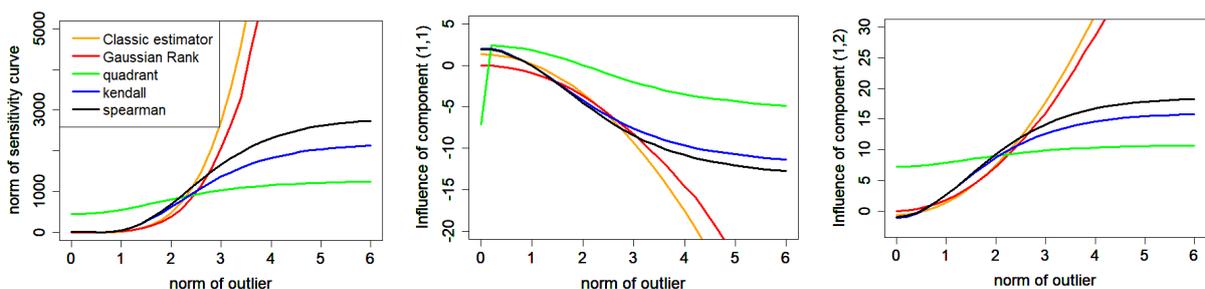
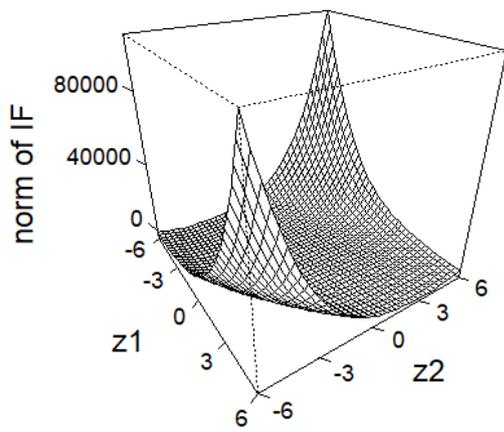
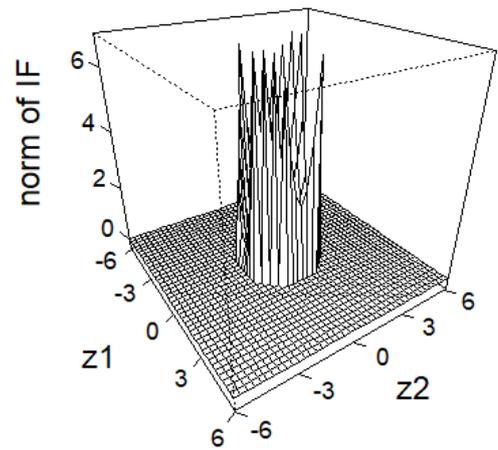


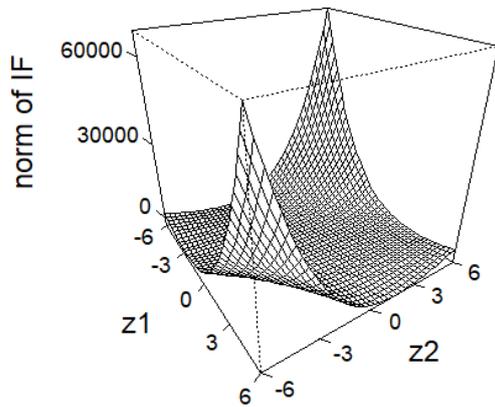
Figure 10 – Sensitivity curves of the graphical lasso based on several covariance matrix estimators.



