

## REMERCIEMENTS

Je remercie en premier lieu, mon dieu qui a bien voulu me donner la force pour effectuer le présent travail.

Mes remerciements vont en second lieu à *M<sup>elle</sup>* **F. Maref** qui a accepté de diriger ce mémoire. Je la remercie de sa disponibilité, de sa patience et de son intérêt pour ce travail.

Je tient également à exprimer ma gratitude en vers *Monsieur* **A. Kandouci** d'avoir accepter d'être le président de ce jury.

Je remercie aussi *M<sup>elle</sup>* **R. Rouane**, *M<sup>elle</sup>* **S. Rahmani** qui ont bien voulu accepter de juger ce travail et de faire parties de ce jury.

# Table des matières

<b>Introduction</b>	<b>3</b>
<b>1 Présentation</b>	<b>6</b>
1.1 Modèle . . . . .	6
1.2 Notations et définitions . . . . .	7
1.3 Outils . . . . .	8
<b>2 Cas i.i.d.</b>	<b>10</b>
2.1 Convergence presque complète . . . . .	10
2.1.1 Hypothèses . . . . .	10
2.1.2 Résultat . . . . .	11
2.1.3 Démonstration . . . . .	11
2.2 Normalité asymptotique . . . . .	16
2.2.1 Hypothèses et résultat . . . . .	16
2.2.2 Démonstration . . . . .	18
<b>3 Cas dépendant</b>	<b>25</b>
3.1 Propriétés asymptotiques . . . . .	25
3.1.1 Démonstration des résultats techniques . . . . .	26
3.2 Préviation en série temporelle . . . . .	30
<b>4 Applications</b>	<b>31</b>
4.1 Application sur des données simulées . . . . .	31
4.1.1 Données . . . . .	31
4.1.2 Algorithme . . . . .	33
4.1.3 Résultat . . . . .	33
4.2 Application sur des données réelles . . . . .	35

# Introduction

Connaître une variable aléatoire  $Y$  sachant une variable aléatoire  $X$  est un sujet très important en statistique. En statistique non paramétrique, la régression est l'outil principal pour répondre à cette question. Cependant, cet outil n'est pas très adapté à certaines situations. Par exemple, la présence de données aberrantes peut amener à des résultats non pertinents. La régression robuste a été introduite pour résoudre ce genre de problèmes. Le but de ce travail est d'étudier l'estimation robuste de la fonction de régression, dans le cas où la variable explicative est à valeurs dans un espace de dimension infinie.

Historiquement, les premiers résultats conséquents, sur la régression robuste, furent établis au début des années soixante par Huber ((1964), [17]), dont il a obtenu la consistance et la normalité asymptotique d'une classe d'estimateurs pour cette fonction. Robinson ((1984), [24]), Härdle ((1984), [14]) et Härdle et Tsybakov ((1989), [15]) ont établi sous des conditions de mélange la normalité asymptotique d'une famille d'estimateurs issue de la méthode du noyau pour la fonction de régression. Parallèlement, Boente et Fraiman ((1989), [6]), (1990, [7])) ont utilisé l'estimateur de Robinson (1984) pour étudier simultanément les deux paramètres de position et d'échelle. La consistance des estimateurs construits est obtenue sous des conditions générales et dans les deux cas indépendants et fortement mélangeants. La convergence uniforme de l'estimateur robuste de la fonction de régression a été obtenu par Collomb et Härdle ((1986), [10]) en considérant des observations  $\phi$ -mélangeantes. Laïb et Ould-Saïd ((2000), [21]) ont adapté l'estimateur de Collomb et Härdle ((1986), [10]) pour le modèle d'auto-régression d'un processus stationnaire ergodique. Ils ont obtenu la convergence uniforme de cet estimateur même lorsque la fonction objective est non bornée. Cai et Ould-Saïd ((2003), [9]) ont utilisé une version robuste de l'estimation par la méthode des polynômes locaux pour la fonction de la régression, ils ont démontré sous des conditions standards et lorsque les observations sont alpha-mélangeantes la normalité asymptotique et la convergence presque sûr de ces estimateurs. On

trouvera également dans cet article un résultat de convergence en moyenne quadratique ainsi que le paramètre de lissage optimisant cette erreur. Récemment, Azzedine et al. ((2008), [5]) ont étudié la convergence presque complète d'estimateurs robustes à noyau. Dans le même cadre, Attouch et al. ((2007), [3] et [4]) ont étudié la normalité asymptotique de ces estimateurs, en considérant les deux types d'observations indépendantes et  $\alpha$ -mélangeantes.

Dans ce travail, on se propose d'étudier l'estimateur à noyau de la fonction de régression robuste dans le cas où la variable explicative est une courbe, en considérant les deux cas : le cas où les observations sont indépendantes identiquement distribuées et le cas des observations fortement mélangeantes.

Ce mémoire est présenté en quatre chapitres, il développe aussi bien des aspects pratiques que théoriques et il est organisé comme suit :

Dans le premier chapitre, on présente notre modèle et on donne une liste des définitions permettant de fixer le vocabulaire utilisé dans la suite de document. Cette liste contient entre autre, les définitions des différents processus de mélange à savoir le  $\phi$ -mélange,  $\rho$ -mélange,  $\alpha$ -mélange, les définitions des différents modes de convergence, . . . On trouvera aussi dans ce chapitre un nombre très important des outils intéressants pour l'élaboration des résultats obtenus.

Dans le deuxième chapitre, on considère le cas des observations indépendantes identiquement distribuées et on établit sous des conditions générales la convergence presque complète et la normalité asymptotique. Ces propriétés asymptotiques sont étroitement liées au phénomène de concentration de la mesure de probabilité de la variable explicative sur des petites boules. Ces résultats nous permettent d'établir une solution originale au problème du fléau de la dimension et de généraliser à la dimension infinie de nombreux résultats asymptotiques existants dans le cas multivarié.

Dans le troisième chapitre, on abordera le cas des observations dépendantes. Dans ce chapitre, nous avons opté pour des conditions un peu plus restrictives, mais, qui donne la même vitesse de convergence du cas i.i.d. On explicitera également comment notre démarche pourrait être appliqué à la prévision en séries temporelles.

Les résultats obtenus dans ce mémoire seront ponctués par des exemples d'application sur des

---

données simulées au réelles. Deux exemples seront étudié dans le dernier chapitre, le premier correspond à des données simulées dont l'objectif est de faire une comparaison entre la régression classique et la régression robuste. L'autre exemple correspond au cas des données réelles.

# Chapitre 1

## Présentation

### 1.1 Modèle

Soit  $(X, Y)$  un couple de variables aléatoires à valeurs dans  $\mathcal{F} \times \mathbb{R}$ , où  $\mathcal{F}$  est un espace semi-métrique. On note  $d$  la semi métrique sur  $\mathcal{F}$ . Pour  $x \in \mathcal{F}$ , on considère une fonction réelle, mesurable, notée  $\psi_x$ . Le paramètre fonctionnel étudié dans ce travail, noté  $\theta_x$ , est la solution (supposée unique) de l'équation en  $t$  définie par :

$$(1.1) \quad \Psi(t, x) = \mathbb{E}(\psi_x(Y - t) / X = x) = 0.$$

En général, la fonction  $\psi_x$  est fixée par le statisticien en fonction de la situation à laquelle il est confronté. Des exemples classiques de  $\psi_x$  conduisent à l'estimation de la moyenne conditionnelle (si  $\psi_x(t) = t$ ) ou de quantiles conditionnels (si  $\psi_x(t) = \mathbb{I}_{[0, +\infty[}(t) - (1 - \alpha)$ ,  $\alpha \in [0, 1]$ ), voir Ferraty et Vieu (2006) et Attouch et al. (2007).

Etant donné  $(X_1, Y_1), \dots, (X_n, Y_n)$  une suite des observations de même loi que le couple  $(X, Y)$ , l'estimateur à noyau de  $\Psi(t, x)$  est donnée par :

$$\widehat{\Psi}(t, x) = \frac{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))\psi_x(Y_i - t)}{\sum_{i=1}^n K(h_K^{-1}d(x, X_i))}, \quad \forall t \in \mathbb{R},$$

où  $K$  est un noyau et  $h_K = h$  est une suite de nombres réels positifs tend vers 0 quand  $n$  tend vers à l'infini.

L'estimateur naturel de  $\theta_x$ , noté  $\widehat{\theta}_x$ , est tel que

$$(1.2) \quad \widehat{\Psi}(\widehat{\theta}_x, x) = 0.$$

Quand  $\psi_x$  est égale à l'identité,  $\widehat{\theta}_x$  est identifié à l'estimateur de Ferraty et Vieu (2006) [12] pour la régression fonctionnelle.

## 1.2 Notations et définitions

Soient  $(\Omega, \mathcal{A}, \mathbb{P})$  un espace de probabilité et  $\{\Delta_i\}_{i \in \mathbb{Z}}$  une famille des variables aléatoires définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans un espace probabilisable  $(E, \xi)$ . On note  $(\sigma_i^j)_{i \neq j \in \mathbb{Z} \cup \{-\infty, +\infty\}}$ , la tribu engendrée par  $\{\Delta_k, i < k < j\}$  et par  $L_2(\sigma_i^j)$  l'espace des variables aléatoires  $\sigma_i^j$ -mesurables et de carrée sommable.

**Définition 1.2.1** *On dit que la famille  $\{\Delta_i, i \in \mathbb{Z}\}$  est  $\varphi$ -mélangeante (resp.  $\rho$ -mélangeante) si la suite*

$$\varphi(n) = \sup_{k \in \mathbb{Z}} \sup_{\{A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{\infty}\}} |\mathbb{P}(B/A) - \mathbb{P}(B)|$$

tend vers 0 quand  $n$  tend vers à l'infinie (resp.

$$\rho(n) = \sup_{k \in \mathbb{Z}} \sup_{\{X \in L_2(\sigma_{-\infty}^k), Y \in L_2(\sigma_{n+k}^{\infty})\}} \frac{\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]}{[\text{Var}(X)\text{Var}(Y)]^2}$$

tend vers 0 quand  $n$  tend vers à l'infinie).

