

N° d'ordre : 00/2014-M/MT

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université des Sciences et de la Technologie Houari Boumediene  
Faculté de Mathématiques



## MÉMOIRE

Présenté pour l'obtention du diplôme de  
**Magister en Mathématiques**  
Option : Probabilité et Statistique

Présenté et soutenu par

**Kenza MEDDOUER**

Sujet

Estimation non paramétrique des données tronquées  
par les polynômes locaux

Soutenu publiquement, le **09/10/2014**, devant le jury composé de :

Mr. K. BOUKHETALA	Professeur	à l'U.S.T.H.B	Président.
Mr. A. TATACHAK	Maître de conférence	à l'U.S.T.H.B	Directeur de mémoire.
Me. Z. GUESSOUM	Maître de conférence	à l'U.S.T.H.B	Examinatrice.

# Table des matières

Table des matières	2
<i>Abréviations et symboles</i>	iii
<i>Introduction générale</i>	I
<i>Chapitre 1 : Estimation non paramétrique</i>	1
1 Introduction	1
2 Estimateur à noyau d'une densité	1
2.1 Propriétés asymptotiques des estimateurs à noyaux . . . . .	4
2.2 Cas d'une densité multivariée . . . . .	6
2.3 Calcul du risque quadratique en un point . . . . .	6
2.4 Calcul du risque quadratique intégré . . . . .	7
3 Sélection de la fenêtre de lissage	8
3.1 Méthode de validation croisée non biaisée [43] . . . . .	8
3.2 Méthode de Ushakov et Ushakov (2012). . . . .	11
3.2.1 Introduction . . . . .	11
3.2.2 Présentation de la méthode . . . . .	11
<i>Chapitre 2 : Régression non paramétrique par la méthode à noyau</i>	15
1 Introduction	15
2 L'ajustement linéaire et polynomial	17
2.1 Exemple : Fan et Gijbels(1996) . . . . .	18
3 Modélisation locale	19
4 Les avantages de l'ajustement polynômial local	21

<b>5</b>	<b>Quelques estimateur à noyau de la fonction de regression</b>	<b>21</b>
5.1	L'estimateur de Nadaraya-Watson (1964) . . . . .	21
5.2	L'estimateur de Gasser et Müller . . . . .	23
5.3	L'ajustement polynômial local et l'estimation des dérivées . . . . .	23
5.3.1	L'ajustement polynômial local . . . . .	24
5.3.2	L'estimation des dérivées . . . . .	25
<b>6</b>	<b>La régression par les polynômes locaux</b>	<b>25</b>
6.1	Le biais et la variance . . . . .	27
6.2	Méthode des noyaux equivalent . . . . .	30
6.3	Le choix optimal de la fenêtre de lissage . . . . .	34
6.4	Le choix de $p$ . . . . .	35
	 <b>Chapitre 3 : Données Incomplètes</b>	 <b>37</b>
<b>1</b>	<b>Introduction</b>	<b>37</b>
<b>2</b>	<b>Censure et Troncature</b>	<b>38</b>
2.1	Censure . . . . .	38
2.2	Troncature . . . . .	38
2.2.1	Un bref historique . . . . .	38
2.2.2	<b>Exemple de troncature à droite.</b> Lagakos et al. (1988)(voir Klein et Moeschberger (1997)[25]). . . . .	39
2.2.3	<b>Exemple de troncature à gauche.</b> (voir Klein et Moeschberger (1997)[25]) . . . . .	39
<b>3</b>	<b>Définitions</b>	<b>39</b>
<b>4</b>	<b>Distributions de la durée de survie</b>	<b>40</b>
4.1	La fonction de survie $S$ . . . . .	40
4.2	La fonction de répartition $F$ . . . . .	40
4.3	La densité de probabilité $f$ . . . . .	40
4.4	Le taux de hasard ou risque instantané $h$ . . . . .	41
4.5	Le taux de hasard cumulé $H$ . . . . .	41
4.6	Quantités associées à la distribution de survie . . . . .	42
4.6.1	Moyenne et variance de la durée de survie . . . . .	42
4.6.2	Le quantile de la durée de survie . . . . .	42

<b>5</b>	<b>Estimation sous le modèle de troncature à gauche</b>	<b>43</b>
5.1	Modèle de troncature à gauche . . . . .	43
5.2	Estimation de la probabilité de troncature . . . . .	48
5.3	Les estimateurs du maximum de vraisemblance de $F$ et $G$ (Lynden-Bell (1971)) . . . . .	49
5.4	Propriétés asymptotiques de l'estimateur de Lynden-Bell . . . . .	50
5.5	Estimation de la densité de la covariable sous le modèle de troncature à gauche dans le cas de la régression . . . . .	50

***Chapitre 4 : Données dépendantes*** **52**

<b>1</b>	<b>Introduction</b>	<b>52</b>
<b>2</b>	<b>Définitions des mesures de dépendances</b>	<b>52</b>
<b>3</b>	<b>Propriétés de base</b>	<b>53</b>
<b>4</b>	<b>Conditions de mélangeance forte</b>	<b>53</b>
4.1	Conditions de mélangeance forte basée sur le passé et le futur . . . . .	53

***Chapitre 5 : Estimation des données tronquées dépendantes par la méthode des polynômes locaux*** **55**

<b>1</b>	<b>Introduction</b>	<b>55</b>
<b>2</b>	<b>Estimateurs</b>	<b>57</b>
<b>3</b>	<b>Méthode des polynômes locaux</b>	<b>59</b>
<b>4</b>	<b>Hypothèses et résultats</b>	<b>61</b>
<b>5</b>	<b>Preuve des résultats principaux</b>	<b>65</b>

***Chapitre 6 : Application et Simulation*** **68**

<b>1</b>	<b>L'estimateurs <math>\hat{m}_{NW}</math></b>	<b>68</b>
<b>2</b>	<b>Estimation de la densité des données tronquées</b>	<b>70</b>

<i>Appendice</i>	i
1 Loi forte des grands nombres (strong law of large numbers)	i
2 Méthode des moindres carrés	i
3 $O(\cdot)$ et $o(\cdot)$ stochastiques	iii
4 Démonstration de $S_{n,j}$	iii
5 Construction des noyaux d'ordre élevés	iv
6 Démonstration du théorème de Ruppert et Wand (1994)	v
 <i>Référence</i>	 ix
Références	ix

*“ A mes chers parents ”*

# Remerciements

*Avant toutes choses, je voudrais remercier le tout puissant **DIEU** de son aide à accomplir ce mémoire pour l'obtention du grade de Magister de l'Université de Houari Boumediene.*

*Je remercie vivement **Mr A. TATACHAK** pour m'avoir proposé le sujet et pour m'avoir encadré durant ces trois années de mémoire.*

*J'exprime aussi ma profonde gratitude à **Mr. K. BOUKHETALA** Professeur à l'USTHB, pour avoir accepté de présider le jury de soutenance. Vous me faites un très grand honneur.*

*Je suis très honorée que **Mme. Z. GUESSOUM** Maître de conférences à l'USTHB aient acceptées de juger ce mémoire. Leur vision m'apportera un nouvel éclairage sur mon travail. Je leur remercie d'avoir fait l'effort de lire mon mémoire et d'examiner mon travail.*

*Par ailleurs, je souhaite aussi remercier mon amie **Melle A. BAOUTA** d'avoir partagé ses connaissances en programmation avec moi, ce que j'ai fortement apprécié. Je ne pourrai jamais assez remercier.*

*Je tiens à exprimer ma reconnaissance à toute ma famille, ma belle famille, mon mari, mes amies, et tous ceux qui m'ont aidé de près ou de loin.*

# Abréviations et symboles

$P$	Mesure de probabilité.
$P(A)$	Probabilité de l'évènement A
$\mathbb{R}$	Ensemble de tous les nombres réels.
$v.a$	Variable aléatoire.
$i.i.d$	indépendantes et identiquement distribuées.
$N$	Loi normale.
$\xrightarrow{p.s}$	Convergence presque sûrement.
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$\xrightarrow{P}$	Convergence en probabilité.
$X \wedge C$	Minimum( $X, C$ ).
$X \vee C$	Maximum( $X, C$ ).
$L^2$	L'espace des fonctions de carré intégrable.
$MSE$	Risque quadratique.
$MISE$	Risque quadratique intégré.
Estimateur $LL$	Estimateur localement linéaire.
Estimateur $NW$	Estimateur de Nadaraya-Watson.
Estimateur $LP$	Estimateur localement polynômial.
$O(\cdot)$	grand $O$ .
$o(\cdot)$	petit $o$ .

# Introduction générale

Le problème de l'analyse des données de survie survient dans un certain nombre de domaines appliqués, tels que la médecine, la biologie, la santé publique, épidémiologie, l'astronomie, l'économie et la démographie. L'analyse des données de survie consiste à étudier le délai de la survenue d'un événement précis, l'évènement étudié est le passage entre deux états  $A$  et  $B$ . L'évènement terminal peut être considérés tels que la survenue d'une maladie en épidémiologie, la survenue d'une panne dans les applications industrielles, l'acceptation d'une offre d'emploi pour une personne au chômage en économie, ou le mariage en démographie.

Les données de survie se caractérisent par l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, à cause des processus de censure et de troncature. Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information.

Si la durée de vie d'un individu n'est observable que dans une période donnée, on parle de censure. Il s'agit de troncature, si l'individu doit survivre longtemps pour être observé, ou lorsque l'individu est retenu par l'étude uniquement s'il a connu l'évènement en question à une date  $t$  donnée.

La régression polynomiale locale est une méthode de régression non paramétrique, elle a été étudiée par Stone (1977) [38], Cleveland (1979) [5], Tsybakov (1986) [43] et Fan (1992, 1993) [7, 8], elle a de nombreuses applications statistiques intéressantes, en particulier pour l'estimation des fonctions statistiques. Pour chaque point de données  $x_0$ , on ajuste un polynôme d'ordre faible (droite) à travers les données au voisinage de  $x_0$ , le résultat sera une fonction lisse sur le support des données.

Cette méthode a de nombreux avantages, elle réduit le biais de l'estimateur de Nadaraya-Watson et la variance de l'estimateur de Gasser-Muller (1979) [16] (voir Chu & Marron, (1991) [4] Fan, (1992) [7]). Elle s'adapte automatiquement aux bornes i.e le biais aux bornes reste de même ordre que dans l'intérieur. (voir Fan et Gijbels (1992) [9] Hastie et Loader, (1993) [21]).

La qualité de la fonction estimée dépend du choix de la fonction de pondération  $K(\cdot)$ , la taille de la fenêtre de lissage  $h$  et l'ordre de polynôme ajusté  $p$ . nous discutons de chacun de ces choix, en accordant une attention particulière à la fenêtre de lissage. La

fenêtre de lissage est introduite pour tenir compte des différents degrés de lissage, ce qui entraîne une réduction du biais au niveau des régions de pic et de la variance au région trop bas. On a deux fenêtres à choisir, la fenêtre locale et la fenêtre globale, la première est choisie par la méthode de validation croisée et une autre méthode proposée par Ushakov et Ushakov (2012) [45], et la deuxième est détaillée par Fan et Gijbels (1992) [9].

Concernant l'estimateur de  $NW$  de la fonction moyenne conditionnelle avec des données complètes dépendantes voir Györfi et al. (1989) [20], Bosq (1998) [1]. Masry et Fan (1997) [32] considère l'estimation de  $E(Y|X = x)$  pour les suites de mélange en utilisant l'estimateur  $LP$ . Pour plus de références sur la technique de régression non paramétrique avec des données dépendantes voir Fan et Yao (2003) [13].

Dans certains domaines, des données censurées à droite ou tronquées à gauche sont souvent rencontrées. Les données tronquées à gauche se produit en astronomie, économie, épidémiologie et en biométrie voir Woodroffe (1985) [48], Feigelson et Babu (1992) [14], Wang et al. (1986) [46], Tsai et al. (1987) [44] et He et Yang (1994) [22]. Recemment Ould-Said et Lemdani (2006) [34] ont construit un nouveau estimateur non paramétrique (de type  $NW$ ) de  $E(Y|X = x)$  pour le modèle tronqué à gauche et ils ont étudié ses propriétés asymptotiques avec des paramètres indépendantes et identiquement distribuées (i.i.d). Voir aussi Liang et Al. (2009) [26] pour des données dépendantes.

Notre travail consiste à étudier l'estimateur localement linéaire de la fonction moyenne conditionnelle et ses dérivées pour les données incomplètes de type tronquées à gauche. La fonction objective est la fonction de régression. Il est supposé que les données de survie observées  $\{(X_k, Y_k, T_k), \{1 \leq k \leq n\}\}$  forment une suite de  $\alpha$ -mélange stationnaire. Ainsi, nous étudions la normalité asymptotique de l'estimateur.

Ce mémoire est composé de six chapitres

**Chapitre 1** : Ce chapitre porte une introduction à l'estimation non paramétrique sur des données complètes, particulièrement, les estimateurs à noyau et les méthodes existants de sélection de la fenêtre de lissage.

**Chapitre 2** : Ce chapitre est consacré à la régression non paramétrique par la méthode à noyau, en détaillant la régression polynomial local.

**Chapitre 3** : Ce troisième chapitre porte une introduction aux données incomplètes et l'estimation sous le modèle de troncature à gauche.

**Chapitre 4** : Dans ce chapitre on introduit des notions de données dépendantes.

**Chapitre 5** : Cet avant dernier chapitre traite l'estimation des données tronquées par la méthode des polynômes locaux, la normalité asymptotique est établi.

**Chapitre 6** : Ce dernier chapitre porte une petite application.

**Mots-clefs** : Estimation non paramétrique, troncature, approximation locale, régression, estimateur de Nadaraya Watson, estimateur localement polynomial,  $\alpha$ -mélange, vitesse de convergence.

# Chapitre

# 1

## Estimation non paramétrique sur des données complètes

### 1 Introduction

La statistique non paramétrique consiste à estimer le paramètre inconnu qui n'est pas supposé d'appartenir à une famille indexée par un petit nombre de paramètres réels. En général, dans la théorie non-paramétrique on suppose que le nombre de paramètres qui décrivent la loi des observations est une fonction croissante du nombre d'observations, ou encore que le nombre de paramètres est infini.

L'estimation non paramétrique de la densité a fait l'objet de multiples travaux par des méthodes diverses durant ces dernières années. L'intérêt de ce type d'estimation est justifié, par le fait qu'elle réduit les hypothèses paramétriques.

Les méthodes les plus utilisées sont les suivantes : les histogrammes et les estimateurs à noyau, les estimateurs par fonctions orthogonales, les méthodes dites du maximum de vraisemblance. Dans ce mémoire on s'intéresse aux estimateurs à noyaux.

### 2 Estimateur à noyau d'une densité

Soient  $X_1, \dots, X_n$  des variables aléatoires indépendantes identiquement distribuées de densité de probabilité  $f$ , et de fonction de répartition  $F(X) = \int_{-\infty}^x f(t)dt$ .

Considérons la fonction de répartition empirique

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(X_i \leq x)}$$

où  $\mathbf{1}_{(\cdot)}$  désigne la fonction indicatrice. D'après la loi forte des grands nombres, presque sûrement,

$$F_n(x) \xrightarrow{P.S} F(x), \text{ quand } n \rightarrow \infty, \forall x \in \mathbb{R}$$

Donc,  $F_n$  est un estimateur convergent (consistant) de  $F$ .

Pour  $h > 0$  assez petit, Rosenblatt (1956) a proposée d'estimer  $f$  par

$$f(x) \approx \frac{F(x+h) - F(x-h)}{2h}$$

En remplaçant ici  $F$  par l'estimateur  $F_n$  on obtient

$$\widehat{f}_n^R(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}$$

On peut aussi l'écrire sous la forme

$$\begin{aligned} \widehat{f}_n^R(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} \mathbf{1}_{(x-h < X_i \leq x+h)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} \mathbf{1}_{(-1 < \frac{X_i - x}{h} \leq 1)} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K_0\left(\frac{X_i - x}{h}\right) \end{aligned}$$

où  $K_0(u) = \frac{1}{2} \mathbf{1}_{(-1 < u \leq 1)}$ . Parzen (1962) a proposé de remplacer  $K_0$  par un noyau plus général.

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$$

C'est l'estimateur à noyau de la densité ou l'estimateur de Parzen - Rosenblatt. La fonction  $K$  est dite noyau et le paramètre  $h$  fenêtre (en anglais "bandwidth") de l'estimateur.

La fenêtre  $h$  dépend de  $n$  et on la notera  $h_n$ . La suite  $(h_n)_{n \geq 1}$  tendant vers 0 lorsque  $n \rightarrow \infty$ . La notation  $h$ , sans indice  $n$ , sera également utilisée afin d'abrégier l'écriture.

En general, les hypothèses suivantes sont imposées sur le noyau :

- 1)  $K$  est symétrique, i.e  $K(u) = K(-u)$
- 2)  $\int_{\mathbb{R}} K(u) du = 1$

$$3) \int_{\mathbb{R}} u^j K(u) du = 0 \text{ pour } j = 1, \dots, k - 1$$

$$4) \int_{\mathbb{R}} u^k K(u) du \neq 0$$

Dans ce dernier cas,  $K$  est appelé noyau d'ordre  $k$ . A cause de la symétrie  $k$  est nécessairement pair. La condition deux garantie que  $\hat{f}_n(x)$  est une densité de probabilité i.e.

$$\begin{aligned} \int \hat{f}_n(x) dx &= \frac{1}{nh} \sum_{i=1}^n \int K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{n} \sum_{i=1}^n \int K(u) du \\ &= 1, \quad x \in \mathbb{R}. \end{aligned}$$

Voici quelques exemples de noyaux classiques :

- $K(u) = \frac{1}{2} \mathbf{1}_{(|u| \leq 1)}$  (noyau rectangulaire ou de Rosenblatt),
- $K(u) = (1 - |u|) \mathbf{1}_{(|u| \leq 1)}$  (noyau triangulaire),
- $K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}_{(|u| \leq 1)}$  (noyau parabolique ou d'Epanechnikov),
- $K(u) = \frac{15}{16} (1 - u^2)^2 \mathbf{1}_{(|u| \leq 1)}$  (noyau quartique ou de Tukey),
- $K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \mathbf{1}_{(u \in \mathbb{R})}$  (noyau gaussien),
- $K(u) = \frac{1}{2} \exp\left(-\frac{|u|}{\sqrt{2}}\right) \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right)$  (noyau de Silverman).