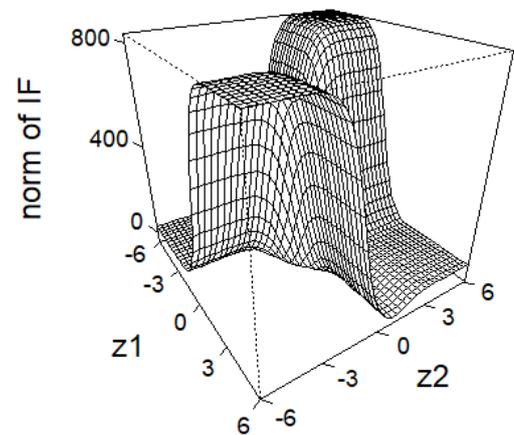
Classical estimator



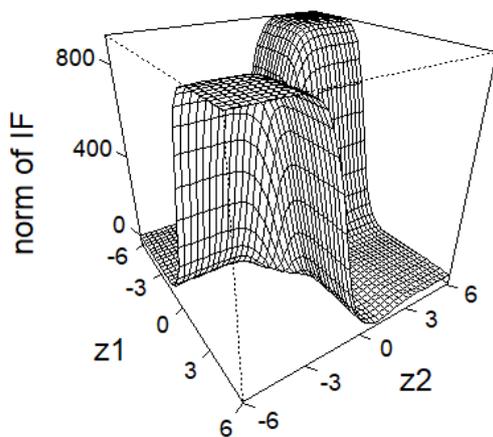
MCD



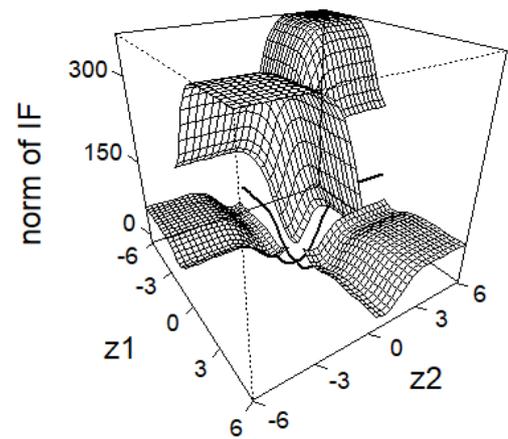
Gaussian rank correlation estimator



Kendall correlation estimator

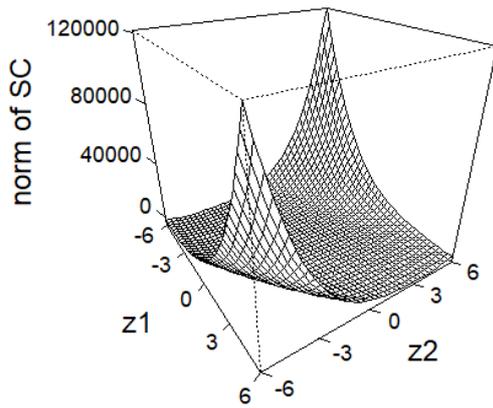


Spearman correlation estimator

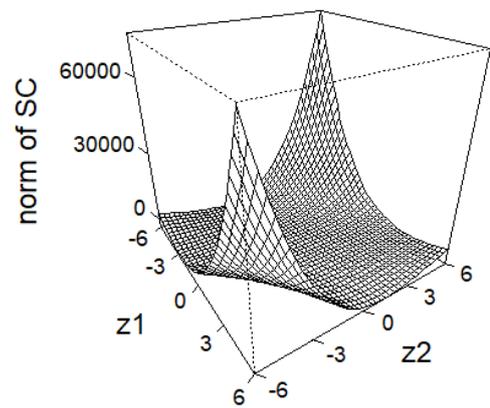


Quadrant correlation estimator

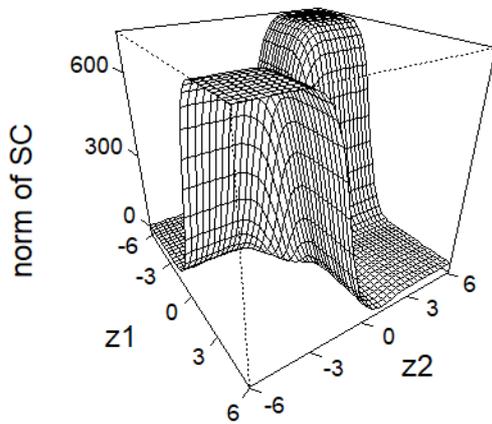
Figure 11 – Influence functions of the graphical lasso based on several covariance matrix estimators



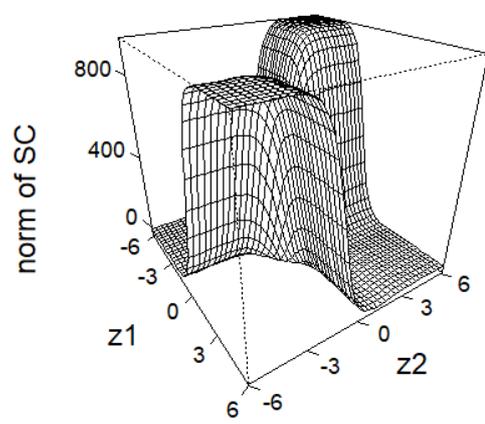
Classical estimator



Gaussian rank correlation estimator



Kendall correlation estimator



Spearman correlation estimator

Figure 12 – Sensitivity curve of the graphical lasso estimator based on several covariance matrix estimators