**Définition 1.2.2** *On dit que la famille  $\{\Delta_i, i \in \mathbb{Z}\}$  est  $\alpha$ -mélangeante si la suite*

$$\alpha(n) = \sup_{\{k \in \mathbb{Z}, A \in \sigma_{-\infty}^k, B \in \sigma_{n+k}^{+\infty}\}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

tend vers 0 quand  $n$  tend vers à l'infinie. La suite  $\alpha_n$  est appelée coefficient de mélange forte.

La relation entre ces trois types des processus est donnée par le lemme suivant :

**Lemme 1.2.1** [12] *Soit  $\{\Delta_i\}_{i \in \mathbb{Z}}$  une famille des variables aléatoires définie sur  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans un espace probabilisable  $(E, \xi)$ . Pour tout  $n \in \mathbb{N}$ , on a*

$$\alpha(n) \leq \rho_n \leq 2\varphi_n^{1/2}.$$

En vertu de ce lemme on peut dire que :

$$\begin{aligned} \text{le processus } \varphi - \text{mélange} &\Rightarrow \text{le processus } \rho - \text{mélange} \\ &\Rightarrow \text{le processus } \alpha - \text{mélange.} \end{aligned}$$

Autrement dit, le processus  $\alpha$ -mélangeant est le plus fort. Ainsi, nous allons modéliser la notion de la dépendance par ce processus  $\alpha$ -mélangeant, car elle englobe les autres processus définis ci-dessus.

**Définition 1.2.3** *On dit que la suite  $(X_n)_{n \in \mathbb{N}}$  converge presque complètement vers  $X$  si*

$$\forall \epsilon > 0, \quad \sum_{n=0}^{\infty} \mathbb{P}(|X_n - X| > \epsilon) < \infty$$

et on note  $X_n \xrightarrow{p.co.} X$ .

**Définition 1.2.4** *On dit que la suite  $X_n = O(Y_n)$  en p.co. s'il existe un  $\epsilon > 0$  vérifiant :*

$$\sum_{n=0}^{\infty} \mathbb{P}(X_n > \epsilon Y_n) < \infty.$$

**Remarque 1.2.1** *Soient  $X_n, Y_n$  deux suites des variables aléatoires. La suite*

$$\left( \frac{X_n}{Y_n} \right)_{n \in \mathbb{N}} \rightarrow 0 \text{ en p.co si}$$

$$X_n \rightarrow 0 \text{ en p.co.}$$

et

$$\exists \delta > 0, \quad \sum_{n=0}^{\infty} \mathbb{P}(|Y_n| < \delta) < \infty.$$

### 1.3 Outils

**Lemme 1.3.1** [16] *Soit  $(\Delta_n)_{n \in \mathbb{N}}$  une suite des variables aléatoires réelles centrées, indépendantes et identiquement distribuées, telles qu'il existe deux réels positifs  $d$  et  $\delta$  vérifiant :*

$$|\Delta_1| \leq d \text{ et } \mathbb{E}\Delta_1^2 \leq \delta^2,$$

alors, pour tout  $\epsilon \in ]0, \frac{\delta^2}{d}[$ , on a

$$\mathbb{P} \left[ n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon \right] \leq 2e^{-\frac{n\epsilon^2}{4\delta^2}}.$$

Le lemme suivant donne l'inégalité de Fuk-Nageav.

**Lemme 1.3.2** [16] *Soit  $\{\Delta_i, i \in \mathbb{N}\}$  une suite des variables aléatoires réelles  $\alpha$ -mélangeante, de coefficient de mélange  $\alpha(n)$  vérifiant :*

$$\exists c \in \mathbb{R}^{*+}, \quad a \in \mathbb{R}^{*+} \quad \alpha(n) \leq cn^{-a}$$

et si  $\forall i, \|\Delta_i\|_\infty < \infty$ , alors, pour tout  $\epsilon > 0$  et  $r > 0$ , on a

$$(1.3) \quad \mathbb{P} \left[ \left| \sum_{k=1}^n \Delta_k \right| > 4\epsilon \right] \leq \left( 1 + \frac{\epsilon^2}{rS_n^2} \right)^{\frac{-r}{2}} + 2ncr^{-1} \left( \frac{2r}{\epsilon} \right)^{a+1}$$

où

$$S_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\text{cov}(\Delta_i, \Delta_j)|.$$

L'inégalité du lemme suivant s'appelle inégalité de covariance et elle est très utile pour le calcul de  $S_n^2$ , définie dans le lemme précédent.

**Lemme 1.3.3** [12] *Soit  $\{\Delta_i, i \in \mathbb{N}\}$  une suite des variables aléatoires réelles  $\alpha$ -mélangeante, de coefficient de mélange  $\alpha(n)$ , telle que  $\|\Delta_i\|_\infty < \infty, \forall i$ . On a, pour tout  $i \neq j$  :*

$$|\text{cov}(\Delta_i, \Delta_j)| \leq 4\|\Delta_i\|_\infty \|\Delta_j\|_\infty \alpha(|i - j|).$$

**Inégalité  $C_r$**  [23]

$$\mathbb{E}[|X + Y|^r] \leq c_r \mathbb{E}[|X|^r] + c_r \mathbb{E}[|Y|^r]$$

où  $c_r = 1$  (resp.  $2^{r-1}$ ) selon que  $r \leq 1$  (resp.  $r \geq 1$ ).

# Chapitre 2

## Cas i.i.d.

Le but principal de ce chapitre est d'étudier l'estimateur à noyau de la fonction de régression robuste, dans le cas où les observations sont indépendantes et identiquement distribuées. Ce chapitre est divisé en deux sections : Dans la première section, on étudie la convergence presque complète. On traite la normalité asymptotique de notre estimateur dans la deuxième section.

### 2.1 Convergence presque complète

#### 2.1.1 Hypothèses

On fixe un point  $x \in \mathcal{F}$ , on note  $B(x, h) = \{x' \in \mathcal{F} / d(x, x') < h\}$  la boule de centre  $x$  et de rayon  $h$  et on introduit les hypothèses suivantes :

$$(H1) \mathbb{P}(X \in B(x, h)) = \phi_x(h) > 0, \forall h > 0.$$

$$(H2) \text{ Il existe } C_1 > 0 \text{ et } b > 0 \text{ tel que } \forall x_1, x_2 \in \mathcal{N}_x, \forall t \in \mathbb{R}$$

$$| \Psi(t, x_1) - \Psi(t, x_2) | \leq C_1 d^b(x_1, x_2).$$

$$(H3) \text{ La fonction } \psi_x \text{ est continument différentiable, strictement monotone par rapport à la deuxième composante et sa dérivée est telle que } | \psi'_x(t) | > C_2 > 0, \forall t \in \mathbb{R}.$$

$$(H4) K \text{ est une fonction continue de support } [0,1] \text{ et tel que } 0 < C < K(t) < C' < \infty.$$

$$(H5) \lim_{n \rightarrow \infty} h_K = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{\log n}{n \phi_x(h_K)} = 0.$$

## Commentaires

1. L'hypothèse (H1) sur la loi marginale de la variable explicative  $X$  n'est pas très restrictive. En effet, c'est une condition sur la fonction de répartition et non pas sur la densité. Dans le cas fini dimensionnel, l'hypothèse (H1) est même moins restrictive que la positivité stricte de la densité de la variable explicative (une condition indispensable dans le cas fini dimensionnel) car on peut avoir une boule centrée en un point et tel que la densité est nulle bien que la probabilité de cette boule soit strictement positive. Prenons l'exemple de la loi du  $\chi^2$  avec un degré de liberté supérieur à deux.
2. L'hypothèse (H2) est une hypothèse de régularité qui caractérise l'espace fonctionnel de notre modèle, ce qui justifie l'emploi des méthodes non paramétriques pour le problème considéré.
3. La condition (H3) contrôle la robustesse de notre modèle.
4. Les hypothèses (H4) – (H5) sont des hypothèses techniques nécessaires à l'obtention de nos résultats.

### 2.1.2 Résultat

On établit le résultat suivant

**Théorème 2.1.1** *Sous les hypothèses (H1) – (H5),  $\hat{\theta}_x$  existe et est unique pour  $n$  assez grand, et on a*

$$(2.1) \quad \hat{\theta}_x - \theta_x = O(h^b) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right), \quad p.co.$$

### 2.1.3 Démonstration

Dans ce qui suit, on note par  $C$  une constante strictement positive et  $K_i = K\left(\frac{d(x, X_i)}{h}\right)$ .

Sous l'hypothèse (H3), nous avons

$$\hat{\Psi}(\hat{\theta}_x, x) = \hat{\Psi}(\theta_x, x) + (\hat{\theta}_x - \theta_x)\hat{\Psi}'(\xi_{x,n}, x)$$

où  $\xi_{x,n}$  est un point entre  $\widehat{\theta}_x$  et  $\theta_x$ .

La dérivabilité de  $\psi_x$  dans (H3), nous permet d'écrire,

$$\begin{aligned} & \exists C_2 > 0, \forall \epsilon_0 > 0, \mathbb{P} \left( |\widehat{\theta}_x - \theta_x| \geq \epsilon_0 \left( h^b + \sqrt{\frac{\log n}{n\phi_x(h)}} \right) \right) \\ & \leq \mathbb{P} \left( |\widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x)| \geq C_2 \epsilon_0 \left( h^b + \sqrt{\frac{\log n}{n\phi_x(h)}} \right) \right). \end{aligned}$$

Alors, il suffit de montrer que :

$$(2.2) \quad \widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x) = O \left( h^b + \sqrt{\frac{\log n}{n\phi_x(h)}} \right), \quad p.co.$$

La preuve de l'équation (2.2) repose sur la décomposition suivante :

$$(2.3) \quad \begin{aligned} \forall t \in \mathbb{R}, \widehat{\Psi}(t, x) - \Psi(t, x) &= \frac{1}{\widehat{\Psi}_D(x)} [(\widehat{\Psi}_N(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)]) - (\Psi(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)])] \\ &\quad - \frac{\Psi(t, x)}{\widehat{\Psi}_D(x)} [\widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)]] \end{aligned}$$

où

$$\widehat{\Psi}_D(x) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^n K_i, \quad \widehat{\Psi}_N(t, x) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^n K_i \psi_x(Y_i - t).$$

Il est clair que  $\widehat{\Psi}(t, x) = \frac{\widehat{\Psi}_N(t, x)}{\widehat{\Psi}_D(x)}$  et  $\mathbb{E}[\widehat{\Psi}_D(x)] = 1$ .