On a programmé l'estimateur à noyau de la densité de la variable  $X$  à l'aide de langage Matlab **R2012a**, et on compare cet estimateur avec la densité normale centrée réduite avec  $n = 500$  et  $h = 0.37$ .

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$$

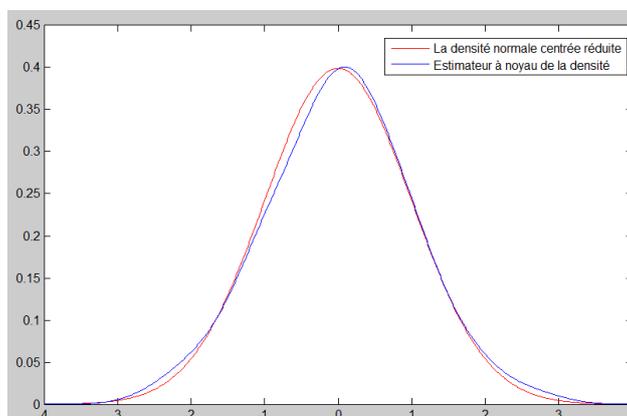


FIG. 1:  $n=500$ ,  $h=0.37$

## 2.1 Propriétés asymptotiques des estimateurs à noyaux

### Calcul du biais

On a :

$$\begin{aligned}
 \mathbb{E} \left[ \widehat{f}_n(x) \right] &= \frac{1}{nh} \mathbb{E} \left[ \sum_{i=1}^n K \left( \frac{X_i - x}{h} \right) \right] \\
 &= \frac{1}{h} \mathbb{E} \left[ K \left( \frac{X - x}{h} \right) \right] \\
 &= \frac{1}{h} \int_{-\infty}^{+\infty} K \left( \frac{t - x}{h} \right) f(t) dt \\
 &= \int_{-\infty}^{+\infty} K(u) f(x + uh) du \text{ en posant } u = \frac{t - x}{h}
 \end{aligned}$$

Comme  $\int_{-\infty}^{+\infty} K(u) du = 1$  le biais peut s'écrire :

$$\mathbb{E}(\widehat{f}_n(x)) - f(x) = \int_{-\infty}^{+\infty} K(u) [f(x + uh) - f(x)] du$$

Prenons un développement de Taylor de  $f$  au voisinage de  $x$  :

$$f(x + uh) = f(x) + uhf'(x) + \frac{(uh)^2}{2!} f''(x) + o(h^2)$$

Le biais s'écrit :

$$\begin{aligned}
 \text{Biais}(\widehat{f}_n(x)) &= \mathbb{E} \left[ \widehat{f}_n(x) \right] - f(x) \\
 &= hf'(x) \int_{-\infty}^{+\infty} uK(u) du + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2) \\
 &= \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2)
 \end{aligned}$$

car  $\int_{-\infty}^{+\infty} uK(u) du = 0$ ,  $K(u)$  est Symétrique.

Donc

$$\text{Biais}(\widehat{f}_n(x)) = \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2)$$

Pour  $h$  petit le biais dépend donc de  $f''(x)$  et du moment d'ordre 2 du noyau. Le biais est du signe de  $f''(x)$ .

### Calcul de la variance

On a :

$$\begin{aligned} \text{Var}(\widehat{f}_n(x)) &= V\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)\right) \\ &= \frac{1}{nh^2} V\left(K\left(\frac{X - x}{h}\right)\right) \\ &= \frac{1}{n} \left\{ \frac{1}{h^2} \mathbb{E}\left(\left[K\left(\frac{X - x}{h}\right)\right]^2\right) - \frac{1}{h} \mathbb{E}^2\left(K\left(\frac{X - x}{h}\right)\right) \right\} \end{aligned}$$

Avec

$$\begin{aligned} \frac{1}{nh^2} \mathbb{E}\left(\left[K\left(\frac{X - x}{h}\right)\right]^2\right) &= \frac{1}{nh^2} \int_{-\infty}^{+\infty} \left[K\left(\frac{t - x}{h}\right)\right]^2 f(t) dt \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} [K(u)]^2 f(x + uh) du \text{ en posant } u = \frac{t - x}{h} \end{aligned}$$

Et

$$\frac{1}{nh} \mathbb{E}^2\left(K\left(\frac{X - x}{h}\right)\right) = \frac{1}{n} \left[ \int_{-\infty}^{+\infty} K(u) f(x + uh) du \right]^2$$

Alors que le terme

$$\frac{1}{nh} \mathbb{E}^2\left(K\left(\frac{X - x}{h}\right)\right) \xrightarrow{n \rightarrow \infty} 0$$

Le terme  $\frac{1}{nh^2} \mathbb{E}\left(\left[K\left(\frac{X - x}{h}\right)\right]^2\right)$  ne tend vers zéro que si  $nh \rightarrow \infty$ . Par conséquent, pour que  $\widehat{f}_n(x)$  converge vers  $f(x)$  en moyenne quadratique il faut que :  $n \rightarrow \infty, h \rightarrow 0, nh \rightarrow \infty$ .

Le terme  $\frac{1}{nh} \mathbb{E}\left(\left[K\left(\frac{X - x}{h}\right)\right]^2\right)$  est d'ordre  $\frac{1}{n}$ , ce que l'on note  $O\left(\frac{1}{n}\right)$ . En utilisant le développement de Taylor :  $f(x + uh) = f(x) + uhf'(x) + o(h)$  On obtient :

$$\frac{1}{nh^2} \mathbb{E} \left( \left[ K \left( \frac{X - x}{h} \right) \right]^2 \right) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O(h)$$

D'où :

$$\text{Var} [\hat{f}_n(x)] = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O(h)$$

## 2.2 Cas d'une densité multivariée

Soit  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  une densité et  $(X_1, Y_1), \dots, (X_n, Y_n)$  un échantillon de densité  $f$ . On utilise un noyau produit et on construit

$$\hat{f}_n(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K \left( \frac{X_i - x}{h} \right) K \left( \frac{Y_i - y}{h} \right).$$

Plus généralement dans  $\mathbb{R}^p$ , on utilise

$$\hat{f}_n(x^1, \dots, x^p) = \frac{1}{nh^p} \sum_{i=1}^n \prod_{j=1}^p K \left( \frac{X_i^j - x^j}{h} \right).$$

## 2.3 Calcul du risque quadratique en un point

$$\begin{aligned} \text{MSE}(\hat{f}_n(x)) &= \mathbb{E} \left\{ \hat{f}_n(x) - f(x) \right\}^2 \\ &= \mathbb{E} \left[ \hat{f}_n(x) - \mathbb{E} [\hat{f}_n(x)] + \mathbb{E} [\hat{f}_n(x)] - f(x) \right]^2 \\ &= \mathbb{E} \left( \hat{f}_n(x) - \mathbb{E} [\hat{f}_n(x)] \right)^2 + \left( \mathbb{E} [\hat{f}_n(x)] - f(x) \right)^2 \\ &\quad + 2 \left( \hat{f}_n(x) - \mathbb{E} [\hat{f}_n(x)] \right) \left( \mathbb{E} [\hat{f}_n(x)] - f(x) \right) \\ &= \text{Var} [\hat{f}_n(x)] + \left( \mathbb{E} [\hat{f}_n(x)] - f(x) \right)^2 \\ &= \text{Var} [\hat{f}_n(x)] + \left( \text{Biais} \{ \hat{f}_n(x) \} \right)^2 \\ &= \frac{h^4}{4} [f''(x)]^2 \left[ \int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + \frac{f(x)}{nh} \int_{-\infty}^{+\infty} [K(u)]^2 du \\ &\quad + O(h^4) + O\left(\frac{1}{h}\right) \end{aligned}$$

Faisant abstraction des termes  $O(h^4) + O(\frac{1}{h})$  négligeables dans les conditions de convergence, on remarque que

Si  $h$  décroît alors le biais décroît et la variance croît.

Si  $h$  augmente alors le biais croît et la variance décroît.

Donc, il faut choisir un  $h$  qui fasse un compromis entre le biais et la variance.

Il existe un optimum (mais valable uniquement au point  $x$ ) est obtenu en dérivant par rapport à  $h$ , soit :

$$h_{opt} = \left[ \frac{f(x) \int_{-\infty}^{+\infty} [K(u)]^{-2} du}{\left( f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du \right)^2} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}$$

et, en remplaçant  $h_{opt}$  dans la formule de l'expression asymptotique de l'e.q.m., Ainsi la convergence est d'ordre  $n^{-4/5}$ .

## 2.4 Calcul du risque quadratique intégré

Jusqu'à présent nous avons raisonné à  $x$  fixé. Il est clair que ce qui nous intéresse est de connaître le comportement de l'estimateur  $\hat{f}_n$  de la fonction  $f$  globalement sur tout  $\mathbb{R}$ . Pour cela on considère, pour une réalisation donnée, son écart à  $f$  intégré sur tout  $\mathbb{R}$  ce qui conduit, en prenant l'espérance mathématique de cet écart intégré, au critère d'erreur quadratique moyenne intégrée (MISE en anglais : mean integrated squared error) :

$$MISE(\hat{f}_n) = \mathbb{E} \left( \int_{-\infty}^{+\infty} \{ \hat{f}_n(x) - f(x) \}^2 dx \right)$$

Celle-ci se calcule aisément à, partir des résultats précédents car, étant donné les conditions de régularité imposées à  $f$  et à  $K$ , il est permis d'invertir les intégrations ce qui conduit à intégrer l'expression  $MSE$  en  $x$  fixé :

$$\begin{aligned} MISE(\hat{f}_n) &= \int_{-\infty}^{+\infty} \mathbb{E} \left( \{ \hat{f}_n(x) - f(x) \}^2 \right) dx \\ &= \int_{-\infty}^{+\infty} MSE(\hat{f}_n(x)) dx \end{aligned}$$

D'où :

$$\begin{aligned}
MISE(\widehat{f}_n(x, h)) &= \frac{h^4}{4} \int_{-\infty}^{+\infty} [f''(x)]^2 dx \left[ \int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 \\
&\quad + \frac{1}{nh} \int_{-\infty}^{+\infty} [K(u)]^2 du + O(h^4) + O\left(\frac{1}{h}\right)
\end{aligned}$$

Dans ce qui suit, nous supposons que la densité à estimer  $f(x)$  et le noyau  $K(x)$  sont des fonctions de carré intégrable, de sorte que le MISE est finie. Pour une fonction  $g(x)$  on définit :

$$R(g) = \int_{-\infty}^{+\infty} (g(x))^2 dx, \quad \mu_2(g) = \int_{-\infty}^{+\infty} x^2 g(x) dx.$$

À condition que les intégrales existent.

Le  $MISE(\widehat{f}_n(x, h))$  est représenté sous la forme :

$$MISE(\widehat{f}_n(x, h)) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh} R(K) + O(h^4) + O\left(\frac{1}{h}\right)$$

Quand  $h \rightarrow 0$ ,  $nh \rightarrow \infty$  on peut définir  $AMISE$  comme :

$$AMISE(\widehat{f}_n(x, h)) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh} R(K)$$

$h_{AMISE}$  est exprimé explicitement :

$$h_{AMISE} = \left[ \frac{R(K)}{R(f'') (\mu_2(K))^2 n} \right]^{\frac{1}{5}}$$

Ainsi, en tant qu'estimateur fonctionnel un estimateur à noyau converge plus vite vers la vraie densité  $f$  que l'histogramme. Mais ce résultat repose sur un choix optimal très théorique (puisque dépendant de l'inconnue  $f''$  et de conditions de convergence artificielles).

### 3 Sélection de la fenêtre de lissage

#### 3.1 Méthode de validation croisée non biaisée [43]

On suppose que le noyau  $K$  est fixé, et on ne s'intéresse qu'au choix de la fenêtre  $h$ . Notons que  $MISE = MISE(h)$  est une fonction de la fenêtre  $h$  et que la valeur idéale

de  $h$  est donnée par

$$h_{id} = \arg \min_{h>0} MISE(h) \quad (1.1)$$

Malheureusement, cette valeur n'est pas accessible car la fonction  $MISE(h)$  dépend de la densité inconnue  $f$ . Les résultats du paragraphe précédent ne permettent pas de construire un estimateur qui approche cette valeur idéale. Pour ce faire, il faut utiliser d'autres méthodes dont la plus courante est celle de la validation croisée.

L'idée principale de la méthode de validation croisée est celle de la minimisation par rapport à  $h$  d'un estimateur du risque intégré ( $MISE$ ). Nous allons essayer de remplacer  $MISE(h)$  par une fonction de  $h$  mesurable par rapport à l'échantillon et dont la valeur, pour chaque  $h > 0$ , est un estimateur sans biais de  $MISE(h)$ . Pour cela, notons que

$$\begin{aligned} MISE(\hat{f}_n) &= E \int \{\hat{f}_n(x) - f(x)\}^2 dx \\ &= E \int \hat{f}_n^2(x) dx - 2E \int f(x) \hat{f}_n(x) dx + \int f^2(x) dx \end{aligned}$$

Comme la dernière intégrale  $\int f^2(x) dx$  ne dépend pas de  $h$ , la valeur  $h_{id}$  de  $h$  qui minimise  $MISE(h)$  est aussi celle qui minimise la fonction

$$J(h) = E \int \hat{f}_n^2(x) dx - 2E \int f(x) \hat{f}_n(x) dx$$

Cherchons un estimateur sans biais de  $J(h)$ . Il suffit pour cela d'estimer sans biais les quantités  $E \int \hat{f}_n^2(x) dx$  et  $E \int f(x) \hat{f}_n(x) dx$ . Or  $E \int \hat{f}_n^2(x) dx$  admet l'estimateur sans biais  $\int \hat{f}_n^2(x) dx$  il reste donc à trouver un estimateur sans biais de  $E \int f(x) \hat{f}_n(x) dx$ .

On remarque que

$$\mathbb{E} \int f(x) \hat{f}_n(x) dx \stackrel{Fubini}{=} \int f(x) \mathbb{E}(\hat{f}_n(x)) dx = \int f(x) \frac{1}{h} \int K\left(\frac{u-x}{h}\right) f(u) du dx$$

On utilise alors le résultat suivant.

**Lemme 3.1.1** *On considère*

$$\hat{T}_n = \frac{1}{n(n-1)h} \sum_{\substack{i=1 \\ j \neq i}}^n \sum_{j=1}^n K\left(\frac{X_i - X_j}{h}\right)$$

Alors  $\widehat{T}_n$  est un estimateur sans biais de

$$\frac{1}{h} \int \int K\left(\frac{u-x}{h}\right) f(u)f(x)du dx$$

**Preuve.** En effet,

$$\begin{aligned} \mathbb{E}(T_n) &= \frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathbb{E}K\left(\frac{X_i - X_j}{h}\right) \\ &\stackrel{X_i \text{ i.i.d.}}{=} \frac{1}{h} \mathbb{E}\left[K\left(\frac{X_1 - X_2}{h}\right)\right] \\ &= \frac{1}{h} \int \int K\left(\frac{u-x}{h}\right) f(u)f(x)du dx \end{aligned}$$

**Remarque 3.1.1** On peut noter que de façon très générale, il est important quand on considère une somme double de la forme  $\sum_{i,j} \phi(X_i - X_j)$ , de la priver de sa diagonale  $i \neq j$ , sinon on augmente le biais. En effet, considérons par exemple

$$\widehat{T}_n = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n K\left(\frac{X_i - X_j}{h}\right)$$

Alors la moyenne de  $T_n$  fait apparaître un terme parasite :

$$\begin{aligned} \mathbb{E}(\widehat{T}_n) &= \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}K\left(\frac{X_i - X_j}{h}\right) \\ &= \frac{1}{nh} K(0) + \frac{n-1}{nh} \mathbb{E}K\left(\frac{X_i - X_j}{h}\right) \end{aligned}$$

Revenons à l'estimation de  $J$ . Ainsi, on définit

$$\widehat{j}(h) = \int \widehat{f}_n^2(x) dx - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n K\left(\frac{X_i - X_j}{h}\right)$$

et d'après ce qui précède, c'est un estimateur sans biais de  $J(h)$ .

Comme la fonction  $\widehat{J}(h)$  est calculable à partir des observations  $X_1, \dots, X_n$ , on peut évaluer

$$h^{CV} = \arg \min_{h>0} \widehat{J}(h)$$

pourvu que le minimum soit atteint. Finalement, on peut définir l'estimateur  $\widehat{f}_n^{CV}(x)$  de

la densité  $f$  par la méthode de la validation croisée :

$$\widehat{f}_n^{CV}(x) = \frac{1}{nh_{CV}} \sum_{i=1}^n K\left(\frac{X_i - x}{h_{CV}}\right)$$

Il s'agit d'un estimateur à noyau dont la fenêtre aléatoire  $h^{CV}$  dépend de l'échantillon  $X_1, \dots, X_n$ . ■

## 3.2 Méthode de Ushakov et Ushakov (2012).

### 3.2.1 Introduction

Après que, Glad, Hjort, et Ushakov (2007) [17] ont obtenu un certain nombre d'inégalités supérieures à  $MISE$  qui étaient des approximations plus précises de la  $MISE$  par rapport à l' $AMISE$ . Ushakov et Ushakov (2012) [46] ont prouvé l'existence et l'unicité de la fenêtre optimale.