### 3.8 Conclusion

In this third chapter, we introduced the graphical lasso, a common method for estimating sparse graphical models. We then introduced the notion of conditional dependence and the estimator used in this convex penalisation method. Afterwards, we gave the expression of the influence function of the MLE of  $\Omega$ , which appears when we consider  $\lambda = 0$ , and then we derive the influence function of the graphical lasso for  $\lambda > 0$ . The final expression depends on the estimation of the precision matrix and also on the influence function of the covariance matrix estimator used in the expression of the graphical lasso. We provided an upper bound for the gross error sensitivity, which also depend on the estimation of the precision matrix and the gross error sensitivity of the covariance matrix estimator. In a next section, we gave several propositions concerning the directions maximising the norm of the influence functions. It appeared that, when  $\lambda = 0$  or  $\lambda$  is big enough such that the estimation is diagonal, the direction of interest corresponds to the eigenvector related to the biggest or the smallest eigenvalue of  $T_\Omega$ . We also provided the asymptotic variance, which again depends on the estimation of the precision matrix and the ASV of the covariance matrix estimator. In the last section, we introduced robust covariance matrix

estimators based on pairwise correlations. We worked with different correlation estimators and provided some numerical experiments based on the influence function of the graphical lasso when using these robust covariance matrix estimators. We looked at the norm of the influence in a specific direction, and we presented three-dimensional plots of the influence functions for contamination in  $z = (z_1, z_2, 0)$ . Overall, we noticed that the influence function of the graphical lasso inherits the boundedness of the influence function of the covariance matrix estimator used.



# Conclusion

En conclusion, nous avons dans le premier chapitre de ce document introduit les outils permettant d'évaluer la robustesse, comme le point de rupture et la fonction d'influence. Nous avons illustré ces outils en utilisant les estimateurs de position que sont la moyenne et la médiane. Nous avons pu ainsi observer que la médiane est théoriquement plus robuste que la moyenne. Nous avons ensuite introduit les M-estimateurs et les L-estimateurs, pour lesquels nous avons livré la preuve de la fonction d'influence. Nous avons de plus défini des estimateurs de dispersion, comme le MAD, très utilisé et hautement robuste. Enfin, nous avons généralisé les différents concepts à des dimensions supérieures, où nous avons aussi introduit un estimateur de matrice de covariance robuste, le MCD.

Dans le second chapitre, nous avons présenté les fonctions de profondeur, plus particulièrement la profondeur de demi-espace et la profondeur de scatter. Nous avons pour la première fois montré qu'il s'agit d'une procédure robuste, car sa fonction d'influence est bornée. Pour la seconde, nous avons livré la preuve de la fonction d'influence. Dans le cas où l'estimateur de position dont elle dépend est connu et fixé, la fonction d'influence obtenue a la même expression que la fonction d'influence de la profondeur de demi-espace. Lorsque la distribution considérée est discrète, nous avons constaté qu'il était difficile d'obtenir une expression générale. Enfin, nous avons obtenu des bornes pour la fonction d'influence lorsque la distribution est absolument continue par rapport à la mesure de Lebesgue. Nous avons pu constater que la dépendance en la fonction d'influence de l'estimateur de position disparaît si la distribution est symétrique. Enfin, nous avons aussi livré l'expression de la fonction d'influence empirique. En vue de continuer ce travail, nous pourrions implémenter numériquement les résultats obtenus afin d'avoir une idée de la valeur des bornes de la fonction d'influence par exemple. Nous pouvons aussi implémenter la variance asymptotique et observer l'impact qu'a l'utilisation de différents estimateurs de position sur sa valeur.

Enfin, nous avons présenté dans le dernier chapitre le graphical lasso, une méthode pour estimer les liens entre les noeuds d'un réseau sparse. Nous avons commencé par introduire la notion de dépendance conditionnelle, et ensuite nous avons illustré l'estimateur utilisé dans ce contexte. Nous avons discuté de sa robustesse au travers de sa fonction d'influence. Nous avons ainsi constaté que cette dernière dépend de la fonction d'influence de l'estimateur de matrice de covariance utilisé. Nous avons de plus discuté de la valeur maximale que prend la norme de la fonction d'influence ainsi que de la direction de la contamination livrant cette plus grande valeur. Nous avons déduit l'efficacité asymptotique et illustré les différents résultats obtenus avec des estimateurs robustes de matrice de covariance. Ceux-ci sont obtenus en utilisant conjointement un estimateur de dispersion et de corrélation. Nous avons ainsi constaté que si l'estimateur possède une fonction d'influence bornée, la fonction d'influence du graphical lasso l'est aussi.