Finalement, la preuve du théorème 2.1.1 est achevée à partir des lemmes suivants.

**Lemme 2.1.1** *Sous les hypothèses (H1) et (H4) – (H5), on a*

$$\widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)] = O \left( \sqrt{\frac{\log n}{n\phi_x(h)}} \right), \quad p.co.$$

**Preuve du lemme 2.1.1**

Nous avons,

$$\widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)] = \frac{1}{n} \sum_{i=1}^n \left( \frac{K_i}{\mathbb{E}K_1} - 1 \right).$$

On pose

$$\Delta_i = \frac{K_i}{\mathbb{E}K_1}.$$

D'après l'hypothèse (H4), on a

$$\mathbb{E}C\mathbb{I}_{B(x,h)}(X_i) < \mathbb{E}K_1 < \mathbb{E}C'\mathbb{I}_{B(x,h)}(X_i),$$

alors,

$$C\phi_x(h) < \mathbb{E}K_1 < C'\phi_x(h)$$

et puisque la fonction  $K$  est bornée, on peut majorer directement  $|\Delta_i|$  par  $\frac{C}{\phi_x(h)}$  et on montre

$$\text{que } \mathbb{E}\Delta_i^2 \leq \frac{C'}{\phi_x(h)}.$$

On applique maintenant l'inégalité de Hoeffding aux variable  $|\Delta_i|$ ,

$$\begin{aligned} \mathbb{P}\left(|\widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)]| > \epsilon\right) &= \mathbb{P}\left[\frac{1}{n} \left| \sum \Delta_i \right| > \epsilon\right] \\ &\leq 2 \exp\left(\frac{-n\epsilon^2 C}{4\phi_x(h)}\right). \end{aligned}$$

On prend

$$\epsilon = \eta \sqrt{\frac{\log n}{n\phi_x(h)}}.$$

On arrive à

$$(2.4) \quad \mathbb{P}\left[|\widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)]| > \eta \sqrt{\frac{\log n}{n\phi_x(h)}}\right] \leq 2n^{-C\epsilon_0^2},$$

donc,

$$\sum_n \mathbb{P} \left[ \left| \widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)] \right| > \eta \sqrt{\frac{\log n}{n\phi_x(h)}} \right] \leq \sum 2n^{-C\eta^2}.$$

Il suffit de choisir  $\eta > \frac{1}{\sqrt{C}}$  pour que la série converge. ■

**Corollaire 2.1.1** *Sous les hypothèses de lemme 2.1.1, on a*

$$\sum_{n \geq 1} \mathbb{P}(|\widehat{\Psi}_D(x)| \leq 1/2) \leq \sum_{n \geq 1} \mathbb{P}(|\widehat{\Psi}_D(x) - 1| > 1/2) < \infty.$$

### Preuve du corollaire 2.1.1

Nous avons,

$$\mathbb{P} \left( |\widehat{\Psi}_D(x)| \leq 1/2 \right) \leq \mathbb{P} \left( |\widehat{\Psi}_D(x) - 1| > 1/2 \right).$$

Notons que  $\widehat{\Psi}_D(x) - 1 = \widehat{\Psi}_D(x) - \mathbb{E}[\widehat{\Psi}_D(x)]$ , on applique le lemme précédent, on peut écrire

$$\sum_{i=1}^n \mathbb{P} \left( \widehat{\Psi}_D(x) < 1/2 \right) < \infty. \blacksquare$$

**Lemme 2.1.2** *Sous les hypothèses (H1) – (H2) et (H4) – (H5), on a pour tout  $t \in \mathbb{R}$ ,*

$$\Psi(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)] = O(h^b).$$

### Preuve du lemme 2.1.2

L'équidistribution des couples  $(X_i, Y_i)$  et la condition (H4) impliquent

$$\Psi(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)] = \frac{1}{\mathbb{E}[K_1]} \mathbb{E}[(K_1 \mathbb{I}_{B(x,h)}(X_1))(\Psi(t, x) - \mathbb{E}[\psi_x(Y_1 - t)/X = X_1])],$$

où  $\mathbb{I}$  est la fonction indicatrice.

L'hypothèse (H2) nous permet d'écrire,

$$K_1 \mathbb{I}_{B(x,h)}(X_1) | \Psi(t, X_1) - \Psi(t, x) | \leq C_1 h^b,$$

alors,

$$\left| \Psi(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)] \right| \leq C_1 h^b. \quad \blacksquare$$

**Lemme 2.1.3** *Sous les hypothèses (H1) – (H3) et (H5), on a pour tout  $t \in \mathbb{R}$ ,*

$$\widehat{\Psi}_N(t, x) - \mathbb{E}[\widehat{\Psi}_N(t, x)] = O\left(\sqrt{\frac{\log n}{n \phi_x(h)}}\right), \quad p.co.$$

### Preuve du lemme 2.1.3

La preuve de ce résultat est similaire que la preuve du lemme 2.1.1.

On prend

$$\Lambda_i = \frac{\{K_i \psi_x(Y_i - t) - \mathbb{E}[K_1 \psi_x(Y_1 - t)]\}}{\mathbb{E}[K_1]}$$

Puisque la fonction  $\psi_x$  est bornée, alors  $|\Lambda_i| \leq C/\phi_x(h)$  et  $\mathbb{E}[\Lambda_i^2] \leq C'/\phi_x(h)$ , pour tout  $i \leq n$ . Comme le lemme 2.1.1, il suffit d'appliquer l'inégalité de Hoeffding pour obtenir le résultat.  $\blacksquare$

**Lemme 2.1.4** *Sous les hypothèses du théorème 2.1.1,  $\widehat{\theta}_x$  existe et est unique presque sûrement pour  $n$  assez grand.*

### Preuve du lemme 2.1.4

Pour tout  $\epsilon > 0$ , la monotonie stricte de  $\psi_x$  implique

$$\Psi(\theta_x - \epsilon, x) < \Psi(\theta_x, x) < \Psi(\theta_x + \epsilon, x).$$

Les lemmes 2.1.1, 2.1.3, 2.1.4 et le corollaire 2.1.1, montrent que

$$\widehat{\Psi}(\theta_x, x) - \Psi(\theta_x, x) = O\left(h^b + \sqrt{\frac{\log n}{n \phi_x(h)}}\right), \quad p.co.$$

pour tout  $t$  un réel fixé, alors, pour  $n$  suffisamment grand,

$$\widehat{\Psi}(\theta_x - \epsilon, x) \leq 0 \leq \widehat{\Psi}(\theta_x + \epsilon, x) \quad p.co.$$

Puisque  $\psi_x$  et  $K$  sont des fonctions continues, donc  $\widehat{\Psi}(t, x)$  est continue, alors il existe  $t_0 = \widehat{\theta}_x \in [\theta_x - \epsilon, \theta_x + \epsilon]$  tel que  $\widehat{\Psi}(\widehat{\theta}_x, x) = 0$ .

Finalement, l'unicité de  $\widehat{\theta}_x$  est une conséquence directe de la monotonie stricte de  $\psi_x$  et la positivité de la fonction  $K$ . ■

**Remarque 2.1.1** *La démonstration du théorème 2.1.1 peut être divisée en deux parties : une partie biais et une partie dispersion. Dans la partie biais, on utilise le fait que les observations sont identiquement distribuées. Par contre, la démonstration de la partie dispersion repose sur le fait que les observations sont indépendantes.*

## 2.2 Normalité asymptotique

La propriété asymptotique abordée dans cette section est la normalité asymptotique, il s'agit d'un sujet très important en statistique. En effet, la normalité asymptotique nous permet de construire les intervalles de confiance et de faire les tests.

### 2.2.1 Hypothèses et résultat

On suppose que  $\lambda_\gamma(u, t) = \mathbb{E}(\psi^\gamma(Y - t)/X = u)$  et  $\Gamma_\gamma(u, t) = \mathbb{E}((\psi')^\gamma(Y - t)/X = u)$  pour  $\gamma \in \{1, 2\}$ , on gardant les mêmes notations que la section précédente, la même hypothèse (H3) et on remplace (H1), (H2), (H4) et (H5) par :

(H1') Il existe une fonction positive différentiable  $\phi$  et une fonction positive  $g$  telle que :

$$\mathbb{P}(X \in B(x, r)) = \phi(r)g(x) + o(\phi(r)).$$

(H2') i La fonction  $\lambda_\gamma(., .)$  satisfait la condition de Lipschitz par rapport à la première composante, c'est-à-dire : il existe une constante  $b_\gamma$  strictement positive tel que :

$$\forall (u_1, u_2) \in N_x \times N_x, \forall t \in \mathbb{R}, |\lambda_\gamma(u_1, t) - \lambda_\gamma(u_2, t)| \leq C_1 d(u_1, u_2)^{b_\gamma}.$$

ii La fonction  $\Gamma_\gamma(., .)$  satisfait la condition de Lipschitz par rapport à la première composante, c'est-à-dire : il existe une constante  $d_\gamma$  strictement positive tel que :

$$\forall (u_1, u_2) \in N_x \times N_x, \forall t \in \mathbb{R}, |\Gamma_\gamma(u_1, t) - \Gamma_\gamma(u_2, t)| \leq C_2 d(u_1, u_2)^{d_\gamma}.$$

(H4') Le noyau  $K$  est une fonction différentiable sur le support  $[0, 1]$ , sa dérivée  $K'$  existe et telle que  $-\infty < C_3 < K'(t) < C_4 < 0$ .

(H5') Le paramètre de lissage  $h$  satisfait :

$$h \downarrow 0, \quad \forall t \in [0, 1], \quad \lim_{h \rightarrow 0} \frac{\phi(th)}{\phi(h)} = \beta(t) \text{ et } n\phi(h) \rightarrow 0 \text{ quand } n \rightarrow \infty.$$

Notre résultat principal est donné par le théorème suivant.