### 3.2.2 Présentation de la méthode

Dans l'estimation à noyau de la densité et la régression, il a été remarqué que la sélection de la fenêtre basée sur la minimisation de la  $AMISE$ , donne souvent un sous-lissage. Ceci est expliqué par les propriétés des méthodes qui sont utilisées pour l'approximation de la  $AMISE$  (validation croisée par exemple). Cependant, le sous-lissage se produit lorsque la densité à estimer a une structure simple ou plus ou moins compliquée (mélange de lois), et même si on utilise l'exacte  $AMISE$ . Ceci est démontré dans la Figure 1. La densité à estimer est un mélange de deux densités normales.

$$\frac{1}{2}N(-1.5, 3) + N(1.5, 0.1)$$

Avec  $N(\mu, \sigma^2)$  désigne la fonction de densité de probabilité de la distribution normale d'espérance  $\mu$  et de variance  $\sigma^2$ . La taille de l'échantillon est  $n = 1000$ . Sur la Figure 1(a), la courbe en trait plein est l'estimateur à noyau avec la fenêtre  $h_{AMISE} = 0,16$ . Elle est nettement sous lissée (l'estimation comporte au moins neuf modes lorsque la densité à être estimée a seulement deux). Un meilleur résultat est obtenu en utilisant la plus grande fenêtre  $h = 0,32$  (la courbe en trait plein sur la Figure 1(b)).

Supposons que la densité à estimée  $f(x)$  a une variance  $\sigma^2$  finie et qui satisfait aux conditions (a) et (b) suivantes :

- a) La dérivée seconde de  $f(x)$  est bornée i.e.  $|f''(x)| \leq m, m \geq 0$ .
- b) Le noyau  $K(x)$  est symétrique et le moment d'ordre deux fini i.e.  $k(x) = k(-x)$  et  $\int x^2 k(x) < \infty$ .

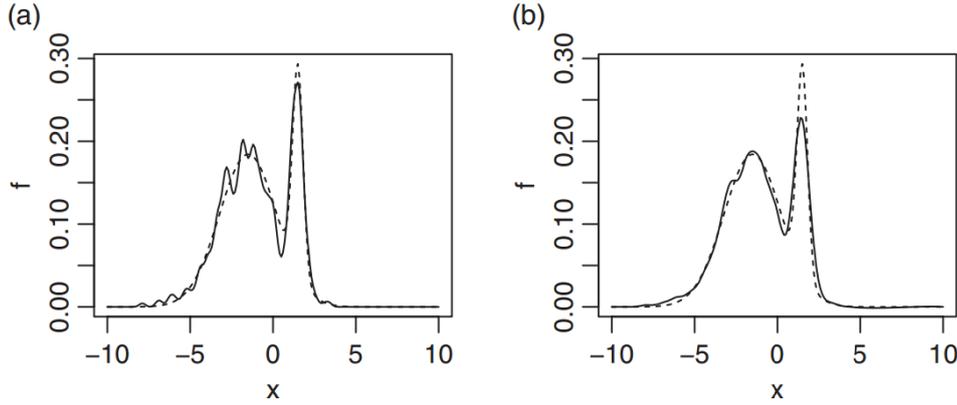


FIG. 2: La ligne continue est l'estimateur, la ligne brisée est la vraie densité. Les fenêtres sont (a)  $h_{AMISE} = 0,16$ , (b)  $h_{AMISE} = 0,32$  la fenêtre optimale est  $h^* = 0,32$ .

Glad, Hjort, et Ushakov (2007) [17] ont obtenu un certain nombre d'inégalités supérieures à  $MISE$  qui étaient des approximations plus précises de la  $MISE$  par rapport à l' $AMISE$ .

$$MISE(f_n(x, h)) < AMISE(f_n(x, h)) - \frac{1}{3\pi\sigma n} \quad (1.2)$$

Le côté droit de la relation (1.2) est aussi simple que  $AMISE$  mais il est essentiellement plus précis, en particulier pour modérer la taille de l'échantillon. Ceci est illustré dans les Figures 2 et 3.

On note par :

$$l_1(h) = AMISE(f_n(x, h))$$

et

$$l_2(h) = AMISE(f_n(x, h)) - \frac{1}{3\pi\sigma n}$$

On remarque sur les courbes de  $l_1(h)$ ,  $l_2(h)$  et l'inégalité 1.2 que les abscisses qui minimisent les deux fonctions  $l_1(h)$ ,  $l_2(h)$  se coïncident. Donc, à première vue, il semble que  $l_2(h)$  et l'équation 1.2 ne donnent pas de nouvelles possibilités pour les comparer avec  $l_1(h)$ . Mais c'est faux. Notez que n'importe quelle courbe entre  $l_1(h)$  et  $l_2(h)$  est une meilleure approximation de  $MISE$ .

Soient  $l^*(h)$  et  $l(h)$  deux courbes parmi les courbes de la bande  $[l_2(h), l_1(h)]$  i.e

$$l_2(h) \leq l^*(h) \leq l_1(h) \text{ et } l_2(h) \leq l(h) \leq l_1(h) \text{ et } h_0 = \arg \min l(h)$$

alors il existe  $h' = \arg \min l^*(h)$  de telle sorte que  $h' \geq h_0$ . On pose :

$$h^* = \max \{ \arg \min l(h) : l_2 \leq l \leq l_1 \} \quad (1.3)$$

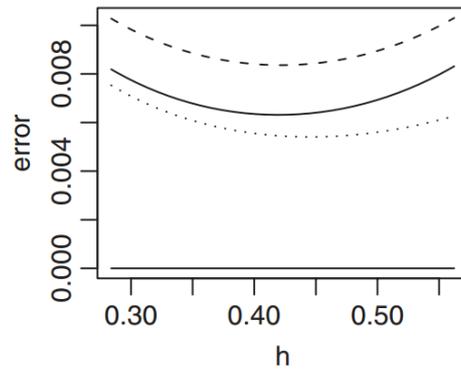


FIG. 3:  $l_1(h)$  est la ligne brisée,  $MISE$  est la ligne pointillée et  $l_2(h)$  en fonction de  $h$ . La densité à estimer et le noyau sont de loi normale standard. La taille de l'échantillon est  $n = 100$ .

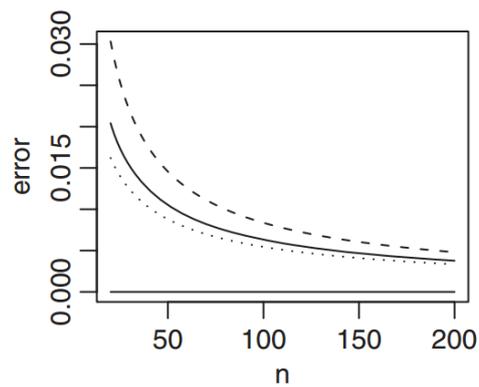


FIG. 4:  $l_1(h)$  est la ligne brisée,  $MISE$  est la ligne pointillée et  $l_2(h)$  en fonction de  $n$ . La densité à estimer et le noyau sont de loi normale standard.  $h = h_{MISE}$ .

Cette valeur  $h^*$  est le choix optimale de la fenêtre. L'existence et l'unicité de  $h^*$  sont prouvées (Voir Ushakov et Ushakov (2012) [46]).

# Chapitre

# 2

## Régression non-paramétrique par la méthode à noyau sur des données complètes

### 1 Introduction

La régression est la technique la plus répandue en statistique pour la prévision. Elle permet d'étudier l'association entre une variable aléatoire, dite explicative  $X$  et une autre variable aléatoire, dite observée  $Y$ . Parmi les méthodes de régression, la régression linéaire simple qui suppose un modèle de la forme :

$$Y_i = a + bX_i + \varepsilon_i$$

Où  $a$  et  $b$  sont des paramètres à estimer (méthode de moindres carrés) et les erreurs aléatoires  $\varepsilon_i$  sont non corrélées, de moyenne nulle et de variance  $\sigma^2$  indépendantes de  $X$ .

Lorsque la linéarité de la relation est mise en défaut, on préfère choisir un modèle plus flexible qui reflète mieux la relation entre  $X$  et  $Y$ . On utilise le modèle de régression non paramétrique.

On dispose d'un échantillon, composé de  $n$  couples indépendants de variables aléatoires  $(X_1, Y_1), \dots, (X_n, Y_n)$ , et on dénote par  $(X, Y)$  un élément générique de cet échantillon.

$$Y_i = m(X_i) + \varepsilon_i, \text{ pour } i = 1, \dots, n, \quad \text{avec } \varepsilon_i \rightsquigarrow N(0, \sigma^2). \quad (*)$$

L'erreur  $\varepsilon_i$  représente la variation de  $Y_i$  autour de  $X_i$ , elle est modélisée par une variable aléatoire gaussienne, qui sera généralement choisie indépendante des observations  $\{X_i :$

$1 \leq i \leq n$ }, et de moyenne nulle. Cette dernière hypothèse simplifie considérablement les calculs et l'expression des propriétés asymptotiques liées à l'estimation de la fonction de régression.

On cherche, dans une famille fixée de fonctions quelle est celle pour laquelle les  $Y$  sont les plus proches de  $m(X)$ . Cette proximité se mesure en général par un risque utilisant l'erreur quadratique moyenne et on essaiera alors de déterminer la fonction  $m^*(X_i)$  qui rendra cette erreur la plus petite possible, c'est-à-dire à trouver une fonction  $m^*(X_i)$  qui minimise l'erreur quadratique moyenne :

$$E |m^*(X) - Y|^2$$

En effet, déterminer  $m^*(X)$  revient à calculer  $\arg \min_m \mathbb{E} [(m(X) - Y)^2 | X = x]$ .

En différenciant  $\mathbb{E} [(m(X) - Y)^2 | X = x]$  par rapport à  $m(\cdot)$ , en égalant le résultat à 0, et finalement en isolant  $m(\cdot)$ , on obtient :

$$\begin{aligned} \frac{\partial}{\partial r} \mathbb{E} [(m(X) - Y)^2 | X = x] &= 0 \\ \Leftrightarrow 2\mathbb{E} [m(X) - Y | X = x] &= 0 \\ \Leftrightarrow 2m(X) - 2\mathbb{E}(Y | X = x) &= 0 \end{aligned}$$

ce qui implique que

$$m^*(x) = \mathbb{E}(Y | X = x)$$

Le fait que la dérivée seconde, qui est égale à 2, soit positive nous permet de conclure que c'est bien un minimum.

Il existe deux cas principaux pour le modèle (\*), dépendants de la nature probabiliste des données  $\{(X_i, Y_i) : 1 \leq i \leq n\}$ .

Le premier cas est le plus simple, et est appelé **dispositif expérimental à effets fixes** (ou "fixed design"). Il correspond à la situation où les  $X_i = x_i$  sont fixés (déterministes).

**Exemple 1.0.1** *Le dispositif expérimental régulier.*[43]

On suppose que  $X_i = x_i = \frac{i}{n}$  et  $m(\cdot)$  une fonction définie de  $[0, 1]$  dans  $\mathbb{R}$  et que les observations  $Y_i$  sont données par

$$Y_i = m\left(\frac{i}{n}\right) + \varepsilon_i, \text{ pour } 1 \leq i \leq n.$$

Le deuxième cas, est le **dispositif expérimental à effets aléatoires** (ou random design) désigne le modèle où les données  $\{X_i : 1 \leq i \leq n\}$  sont strictement aléatoires. Ce modèle est clairement plus général.

Nous allons présenter la fonction de régression, dans le cadre du modèle aléatoire univarié.

Soit  $(X, Y)$  un couple de variables aléatoires réelles admettant une densité jointe sur  $\mathbb{R}^2$  notée  $f_{X,Y}$  et une densité marginale  $f_X$ . La variable  $Y$  est supposée intégrable, i.e.  $\mathbb{E}(|Y|) < \infty$ . On définit la fonction de régression ou l'espérance conditionnelle de  $Y$  sachant que  $X = x$ , par

$$\forall x \in \mathbb{R},$$

$$\begin{aligned} m(x) &= \mathbb{E}(Y|X = x) \\ &= \int_{\mathbb{R}} y f_{Y|X=x}(y) dy \\ &= \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{\int_{\mathbb{R}} f_{X,Y}(x, y) dy} = \frac{r(x)}{f_X(x)} \quad \text{si } f_X(x) \neq 0 \end{aligned}$$

**Définition 1.0.1** Un estimateur  $\hat{m}_n(x)$  de  $m(x)$  est dit **estimateur linéaire** de la régression non-paramétrique si

$$\hat{m}_n(x) = \sum_{i=1}^n Y_i W_{n_i}(x)$$

où la fonction de poids  $W_{n_i}(\cdot)$  ne dépend pas des observations  $Y_i$ .

## 2 L'ajustement linéaire et polynomial

Connaissant un ensemble de points expérimentaux (nuage de points), le problème est de trouver la forme mathématique de la liaison et les bons coefficients (ou paramètres) numériques. La courbe  $y = f(x)$  que l'on cherche à déterminer sera la courbe qui passera au mieux au milieu du nuage des points. La courbe ainsi déterminée, sera appelée courbe d'ajustement (aussi courbe de régression), dans le cas particulier affine, droite d'ajustement (aussi droite de régression).

Si  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$  est le nuage de points, la méthode standard due à Gauss et Legendre est la méthode des moindres carrés : elle consiste à déterminer les valeurs des paramètres, comme celles qui minimisent la somme des carrés des écarts

$$\sum_{i=1}^n (y_i - f(x_i))^2$$

## 2.1 Exemple : Fan et Gijbels(1996)

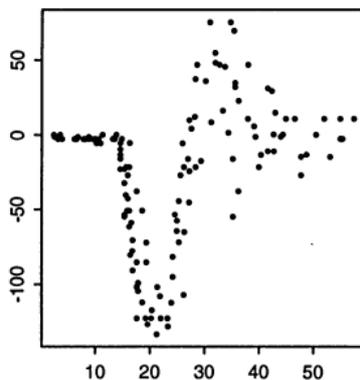


FIG. 5: Nuage de points concerne 133 observations des données de moto. L'axe des X est le temps et l'axe des Y est l'accélération. (Schmidt, Mattern and Schiiler (1981)).

Le nuage de points apparaît non linéaire dans la figure suivante.

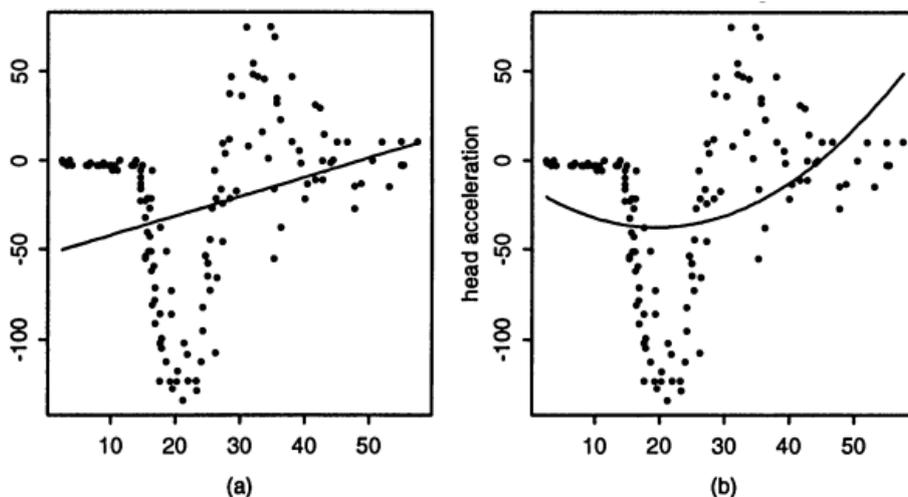


FIG. 6: Ajustement polynomial (linéaire et quadratique) pour les données de moto.

La régression linéaire créera un biais de modélisation très grand. L'approche la plus populaire est d'augmenter le nombre de paramètres en utilisant la régression polynomiale.

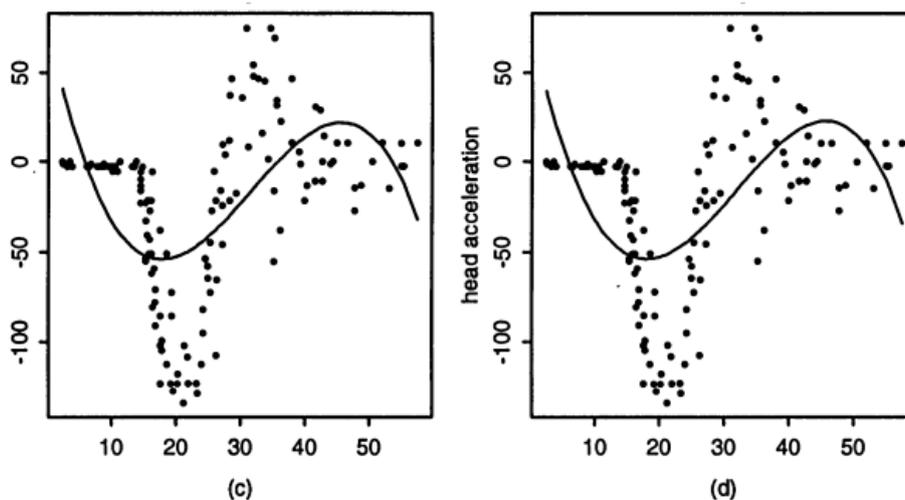


FIG. 7: Ajustement polynomial (cubique et de degré 4) pour les données de moto.

Les figure 5 et 6 montrent que l'ajustement polynomial possède un biais de modélisation très grand.

La régression polynomiale est largement utilisée mais souffre de quelques inconvénients :

- Les fonctions polynomiales ne sont pas très flexibles pour modéliser de nombreux problèmes rencontrés en pratique.
- Les observations peuvent avoir une grande influence sur les régions éloignées de la courbe.
- Le degré du polynôme ne peut pas être contrôlé.

### 3 Modélisation locale

Il y'a plusieurs façon de réparer les inconvénients de l'ajustement polynomial, on cite la méthode de spline, la méthode des séries orthogonales et la régression linéaire locale. Dans notre étude on s'intéresse à la régression linéaire locale.