# Bibliographie

- [1] JURECKOVA, J., PICEK, J., *Robust statistical method with R*, 2005
- [2] R. HAMPEL, F., et al., *Robust statistics : The approach based on influence functions*, 1986
- [3] J. HUBERT, P., M. RONCHETTI, E., *Robust statistics*, 2009
- [4] ANDRÉ, L., *Outlier robust unit root analysis*, Thesis Publishers, 1996
- [5] HAESBROECK, G., *Moyenne ou médiane ?*, slides pour le Congrès de la SBPMef, 2017
- [6] CROUX, C., *Limit behaviour of the empirical influence function of the median*, *Statistics & Probability Letters*, Volume 37, 331-340, 1998
- [7] HAESBROECK, G., CROUX, C., *Influence function and efficiency of the minimum covariance determinant scatter matrix estimator*, *Journal of Multivariate Analysis*, 71, 161-190, 1999
- [8] TYLER, D., *A distribution-free M-estimator of multivariate scatter*, *The Annals of Statistics*, 15, 1, 234-251, 1987
- [9] ZUO, Y., SERFLING, R., *General notions of statistical depth function*, *Journal of Multivariate Analysis*, 28, 2, 461-482, 2000
- [10] ROMANAZZI, M., *Influence Function of Halfspace Depth*, *Journal of Multivariate Analysis*, 77, 138-161, 2001
- [11] ZHIQIANG C., TYLER., D., *The influence function and maximum bias of Tukey's median*, *Ann. Statist.*, 30 (6), 1737 - 1759, 2002
- [12] CHEN, M., GAO, C., REN, Z., *Robust Covariance and Scatter Matrix Estimation under Huber's Contamination Model*, *The Annals of Statistics*, 46, 5, 1932-1960, 2018
- [13] PAINDAVEINE, D., VAN BEVER, G., *Halfspace depths for scatter, concentration and shape matrices*, *The Annals of Statistics*, 2017
- [14] BOOS, D., *A converse to Scheffé's theorem*, *Ann. Statist.*, 13 (1), 423-427, 1985
- [15] SWEETING, T.J., *On a converse to Scheffé's theorem*, *Ann. Statist.*, 14 (3), 1252-1256, 1986
- [16] LIVSHYTS, G., PAOURIS, G., PIVOVAROV, P., *On sharp bounds for marginal densities of product measures*, arXiv, 2015
- [17] DANN, S., PAOURIS, G., PIVOVAROV, P., *Bounding marginal densities via affine isoperimetry*, arXiv, 2015
- [18] FRIEDMAN, J., HASTIE, T., AND TIBSHIRANI, R., *Sparse inverse covariance estimation with the graphical lasso*, *Biostatistics* 9 (3), 432-441, 2008

- [19] YUAN, M. AND LIN, Y., *Model selection and estimation in the Gaussian graphical model*, *Biometrika*, 94, 19-35, 2007
- [20] ROTHMAN, A. J., BICKEL, P. J., LEVINA, E. AND ZHU, J., *Sparse permutation invariant covariance estimation*, *Electronic Journal of Statistics*, 2, 494-515, 2008
- [21] BANERJEE, O., EL GHAOUI, L., D'ASPREMONT, A., *Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data*, *The Journal of Machine Learning, Research* 9, 485-516, 2008
- [22] TARR, G., S. MULLER, AND N. C. WEBER, *Robust estimation of precision matrices under cellwise contamination*, *Computational Statistics & Data Analysis* 93, 404-420, 2016
- [23] OLLERER, V. AND C. CROUX, *Robust High-Dimensional Precision Matrix Estimation*, pp. 325-350. Cham : Springer International Publishing, 2015
- [24] PO-LING LOH AND XIN LU TAN, *High-dimensional robust precision matrix estimation : Cellwise corruption under  $\epsilon$ -contamination*, arXiv, 2015
- [25] MEDINA, A., ANDRÉS, M., *Robust penalized M-estimators for generalized linear and additive models*, thèse de doctorat : Univ. Genève, 2016, no. GSEM 20
- [26] KOLLO, T., VON ROSEN, D., *Advanced Multivariate Statistics with Matrices*, Springer, 2005
- [27] OLIVEIRA, O., *The Implicit Function Theorem for Maps that are Only Differentiable : An Elementary Proof*, *Real Analysis Exchange*, Vol. 43(2), pp. 429-444, 2018
- [28] SRINIVASAN, S., PANDA, N., *What is the gradient of a scalar function of a symmetric matrix*, May 2020
- [29] NEUDECKER, H., WESSELMAN, A.M., *The Asymptotic Variance Matrix of the Sample Correlation Matrix*, *Linear Algebra and its Applications* 127, 589-599, 1990
- [30] ROUSSEEUW, P.J., CROUX, C., *Alternatives to the median absolute deviation*, *Journal of the American Statistical Association*, 88(424) :1273-1283, 1993
- [31] CROUX, C., DEHON, C., *Influence functions of the Spearman and Kendall correlation measures*, *Stat Methods Appl* 19, 497-515, 2010