**Théorème 2.2.1** *Sous les hypothèses (H1') – (H2'), (H3) et (H4') – (H5'),  $\hat{\theta}_x$  existe et est unique avec une probabilité tend vers 1, et pour tout  $x \in \mathcal{A}$ , on a*

$$(2.5) \quad \left( \frac{n\phi(h)}{\sigma^2(x, \theta_x)} \right)^{1/2} (\hat{\theta}_x - \theta_x - B_n(x)) \xrightarrow{D} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

où

$$B_n(x) = \frac{h}{\phi(h)\alpha_1\Gamma_1(x, \theta_x)} \int_0^1 K(t)\varphi_x(th)\phi'(th)dt + o(1)$$

(avec  $\varphi_x(s) = \mathbb{E}[\psi(Y, \theta_x)/d(X, x) = s]$ ),

$$\sigma^2(x, \theta_x) = \frac{\alpha_2\lambda_2(x, \theta_x)}{\alpha_1^2g(x)(\Gamma_1(x, \theta_x))^2}$$

(avec  $\alpha_j = \int_0^1 (K^j)'(s)\beta(s)ds$ , pour  $j = 1, 2$ ),

$$\mathcal{A} = \{x \in \mathcal{F}, g(x)\lambda_1(x, \theta_x)\Gamma_1(x, \theta_x) \neq 0\}$$

et  $\xrightarrow{D}$  denote la convergence en loi.

**Corollaire 2.2.1** *Sous les hypothèses du théorème 2.2.1 et si le paramètre de lissage  $h$  satisfait  $nh^{2b_1}\phi(h) \rightarrow 0$  quand  $n \rightarrow \infty$ , on a*

$$(2.6) \quad \left( \frac{n\phi(h)}{\sigma^2(x, \theta_x)} \right)^{1/2} (\hat{\theta}_x - \theta_x) \xrightarrow{D} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

## 2.2.2 Démonstration

Pour simplifier les notations, on pose, pour  $i = 1 \dots, n$ ,  $K_i = K\left(\frac{d(x, X_i)}{h}\right)$ ,  $\psi_i(t) = \psi(Y_i, t)$

et

$$\widehat{\Psi}_N(x, t) = \frac{1}{n\mathbb{E}[K_1]} \sum_{i=1}^n K_i \psi_i(t).$$

On utilise le développement de Taylor d'ordre 1 autour de  $\theta_x$ , on obtient

$$\widehat{\Psi}_N(x, \widehat{\theta}_x) = \widehat{\Psi}_N(x, \theta_x) + (\widehat{\theta}_x - \theta_x) \widehat{\Psi}'_N(x, \xi_n).$$

avec  $\xi_n \in (\widehat{\theta}_x, \theta_x)$ .

Par définition de  $\widehat{\theta}_x$ , on a

$$\widehat{\theta}_x - \theta_x = \frac{-\widehat{\Psi}_N(x, \theta_x)}{\widehat{\Psi}'_N(x, \xi_n)}.$$

Finalement, on utilise la décomposition suivante

$$\sqrt{n\phi(h)}(\widehat{\theta}_x - \theta_x) = \frac{-\sqrt{n\phi(h)}\left(\widehat{\Psi}_N(x, \theta_x) - \mathbb{E}[\widehat{\Psi}_N(x, \theta_x)]\right)}{\widehat{\Psi}'_N(x, \xi_n)} - \frac{\sqrt{n\phi(h)}\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)]}{\widehat{\Psi}'_N(x, \xi_n)}$$

La preuve du théorème 2.2.1 est basée sur les lemmes suivants.

**Lemme 2.2.1** *Sous les hypothèses (H1') – (H2'), (H3) et (H4') – (H5'), on a pour tout  $x \in \mathcal{A}$*

$$\left(\frac{n\phi(h)\alpha_1^2 g(x)}{\alpha_2 \lambda_2(x, \theta_x)}\right)^{1/2} \left(\widehat{\Psi}_N(x, \theta_x) - \mathbb{E}[\widehat{\Psi}_N(x, \theta_x)]\right) \xrightarrow{D} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

### Preuve du lemme 2.2.1

La démonstration de ce lemme est basée sur le théorème central limite qui exige comme condition suffisante l'existence d'une constante  $\delta > 0$  vérifiant :

$$\frac{\sum_{i=1}^n \mathbb{E} \left[ |L_i(x) - \mathbb{E}[L_i(x)]|^{2+\delta} \right]}{\left( \text{Var} \left( \sum_{i=1}^n L_i(x) \right) \right)^{(2+\delta)/2}} \rightarrow 0$$

où

$$L_i(x) = \frac{1}{n\mathbb{E}[K_1]} [K_i \psi_i(\theta_x)].$$

D'où la nécessité de calculer  $n\phi(h)\text{Var}(\widehat{\Psi}_N(x, t))$ , par définition de  $\widehat{\Psi}_N(x, t)$ , on a

$$\begin{aligned} \text{Var}(\widehat{\Psi}_N(x, t)) &= \frac{1}{(n\mathbb{E}[K_1])^2} \sum_{i=1}^n \text{Var}[K_i \psi_i(t)] \\ &= \frac{1}{n\mathbb{E}^2[K_1]} \text{Var}[K_1 \psi_1(t)] \\ &= \frac{1}{n\mathbb{E}^2[K_1]} \mathbb{E}[K_1^2 \psi^2(Y, t)] - \frac{1}{n\mathbb{E}^2[K_1]} (\mathbb{E}[K_1 \psi(Y, t)])^2 \\ &= \frac{\mathbb{E}[K_1^2]}{n\mathbb{E}^2[K_1]} \mathbb{E} \left[ \frac{K_1^2 \psi^2(Y, t)}{\mathbb{E}[K_1^2]} \right] - \frac{1}{n} \left( \mathbb{E} \left[ \frac{K_1 \psi(Y, t)}{\mathbb{E}[K_1]} \right] \right)^2. \end{aligned}$$

Donc

$$(2.7) \quad n\phi(h)\text{Var}(\widehat{\Psi}_N(x, t)) = \frac{\phi(h)\mathbb{E}[K_1^2]}{n\mathbb{E}^2[K_1]} \left( \mathbb{E} \left[ \frac{K_1^2 \psi^2(Y, t)}{\mathbb{E}[K_1^2]} \right] \right) - \phi(h) \left( \mathbb{E} \left[ \frac{K_1 \psi(Y, t)}{\mathbb{E}[K_1]} \right] \right)^2.$$

D'une part, nous évaluons la limite de seconde terme de la côté à droite de l'équation (2.7). Puisque  $\phi(h) \rightarrow 0$ , alors, il suffit de montrer que

$$(2.8) \quad \forall t \in \mathbb{R}, \quad \mathbb{E} \left[ \frac{K_1 \psi(Y, t)}{\mathbb{E}[K_1]} \right] \rightarrow \lambda_1(x, t).$$

En effet,

$$\left| \mathbb{E} \left[ \frac{K_1 \psi(Y, t)}{\mathbb{E}[K_1]} \right] - \lambda_1(x, t) \right| = \frac{1}{\mathbb{E}[K_1]} |\mathbb{E}[K_1] \mathbb{I}_{B(x, h)}(X_1) \times (\lambda_1(X_1, t) - \lambda_1(x, t))|.$$

On utilise l'hypothèse (H3') (avec  $\gamma = 1$ ), on obtient

$$\mathbb{I}_{B(x, h)}(X_1) |\lambda_1(X_1, t) - \lambda_1(x, t)| \leq ch^{b_1},$$

Alors,  $\left| \mathbb{E} \left[ \frac{K_1 \psi(Y, t)}{\mathbb{E}[K_1]} \right] - \lambda_1(x, t) \right| \leq ch^{b_1} \rightarrow 0$ .

D'une autre part, pour le premier terme de la partie à gauche de l'équation (2.7), on va montrer que :

$$(2.9) \quad \forall t \in \mathbb{R}, \quad \mathbb{E} \left[ \frac{K_1^2 \psi_x^2(Y, t)}{\mathbb{E}[K_1^2]} \right] \rightarrow \lambda_2(x, t).$$

En effet, l'hypothèse (H2), nous permet d'obtenir

$$\left| \mathbb{E} \left[ \frac{K_1^2 \psi_x^2(Y, t)}{\mathbb{E}[K_1^2]} \right] - \lambda_2(x, t) \right| = \frac{1}{\mathbb{E}[K_1^2]} |\mathbb{E}[K_1^2] \mathbb{I}_{B(x, h)}(X_1) \times (\lambda_2(X_1, t) - \lambda_2(x, t))|.$$

Finalement, l'hypothèse (H3') (avec  $\gamma = 2$ ) donne,

$$\frac{1}{\mathbb{E}[K_1^2]} |\mathbb{E}[K_1^2] \mathbb{I}_{B(x, h)}(X_1) \times (\lambda_2(X_1, t) - \lambda_2(x, t))| \leq Ch^{b_1},$$

On va calculer maintenant les deux termes  $\mathbb{E}K_1$  et  $\mathbb{E}K_1^2$ . En effet, on a

$$\mathbb{E}K \left( \frac{d(x, X_1)}{h} \right) = \int_0^1 K(u) d\mathbb{P}^{\frac{d(x, X_1)}{h}}(u).$$

Il est clair que  $\int_0^u K'(t) dt = K(u) - K(0)$ , donc

$$\begin{aligned} \mathbb{E}K \left( \frac{d(x, X_1)}{h} \right) &= \int_0^1 K(0) d\mathbb{P}^{\frac{d(x, X_1)}{h}}(u) + \int_0^1 \int_0^u K'(t) dt d\mathbb{P}^{\frac{d(x, X_1)}{h}}(u) \\ &= K(0) \phi_x(h) + \int_0^1 \int_0^1 K'(t) \mathbb{I}_{[0, u]}(t) dt d\mathbb{P}^{\frac{d(x, X_1)}{h}}(u) \\ &= K(0) \phi_x(h) + \int_0^1 \int_0^1 K'(t) \mathbb{I}_{[t, 1]}(u) dt d\mathbb{P}^{\frac{d(x, X_1)}{h}}(u) \\ &= K(0) \phi_x(h) + \int_0^1 K'(u) \mathbb{P}(t < \frac{d(x, X_1)}{h} < 1) dt \\ &= K(0) \phi_x(h) + \phi_x(h) K(1) - K(0) \phi_x(h) - g(x) \int K^{(1)}(t) \phi_x(th) dt + o(\phi(h)) \\ &= -g(x) \int K^{(1)}(t) \phi_x(ht) dt + o(\phi(h)) \quad (\text{car } K(1) = 0). \end{aligned}$$