Pour un point donné  $x$ , on a la fonction linéaire  $m(\cdot)$  au voisinage de  $x$ , on applique la technique de régression linéaire à un groupe de données au voisinage de  $x$ . cette approche s'appelle l'approche de modélisation locale (linéaire).

La taille du voisinage locale s'appelle fenêtre (bandwidth).

Ainsi, au lieu de résoudre un problème paramétrique avec plusieurs paramètres, on résout plusieurs problèmes de régression linéaire avec seulement deux paramètres.

Pour obtenir la courbe de régression en un point donné  $x$ , on applique la technique de régression linéaire sur un groupe de données au voisinage de  $x$ .

Soit  $h$  la taille de la fenêtre locale. Nous modélisons le groupe de données au voisinage de  $x$  par :

$$Y_i = a(x) + b(x)X_i + \varepsilon_i, \quad \text{pour } X_i \in x \pm h$$

Où les paramètres  $a$  et  $b$  dépendent de  $x$ . On peut attribuer un poids de  $K\left(\frac{X_i - x}{h}\right)$  au point  $(X_i, Y_i)$  au voisinage de  $x$ , en conduisant au problème de moindres carrés pondérés suivant :

$$\sum_{i=1}^n (Y_i - a(x) - b(x)X_i)^2 K\left(\frac{X_i - x}{h}\right) \mathbf{1}_{\left\{\left|\frac{X_i - x}{h}\right| \leq 1\right\}} \quad (**)$$

Si  $K$  est à support compact  $[-1, 1]$ , le problème  $(**)$  peut s'écrire comme suit :

$$\sum_{i=1}^n (Y_i - a(x) - b(x)X_i)^2 K\left(\frac{X_i - x}{h}\right)$$

La fonction  $K$  est la fonction noyau, on note par  $\hat{a}(x)$  et  $\hat{b}(x)$  les estimateurs de moindres carrés pondérés au voisinage de  $x$ . alors, l'estimateur de la courbe de régression au point  $x$  est :

$$\hat{m}(x) = \hat{a}(x) + \hat{b}(x)x$$

**Exemple :** Fan et Gijbels (1996)

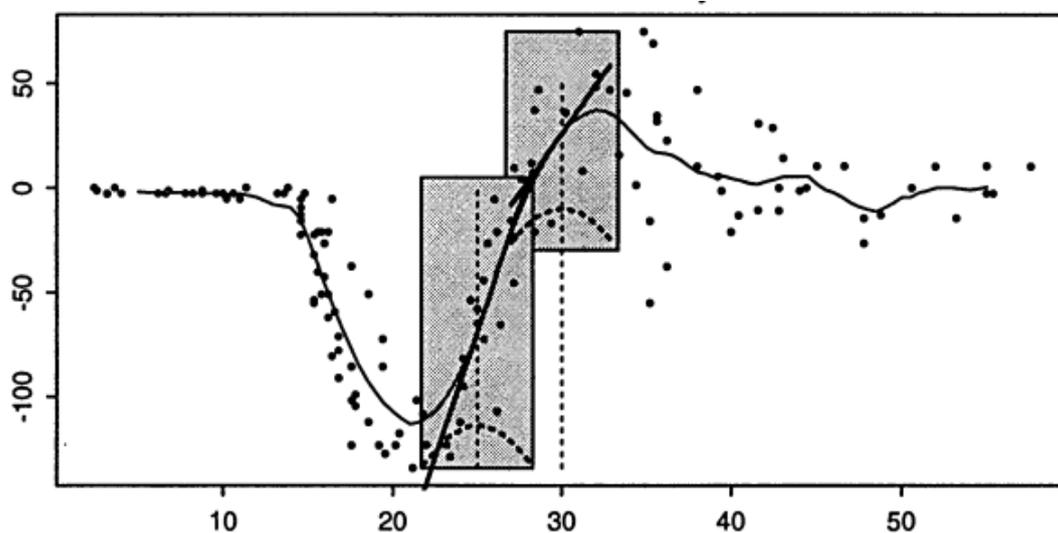


FIG. 8: Ajustement linéaire locale des données de moto. (L'axe des X est le temps et l'axe des Y est l'accélération). Illustration de l'approche de modélisation locale. Pour  $x$  donné, on ajuste un modèle linéaire pondéré pour le groupe de points contenu dans la bande  $x \pm 3.3$ .

Pour obtenir la courbe de régression en un point donné  $x = 25$ , on ajuste une droite à travers les données au voisinage de  $25 \pm 3.3$ . Avec la fonction noyau  $K(t) = (1 - t^2)$  on

trouve la valeur estimée  $\hat{m}(25) = -69,01$ . Maintenant, si on veut estimer la fonction de régression en un autre point  $x = 30$ , nous avons ajuster une droite en utilisant les données au voisinage de  $30 \pm 3,3$ . Donc, la courbe est obtenue en exécution de 101 régressions linéaires associées à 101 bandes de données.

## 4 Les avantages de l'ajustement polynômial local

- Il s'adapte au modèle aléatoire et au modèle fixé.
- Il y'a une absence des effets de bords : sans utilisation de noyaux spécifiques aux bords, le biais reste du même ordre que dans l'intérieur.
- Il ne nécessite pas la modification des bords, et c'est un mérite important, surtout lorsqu'il s'agit de situations multidimensionnelles, car c'est très difficile.
- Les estimateurs par les polynômes locaux ont de belles propriétés d'efficacité minimax : l'efficacité minimax asymptotique pour les commandes fréquemment utilisées est de 100% chez tous les estimateurs linéaires et seulement une petite perte doit être tolérée au-delà de cette classe.

## 5 Quelques estimateur à noyau de la fonction de régression

### 5.1 L'estimateur de Nadaraya-Watson (1964)

Soit un  $n$ -échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  de variables aléatoires à valeurs réelles, de même loi que le couple  $(X, Y)$ . L'estimateur  $\hat{m}_n(x)$  de la fonction de régression est construit à partir des couples d'observations  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ . l'estimateur à noyau de Nadaraya-Watson noté estimateur  $NW$ . Il est construit à partir d'une fonction noyau  $K(\cdot)$  et d'une fenêtre  $h$ .

Si le noyau  $K$  est symétrique (ou d'ordre 1) alors, l'estimateur  $NW$  se présente sous la forme d'une moyenne locale pondérée des valeurs  $Y_i$ , il est défini par

$$\forall x \in \mathbb{R}, \hat{m}_{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \times \mathbf{1}_{\left\{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \neq 0\right\}}$$

où  $\mathbf{1}_{\{\cdot\}}$  désigne la fonction indicatrice.

**Preuve.** On a

$$\forall x \in \mathbb{R}, m(x) = \mathbb{E}(Y|X = x) = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{\int_{\mathbb{R}} f_{X,Y}(x, y) dy} = \frac{r(x)}{f_X(x)}$$

Comme les densités  $f_{X,Y}$  et  $f_X$  sont inconnues, on peut les estimer via un estimateur à noyau. On considère donc

$$\forall (x, y) \in \mathbb{R}, \hat{f}_{X,Y}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right),$$

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right),$$

En effet, pour tout  $x \in \mathbb{R}$  tel que  $\hat{f}_X(x) \neq 0$ , on a

$$\hat{m}(x) = \frac{\int y \hat{f}_{X,Y}(x, y) dy}{\hat{f}_X(x)} = \frac{\frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int y K\left(\frac{Y_i - y}{h}\right) dy}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

Si  $K$  est un noyau d'ordre 1, alors,  $\int y K\left(\frac{Y_i - y}{h}\right) dy = h \int (Y_i - uh) K(u) du = h Y_i$ . Donc on obtient bien

$$\hat{m}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

L'estimateur de Nadaraya-Watson est une moyenne pondérée des observations  $Y_i$ . On a

$$\forall x \in \mathbb{R}, \hat{m}(x) = \sum_{i=1}^n w_{n,i}(x) Y_i$$

où les poids  $w_{n,i}(x)$  vérifient

$$w_{n,i}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right)} \times \mathbf{1}_{\left\{\sum_{j=1}^n K\left(\frac{X_j - x}{h}\right) \neq 0\right\}}$$

Ces poids ne dépendent pas des observations  $Y_i$ . En particulier,  $\hat{m}_{NW}$  est un estimateur

linéaire de la régression non paramétrique des  $Y_i$  sur les  $X_i$ . Le noyau  $K$  détermine la forme du voisinage autour du point  $x$  alors que la fenêtre  $h$  contrôle la taille de ce voisinage c'est-à-dire le nombre d'observations prises pour effectuer la moyenne locale.

L'estimateur de Nadaraya-Watson est un cas particulier d'une classe plus générale : les estimateurs par polynômes locaux, que nous allons introduire et étudier dans la section suivante. ■

## 5.2 L'estimateur de Gasser et Müller

Le dénominateur aléatoire de l'estimateur de Nadaraya-Watson est un inconvénient, notamment pour l'étude des dérivés de cet estimateur et lors du calcul de ses propriétés asymptotiques.

Supposons que les  $X_i$  sont ordonnés, Gasser et Muller (1979) ont proposé l'estimateur suivant :

$$\hat{m}_{GM} = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(u-x) du Y_i$$

où  $s_i = \frac{X_i + X_{i+1}}{2}$ ,  $X_0 = -\infty$  et  $X_{n+1} = +\infty$ . L'estimateur de Gasser et Müller est une modification d'un estimateur proposé par Priestley et Chao (1972)[35]. Müller (1988)[33] donne plus de détail sur l'estimateur de Gasser et Müller, les propriétés asymptotiques de cet estimateur sont développées par Mack and Muller (1989)[39] et Chu and Marron (1991)[4].

## 5.3 L'ajustement polynômial local et l'estimation des dérivées

L'idée de régression par les polynômes locaux existe depuis longtemps. Elle a été systématiquement étudiée par Stone (1977, 1980, 1982)[38], [39], [40] et Cleveland (1979)[5]. Des travaux récents sur l'ajustement par polynômes locaux comprend Fan (1992, 1993a)[7][8], Fan et Gijbels (1992)[9] et Ruppert et Wand (1994)[37], Fan et Gijbels (1996)[11]. Ces articles donnent une idée détaillée des avantages de ajustement polynomial local.

Nous considérons l'ajustement polynomial locale pour estimer une fonction de régression non paramétrique et de ses dérivés.

**Remarque 5.3.1** Si  $K$  est un noyau positif, alors  $\hat{m}_{NW}$  vérifie

$$\forall x \in \mathbb{R}, \hat{m}_{NW}(x) = \arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n (Y_i - \theta)^2 K \left( \frac{X_i - x}{h} \right).$$

En effet, si on cherche les points singuliers correspondants, on obtient

$$-2 \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) (Y_i - \theta) = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

et il s'agit bien d'un minimum si  $K \geq 0$ .

c'est à dire que l'estimateur  $\hat{m}_{NW}$  est obtenu par une approximation des moindres carrés localement constante.

### 5.3.1 L'ajustement polynômial local

L'estimation localement polynômiale est basée sur l'ajustement local d'un polynôme de degré  $p$  aux données  $(X_i, Y_i)_{i=1, \dots, n}$ . Pour  $p \in \mathbb{N}$ , nous cherchons à ajuster le polynôme

$$\beta_0 + \beta_1(\cdot - x) + \dots + \beta_p(\cdot - x)^p$$

aux données  $(X_i, Y_i)_{i=1, \dots, n}$  par la méthode des moindres carrés. Pour cela, on doit supposer l'existence de la  $(p+1)$ ème dérivée de la fonction de régression  $m(\cdot)$  au point  $x$  pour pouvoir faire un développement de Taylor, alors pour  $u$  appartenant à un voisinage de  $x$  on a

$$\begin{aligned} m(u) &\approx m(x) + m'(x)(u - x) + \frac{m''(x)}{2}(u - x)^2 + \dots + \frac{m^{(p)}(x)}{p!}(u - x)^p \quad (2.1) \\ &\approx \sum_{j=0}^p \frac{m^{(j)}(x)}{j!} (u - x)^j \\ &= \sum_{j=0}^p \beta_j (u - x)^j. \end{aligned}$$

Pour ajuster localement le polynôme (2.1) aux données  $(X_i, Y_i)_{i=1, \dots, n}$  par la méthode des moindres carrés pondérés, on doit minimiser par rapport au vecteur  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$  la quantité

$$\min_{\beta} \sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x)^j \right\}^2 K_h(X_i - x) \quad (2.2)$$

Où  $h$  est la fenêtre (bandwidth), elle contrôle la taille du voisinage locale.

et  $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ , on suppose que le noyau  $K$  est un noyau symétrique à support compact.

### 5.3.2 L'estimation des dérivées

Soit  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^t \in \mathbb{R}^{p+1}$  le vecteur qui minimise le problème 2.2. D'après l'égalité 2.1, la dérivée  $\nu^{\text{ème}}$ ,  $m^{(\nu)}(x)$ , peut être estimée par  $\hat{\beta}_\nu \cdot \nu!$  pour  $\nu = 0, \dots, p$ .

**Définition 5.3.1** *La statistique*

$$\hat{m}_{LP}^{(\nu)}(x, p) = \hat{\beta}_\nu \cdot \nu! \quad 0 \leq \nu \leq p \quad (2.3)$$

est l'estimateur localement polynômial d'ordre  $p$  de la dérivée  $\nu^{\text{ème}}$  de la fonction de régression  $m(x)$ , c-à-d  $m^{(\nu)}(x)$ .

**Remarque 5.3.2** *Lorsque  $\nu = p = 0$ , on retrouve l'estimateur de Nadaraya-Watson, i.e.  $\hat{m}_{LP}^0(x) = \hat{m}_{NW}(x)$*

## 6 La régression par les polynômes locaux

Soient  $(X_1, Y_1), \dots, (X_n, Y_n)$ , des paires de variables aléatoires indépendantes identiquement distribuées.

Le but est d'estimer  $m(x) = \mathbb{E}(Y|X = x)$  et ses dérivées  $m'(x), m''(x), \dots, m^{(p)}(x)$

D'après la notation de Fan et Gijbels (1996), On considère le modèle :

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i, \quad (2.4)$$

Où  $\varepsilon_i \rightsquigarrow N(0, 1)$  et  $X$  indépendant du  $\varepsilon$ .

On note par  $\sigma^2(x)$  la variance de  $Y$  sachant  $X = x$ .

On suppose que la  $(p+1)^{\text{ème}}$  dérivée de  $m(\cdot)$  au point  $x$  existe. Alors, on doit ajuster localement la fonction de régression inconnue  $m(x)$  par un polynôme de degré  $p$ .

Soit le développement de Taylor

$$m(u) \approx m(x) + m'(x)(u-x) + \frac{m''(x)}{2}(u-x)^2 + \dots + \frac{m^{(p)}(x)}{p!}(u-x)^p \quad (2.5)$$

où  $u$  appartenant à un voisinage de  $x$ .

Pour ajuster localement le polynôme (2.5) aux données  $(X_i, Y_i)_{i=1, \dots, n}$  par la méthode des moindres carrés pondérés par  $K_h(\cdot)$ , on doit minimiser par rapport au vecteur  $\beta = (\beta_0, \beta_1, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$  la quantité

$$\min_{\beta} \sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p \beta_j (X_i - x)^j \right\}^2 K_h(X_i - x) \quad (2.6)$$

Où  $h$  :la fenêtre (bandwidth), elle contrôle la taille du voisinage locale.

et  $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ , on suppose que le noyau  $K$  est un noyau symétrique et a support compact.

Soit  $\hat{\beta}_j, j = 1, \dots, p$  la solution de moindres carrés du problème (2.6).

D'après (2.5),

$$\hat{m}^{(\nu)}(x) = \hat{\beta}_\nu \cdot \nu!$$

est l'estimateur de  $m^{(\nu)}(x)$  pour  $0 \leq \nu \leq p$ .

Il est plus pratique de travailler avec la notation matricielle. Soit  $X$  la matrice définie par

$$X = \begin{pmatrix} 1 & (X_1 - x) & \cdots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x) & \cdots & (X_n - x)^p \end{pmatrix}_{n \times (p+1)}$$

Nous posons

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}_{n \times 1}, \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}_{(p+1) \times 1}$$

On désigne par  $W$  la matrice diagonale  $n \times n$  des poids :

$$W = \begin{pmatrix} K_h(X_1 - x) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & K_h(X_n - x) \end{pmatrix}_{n \times n}$$

Alors le problème de moindres carrés pondérés (2.6) peut être écrit comme suit :

$$\min_{\beta \in \mathbb{R}^{p+1}} (Y - X\beta)^t W (Y - X\beta) \quad (2.7)$$

où l'exposant  $t$  dénote la transposition, pour un vecteur ou une matrice. On suppose dorénavant l'inversibilité de la matrice carrée  $X^t W X$ .

la solution matricielle est donnée par :

$$\begin{aligned} \hat{\beta} &= (X^t W X)^{-1} X^t W Y \\ &= S_n^{-1} T_n \end{aligned} \quad (2.8)$$

$$\mathbf{S}_n = \begin{pmatrix} S_{n,0} & \cdots & S_{n,p} \\ \vdots & \ddots & \vdots \\ S_{n,p} & \cdots & S_{n,2p} \end{pmatrix}_{(p+1) \times (p+1)}, \quad \mathbf{T}_n = \begin{pmatrix} T_{n,0} \\ \vdots \\ T_{n,p} \end{pmatrix}_{(p+1) \times 1}$$

$$S_{n,j} = \sum_{i=1}^n K_h(X_i - x)(X_i - x)^j, \quad j = 0, \dots, 2p.$$

$$T_{n,j} = \sum_{i=1}^n K_h(X_i - x)(X_i - x)^j Y_i, \quad j = 0, \dots, p.$$

Cette dernière égalité (2.8) permet de formuler aisément le biais et la variance conditionnels de l'estimateur  $\widehat{\beta}$ . Nous rappelons la définition du vecteur  $\beta$ ,

$$\beta = \left( m(x), \frac{m^{(1)}(x)}{1!}, \dots, \frac{m^{(p)}(x)}{p!} \right)^t$$

## 6.1 Le biais et la variance

Dans le cas unidimensionnel ( $X_i \in \mathbb{R}$ ) et multidimensionnel ( $X_i \in \mathbb{R}^d, d \geq 1$ ) de l'ajustement polynômial local, Ruppert et Wand (1994)[37] ont établi les expressions de biais et de variance conditionnels de l'estimateur de la régression polynômiale locale, et ils ont montré que cet estimateur s'adapte automatiquement aux bords.