De même, pour le deuxième terme,

$$\mathbb{E}K_1^2 = -g(x) \int (K^2)'(t)\phi_x(th)dt + o(\phi(h)).$$

Alors,

$$(2.10) \quad \frac{\phi(h)\mathbb{E}[K_1^2]}{\mathbb{E}^2[K_1]} \rightarrow \frac{\alpha_2}{\alpha_1 g(x)} \quad \text{quand } n \rightarrow \infty.$$

Les équations (2.8), (2.9) et (2.10), nous permet d'obtenir,

$$(2.11) \quad n\phi(h)\text{Var}(\Psi_N(x, t)) \rightarrow \frac{\alpha_2\lambda_2(x, t)}{\alpha_1 g(x)}.$$

Il reste maintenant à évaluer la limite de numérateur, pour laquelle on utilise l'inégalité Cr, ce qui donne

$$(2.12) \quad \begin{aligned} (n\phi(h))^{1+\delta/2} \sum_{i=1}^n \mathbb{E}[|L_i(x) - \mathbb{E}[L_i(x)]|^{2+\delta}] &\leq 2^{1+\delta}(n\phi(h))^{(2+\delta)/2} \sum_{i=1}^n \mathbb{E}[|L_i(x)|^{2+\delta}] \\ &\quad + 2^{1+\delta}(n\phi(h))^{(2+\delta)/2} \sum_{i=1}^n |\mathbb{E}[L_i(x)]|^{2+\delta}. \end{aligned}$$

On remarque que  $\forall j > 0, \mathbb{E}[K_1^j] = O(\phi(h))$ , alors, la positivité de la fonction  $\psi$  et l'hypothèse (H2), nous permet d'obtenir

$$\begin{aligned} 2^{1+\delta}(n\phi(h))^{1+\delta/2} \sum_{i=1}^n \mathbb{E}[|L_i(x)|^{2+\delta}] &= 2^{1+\delta}n^{-\delta/2}(\phi(h))^{(-1-\delta)/2} \left(\frac{\phi(h)}{\mathbb{E}(K_1)}\right)^{2+\delta} \mathbb{E}[K_1^{2+\delta}|\psi_1(\theta_x)|^{2+\delta}] \\ &\leq C(n\phi(h))^{-\delta/2} \left(\frac{\phi(h)}{\mathbb{E}(K_1)}\right)^{2+\delta} \mathbb{E}[K_1^{2+\delta}/\phi(h)] \rightarrow 0. \end{aligned}$$

En ce qui concerne le deuxième terme de l'équation (2.12), on a

$$\begin{aligned} 2^{1+\delta}(n\phi(h))^{1+\delta/2} \sum_{i=1}^n |\mathbb{E}[L_i(x)]|^{2+\delta} &\leq 2^{1+\delta}n^{-\delta/2}(\phi(h))^{(1+\delta)/2}\mathbb{E}(K_1)^{2+\delta}|\mathbb{E}[K_1\psi_1(\theta_x)]|^{2+\delta} \\ &\leq Cn^{-\delta/2}(\phi(h))^{1+\delta/2} \rightarrow 0. \end{aligned}$$

■

**Lemme 2.2.2** *Sous les hypothèses (H1') et (H5'), on a*

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{h}{\phi(h)\alpha_1} \int_0^1 K(t)\varphi_x(th)\phi'(th)dt + o(1).$$

**Preuve du lemme 2.2.2**

On a

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{\mathbb{E}[K_1\psi_1(\theta_x)]}{\mathbb{E}[K_1]}.$$

Par conditionnement à  $d(x, X_1)$ , on obtient

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{\mathbb{E}[K_1 \mathbb{E}[\psi_1(\theta_x)/d(x, X_1)]]}{\mathbb{E}[K_1]}.$$

L'intégration par rapport à la première composante montre que

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{hg(x) \int_0^h K(h^{-1}t)\varphi_x(t)\phi'(t)dt + O(\phi(h))}{-g(x) \int_0^h K'(t)\phi(th)dt + O(\phi(h))}.$$

On prend le changement de variable  $h^{-1}t = s$

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{hg(x) \int_0^1 K(s)\varphi_x(hs)\phi'(hs)ds + O(\phi(h))}{-g(x) \int_0^h K'(t)\phi(th)dt + O(\phi(h))}.$$

Clairement d'après la définition de  $\alpha_1$ , le dénominateur normalisé par  $g(x)\phi(h)$  converge vers  $\alpha_1$  ce qui implique

$$\mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] = \frac{hg(x)}{\phi(h)g(x)\alpha_1 + O(\phi(h))} \left( \int_0^1 K(s)\varphi_x(hs)\phi'(hs)ds + O(\phi(h)) \right). \quad \blacksquare$$

**Lemme 2.2.3** *Sous les hypothèses (H1') – (H2'), (H3) et (H4') – (H5'), on a pour tout  $t \in \mathbb{R}$*

$$\mathbb{E}[\widehat{\Psi}_N(x, t)] = \lambda_1(x, \theta_x) + O(h^{b_1}).$$

**Preuve du lemme 2.2.3**

Nous avons,

$$\left| \mathbb{E}[\widehat{\Psi}_N(x, t)] - \lambda_1(x, t) \right| \leq \left| \mathbb{E}\left[\frac{1}{\mathbb{E}[K_1]} K_1 \mathbb{I}_{B(x, h)}(X_1) (\lambda_1(X_1, t) - \lambda_1(X, t))\right] \right|.$$

Sous  $(H2')$ , on obtient

$$\left| \mathbb{E}\left[\frac{1}{\mathbb{E}[K_1]} K_1 \mathbb{I}_{B(x, h)}(X_1) (\lambda_1(X_1, t) - \lambda_1(X, t))\right] \right| \leq Ch^{b_1}. \quad \blacksquare$$

**Corollaire 2.2.2** *Sous les hypothèses du lemme 2.2.3 et si le paramètre de lissage  $h$  satisfait  $nh^{2b_1}\phi(h) \rightarrow 0$ , quand  $n \rightarrow \infty$ , on a*

$$\sqrt{n\phi(h)} \mathbb{E}[\widehat{\Psi}_N(x, \theta_x)] \rightarrow 0, \quad \text{quand } n \rightarrow \infty.$$

**Preuve du corollaire 2.2.2**

Le lemme 2.2.3 permet d'obtenir facilement

$$\sqrt{n\phi(h)} \mathbb{E}[\Psi_N(x, \theta_x)] = \sqrt{n\phi(h)h^{2b_1}}. \quad \blacksquare$$

**Proposition 2.2.1** *(voir Azzedine et al. (2008))*

*Supposons que  $(H1') - (H2')$ ,  $(H3)$  et  $(H4') - (H5')$  sont satisfaites, donc  $\widehat{\theta}_x$  existe et il est unique avec une probabilité tend vers 1, et on a*

$$(2.13) \quad \widehat{\theta}_x - \theta_x \rightarrow 0 \text{ en probabilité quand } n \rightarrow \infty.$$

**Lemme 2.2.4** *Sous les hypothèses du théorème 2.2.1, on a*

$$\widehat{\Psi}'_N(x, \xi_n) \rightarrow \Gamma_1(x, \theta_x) \text{ en probabilité quand } n \rightarrow \infty.$$

**Preuve du lemme 2.2.4**

On utilise la décomposition suivante

$$(2.14) \quad \left| \Psi'_N(x, \xi_n) - \Gamma_1(x, \theta_x) \right| \leq \left| \Psi'_N(x, \xi_n) - \Psi'_N(x, \theta_x) \right| + \left| \Psi'_N(x, \theta_x) - \Gamma_1(x, \theta_x) \right|.$$

Concernant le premier terme, on remarque que

$$\left| \Psi'_N(x, \xi_n) - \Psi'_N(x, \theta_x) \right| \leq \sup_{y \in \mathbb{R}} \left| \frac{\partial \psi(y, \xi_n)}{\partial t} - \frac{\partial \psi(y, \theta_x)}{\partial t} \right| \widehat{\Psi}_D(x),$$

et puisque  $\frac{\partial \psi(y, t)}{\partial t}$  est continue au  $\theta_x$  uniformément à  $y$ . La proposition 2.2.1 et la convergence en probabilité de  $\widehat{\Psi}_D(x)$  vers 1 montre que le premier terme de (2.14) converge en probabilité vers 0. Cependant, la limite de seconde terme est obtenue par évaluer séparément le biais et la variance de  $\Psi'_N(x, \theta_x)$ . Clairement, des arguments similaires de la preuve le lemme 2.1.3 peuvent être utilisés pour obtenir

$$\mathbb{E}[\Psi'_N(x, \theta_x)] \rightarrow \Gamma_1(x, \theta_x).$$

De plus, on utilise des arguments analogues que 2.11, on peut montrer,

$$n\phi(h)\text{Var}[\Psi'_N(x, \theta_x)] \rightarrow \frac{\alpha_2\Gamma_2(x, \theta_x)}{\alpha_1^2g(x)}.$$

Finalement, sous les hypothèses  $(H1')$  et  $(H4')$  la preuve est achevée. ■

# Chapitre 3

## Cas dépendant

Le but de ce chapitre est la généralisation du résultat donné dans le chapitre précédent à des observations mélangantes. Notant que cette généralisation est très utiles dans la prévision en série temporelle. Ce chapitre est aussi divisé en deux sections. on établit la convergence presque complète de l'estimateur à noyau de la fonction de régression robuste dans la première section. Application à la prévision en série temporelle se trouve dans la deuxième section.

### 3.1 Propriétés asymptotiques

Afin d'établir la convergence presque complète de l'estimateur à noyau défini dans le premier chapitre, on garde les mêmes hypothèses, ainsi les mêmes notations du chapitre précédent et on ajoute les hypothèses suivantes.

(H6)  $(X_i, Y_i)_{i \geq 1}$  est une suite  $\alpha$ -mélangeante avec ces coefficients satisfaisent

$$\alpha(n) = O(n^{-a}) \text{ pour } a > 0$$

(H7)  $0 < \sup_{i \neq j} \mathbb{P}((X_i, X_j) \in B(x, h) \times B(x, h)) = O\left(\frac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}}\right)$ .

(H8) Il existe  $\eta > 0$ , tel que  $Cn^{\frac{3-\alpha}{a+1}+\eta} \leq \phi_x(h) \leq C'n^{\frac{1}{1-a}+\eta}$  avec  $a > \frac{5+\sqrt{17}}{2}$ .

**Remarque 3.1.1** *Les hypothèses (H6) – (H8) sont ajoutées pour éviter l'expression de covariance dans la vitesse de convergence. Autrement dit, on peut démontrer la convergence presque*

complète sans ces hypothèses. Cependant, la vitesse de convergence sera donné en fonction de covariance des observations et elle sera lente par rapport à la vitesse du cas indépendant. Ainsi, nous établissons la convergence presque complète avec la même précision, mais, sous des conditions un plus fort que le cas *i.i.d.*

**Théorème 3.1.1** *Sous les hypothèses (H1) – (H8) et si  $\Gamma(x, \theta_x) \neq 0$ , alors  $\hat{\theta}_x$  existe et il est unique p.s pour  $n$  assez grand, et nous avons*

$$\hat{\theta}_x - \theta_x = O(h^{b_1}) + O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) \text{ p.co quand } n \rightarrow \infty$$

### 3.1.1 Démonstration des résultats techniques

La démonstration est essentiellement basée sur les mêmes arguments analytiques utilisés dans la démonstration du théorème 2.1.1 à savoir le développement du Taylor de l'estimateur et la décomposition (2.3) et en vertu de la remarque 2.1.1, on peut dire que la propriété de l'indépendance des observations n'aucune influence sur la partie biais de la vitesse. Autrement dit, la vitesse de convergence de la partie biais de théorème 2.1.1 sera la même dans le cas de mélange. Cependant, la partie dispersion est basée sur les deux lemmes suivants.

**Lemme 3.1.1** *Sous les hypothèses (H1), (H3) – (H8), on a*

$$\hat{\Psi}_D(x) - \mathbb{E}[\hat{\Psi}_D(x)] = O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right), \text{ p.co quand } n \rightarrow \infty.$$

#### Preuve du lemme 3.1.1

On pose  $\Delta_i = K_i(x) - \mathbb{E}[K_i(x)]$ , alors,

$$\hat{\Psi}_D(x) - \mathbb{E}[\hat{\Psi}_D(x)] = \frac{1}{n\mathbb{E}[K_1(x)]} \sum_{i=1}^n \Delta_i(x)$$

On applique l'inégalité de Fuck-Nagaev (Rio, 2000, p. 87), on obtient pour tout  $r > 0$  et  $\varepsilon > 0$ ,

$$\mathbb{P}\left[|\mathbb{E}(\hat{\Psi}_D(x)) - \hat{\Psi}_D(x)| > \varepsilon\right] \leq \mathbb{P}\left[\left|\sum_{i=1}^n \Delta_i(x)\right| > \varepsilon n\mathbb{E}[K_1(x)]\right]$$

$$(3.1) \quad \leq C(A_1(x) + A_2(x)),$$

où

$$A_1(x) = \left(1 + \frac{\varepsilon^2 n^2 (\mathbb{E}[K_1(x)])^2}{r S_n^2}\right)^{-r/2}, \quad A_2(x) = nr^{-1} \left(\frac{r}{\varepsilon n \mathbb{E}[K_1(x)]}\right)^{a+1}$$

et

$$(3.2) \quad S_n^2(x) = \sum_{i=1}^n \sum_{j=1}^n Cov(\Delta_i(x), \Delta_j(x)) = S_n^{2*}(x) + nVar[\Delta_1(x)]$$

avec

$$(3.3) \quad S_n^{2*}(x) = \sum_{i \neq j} Cov(\Delta_i(x), \Delta_j(x)).$$

On utilise les techniques de Masry (1986) et on partage cette somme sur les deux ensembles

$$E_1 = \{(i, j) \text{ tels que } 1 \leq |i - j| \leq m_n\}$$

et

$$E_2 = \{(i, j) \text{ tels que } m_n + 1 \leq |i - j| \leq n - 1\}$$

où  $m_n \rightarrow \infty$ , quand  $n \rightarrow \infty$  où  $m_n$  est une suite des entiers tend vers l'infinie quand  $n$  tend vers à l'infinie. Alors,

$$S_n^{2*}(x) = \sum_{i \neq j} Cov(\Delta_i(x), \Delta_j(x)) = \sum_{E_1} Cov(\Delta_i(x), \Delta_j(x)) + \sum_{E_2} Cov(\Delta_i(x), \Delta_j(x))$$

En ce qui concerne la première partie, on a

$$J_{1,n} = \sum_{E_1} |Cov(\Delta_i(x), \Delta_j(x))| \leq \sum_{E_1} |\mathbb{E}[K_i(x)K_j(x)] - \mathbb{E}^2[K_1(x)]|.$$

On utilise les hypothèses (H1) et (H7) – (H8), on obtient

$$J_{1,n} \leq Cnm_n \phi_x(h) \left( \left( \frac{\phi_x(h)}{n} \right)^{1/a} + \phi_x(h) \right).$$

Pour  $E_2$ , on a d'après l'inégalité de covariance

$$| \text{Cov}(K_i(x), K_j(x)) | \leq C\alpha(|i - j|).$$

Par conséquent,

$$J_{2,n} = \sum_{E_2} | \text{Cov}(K_i(x), K_j(x)) | \leq n^2 m_n^{-a}.$$

Choisissons  $m_n = \left( \frac{\phi_x(h)}{n} \right)^{1/a}$  permet d'obtenir, sous (H8)

$$(3.4) \quad S_n^{2*} = J_{1,n} + J_{2,n} = O(n\phi_x(h)).$$

Concernant la variance, l'hypothèse (H1) nous permet d'obtenir

$$(3.5) \quad \text{Var}(\Delta_1(x)) \leq C(\phi_x(h) + (\phi_x(h))^2).$$

Finalement, par les équations (3.10), (3.3), (3.11) et (3.12), on obtient

$$(3.6) \quad S_n^2(x) = O(n\phi_x(h)).$$

Maintenant, on applique (3.1) avec

$$(3.7) \quad \varepsilon = \lambda \frac{\sqrt{n \log n \phi_x(h)}}{n \mathbb{E}[K_1(x)]} \text{ et } r = C(\log n)^2.$$

Sous (H8), nous avons,

$$A_2(x) \leq C n^{-1-\eta(a+1)/2} (\log n)^{(3a-1)/2}.$$

Alors, il existe  $\nu > 0$  tel que

$$(3.8) \quad A_2(x) \leq C n^{-1-\nu}.$$

On utilise (3.6) et (3.7), on obtient

$$A_1(x) \leq C \exp\left(-\lambda^2 \frac{\log n}{2}\right) = C n^{-\lambda^2/2}.$$

Donc, pour  $\lambda$  assez grand

$$(3.9) \quad \exists \nu' > 0, \quad A_1(x) \leq C n^{-\lambda^2/2} \leq C n^{-1-\nu'}.$$

Finalement, le résultat est déduit facilement par (3.1), (3.14) et (3.9). ■

**Lemme 3.1.2** *Sous les hypothèses (H1), (H3) – (H8), on a*

$$\widehat{\Psi}_N(x) - \mathbb{E}[\widehat{\Psi}_N(x)] = O\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right), p.co \text{ quand } n \rightarrow \infty.$$

**Preuve du lemme 3.1.2**

On va calculer

$$(3.10) \quad S_n^2(x) = \sum_{i=1}^n \sum_{j=1}^n Cov(\Delta_i(x), \Delta_j(x)) = \sum_{ij}^n Cov(\Delta_i(x), \Delta_j(x)) + nVar[\Delta_1(x)]$$

avec  $\Delta_i = K_i(x)\psi(Y_i - t) - \mathbb{E}[K_i(x)\psi_x(Y_i - t)]$ . On partage cette somme sur les deux ensembles

$$S_1 = \{(i, j) \text{ tels que } 1 \leq |i - j| \leq u_n\}$$

et

$$S_2 = \{(i, j) \text{ tels que } u_n + 1 \leq |i - j| \leq n - 1\}$$

Alors,

$$J'_{1,n} = \sum_{S_1} |Cov(\Delta_i(x), \Delta_j(x))| \leq \sum_{E_1} |\mathbb{E}[K_i(x)K_j(x)] - \mathbb{E}^2[K_1(x)]|.$$

On utilise les hypothèses (H1), (H7) et (H8), on obtient

$$J'_{1,n} \leq Cnu_n\phi_x(h) \left( \left( \frac{\phi_x(h)}{n} \right)^{1/a} + \phi_x(h) \right).$$

et on a,

$$J'_{2,n} = \sum_{S_2} |Cov(K_i(x), K_j(x))| \leq n^2 m_n^{-a}.$$

On prend  $u_n = \left( \frac{\phi_x(h)}{n} \right)^{1/a}$ , sous (H8), on obtient,

$$(3.11) \quad \sum_{i=1}^n \sum_{ij}^n Cov(\Delta_i(x), \Delta_j(x)) = O(n\phi_x(h)).$$

Concernant la variance, l'hypothèse (H1) nous permet d'obtenir

$$(3.12) \quad \text{Var}(\Delta_1(x)) \leq C(\phi_x(h) + (\phi_x(h))^2).$$

On applique l'inégalité de Fuk-Nagaev, on obtient,  $\forall r > 0$  et  $\forall \varepsilon > 0$ ,

$$\begin{aligned} \mathbb{P} \left[ |\mathbb{E}(\widehat{\Psi}_N(x)) - \widehat{\Psi}_N(x)| > \varepsilon \right] &\leq \mathbb{P} \left[ \left| \sum_{i=1}^n \Delta_i(x) \right| > \varepsilon n \mathbb{E}[K_1(x)] \right] \\ &\leq C(A'_1(x) + A'_2(x)) \end{aligned}$$

où

$$A'_1(x) = \left( 1 + \frac{\varepsilon^2 n^2 (\mathbb{E}[K_1(x)])^2}{r S_n^2} \right)^{-r/2}, \quad A'_2(x) = nr^{-1} \left( \frac{r}{\varepsilon n \mathbb{E}[K_1(x)]} \right)^{a+1}.$$

On prend

$$(3.13) \quad \varepsilon = \lambda \frac{\sqrt{n \log n \phi_x(h)}}{n \mathbb{E}[K_1(x)]} \text{ et } r = C(\log n)^2.$$

on obtient,

$$A'_2(x) \leq C n^{-1-\eta(a+1)/2} (\log n)^{(3a-1)/2}.$$

Alors, il existe  $\nu' > 0$  tel que

$$(3.14) \quad A'_2(x) \leq C n^{-1-\nu'}.$$

Finalement, les équations (3.1), (3.14) et (3.9) nous permet de conclure le résultat. ■

## 3.2 Prédiction en série temporelle

On peut obtenir aisément un échantillon de variables fonctionnelles à partir d'un processus à temps continu. En effet, soit  $(Z_t)_{t \in [0, b]}$  un processus réel stationnaire à temps continu. A partir de  $Z_t$ , on construit  $N$  variables fonctionnelles  $(X_i)_{i=1, \dots, N}$  définies de la façon suivante :

$$\forall t \in [0, b], X_i(t) = Z_{N^{-1}((i-1)b+t)}.$$

Si on souhaite prédire une caractéristique réelle (notée  $Y$ ) de  $X_N$  sachant la courbe  $X_{N-1}$ , on considère les observations  $(X_i, Y_i)$  où  $Y_i$  est la caractéristique qu'on veut prévoir à l'instant  $i$ . En utilisant la régression robuste comme outil de prédiction, on peut prévoir  $Y_N$  par  $\hat{\theta}(X_{N-1})$ .