Lorsque vous traitez le problème de sélection de la fenêtre de lissage, une question clé est d'avoir un bon aperçu de biais et de la variance des estimateurs. Dans cette section, nous allons donner l'expression exacte et asymptotique pour le biais et la variance.

Le biais et la variance conditionnels de l'estimateur  $\widehat{\beta}$  sont obtenus à partir de sa définition en (2.8).

Nous définissons,  $m = (m(X_1), \dots, m(X_n))^t$ ,  $r$  est le vecteur des résidus et

$$\begin{aligned} m &= \mathbb{E}(Y|X) \\ &= X\beta + \mathbb{E}(\varepsilon|X) \\ &= X\beta + r \end{aligned}$$

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}/X) &= (X^t W X)^{-1} X^t W \mathbb{E}(Y|X) \\
&= (X^t W X)^{-1} X^t W m \\
&= (X^t W X)^{-1} X^t W (X\beta + r) \\
&= (X^t W X)^{-1} (X^t W X) \beta + (X^t W X)^{-1} X^t W r \\
&= \beta + (X^t W X)^{-1} X^t W r.
\end{aligned}$$

Alors,

$$\text{Biais}(\widehat{\beta}/X) = (X^t W X)^{-1} X^t W r.$$

Soit  $\Sigma = \text{diag}[K_h^2(X_i - x)\sigma^2(X_i)]$  ou  $\sigma^2(x) = \text{Var}(Y|X = x)$

On rappelle que  $\text{Var}(AX) = A^t(\text{Var}(X))A$

Puisque  $X^t W X$  est symétrique alors,

$$\begin{aligned}
\text{Var}(\widehat{\beta}|X) &= (X^t W X)^{-1} (X^t \text{var}(WY|X) X) (X^t W X)^{-1} \\
&= (X^t W X)^{-1} (X^t \Sigma X) (X^t W X)^{-1}
\end{aligned} \tag{2.9}$$

On note par

$$\begin{aligned}
S_n &= X^t W X \\
S_n^* &= X^t \Sigma X
\end{aligned}$$

et

$$S_{n,j}^* = \sum_{i=1}^n (X_i - x)^j K_h^2(X_i - x) \sigma^2(X_i)$$

La variance conditionnelle dans (2.9) peut être ré-exprimée par  $S_n^{-1} S_n^* S_n^{-1}$ .

Ces expressions exactes du biais et de variance ne sont pas directement utilisables, car ils dépendent des quantités inconnues : le résidu  $r$  et la matrice diagonale  $\Sigma$ . Donc, il faut approximer le biais et la variance. Les expressions asymptotiques du biais et de la variance de l'estimateur  $\widehat{m}_\nu(x) = \nu! \widehat{\beta}_\nu$  sont données dans le prochain théorème.

Avant d'énoncer le théorème de Ruppert et Wand (1994), on rappelle quelques notations utiles. Les moments de  $K$  et de  $K^2$  sont désignés respectivement par

$$\mu_j = \int u^j K(u) du \quad \text{et} \quad \nu_j = \int u^j K^2(u) du$$

Certaines matrices et vecteurs des moments apparaissent dans les expressions asymptotiques. soient

$$\begin{aligned}
S &= (\mu_{j+l})_{0 \leq j, l \leq p} \\
\tilde{S} &= (\mu_{j+l+1})_{0 \leq j, l \leq p} \\
S^* &= (\nu_{j+l})_{0 \leq j, l \leq p} \\
c_p &= (\mu_{p+1}, \dots, \mu_{2p+1})^t \\
\tilde{c}_p &= (\mu_{p+2}, \dots, \mu_{2p+2})^t.
\end{aligned}$$

$$\begin{aligned}
S &= \begin{pmatrix} \mu_0 & \cdots & \mu_p \\ \vdots & \ddots & \vdots \\ \mu_p & \cdots & \mu_{2p} \end{pmatrix}_{(p+1) \times (p+1)}, \quad \tilde{S} = \begin{pmatrix} \mu_1 & \cdots & \mu_{p+1} \\ \vdots & \ddots & \vdots \\ \mu_{p+1} & \cdots & \mu_{2p+1} \end{pmatrix}_{(p+1) \times (p+1)}, \\
S^* &= \begin{pmatrix} \nu_0 & \cdots & \nu_p \\ \vdots & \ddots & \vdots \\ \nu_p & \cdots & \nu_{2p} \end{pmatrix}_{(p+1) \times (p+1)}, \\
c_p &= \begin{pmatrix} \mu_{p+1} \\ \vdots \\ \mu_{2p+1} \end{pmatrix}_{(p+1) \times 1}, \quad \tilde{c}_p = \begin{pmatrix} \mu_{p+2} \\ \vdots \\ \mu_{2p+2} \end{pmatrix}_{(p+1) \times 1}.
\end{aligned}$$

De plus, on considère le vecteur unité

$$e_{\nu+1} = (0, \dots, 0, 1, 0, \dots, 0)^t$$

avec 1 dans la  $(\nu + 1)$ <sup>ème</sup> position.

On rappelle que  $(X_1, Y_1), \dots, (X_n, Y_n)$  est un échantillon indépendant identiquement distribué d'une population  $(X, Y)$  et que  $o(1)$  une quantité qui tend vers zéro en probabilité.

**Théorème 6.1.1** (*Ruppert et Wand (1994)*).

On suppose que  $f(x) > 0$  et  $f(\cdot)$ ,  $m^{(p+1)}(\cdot)$  et  $\sigma^2(\cdot)$  sont continues au voisinage de  $x$ . De plus on suppose que la fenêtre  $h$  vérifie  $h \rightarrow 0$  et  $nh \rightarrow \infty$ . Alors, La variance conditionnelle asymptotique de  $\hat{m}_\nu(x)$  est donnée par

$$\text{Var}(\hat{m}_\nu(x)|X) = e_{\nu+1}^t S^{-1} S^* S^{-1} e_{\nu+1} \frac{\nu!^2 \sigma^2(x)}{f(x) n h^{1+2\nu}} + o\left(\frac{1}{n h^{1+2\nu}}\right) \quad (2.10)$$

**Lorsque  $p - \nu$  est impair**, le biais conditionnel asymptotique est donnée par

$$\text{Biais}(\hat{m}_\nu(x)|X) = e_{\nu+1}^t S^{-1} c_p \frac{\nu!}{(p+1)!} m^{(p+1)}(x) h^{p+1-\nu} o(h^{p+1-\nu}) \quad (2.11)$$

**Lorsque  $p - \nu$  est pair**, en supposant  $f'(\cdot)$  et  $m^{(p+2)}(\cdot)$  continues dans un voisinage du point  $x$  ainsi que  $nh^3 \rightarrow \infty$ , le biais conditionnel asymptotique est donné par,

$$\begin{aligned} \text{Biais}(\widehat{m}_\nu(x)|X) &= e_{\nu+1}^t S^{-1} \tilde{c}_p \frac{\nu!}{(p+2)!} \left\{ m^{(p+2)}(x) + (p+2)m^{(p+1)}(x) \frac{f'(x)}{f(x)} \right\} \\ &\quad \times h^{p+2-\nu} o(h^{p+2-\nu}) \end{aligned} \quad (2.12)$$

**Preuve.** Voir Appendice

D'après le théorème ci-dessus, on remarque qu'il existe une différence entre le cas  $p - \nu$  pair et le cas  $p - \nu$  impair. Lorsque  $p - \nu$  pair, le terme de biais principal en  $O(h^{p+1})$  s'annule via la symétrie de noyau  $K$ . Par contre, lorsque  $p - \nu$  impair, le terme de biais asymptotique a une expression simple où ne figure pas de termes de dérivées tels  $f'_X(x)$ .

**Preuve.** Donc, ici, nous avons quelques raisons théoriques de la distinction entre les cas  $p - \nu$  impair et  $p - \nu$  pair. Il va apparaître clairement que le polynôme ajusté correspond à  $p - \nu$  impair plus performant que celles avec  $p - \nu$  pair. Cette question est développée en détail dans la suite. Il nous donne une explication intuitive de cette observation importante, et de justifier le cas  $p - \nu$  impair devrait être appelé le naturel. Par conséquent, dans le reste de la présente section, nous prenons  $p - \nu$  impair. ■

## 6.2 Méthode des noyaux équivalant

La meilleure représentation des estimateurs localement polynomiaux est obtenue par la méthode des noyaux équivalents, Nous montrons comment la méthode d'ajustement polynomial local affecte une pondération à chaque point de données.

On rappelle la notation

$$S_{n,j} = \sum_{i=1}^n K_h(X_i - x)(X_i - x)^j \quad (2.13)$$

et  $S_n = X^t W X$ , la matrice carrée de dimension  $p + 1$  définie également par

$$S_n = (S_{n,j+l})_{0 \leq j, l \leq p}.$$

D'après (2.8),

$$\begin{aligned}
\hat{\beta}_\nu &= e_{\nu+1}^t \hat{\beta} \\
&= e_{\nu+1}^t (X^t W X)^{-1} X^t W Y \\
&= e_{\nu+1}^t S_n^{-1} X^t W Y \\
&= \sum_{i=1}^n W_\nu^n \left( \frac{X_i - x}{h} \right) Y_i, \quad 0 \leq \nu \leq p.
\end{aligned} \tag{2.14}$$

On remarque que

$$X^t W = \begin{pmatrix} K_h(X_1 - x) & \cdots & K_h(X_n - x) \\ \vdots & \ddots & \vdots \\ (X_1 - x)^p K_h(X_1 - x) & \cdots & (X_n - x)^p K_h(X_n - x) \end{pmatrix}_{(p+1) \times n}$$

Il s'ensuit

$$X^t W Y = \begin{pmatrix} \sum_{i=1}^n Y_i K_h(X_i - x) \\ \vdots \\ \sum_{i=1}^n Y_i (X_i - x)^p K_h(X_i - x) \end{pmatrix}_{(p+1) \times 1}$$

En effet,

$$\begin{aligned}
W_\nu^n \left( \frac{X_i - x}{h} \right) &= e_{\nu+1}^t S_n^{-1} \begin{pmatrix} K_h(X_i - x) \\ \vdots \\ (X_i - x)^p K_h(X_i - x) \end{pmatrix} \\
&= e_{\nu+1}^t S_n^{-1} \begin{pmatrix} 1 \\ (X_i - x) \\ \vdots \\ (X_i - x)^p \end{pmatrix} K_h(X_i - x) \\
&= e_{\nu+1}^t S_n^{-1} \begin{pmatrix} 1 \\ h \left( \frac{X_i - x}{h} \right) \\ \vdots \\ h^p \left( \frac{X_i - x}{h} \right)^p \end{pmatrix} \frac{1}{h} K \left( \frac{X_i - x}{h} \right)
\end{aligned}$$

on pose  $t = \frac{X_i - x}{h}$ , D'où

$$W_\nu^n(t) = e_{\nu+1}^t S_n^{-1} [1, th, \dots, (th)^p]^t \frac{1}{h} K(t), \quad 0 \leq \nu \leq p$$

L'expression ci-dessus montre que l'estimateur  $\widehat{\beta}_\nu$  ressemble beaucoup à un estimateur à noyau classique, sauf que le noyau  $W_\nu^n$  dépend des points  $X_i$  et de leur localisation. Ceci explique intuitivement pourquoi l'estimation localement polynomiale s'adapte aux différents dispositifs expérimentaux ainsi qu'à l'estimation aux bornes du support de la densité. Nous énonçons à présent une propriété fondamentale des estimateurs LP(p).

On peut montrer que la fonction de poids  $W_\nu^n$  satisfait la condition de moment discret suivante :

$$\sum_{i=1}^n (X_i - x)^q W_\nu^n \left( \frac{X_i - x}{h} \right) = \delta_{\nu,q} \quad 0 \leq \nu, q \leq p. \quad (2.15)$$

Ci-dessus  $\delta_{\nu,q}$  dénote le symbole de Kronecker.

En effet,

$$\begin{aligned} \sum_{i=1}^n (X_i - x)^q W_\nu^n \left( \frac{X_i - x}{h} \right) &= \sum_{i=1}^n (X_i - x)^q \mathbf{e}_{\nu+1}^t S_n^{-1} X^t W \\ &= \mathbf{e}_{\nu+1}^t S_n^{-1} \sum_{i=1}^n (X_i - x)^q \begin{pmatrix} 1 \\ (X_i - x) \\ \vdots \\ (X_i - x)^p \end{pmatrix} K_h(X_i - x_0) \\ &= \mathbf{e}_{\nu+1}^t \mathbf{S}_n^{-1} \mathbf{S}_n \mathbf{e}_{q+1}^t \\ &= \mathbf{e}_{\nu+1}^t \times \mathbf{e}_{q+1}^t = \delta_{\nu,q}. \end{aligned}$$

Comme conséquence de cette relation, le biais de l'estimateur  $\widehat{\beta}_\nu$  est nul lorsque l'échantillon est fini et la fonction  $m^{(\nu)}(\cdot)$  à estimer est un polynôme de degré inférieur ou égal à  $p$ . Cette propriété met en évidence un des avantages pratiques de l'estimation par la méthode des polynômes locaux pour la réduction du biais, en comparaison avec l'utilisation de noyaux d'ordres élevés. En effet, le biais est nul à  $n$  fixé et non asymptotiquement.

En continuant l'investigation des propriétés de la fonction de poids  $W_\nu^n$ . Lorsque  $h \rightarrow 0$  et  $nh \rightarrow \infty$ , on note par (voir l'appendice)

$$S_{n,j} = nh^j \mu_j f(x) \{1 + o(1)\} \quad (2.16)$$

Puisque

$$HSH = (h^j \mu_j)_{0 \leq j \leq 2p}$$

Alors

$$S_n = n f(x) HSH \{1 + o_p(1)\} \quad (2.17)$$

où

$$H = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & h & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h^p \end{pmatrix}_{(p+1) \times (p+1)} \quad \text{et } S = \begin{pmatrix} \mu_0 & \cdots & \mu_p \\ \vdots & \ddots & \vdots \\ \mu_p & \cdots & \mu_{2p} \end{pmatrix}_{(p+1) \times (p+1)}$$

Par conséquent en remplaçant 2.17 dans la définition de  $W_\nu^n$ , Nous constatons que

$$W_\nu^n(t) = \frac{1}{nh^{\nu+1}f(x)} e^t_{\nu+1} S^{-1} [1, t, \dots, t^p]^t K(t) [1 + o(1)] \quad (2.18)$$

Il en découle,

$$\hat{\beta}_\nu = \frac{1}{nh^{\nu+1}f(x)} \sum_{i=1}^n K_\nu^*\left(\frac{X_i - x}{h}\right) Y_i [1 + o(1)] \quad (2.19)$$

où

$$\begin{aligned} K_\nu^*(t) &= e^t_{\nu+1} S^{-1} [1, t, \dots, t^p]^t K(t) \\ &= \left( \sum_{l=0}^p S^{\nu l t^l} \right) K(t) \quad \text{avec } S^{-1} = (S^{jl})_{0 \leq j, l \leq p}. \end{aligned} \quad (2.20)$$

Le noyau  $K_\nu^*(t)$  est appelé noyau équivalent (equivalent kernel) et il est très utile pour exprimer les propriétés asymptotiques de l'estimateur LP( $p$ ). Le noyau (2.20) vérifie les conditions de moments suivantes :

$$\int t^q K_\nu^*(t) dt = \delta_{\nu, q}, \quad 0 \leq \nu, q \leq p. \quad (2.21)$$

qui est une version asymptotique des conditions de moments discrets présentés dans 2.15.

Le noyau équivalent  $K_\nu^*(t)$  est simplement un noyau d'ordre  $(\nu, p + 1)$ . Voir la définition de ces noyaux en Gasser, Müller et mammitzsch (1985)[15] (voir appendice). Les noyaux équivalents ont déjà été utilisés pour analyser l'ajustement polynomial par Lejeune (1985)[29] et Muller (1987)[33] d'une manière légèrement différente.

Pour souligner la dépendance de  $p$ , on note par  $K_{\nu, p}^*$  le noyau équivalent donné par (2.20) Le tableau suivant donne quelque formes de la fonction de noyau équivalent. Comme illustration, Fan et Gijbels (1996)[11] ont tracé le noyau Epanechnikov  $K(z) = \frac{3}{4}(1 - z^2)$ , comme étant un meilleur noyau équivalent  $K_{\nu, p}^*$  pour quelque valeurs de  $p$  et  $\nu$ .

$\nu$	$p$	Les fonctions du noyau équivalent $K_{\nu,p}^*(t)$
0	1	$k(t)$
0	3	$(\mu_4 - \mu_2^2)^{-1} (\mu_4 - \mu_2 t^2) k(t)$
1	2	$\mu_2^{-1} t k(t)$
2	3	$(\mu_4 - \mu_2^2)^{-1} (t^2 - \mu_2) k(t)$

Tableau 1. Les fonctions du noyau équivalent  $K_{\nu,p}^*(t)$ .

La variance et le biais conditionnels de l'estimateur  $\widehat{m}_\nu(x)$ , spécifiés en 2.10 et (2.11) respectivement, peuvent être exprimés en fonction du noyau équivalent  $K_\nu^*(\cdot)$ , nous conduisant aux expressions asymptotiques suivantes :

$$Biais(\widehat{m}_\nu(x)|X) = \left\{ \int t^{p+1} K_\nu^*(t) dt \right\} \frac{\nu!}{(p+1)!} m^{(p+1)}(x) \times h^{p+1-\nu} + o(h^{p+1-\nu}) \quad (2.22)$$

$$Var(\widehat{m}_\nu(x)|X) = \int K_\nu^{*2}(t) dt \frac{(\nu!)^2 \sigma^2(x)}{f(x) n h^{1+2\nu}} + o\left(\frac{1}{n h^{1+2\nu}}\right) \quad (2.23)$$

Ces deux expressions peuvent être obtenues à partir de (2.19) et (2.21).

### 6.3 Le choix optimal de la fenêtre de lissage

Le choix de la fenêtre de lissage est assez crucial, une fenêtre locale (constante)  $h(x)$  varie avec la localisation du point  $x$ , tandis que la fenêtre globale  $h(X_i)$  varie avec les points de données.

La fenêtre de lissage est introduite pour tenir compte des différents degrés de lissage, ce qui entraîne une réduction du biais au niveau des régions de pic et de la variance dans les régions trop bas.

Le choix théorique de la fenêtre globale est détaillé par Fan et Gijbels (1992)[9]. Ici, on concentre sur le choix de la fenêtre locale.

La fenêtre optimale locale pour estimer  $m^{(\nu)}(x)$  est obtenue par la minimisation de l'erreur quadratique moyenne conditionnelle  $MSE$  donnée par

$$[Biais(\widehat{m}_\nu(x)|X)]^2 + Var(\widehat{m}_\nu(x)|X) \quad (2.24)$$

La fenêtre optimale au sens du critère local de minimisation de l'AMSE est obtenue à partir de (2.22) et (2.23) :

$$h_{opt}(x) = C_{\nu,p}(K) \left[ \frac{\sigma^2(x)}{[m^{(p+1)}(x)]^2 f(x)} \right]^{1/(2p+3)} n^{-1/(2p+3)} \quad (2.25)$$

avec

$$C_{\nu,p}(K) = \left[ \frac{(p+1)!^2(2\nu+1) \int K_\nu^{*2}(t) dt}{2(p+1-\nu) \left[ \int t^{p+1} K_\nu^*(t) dt \right]^2} \right]^{1/(2p+3)} \quad (2.26)$$

Minimisation de l'erreur quadratique moyenne intégrée pondérée (*MISE*) de  $\widehat{m}_\nu(x)$  sachant que  $X$

$$\int \{ [\text{Biais}(\widehat{m}_\nu(x)|X)]^2 + \text{Var}(\widehat{m}_\nu(x)|X) \} w(x) dx \quad (2.27)$$

avec  $w(x)$  la fonction de poids, conduit à une fenêtre théorique constante optimale. En utilisant à nouveau les expressions asymptotiques en 2.19 et 2.20, nous trouvons une fenêtre constante asymptotiquement optimale donnée par

$$h_{opt} = C_{\nu,p}(K) \left[ \frac{\int \sigma^2(x)w(x)/f(x)dx}{\int [m^{(p+1)}(x)]^2 w(x)dx} \right]^{1/(2p+3)} n^{-1/(2p+3)} \quad (2.28)$$

Les intégrales sont finies et le dénominateur ne s'annule pas.

Ces fenêtres optimales dépendent des quantités inconnues telles que la densité  $f(\cdot)$ , la variance conditionnelle  $\sigma^2(\cdot)$ , et la fonction dérivée  $m^{(p+1)}(\cdot)$ , et donc des travaux supplémentaires sont nécessaires pour atteindre la fenêtre pratique.

## 6.4 Le choix de $p$

L'ajustement polynomial d'ordre grand, réduit le biais, mais augmente la variance, à cause de l'introduction de plusieurs paramètres locaux. La variance asymptotique de  $\widehat{m}_\nu(x)$  augmente seulement quand  $p$  va de l'ordre impair en ordre paire. Il n'y a aucune perte lors du passage de  $p = 2$  à  $p = 3$ , mais allant de  $p = 1$  à  $p = 2$  va augmenter la variance asymptotique. Cela suggère fortement de travaillé avec les polynômes d'ordre impairs. Le Tableau 2 montre comment la variance augmente avec l'ordre du polynôme ajusté  $p$ . ce résultat a été obtenu par Fan et Gijbels (1995b)[10].

$p$	Gaussien	Uniforme	Epanechnikov
1	1	1	1
2	1.6876	2.2500	2.0833
3	1.6876	2.2500	2.0833
4	2.2152	3.5156	3.1550
5	2.2152	3.5156	3.1550
6	2.6762	4.7852	4.2222
7	2.6762	4.7852	4.2222
8	3.1224	6.0562	5.2872
9	3.1224	6.0562	5.2872
10	3.5704	7.3281	6.3509

Tableau 2. L'augmentation de la variance avec l'ordre du polynôme ajusté  $p$ .

On constate que l'ordre impair est préférable. En effet, un ajustement polynomiale d'ordre impair  $2q + 1$  introduit des paramètres supplémentaire par rapport à l'ajustement polynomiale d'ordre pair  $2q$ , mais, il n'y'a pas d'augmentation de la variance.

# Chapitre

# 3

## Données Incomplètes

### 1 Introduction

Le problème de l'analyse des données de survie survient dans un certain nombre de domaines appliqués, tels que la médecine, la biologie, la santé publique, l'épidémiologie, l'astronomie, l'économie et la démographie. L'analyse des données de survie consiste à étudier le délai de la survenue d'un événement précis, l'évènement étudié est le passage entre deux états  $A$  et  $B$ . L'évènement terminal peut être considéré tels que la survenue d'une maladie en épidémiologie, la survenue d'une panne dans les applications industrielles, l'acceptation d'une offre d'emploi pour une personne au chômage en économie, ou le mariage en démographie.

Les données de survie se caractérisent par l'existence d'observations incomplètes. En effet, les données sont souvent recueillies partiellement, à cause des processus de censure et de troncature. Les données censurées ou tronquées proviennent du fait qu'on n'a pas accès à toute l'information.

Si la durée de vie d'un individu n'est observable que dans une période donnée, on parle de censure. Il s'agit de troncature, si l'individu doit survivre longtemps pour être observé, ou lorsque l'individu est retenu par l'étude uniquement s'il a connu l'évènement en question à une date  $t$  donnée.

## 2 Censure et Troncature

### 2.1 Censure

**Définition 2.1.1** Soit  $Y$  la variable d'intérêt, si au lieu d'observer  $Y$  on observe une constante  $C$  et si on sait que  $Y > C$ , on dira alors qu'il y a censure à droite.

Il arrive souvent que la constante de censure soit elle même une variable aléatoire, dans ce cas, on parle de censure aléatoire à droite et on a :

**Définition 2.1.2** Etant donnée une variable aléatoire  $Y$ , on dit qu'il y a censure aléatoire à droite s'il existe une variable aléatoire  $C$ , telle qu'au lieu d'observer  $Y$  on observe :

$$\{(Z, \delta)/Z = \min(Y, C), \delta = \mathbf{1}_{\{Y \leq C\}}\}$$

### 2.2 Troncature

**Définition 2.2.1** Soit  $Y$  la variable d'intérêt et  $T$  une autre variable aléatoire, si  $Y$  et  $T$  sont observables uniquement si  $Y > T$ , et rien sinon, on dira que la variable  $Y$  est aléatoirement tronquée à gauche par la variable de troncature  $T$ . Dans le cas contraire, on parle de troncature aléatoire à droite.

#### 2.2.1 Un bref historique

Le modèle de troncature est apparu tous d'abord en astronomie. Les luminosités absolues et apparentes d'un objet astral sont observées à une distance fixe depuis la terre.

Dans certains modèles, le décalage vers le rouge  $z$  et la magnitude absolue  $M$  d'objets astronomiques sont supposés étant des variables aléatoires indépendantes qui sont liées à la magnitude apparente  $m$  par l'équation

$$m = f(z) + M$$

Où  $f$  est une fonction connue.

On ne peut observer que les objets qui sont suffisamment brillants, c'est-à-dire  $m \leq m^*$ . On note  $Y = \exp(-f(z))$  et  $T = \exp(M - m^*)$  on observe que si  $Y \geq T$ .

Dans d'autres applications, le  $Y_e$  peut être la taille des objets cachés pour lesquels on recherche et  $Z_e = T_e/Y_e$  pourrait être le moment où l'on trouverait l'objet  $e$ . Alors, la probabilité conditionnelle de trouver objet  $e$  sachant  $Y_e$  est  $G(Y_e)$ , une fonction inconnue mais croissante de  $Y_e$ .

De nombreux travaux ont été effectués sur l'analyse de données tronquées (Voir Klein et Moeschberger, (1997)[25]).

### 2.2.2 Exemple de troncature à droite. Lagakos et al. (1988)(voir Klein et Moeschberger (1997)[25]).

Présentant les données sur l'infection et le temps d'induction pour 258 adultes et 37 enfants qui ont été infectés par le virus du sida et qui ont développé le sida avant le 30 Juin, 1986. Les données se composent du temps au cours des années, mesuré à partir du 1er Avril 1978, quand les adultes ont été infectés par le virus lors d'une transfusion de sang contaminé, et le temps d'attente pour le développement du sida, mesuré à partir de la date d'infection. Pour la population pédiatrique, les enfants ont été infectés à la naissance, et le temps de l'infection est le nombre d'années à partir du 1er Avril 1978 à la naissance.

Dans ce plan d'échantillonnage, seules les personnes qui ont développé le SIDA avant la fin de la période d'étude sont incluses dans l'étude. Les personnes infectées qui n'ont pas encore développé le sida ne sont pas inclus dans l'échantillon. Ce type de données est appelé données tronquées à droite.

### 2.2.3 Exemple de troncature à gauche. (voir Klein et Moeschberger (1997)[25])

Une étude de survie des résidents du centre de retraite en Californie est décrite. On enregistre l'âge au décès et l'âge d'entrée au centre. Un individu doit survivre à un âge suffisant pour entrer dans le centre, tous les individus qui sont morts ne seront pas entré dans le centre et sont donc hors de la connaissance de l'enquêteur, ces personnes n'ont aucune chance d'être dans l'étude et sont considérées comme tronquées à gauche.

## 3 Définitions

Quelques définitions sont couramment utilisées dans les études de survie.

**Durée de survie :** La durée de survie noté par la variable aléatoire  $X$ , (souvent positive) est le temps qui s'écoule depuis une date d'origine (début du traitement,...) jusqu'à la survenue d'un événement d'intérêt (décès, guérison,...).

**Date d'origine :** Elle correspond à l'origine de la durée étudiée. Elle peut être la date de naissance, le début d'une exposition à un facteur de risque, la date d'une opération chirurgicale, la date de début d'une maladie ou la date d'entrée dans l'étude. Chaque individu peut donc avoir une date d'origine différente (pas important car c'est la durée qui nous intéresse).

**Date de point** : C'est la date au-delà de laquelle on arrêtera l'étude et on ne tiendra plus compte des informations sur les sujets.

**Date des dernières nouvelles** : C'est la date la plus récente où des informations sur un sujet ont été recueillies.

## 4 Distributions de la durée de survie

Supposons que la durée de survie  $X$  soit une variable positive ou nulle, et absolument continue, alors sa loi de probabilité peut être définie par l'une des cinq fonctions équivalentes suivantes (chacune des fonctions ci-dessous peut être obtenue à partir de l'une des autres fonctions) :

### 4.1 La fonction de survie $S$

Pour  $t$  fixé, la fonction de survie est la probabilité de survivre jusqu'à l'instant  $t$ , c'est-à-dire

$$S(t) = P\{X \geq t\}, t \geq 0 \quad (3.1)$$

### 4.2 La fonction de répartition $F$

Pour  $t$  fixé, la fonction de répartition est la probabilité de mourir avant l'instant  $t$ .

$$F(t) = P\{X < t\} = 1 - S(t) \quad (3.2)$$

### 4.3 La densité de probabilité $f$

C'est une fonction  $f(t) \geq 0$  telle que pour tout  $t \geq 0$

$$F(t) = \int_0^t f(s) ds \quad (3.3)$$

Si la fonction de répartition a une dérivée au point  $t$  alors

$$f(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq X < t + dt)}{dt} = F'(t) = -S'(t) \quad (3.4)$$

Pour  $t$  fixé, la densité de probabilité caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant  $t$ .

#### 4.4 Le taux de hasard ou risque instantané $h$

Le risque instantané est aussi très souvent appelé "le taux de hasard", c'est défini comme

$$h(t) = \begin{cases} 0 & \text{si } S(t) = 0 \\ \frac{f(t)}{S(t)} & \text{si } S(t) \neq 0 \end{cases} \quad (3.5)$$

Le taux de hasard se justifie par le fait que si  $f$  est continue, alors, pour  $t > 0$  tel que  $p(T > t) > 0$  on a

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq X < t + dt | X \geq t)}{dt}$$

En effet  $P(t \leq X < t + dt | X \geq t) = \frac{P(t \leq X < t + dt)}{S(t)}$

$$\begin{aligned} \lim_{dt \rightarrow 0} \frac{P(t \leq X < t + dt | X \geq t)}{dt} &= \lim_{dt \rightarrow 0} \frac{1}{dt} \frac{P(t \leq X < t + dt)}{S(t)} \\ &= \lim_{dt \rightarrow 0} \frac{1}{dt} \frac{F(t + dt) - F(t)}{S(t)} \\ &= \frac{F'(t)}{S(t)} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

Pour  $t$  fixé,  $h(t)$  caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant  $t$ , conditionnellement au fait d'avoir survécu jusqu'à l'instant  $t$ . Aussi cela signifie-t-il le risque de mort instantané pour ceux qui ont survécues.

#### 4.5 Le taux de hasard cumulé $H$

Le taux de hasard cumulé est l'intégrale du taux de hasard  $h$  :

$$H(t) = \int_0^t h(u) du = -\ln \{S(t)\} \quad (3.6)$$

On peut déduire la fonction de survie du taux de hasard cumulé grâce à la relation :

$$S(t) = \exp\{-H(t)\} = \exp\left\{-\int h(u) du\right\} \quad (3.7)$$

N'importe laquelle des fonctions ci-dessus peut être obtenue à partir de l'une quelconque des autres.

## 4.6 Quantités associées à la distribution de survie

### 4.6.1 Moyenne et variance de la durée de survie

Le temps moyen de survie  $E(X)$  ainsi que sa variance  $Var(X)$  sont des quantités importantes :

$$E(X) = \int_0^{\infty} S(t)dt, \quad Var(X) = 2 \int_0^{\infty} tS(t)dt - [E(X)]^2.$$

La moyenne et la variance peuvent être déduites de n'importe laquelle des cinq fonctions ci-dessus ( $F, S, f, h, H$ ), mais pas vice versa.

### 4.6.2 Le quantile de la durée de survie

La médiane de la durée de survie est le temps  $t$  pour lequel la probabilité de survie  $S(t)$  est égale à 0.5, c'est-à-dire, la valeur  $t_m$  qui satisfait  $S(t_m) = 0,5$ .

Dans le cas où l'estimateur est une fonction en escalier (ex : Kaplan-Meier), il se peut qu'il y ait un intervalle de temps vérifiant  $S(t_m) = 0,5$ . Il faut alors être prudent dans l'interprétation, notamment si les deux événements encadrant le temps médian sont éloignés.

Il est possible d'obtenir un intervalle de confiance du temps médian. Soit  $[B_s; B_i]$  un intervalle de confiance de niveau  $\alpha$  de  $S(t_m)$ ; alors un intervalle de confiance de niveau  $\alpha$  du temps médian  $t_m$  est

$$[S^{-1}(B_s), S^{-1}(B_i)]$$

Pour  $0 < p < 1$ , on définit le quantile  $t_p$  par la fonction  $q(p)$ ,  $p \in (0, 1)$  comme

$$t_p \equiv q(p) = \inf\{t : F(t) \geq p\} \quad (3.8)$$

Quand  $F(t)$  est strictement croissante et continue alors

$$t_p = q(p) = F^{-1}(p), 0 < p < 1 \quad (3.9)$$

Pour  $p$  fixé, le quantile  $t_p$  est le temps auquel une proportion  $p$  de la population a disparue.

## 5 Estimation sous le modèle de troncature à gauche

### 5.1 Modèle de troncature à gauche

Soient  $\{(Y_i, T_i), 1 \leq i \leq N\}$  une suite de variable aléatoire indépendante et identiquement distribuée ou  $T$  est la variable de troncature. On suppose que  $T$  est indépendante de  $Y$  et que  $T$  et  $Y$  ont les fonctions de distributions  $G(\cdot)$  et  $F(\cdot)$  respectivement.

Soit  $H(\cdot, \cdot)$  la fonction de distributions conjointe de  $Y$  et  $T$ .

Puisque  $T$  est indépendant avec  $Y$  alors,

$$\begin{aligned} H(y, t) &= P(Y \leq y, T \leq t) \\ &= F(y)G(t) \end{aligned} \quad (3.10)$$

Dans le modèle de troncature à gauche la durée de vie de  $Y_i$  est interférée par la variable de troncature aléatoire  $T_i$  en telle sorte que  $Y_i$  et  $T_i$  sont observables que si  $Y_i \geq T_i$  tandis que ni l'un ni l'autre sont observés si  $Y_i < T_i$  pour  $i = 1, \dots, N$ , où  $N$  est la taille de l'échantillon. A cause de l'occurrence de la troncature,  $N$  est inconnue.

Soient  $\{(Y_i, T_i), 1 \leq i \leq n\}$  l'échantillon réellement observé (i.e  $Y_i \geq T_i$ ) avec  $n \leq N$ .

Soit  $\theta = P(Y \geq T)$  la probabilité que la variable aléatoire  $Y$  est observable. si  $\theta = 0$  alors les données ne sont pas observables, nous supposons que  $\theta > 0$ .