# Chapitre 4

## Applications

Le but de ce chapitre est de faire, à l'aide des données simulées puis réelles, une comparaison entre la régression classique et la régression robuste. Ce chapitre est divisée en deux section intitulés respectivement : application sur des données simulées et application sur des données réelles.

### 4.1 Application sur des données simulées

#### 4.1.1 Données

Nous récoltons des observations fonctionnelles générées à l'aide du processus suivant :

$$X_i(t) = \exp(2Wt), \forall t \in [0, 1]$$

tel que  $W$  est une variable aléatoire uniforme. Ces variables fonctionnelles  $X$  sont représentées dans la Figure 4.1

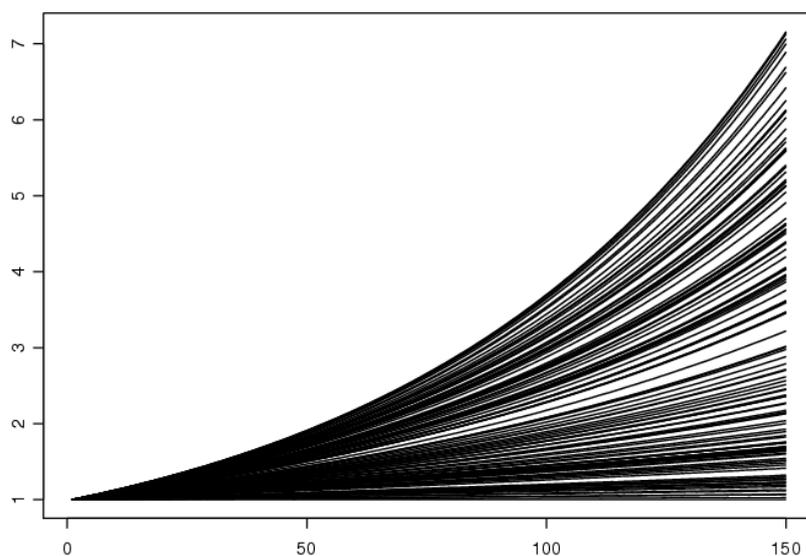


FIGURE 4.1 – 120 courbes,  $X_i(t), t \in [0, 1], i = 1 \dots, 120$ .

Concernant les variables réponses  $Y_i$ , nous les récoltons selon le modèle de régression suivant :

$$Y = R(X) + \epsilon \quad \text{où} \quad R(X) = 4 \int (X'(t))^2 dt$$

et  $\epsilon$  est de loi normale perturbée de la forme suivante

$$(1 - \epsilon_0)\mathcal{N}(0, 1) + \epsilon_0 \exp(0.05)$$

$\epsilon_0$  est le degré de perturbation compris entre 0 et 1, et  $\exp(0.05)$  est la loi exponentielle de paramètre 0.05.

On se propose de comparer deux méthodes à savoir : la régression classique  $r(x) = \mathbb{E}(Y/X = x)$

dont un estimateur est donné par (i.e  $\psi(t) = t$ )

$$\hat{r}(x) = \frac{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right) Y_i}{\sum_{i=1}^n K\left(\frac{d(x, X_i)}{h_K}\right)}$$

et la régression robuste  $\theta_x$  associée à  $\psi(t) = \frac{t}{\sqrt{1+t^2/2}}$ .

### 4.1.2 Algorithme

Dans notre simulation, on a considéré 120 observations, un noyau quadratique  $K(x) = \frac{3}{2}(1-x^2)\mathbb{I}_{[0,1]}$  et le paramètre de lissage est choisi par la méthode de la validation croisée sur le nombre des voisins les plus proches.

On procède par l'algorithme suivant :

**Etape1.** On divise nos observations en deux paquets :

- $(X_j, Y_j)_{j \in (1:80)}$  échantillon d'apprentissage,
- $(X_i, Y_i)_{i \in (81:120)}$  échantillon de test.

**Etape2.** On calcule  $\hat{\theta}_{X_j}$  pour chaque  $j \in (1 : 80)$

**Etape3.** Pour chaque  $X_i$  dans l'échantillon test, on calcule  $i_* = \text{Argmin}_{j \in (1:80)} d(X_i, X_j)$ .

**Etape4.** Pour tout  $i \in (81 : 120)$  on pose  $\hat{Y} = \hat{\theta}_{X_{i_*}}$ .

**Etape5.** On calcule la moyenne des erreurs absolues  $\frac{1}{40} \sum_{i=81}^{120} |Y_i - \hat{T}(X_i)|$ , où  $\hat{T}$  désigne l'estimateur utilisé : classique ou régression robuste.

### 4.1.3 Résultat

La première illustration est donnée dans la Fig.4.2 où on montre qu'en absence de données aberrantes, les deux méthodes sont équivalentes.

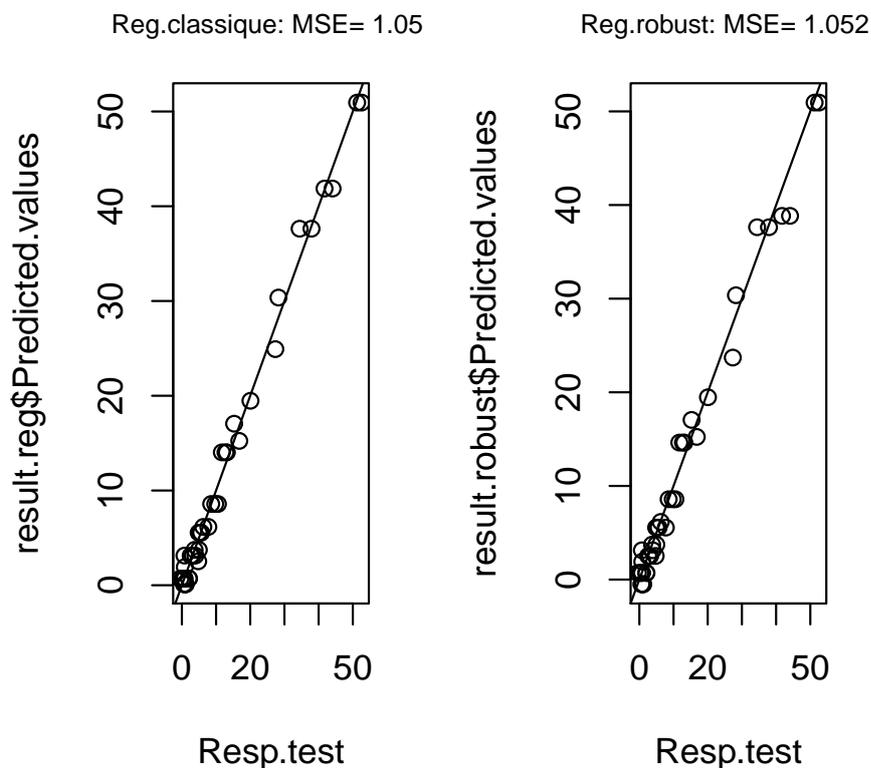


FIGURE 4.2 – Comparaison entre les deux méthodes en absence de données aberrantes

La deuxième illustration est donnée dans la table suivante, où on observe des valeurs aberrantes.

degré de perturbation	MEA. régression classique	MEA. régression robuste
$\epsilon_0 = 0.2$	3.074783	2.589599
$\epsilon_0 = 0.4$	16.00928	7.444895
$\epsilon_0 = 0.6$	8.586478	8.539707
$\epsilon_0 = 0.7$	15.13026	8.939606
$\epsilon_0 = 0.9$	13.08279	10.34014

On remarque que, la régression robuste donne des meilleurs résultats que la méthode classique, Dans le sens que, même si la Moyenne des Erreurs Absolues (MEA), augmente avec le nombre de points perturbés, mais elle reste moins élevée pour la méthode robuste.

## 4.2 Application sur des données réelles

Le but de cette section est d'appliquer notre modèle à des données réelles. Plus précisément, notre but est de comparer la régression classique et la régression robuste en présence de données aberrantes.

La spectroscopie dans le proche infrarouge est une technique analytique basée sur le principe d'absorption des rayonnements (infrarouge) par la matière organique en utilisant un spectromètre. Cette méthode d'absorption est utilisée pour estimer la composition chimique des aliments (fourrages, viandes, ...).

On se donne un échantillon de 215 morceaux de viande, le but est d'estimer le taux de matière grasse. Pour cela, on observe le spectre pour 100 longueurs d'onde réparties entre 250 et 1050 nanomètres ( les données sont disponibles sur le site web (<http://lib.stat.cmu.edu/datasets/tecator>)). On observe alors pour chaque morceau de viande  $i$ , la variable fonctionnelle  $X_i(t), t \in [850, 1050]$  qui est la courbe spectrométrique du morceau de viande  $i$ , la répartition graphique des 215 spectres montre l'aspect fonctionnel des données.

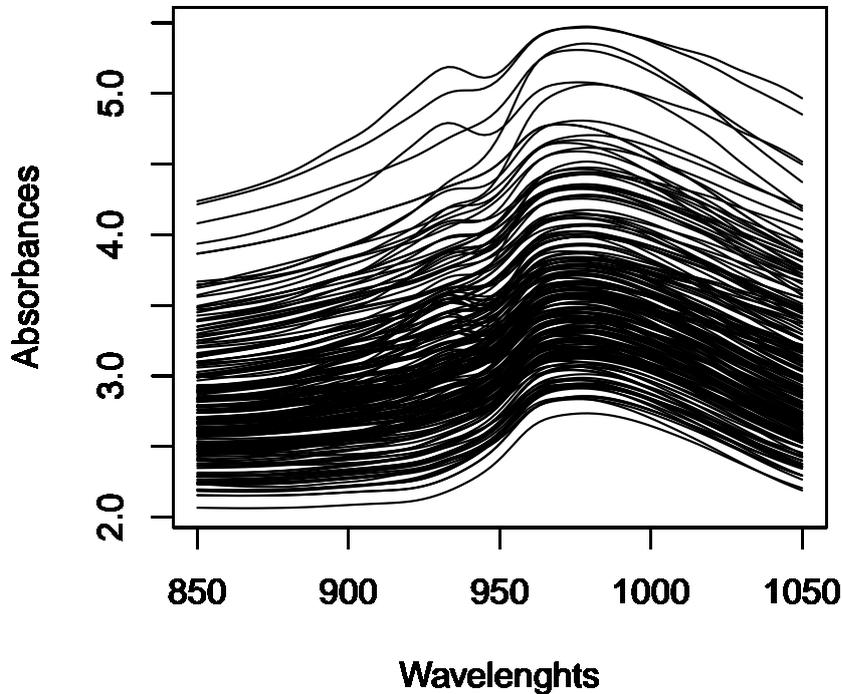


FIGURE 4.3 – 215 courbes spectrométriques,  $X_i(t), t \in [850, 1050], i = 1 \dots, 215$ .

Une des questions qu'on peut se poser est la suivante : étant donné le taux de lipide  $Y_i$  contenu dans un morceau de viande  $i$  obtenu à partir de procédé classique fiable mais plus coûteux que l'étude par spectromètre, peut-on modéliser pour  $i = 1 : 215$ , le lien entre la variable réelle  $Y_i$  et la variable fonctionnelle  $X_i$ ?

Afin de rechercher les courbes aberrantes pour ces données fonctionnelles, on utilise les sous-routines du langage de programmation R proposer par Febrero M. sur son site web. les courbes atypiques sont tracés en ligne gras continue, et les courbes non-atypiques en pointillés.

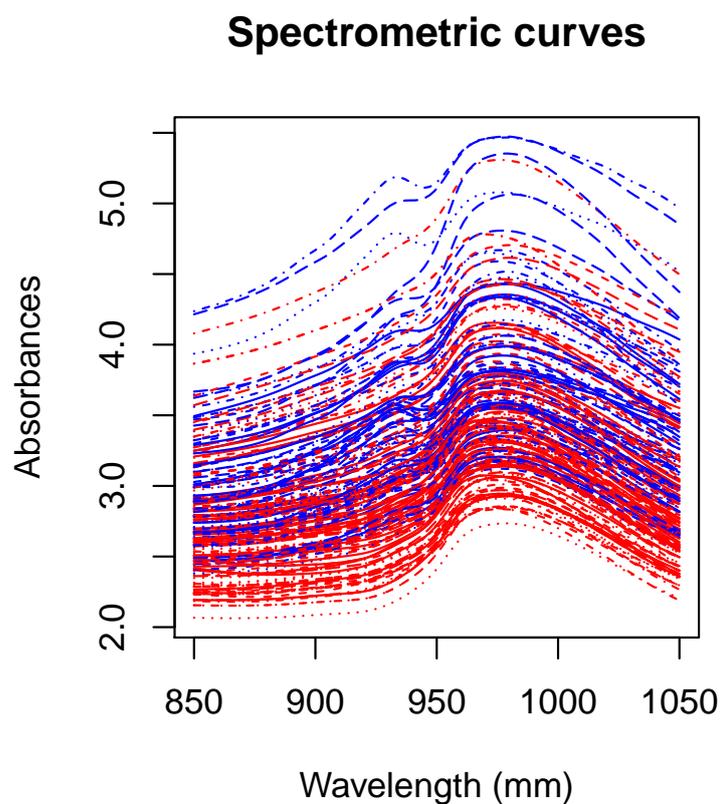


FIGURE 4.4 – Courbes atypiques

Considérons les 170 premières observations comme échantillon d'apprentissage et les 45 dernières comme échantillon de test. Le paramètre de lissage et le noyau sont choisis de la même manière que dans la première section.

- On calcule la moyenne des erreurs absolues  $\frac{1}{45} \sum_{i=171}^{215} |Y_i - \hat{T}(X_i)|$ , où  $\hat{T}$  désigne l'estimateur utilisé : classique ou régression robuste.
- On compare les deux méthodes en présentant notre résultat par un graphe de type boîte à moustache.

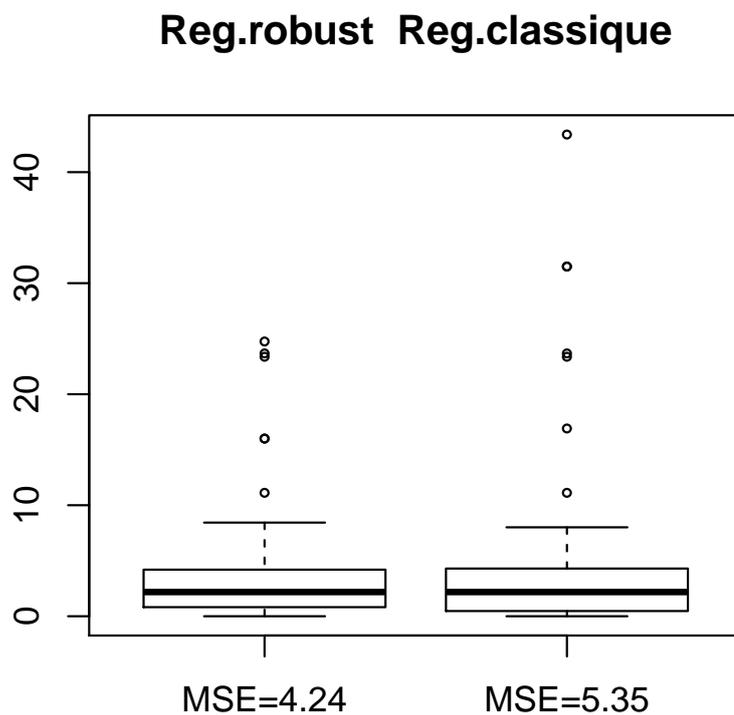


FIGURE 4.5 – Comparaison entre les deux méthodes en présence de données aberrantes.

A partir de résultat obtenu, nous pouvons conclure que, la méthode de régression robuste est meilleure que la méthode de régression classique.

# Bibliographie

- [1] D. F. Andrews. (1974). A robust method for multiple linear regression. *Thechnometrics*, 16, 523-531.
- [2] M. Attouch, A. Laksaci, E. Ould-Saïd. (2007a). Strong uniform convergence rate of robust estimator of the regression function for functional and dependent processes. Preprint, LMPA., Univ. du Littoral, 340, submitted.
- [3] M. Attouch, A. Laksaci, E. Ould-Saïd. (2008a). Asymptotic distribution of robust estimator for functional nonparametric models. *Comm. Statist. Theory and Methods*, in press.
- [4] M. Attouch, A. Laksaci, E. Ould-Saïd. (2008b). Asymptotic distribution of robust estimator for functional dependent data. Preprint, LMPA No. 378, Janvier 2008. Univ. du Littoral C ?te d'Opale. submitted.
- [5] N. Azzeddine, A. Laksaci, E. Ould-Saïd. (2008). On the robust nonparametric regression estimation for functional regressor. *Statist. Probab. Lett.*, in press.
- [6] G. Boente, R. Fraiman. (1989) Nonparametric regression estimation. *J. Multivariate Anal.*, 29, 180-198.
- [7] G. Boente, R. Fraiman. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.*, 18, 891-906.
- [8] G. E. P. Box. (1954). Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems, I. Effect of Inequality of Variance in the One-Way Classification. *The Annals of Mathematical Statistics*, 25, No.2, 290-302.
- [9] Z. Cai, E. Ould Saïd. (2003). Local M-estimator for nonparametric time series. *Statist. and Probab. Lett.*, 65, 433-449.

- 
- [10] G. Collomb, W. Härdle. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction : Kernel regression estimation from dependent observations. *Stoch. Proc. Appl.*, 23, 77-89.
- [11] J. Fan, T.C. Hu, Y.K. Truong. (1994). Robust nonparametric function estimation. *Scand. J. Statist.*, 21 ,433-446.
- [12] Ferraty, F., Vieu, Ph., *Modèle de régression pour variables aléatoires uni, multi et  $\infty$ -dimensionnées* (2004). Cours de DEA.
- [13] F. R. Hampel. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.*, 42, 1887-1896.
- [14] W. Härdle. (1984). Robust and nonlinear time series analysis. *Lecture Notes in Statistics*, Springer- Verlag, New York, 26.
- [15] W. Härdle, A. B. Tsybakov. (1988). Robust nonparametric regression with simultaneous scale curve estimation. *Ann. Statist.*, 16, No.1, 120-135.
- [16] Hoeffding,W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58, 15-30.
- [17] P.J. Huber. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, 35, 73-101.
- [18] P.J. Huber. (1973). Robust regression. Asymptotics, conjectures and monte Carlo. *Ann. Stat.*, 43, 799-821.
- [19] P.J. Huber. (1981). *Robust statistics*. John Wiley.
- [20] W. S. Krasker, R. E. Welsch. (1982). Efficient bounded influence regression estimation. *J. A. S. A.*, 77, 595-604.
- [21] N. Laïb, E. Ould-Saïd. (2000). A robust nonparametric estimation of the autoregression function under ergodic hypothesis. *Canad. J. Statist.*, 28, 817-828.
- [22] P. S. Laplace. (1818). *Deuxième supplément de Laplace*.
- [23] M. Loève. (1963). *Probability Theory*. Third Edition. Van Nostrand Princeton.
- [24] R. Robinson. (1984). Robust nonparametric autoregression. *Lecture Notes in Statistics*, 26, 247-255. Springer-Verlag, New York.