Utilisant la loi forte des grands nombres, on a

$$\hat{\theta}_n = \frac{n}{N} \xrightarrow{N \rightarrow \infty} \theta, \quad P - p.s. \quad (3.11)$$

Les données réellement observées  $\{(Y_i, T_i), \{1 \leq i \leq n\}\}$  sont de fonction de distributions conjointe  $H^*(y, t)$

On pose  $\theta = p(Y \geq T)$

$$\begin{aligned}
H^*(y, t) &= \mathbf{P}(Y \leq y, T \leq t | Y \geq T) & (3.12) \\
&= \frac{\mathbf{P}(Y \leq y, T \leq t, Y \geq T)}{p(Y \geq T)} \\
&= \theta^{-1} \mathbf{E}(\mathbf{1}_{\{Y \leq y, T \leq t, Y \geq T\}}) \\
&= \theta^{-1} E \left[ \mathbf{E}(\mathbf{1}_{\{Y \leq y, T \leq t, Y \geq T\}} | Y) \right] \\
&= \theta^{-1} \int_0^{+\infty} \mathbf{E}(\mathbf{1}_{\{Y \leq y, T \leq t, Y \geq T\}} | Y = u) dF(u) \\
&= \theta^{-1} \int_0^{+\infty} \mathbf{P}(Y \leq y, T \leq t, Y \geq T | Y = u) dF(u) \\
&= \theta^{-1} \int_0^y \mathbf{P}(T \leq t, T \leq Y | Y = u) dF(u) \\
&= \theta^{-1} \int_0^y \mathbf{P}(T \leq t \wedge u) dF(u) \\
&= \theta^{-1} \int_0^y G(t \wedge u) dF(u)
\end{aligned}$$

Stute (1993)[41] a prouvé que, en raison de l'indépendance entre  $T$  et  $Y$ , Les fonctions de répartitions conditionnelles de  $Y$  et  $T$  sont les suivantes :

$$\begin{aligned}
F^*(y) &= \mathbf{P}(Y \leq y | Y \geq T) \\
&= \theta^{-1} \mathbf{P}(Y \leq y, Y \geq T) \\
&= \theta^{-1} \mathbf{E}(\mathbf{1}_{\{Y \leq y, Y \geq T\}}) \\
&= \theta^{-1} \mathbf{E}[\mathbf{E}(\mathbf{1}_{\{Y \leq y, Y \geq T\}} | Y)] \\
&= \theta^{-1} \int_{\mathbb{R}} \mathbf{E}(\mathbf{1}_{\{Y \leq y, Y \geq T\}} | Y = u) dF(u) \\
&= \theta^{-1} \int_{\mathbb{R}} \mathbf{P}(Y \leq y, Y \geq T | Y = u) dF(u) \\
&= \theta^{-1} \int_0^y \mathbf{P}(T \leq u) dF(u) \text{ car } Y \text{ et } T \text{ sont indépendants} \\
&= \theta^{-1} \int_0^y G(u) dF(u)
\end{aligned}$$

$$\begin{aligned}
F^*(y) &= \lim_{t \rightarrow +\infty} H^*(y, t) \\
&= H^*(y, +\infty) \\
&= \theta^{-1} \int_0^y G(u) dF(u)
\end{aligned}$$

Et

$$\begin{aligned}
G^*(y) &= \mathbf{P}(T \leq y | Y \geq T) & (3.14) \\
&= \theta^{-1} \mathbf{P}(T \leq y, Y \geq T) \\
&= \theta^{-1} \mathbf{E}(\mathbf{1}_{\{T \leq y, Y \geq T\}}) \\
&= \theta^{-1} E \left[ \mathbf{E}(\mathbf{1}_{\{T \leq y, Y \geq T\}} | Y) \right] \\
&= \theta^{-1} \int_0^{+\infty} \mathbf{E}(\mathbf{1}_{\{T \leq y, Y \geq T\}} | Y = u) dF(u) \\
&= \theta^{-1} \int_0^{+\infty} \mathbf{P}(T \leq y, Y \geq T | Y = u) dF(u) \\
&= \theta^{-1} \int_0^{+\infty} \mathbf{P}(T \leq y, T \leq u) dF(u) \\
&= \theta^{-1} \int_0^{+\infty} G(y \wedge u) dF(u)
\end{aligned}$$

$$\begin{aligned}
G^*(y) &= \lim_{y \rightarrow +\infty} H^*(y, t) & (3.15) \\
&= H^*(+\infty, t) \\
&= \theta^{-1} \int_0^{+\infty} G((t \wedge u)) dF(u)
\end{aligned}$$

Les estimateurs empiriques de  $F^*$  et  $G^*$  sont :

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{Y_i \leq y\}} \quad (3.16)$$

$$G_n^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq t\}} \quad (3.17)$$

### Remarque

$F_n^*$  et  $G_n^*$  sont les estimateurs de  $F^*$  et  $G^*$  mais pas de  $F$  et  $G$ .

Soient  $a_W = \inf\{y : W(y) > 0\}$  et  $b_W = \sup\{y : W(y) < 1\}$  La borne inférieure et la borne supérieure de la fonction  $W$  respectivement.

Puisque

$$\begin{aligned}
C(y) &= P(T \leq y \leq Y) & (3.18) \\
&= \mathbf{P}(T \leq y \leq Y | Y \geq T) \\
&= \frac{\mathbf{P}(T \leq y \leq Y, Y \geq T)}{\mathbf{P}(Y \geq T)} \\
&= \theta^{-1} \mathbf{P}(T \leq y \leq Y)
\end{aligned}$$

Posons  $A = \{T \leq y\}$  et  $B = \{Y < y\}$  alors  $p(A \cap \bar{B}) = p(A) - p(A \cap B)$

Donc

$$\begin{aligned}
C(y) &= \theta^{-1} \mathbf{P}(T \leq y \leq Y) & (3.19) \\
&= \theta^{-1} [\mathbf{P}(T \leq y) - \mathbf{P}(T \leq y, Y < y)] \\
&= \theta^{-1} [\mathbf{P}(T \leq y) - \mathbf{P}(T \leq y) \mathbf{P}(Y < y)] \\
&= \theta^{-1} [G(y) - G(y)F(y)] \\
&= \theta^{-1} G(y) [1 - F(y)] \\
&= \theta^{-1} G(y) \bar{F}(y)
\end{aligned}$$

on peut montrer que  $C(y) = G^*(y) - F^*(y)$

L'estimateur empirique de  $C(y)$  est défini par :

$$C_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq y \leq Y_i\}} \quad (3.20)$$

En rappelant que la fonction de hasard cumulative d'une fonction de distribution  $F$  est définie par :

$$\begin{aligned}
\Omega(y) &= \int_0^y h(t) dt & (3.21) \\
&= \int_0^y \frac{dF(t)}{1 - F(t)} \quad 0 < y < \infty \\
&= \frac{\gamma}{\gamma} \int_{a_F}^y \frac{G(t) dF(t)}{G(t) \bar{F}(t)}
\end{aligned}$$

On pose  $C(t) = \frac{1}{\gamma} G(t) \bar{F}(t)$  et  $F^*(y) = \frac{1}{\gamma} \int G(t) dF(t)$  et donc  $dF^*(y) = \frac{1}{\gamma} G(t) dF(t)$ . alors

$$\begin{aligned}
\Omega(y) &= \frac{1}{\gamma} \int_{a_F}^y \frac{G(t) dF(t)}{C(t)} & (3.22) \\
&= \int_{a_F}^y \frac{dF^*(t)}{C(t)}
\end{aligned}$$

Ceci suggère d'estimer  $\Omega$  par

$$\hat{\Omega}_n(y) = \int_{a_F}^y \frac{dF_n^*(u)}{C_n(u)} = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{1}_{\{Y_i \leq y\}}}{C_n(Y_i)} \quad (3.23)$$

Où  $F_n^*$  et  $C_n$  sont les estimateurs empiriques de  $F^*$  et  $C$  déjà définis.

## 5.2 Estimation de la probabilité de troncature

Soit  $\theta = P(Y \geq T)$  la probabilité que la variable aléatoire  $Y$  est observable. si  $\theta = 0$  alors les données ne sont pas observable, nous supposons que  $\theta > 0$ .

un estimateur de  $\theta$  est donné par  $\frac{n}{N}$ , mais cet estimateur ne peut pas être calculé vu que  $N$  est inconnu.

$$\begin{aligned} \theta &= \mathbb{P}(Y \geq T) \\ &= \mathbb{E}(\mathbf{1}_{\{Y \geq T\}}) \\ &= \mathbb{E}[\mathbb{E}(\mathbf{1}_{\{Y \geq T\}}|Y)] \\ &= \int_{a_F}^{+\infty} \mathbb{E}(\mathbf{1}_{\{Y \geq T\}}|Y = y)dF(y) \\ &= \int_{a_F}^{+\infty} \mathbb{P}(Y \geq T|Y = y)dF(y) \\ &= \int_{a_F}^{+\infty} \mathbb{P}(T \leq y)dF(y) \text{ car } Y \text{ et } T \text{ sont indépendants.} \\ &= \int_{a_F}^{+\infty} G(y)dF(y) . \end{aligned}$$

Keiding et Gill (1990)[24] suggèrent d'estimer  $\theta$  par

$$\theta_n = \int_{a_F}^{+\infty} G_n(y)dF_n(y)$$

Où  $G_n$  et  $F_n$  sont les estimateurs du maximum de vraisemblance de Lynden-Bell (1971)[28] de  $F$  et  $G$  respectivement.

on a  $C(y) = P(T \leq y \leq Y) = \theta^{-1}G(y) [1 - F(y)]$  donc

$$\theta = \frac{G(y) [1 - F(y)]}{C(y)}, \forall y \in ]a_F, b_F[ \text{ et } C(y) > 0$$

L'estimateur de  $\theta$  est défini ( He et Yang (1998)) par  $\theta_n = G_n(y)[1 - F_n(y)]C_n^{-1}(y)$ . He et Yang (1998)[23] ont prouvé que  $\theta_n$  ne dépend pas de  $y$  et sa valeur peut alors être

obtenue pour toute  $y$  tel que  $C_n(y) \neq 0$ . de plus, ils ont montré (corolaire 2.5) que

$$\theta_n \xrightarrow{p.s} \theta \text{ quand } n \longrightarrow \infty.$$

### 5.3 Les estimateurs du maximum de vraisemblance de $F$ et $G$ (Lynden-Bell (1971))

Dans le cas absolument continu, l'estimateur de  $F$  s'obtient par :

$$\begin{aligned} \Omega_n(y) &= \int_0^y h(t) dt & (3.24) \\ &= \int_0^y \frac{dF_n(t)}{1 - F_n(t)} \\ &= -\log(1 - F_n(y)) \end{aligned}$$

donc

$$\begin{aligned} F_n(y) &= 1 - \exp[-\Omega_n(y)] & (3.25) \\ &= 1 - \exp\left[-\frac{1}{n} \sum_{i=1}^n \frac{I_{\{Y_i \leq y\}}}{C_n(Y_i)}\right] \\ &= 1 - \prod_{i=1}^n \exp\left[-\frac{1}{n} \frac{I_{\{Y_i \leq y\}}}{C_n(Y_i)}\right] \\ &= 1 - \prod_{\substack{i=1 \\ Y_i \leq y}}^n \exp\left[-\frac{1}{nC_n(Y_i)}\right] \end{aligned}$$

Puisque  $\exp\left[-\frac{1}{nC_n(Y_i)}\right] \simeq 1 - \frac{1}{nC_n(Y_i)} = \frac{nC_n(Y_i) - 1}{nC_n(Y_i)}$  alors

$$F_n(y) = 1 - \prod_{\substack{i=1 \\ Y_i \leq y}}^n \left(1 - \frac{1}{nC_n(Y_i)}\right) \quad (3.26)$$

On peut montrer que l'estimateur de maximum de vraisemblance de  $G$  est :

$$G_n(y) = \prod_{\substack{i=1 \\ T_i > y}}^n \left(1 - \frac{1}{nC_n(T_i)}\right) \quad (3.27)$$

## 5.4 Propriétés asymptotiques de l'estimateur de Lynden-Bell

Woodroffe (1985)[49] (dans le théorème 2) a établi des résultats de la convergence uniforme :

$$\sup_y |F_n(y) - F_0(y)| \xrightarrow{P-p.s.} 0$$

et

$$\sup_y |G_n(y) - G_0(y)| \xrightarrow{P-p.s.} 0 \quad (3.28)$$

Où  $F_0$  dénote la distribution conditionnelle de  $Y$  sachant que  $Y \geq a_G$  et  $G_0$  dénote la distribution de  $T$  sachant que  $T \leq b_F$ . Donc,  $F$  est identifiable ( $F = F_0$ ) que si  $a_G \leq a_F$ , tandis que  $G$  est identifiable ( $G = G_0$ ) si  $b_G \leq b_F$ . C'est des conditions nécessaires mais ne sont pas suffisantes pour l'identifiable.

Par conséquent,  $\theta$  est identifiable si et seulement si  $a_G \leq a_F$  et  $b_G \leq b_F$ .

L'estimateur  $\hat{\theta}_n$  défini dans l'équation 3.11 ne peut pas être calculé (car  $N$  est inconnu) un autre estimateur est utilisé appelé

$$\theta_n = \frac{G_n(y) [1 - F_n(y^-)]}{C_n(y)} \quad \text{si } C_n(y) \neq 0. \quad (3.29)$$

He et Yang (1998)[23] ont prouvé que  $\theta_n$  ne dépend pas de  $y$ . de plus, ils ont prouvé (corolaire 2.5) que

$$\theta_n \xrightarrow[n \rightarrow \infty]{} \theta, P - p.s.$$

## 5.5 Estimation de la densité de la covariable sous le modèle de troncature à gauche dans le cas de la régression

Maintenant, en plus de deux variables considérées auparavant  $Y$  et  $T$ , nous considérons un vecteur aléatoire  $X \in \mathbb{R}^d$  de covariables, supposé absolument continu de fonction de répartition  $L(\cdot)$  et une densité continue  $l(\cdot)$ .

On note  $(X_i, Y_i, T_i), 1 \leq i \leq n$  l'échantillon observé (i.e  $Y_i \geq T_i$ ). On suppose que  $T$  est indépendante de  $(X, Y)$ .

Ould-Said et Lemdani (2006)[34] ont construit les estimateurs de  $L(\cdot)$  et  $l(\cdot)$ .

Soient  $X_1, X_2, \dots, X_n$  un échantillon de vecteurs aléatoires de  $d$ -variable aléatoire tirée d'une distribution commune décrite par la fonction de densité  $l$ .

L'estimateur de la densité conditionnelle  $l^*(x)$  (sachant que  $Y_i \geq T_i$ ) est défini par

$$l_n^*(x) = \frac{1}{nh^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h} \right) \quad (3.30)$$

où  $K_d : \mathbb{R}^d \rightarrow \mathbb{R}$  est un noyau fixé avec  $\int_{\mathbb{R}^d} K_d = 1$

Puisque  $T$  est indépendante de  $(X, Y)$ , la distribution conditionnelle conjointe de  $(X, Y, T)$  satisfait :

$$\begin{aligned} H^*(x, y, t) &= \mathbb{P}(X \leq x, Y \leq y, T \leq t | Y \geq T) \\ &= \mathbb{P}(X \leq x, Y \leq y | Y \geq T) \mathbb{P}(T \leq t | Y \geq T) \\ &= \theta^{-1} \mathbb{P}(X \leq x, Y \leq y | Y \geq T) \mathbb{P}(T \leq t, Y \geq T) \\ &= \theta^{-1} \int_{s \leq x} \int_{a_G \leq \nu \leq y} G(\nu \wedge t) F(ds, d\nu). \end{aligned}$$

On prend  $t = +\infty$ , la fonction de répartition du couple observé  $(X, Y)$  sachant que  $Y \geq T$  noté  $F^*(., .)$  est :

$$F^*(x, y) = H^*(x, y, \infty) = \theta^{-1} \int_{s \leq x} \int_{a_G \leq \nu \leq y} G(\nu) F(ds, d\nu). \quad (3.32)$$

Par différentiation, on obtient

$$F(dx, dy) = \frac{\theta}{G(y)} F^*(dx, dy) \text{ pour } y > a_G \quad (3.33)$$

D'où

$$f(x, y) = \frac{\theta}{G(y)} f^*(x, y) \quad (3.34)$$

En intégrant la fonction  $f(x, y)$  précédente par rapport à  $y$ , on obtient la distribution de  $X$

$$L(x) = \theta \int_{s \leq x} \int_{y \geq a_G} \frac{1}{G(y)} F^*(ds, dy) \quad (3.35)$$

Un estimateur naturel de  $L(x)$  est donné par

$$L_n(x) = \frac{\theta_n}{n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} \mathbf{1}_{\{X_i \leq x\}}. \quad (3.36)$$

Notons que 3.36 est vraie si  $G_n(Y_i) \neq 0$ , Finalement, 3.36 donne l'estimateur de la densité de  $X$  de la forme

$$\begin{aligned} l_n(x) &= \frac{1}{h^d} \int K\left(\frac{x-u}{h}\right) dL_n(u) \\ &= \frac{\theta_n}{nh} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{x-X_i}{h}\right) \end{aligned} \quad (3.37)$$

# Chapitre

# 4

## Données dépendantes

### 1 Introduction

Souvent, en statistique, on rencontre des données dépendantes, c'est pourquoi, plusieurs auteurs se sont intéressés à ce type de données. Dans leurs études, les statisticiens et les probabilistes considèrent en général des conditions de dépendance faible appelées conditions de mélanges. Afin de mesurer cette dépendance entre les  $\sigma$ -algèbres, différents coefficients sont utilisés.

### 2 Définitions des mesures de dépendances

Soit  $(\Omega, F, P)$  un espace de probabilité, prenant  $A$  et  $B$  deux sous  $\sigma$ -algèbres de  $F$ . Plusieurs mesures de dépendance entre  $A$  et  $B$  sont définies comme suit :

**1) Le coefficient de mélange fort**

$$\alpha(A, B) = \sup \{|P(a \cap b) - P(a)P(b)|, a \in A, b \in B\} \text{ (Rosenblatt(1956))}$$

**2) Le coefficient de régularité absolue**

$$\beta(A, B) = \sup \left\{ \frac{1}{2} \sum_{i \in I} \sum_{j \in J} |P(a_i \cap b_j) - P(a_i)P(b_j)| \right\} \text{ (Volkonskii et Rozanov(1959))}$$

où le sup est pris sur toutes les partitions  $a_i$  et  $b_j$  de  $\Omega$  telles que  $a_i \in A$  et  $b_j \in B$ .

**3) Le coefficient mélange uniforme**

$$\phi(A, B) = \sup \{|P(b/a) - P(b)|, a \in A, b \in B \text{ et } p(a) > 0\} \text{ (Ibragimov(1959))}$$

**4) Le coefficient  $\Psi$ -mélangeant**

$$\Psi(A, B) = \sup \left\{ \left| \frac{P(a \cap b)}{P(a)P(b)} - 1 \right|, a \in A, b \in B, P(a) > 0, P(b) > 0 \right\} \text{ (Blum et al (1963))}$$

### 5) Le coefficient $\rho$ -mélangeant

$\rho(A, B) = \sup \{ |corr(f, g)|; f \in L^2(A), g \in L^2(B) \}$ , (Kolmogorov et Rozanov(1960))  
où  $L^2(A)$  est l'espace des variables aléatoires de carré intégrable et  $A$ -mesurables.

## 3 Propriétés de base

1) On a les inégalités suivantes des mesures de dépendance mentionnées précédemment

1.  $0 \leq \alpha(A, B) \leq 1/4$ ,
2.  $0 \leq \phi(A, B) \leq 1$ ,
3.  $0 \leq \Psi(A, B) \leq \infty$ ,
4.  $0 \leq \rho(A, B) \leq 1$ ,
5.  $0 \leq \beta(A, B) \leq 1$ .

2) Si les tribus  $A$  et  $B$  sont indépendantes, alors, on a

$$\alpha(A, B) = 0, \phi(A, B) = 0, \Psi(A, B) = 0, \rho(A, B) = 0, \beta(A, B) = 0.$$

3) Les mesures de dépendances satisfont les inégalités

1.  $2\alpha(A, B) \leq \beta(A, B) \leq \phi(A, B) \leq \frac{1}{2}\Psi(A, B)$
2.  $4\alpha(A, B) \leq \rho(A, B) \leq \Psi(A, B)$
3.  $\alpha(A, B) \leq 2[\phi(A, B)]^{1/2}[\phi(B, A)]^{1/2} \leq 2[\phi(A, B)]^{1/2}$ .

## 4 Conditions de mélangeance forte

### 4.1 Conditions de mélangeance forte basée sur le passé et le futur

Soit  $X = (X_k, k \in \mathbb{Z})$  (pas nécessairement stationnaire), une suite de variables aléatoires.

Pour  $-\infty \leq J \leq L \leq \infty$ , on définit la  $\sigma$ -algèbre

$$F_J^L = \sigma(X_k, J \leq k \leq L, (k \in \mathbb{Z})).$$

Pour tout  $n \geq 1$ , définissons les coefficients de dépendance suivants :

$$\alpha(n) = \sup_{j \in \mathbb{Z}} \alpha(F_{-\infty}^j, F_{j+n}^{\infty})$$

$$\phi(n) = \sup_{j \in \mathbb{Z}} \phi(F_{-\infty}^j, F_{j+n}^{\infty})$$

$$\Psi(n) = \sup_{j \in \mathbb{Z}} \Psi(F_{-\infty}^j, F_{j+n}^{\infty})$$

$$\beta(n) = \sup_{j \in \mathbb{Z}} \beta(F_{-\infty}^j, F_{j+n}^{\infty})$$

$$\rho(n) = \sup_{j \in \mathbb{Z}} \rho(F_{-\infty}^j, F_{j+n}^{\infty})$$

Le processus aléatoire est dit :

1. " $\alpha$ -mélangeant" si  $\alpha(n) \rightarrow 0$  quand  $n \rightarrow \infty$ ,
2. " $\phi$ -mélangeante" si  $\phi(n) \rightarrow 0$  quand  $n \rightarrow \infty$ ,
3. " $\Psi$ -mélangeante" si  $\Psi(n) \rightarrow 0$  quand  $n \rightarrow \infty$ ,
4. " $\rho$ -mélangeante" si  $\rho(n) \rightarrow 0$  quand  $n \rightarrow \infty$ ,
5. "absolument régulière" ou  $\beta$ -mélangeante si  $\beta(n) \rightarrow 0$  quand  $n \rightarrow \infty$ .

**Remarque.**

Lorsque le processus  $X$  est strictement stationnaire, on peut définir les coefficients de dépendance par

$$\alpha(n) = \alpha(F_{-\infty}^0, F_n^{\infty})$$

et la même chose pour les autres coefficients.

Pour plus de détails sur le sujet, on peut consulter par exemple, Bradley (2005)[2].

# Chapitre 5

## Estimation de la moyenne conditionnelle des données tronquées dépendantes par la méthode des polynômes locaux

Dans ce chapitre nous allons reprendre les résultats obtenus par Liang, de Uña-Álvarez et Iglesias-Pérez (2011)[27], se basant sur l'idée de lissage à noyau de Nadaraya-Watson (*NW*) et de la technique de lissage localement linéaire (*LL*), on reprend les estimateurs de *NW* et *LL* de la fonction moyenne conditionnelle et ses dérivées pour le modèle tronqué à gauche. La fonction objective est la fonction de régression, le moment conditionnel et la fonction de distribution conditionnelle comme cas particulier. Il est supposé que les données de survie observées  $\{(X_k, Y_k, T_k), \{1 \leq k \leq n\}\}$  forment une suite de  $\alpha$ -mélange stationnaire. La normalité asymptotique de l'estimateur est établie.

### 1 Introduction

L'étude de la relation entre une variable d'intérêt  $Y$  et la covariable  $X$  est l'un des problèmes les plus importants en statistiques. Cela peut être fait par l'intermédiaire d'estimation de la fonction  $m(x) = E(Y|X = x)$ , Si nous avons  $n$  observations  $(X_i, Y_i)$  de  $(X, Y)$ , plusieurs méthodes de lissage à noyau consistent à estimer  $m(x)$  en calculant une moyenne locale pondérée des valeurs  $Y_i$ . Cela peut être écrit comme

$$\sum_{i=1}^n w_i(x) Y_i \tag{5.1}$$

Où  $w_i(x)$  est une fonction de poids donnée décrivant le degré de lissage sur la variable. Cas particuliers de 5.1 comprennent l'estimateur de Nadaraya-Watson (*NW*) et l'estimateur par polynômes locaux (*LP*). Pour les propriétés statistiques de ces deux estimateurs, dans le cas de données indépendantes, voir le livre de Fan et Gijbels (1996)[11].

Nous considérons des observations dépendantes, alors  $\{(X_k, Y_k), k \geq 1\}$  désigne un processus stationnaire de  $\mathbb{R} \times \mathbb{R}$  de  $(X, Y)$ .

Soit  $\{(X_k, Y_k, T_k), k \geq 1\}$  une suite de vecteurs aléatoire de  $(X, Y, T)$  où  $Y$  est la variable d'intérêt,  $T$  est la variable de troncature et  $X$  le vecteur des covariables. On suppose que  $T$  est indépendante de  $(X, Y)$  et que  $T, X$  et  $Y$  ont les fonctions de distributions  $G(\cdot)$ ,  $L(\cdot)$  et  $F(\cdot)$  continues respectivement.

Soit  $(X, Y)$  un couple de variable aléatoire admettant une fonction de distributions conjointe  $F(\cdot, \cdot)$  et une densité  $f(\cdot, \cdot)$ , comme toujours avec l'analyse de survie, on suppose que  $Y$  et  $T$  sont des variables non négatives. Dans le modèle de troncature à gauche la durée de vie de  $Y_i$  est interférée par la variable de troncature aléatoire  $T_i$  en telle sorte que  $Y_i$  et  $T_i$  ne soient observables que si  $Y_i \geq T_i$  tandis que ni l'un ni l'autre sont observées si  $Y_i < T_i$  pour  $i = 1, \dots, N$ , où  $N$  est la taille potentielle de l'échantillon. A cause de l'occurrence de la troncature,  $N$  est inconnue, et  $n$  ( la taille de l'échantillon actuellement observé) est connue et aléatoire avec  $n \leq N$ .

Se basant sur l'échantillon observé, la fonction de régression est

$$m(x) = E(Y|X = x, Y \geq T)$$

Soit  $\theta = P(Y \geq T)$  la probabilité que la variable aléatoire  $Y$  est observable. si  $\theta = 0$  alors les données ne sont pas observables, nous supposons que  $\theta > 0$ . Puisque  $N$  est inconnu et  $n$  est connu (même si aléatoire), les résultats ne seront pas obtenus par rapport à la mesure de probabilité  $\mathbb{P}$  (liée à  $N$ -échantillon), mais par rapport à la probabilité conditionnelle  $P$  par rapport à  $n$ -échantillon effectivement observé. Aussi  $\mathbb{E}$  et  $E$  dénotent les opérateurs d'espérance sous  $\mathbb{P}$  et  $P$ , respectivement.

On rappelle qu'une suite  $\{\zeta_k, k \geq 1\}$  est dite  $\alpha$ -mélange on l'appelle aussi mélange fort, si le coefficient  $\alpha$ -mélange :

$$\alpha(n) = \sup_{k \geq 1} \{ |p(A \cap B) - p(A)p(B)| : A \in F_{n+k}^\infty, B \in F_1^k \} \xrightarrow{n \rightarrow \infty} 0.$$

où  $F_a^b = \sigma\{\zeta_i, a \leq i \leq b\}$  désigne sigma algèbre engendrée par  $\zeta_a, \zeta_{a+1}, \dots, \zeta_b$ . Parmi les diverses conditions de mélange utilisées dans la littérature, Gorodetskii (1977)[18] et Withers (1981)[48] ont tiré des conditions sous lesquelles un processus Linéaire est  $\alpha$ -mélange.

## 2 Estimateurs

Stute (1993)[41] a prouvé que, en raison de l'indépendance entre  $T$  et  $Y$ , Les fonctions de répartitions conditionnelles de  $Y$  et  $T$  sont les suivantes (voir chapitre 3) :

$$\begin{aligned} F^*(y) &= P(Y \leq y) \\ &= \mathbb{P}(Y \leq y | Y \geq T) \\ &= \theta^{-1} \int_0^y G(u) dF(u) \end{aligned}$$

Et

$$\begin{aligned} G^*(y) &= \mathbb{P}(T \leq y | Y \geq T) \\ &= \theta^{-1} \int_0^{+\infty} G(y \wedge u) dF(u) \end{aligned}$$

Les estimateurs empiriques de  $F^*(y)$  et  $G^*(y)$  sont définis par

$$F_n^*(y) = n^{-1} \sum_{i=1}^n \mathbf{1}_{(Y_i \leq y)}$$

et

$$G_n^*(y) = n^{-1} \sum_{i=1}^n \mathbf{1}_{(T_i \leq y)}$$

respectivement avec  $\mathbf{1}_{(\cdot)}$  est la fonction indicatrice.

Puisque

$$\begin{aligned} C(y) &= P(T \leq y \leq Y) \\ &= \mathbb{P}(T \leq y \leq Y | Y \geq T) \\ &= \theta^{-1} \mathbb{P}(T \leq y \leq Y, Y \geq T) \\ &= \theta^{-1} \mathbb{P}(T \leq y \leq Y) \end{aligned}$$

Posons  $A = \{T \leq y\}$  et  $B = \{Y < y\}$  alors  $p(A \cap \bar{B}) = p(A) - p(A \cap B)$

Donc

$$\begin{aligned}
C(y) &= \theta^{-1} \mathbb{P}(T \leq y \leq Y) \\
&= \theta^{-1} [\mathbb{P}(T \leq y) - \mathbb{P}(T \leq y, Y < y)] \\
&= \theta^{-1} [\mathbb{P}(T \leq y) - \mathbb{P}(T \leq y) \mathbb{P}(Y < y)] \\
&= \theta^{-1} [G(y) - G(y)F(y)] \\
&= \theta^{-1} G(y) [1 - F(y)] \\
&= G^*(y) - F^*(y)
\end{aligned}$$

L'estimateur empirique de  $C(y)$  est défini par :

$$C_n(y) = n^{-1} \sum_{i=1}^n I(T_i \leq y \leq Y_i) = G_n^*(y) - F_n^*(y-)$$

où  $F_n^*(y-)$  dénote la limite gauche de  $F_n^*$  en  $y$ .

Suite à l'idée de Lynden-Bell (1971)[28], les estimateurs du maximum de vraisemblance de  $F$  et  $G$  sont donnés par :

$$1 - F_n(y) = \prod_{i: Y_i \leq y} \left(1 - \frac{1}{nC_n(Y_i)}\right) \quad \text{et} \quad G_n(y) = \prod_{i: T_i > y} \left(1 - \frac{1}{nC_n(T_i)}\right)$$

L'estimateur de  $\theta$  est défini (He et Yang 1998) par  $\theta_n = G_n(y)[1 - F_n(y)]C_n^{-1}(y)$ . He et Yang (1998)[23] ont prouvé que  $\theta_n$  ne dépend pas de  $y$  et sa valeur peut alors être obtenue pour toute  $y$  tel que  $C_n(y) \neq 0$ .

Pour une fonction de distribution  $W$  Woodroffe (1985)[49] a posé la borne inférieure et la borne supérieure  $a_W = \inf\{y : W(y) > 0\}$  et  $b_W = \sup\{y : W(y) < 1\}$  respectivement.

Puisque  $T$  une variable indépendante de  $(X, Y)$ , la distribution conditionnelle conjointe de  $(X, Y, T)$  satisfaite (voir chapitre 3) :

$$\begin{aligned}
H^*(x, y, t) &= P(X \leq x, Y \leq y, T \leq t) \\
&= \mathbb{P}(X \leq x, Y \leq y, T \leq t | Y \geq T) \\
&= \theta^{-1} \int_{s \leq x} \int_{a_G \leq \nu \leq y} G(\nu \wedge t) F(ds, d\nu).
\end{aligned}$$

On prend  $t = +\infty$ , la fonction de répartition du couple observé  $(X, Y)$  sachant que  $Y \geq T$  noté  $F^*(\cdot, \cdot)$  est :

$$F^*(x, y) = H^*(x, y, \infty) = \theta^{-1} \int_{s \leq x} \int_{a_G \leq \nu \leq y} G(\nu) F(ds, d\nu).$$

Qui donne :

$$F(dx, dy) = [\theta^{-1}G(y)]^{-1} F^*(dx, dy) \quad \text{pour } y > a_G \quad (5.2)$$

En utilisant (A1) et (A3) dans la Section 4, si  $m(\cdot)$  est continue en  $x$ , alors depuis 5.2 nous avons :

$$\begin{aligned} & \theta E \{ K_{h_n}(X - x) G^{-1}(Y) [Y - m(x)] \} \\ &= \frac{\theta}{h_n} \int_{\mathbb{R}} \int_{\mathbb{R}} K \left( \frac{s - x}{h_n} \right) G^{-1}(y) [y - m(x)] F^*(ds, dy) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K(t) [y - m(x)] f(x + h_n t, y) dt dy \\ &= \int_{\mathbb{R}} K(t) l(x + h_n t) [m(x + h_n t) - m(x)] dt \quad \xrightarrow{n \rightarrow \infty} 0 \end{aligned} \quad (5.3)$$

Où  $K_{h_n}(\cdot) = K(\cdot/h_n)/h_n$  avec  $K(\cdot)$  une fonction noyau sur  $\mathbb{R}$  et  $0 < h_n \rightarrow 0$  quand  $n \rightarrow \infty$ .

L'équation 5.3 suggère que l'estimation de  $m(x)$  peut être considérée comme une régression non paramétrique de  $Y_i$  sur  $X_i$  et lié à  $G^{-1}(Y_i)$ . Par conséquent, fondée sur l'idée du lissage à noyau de NW, l'estimateur de NW de  $m(x)$  sous le modelé de troncature à gauche est défini par :

$$\begin{aligned} \hat{m}_{NW}(x) &= \sum_{i=1}^n Y_i w_i^{NW}(x) \quad \text{avec} \\ w_i^{NW}(x) &= \frac{K_{h_n}(X_i - x) G_n^{-1}(Y_i)}{\sum_{j=1}^n K_{h_n}(X_j - x) G_n^{-1}(Y_j)} \end{aligned} \quad (5.4)$$

### Remarque

Quand  $m(x) = E(Y|X = x) = l(x)^{-1} \int_{\mathbb{R}} y f(x, y) dy = \frac{\psi(x)}{l(x)}$ . En utilisant les estimateurs de  $\psi(x)$  et  $l(x)$ , Ould-Saïd et Lemdani (2006)[34] ont construit pour la première fois l'estimateur  $\hat{m}_{NW}(x)$  de  $m(x)$  ayant discuté de ses propriétés asymptotiques dans le cas *i.i.d.*

## 3 Méthode des polynômes locaux

On considère  $(X_1, Y_1), \dots, (X_n, Y_n)$ , des paires de variables aléatoires indépendantes identiquement distribuées.

Le but est d'estimer  $m(x) = \mathbb{E}(Y|X = x)$  et ses dérivées  $m'(x), m''(x), \dots, m^{(p)}(x)$

D'après la notation de Fan et Gijbels (1996)[11], On considéré le modèle :

$$Y_i = m(X_i) + \sigma(X_i)\varepsilon_i,$$

Où  $\varepsilon_i \rightsquigarrow N(0, 1)$  et  $X$  independant avec  $\varepsilon$ .

On note par  $\sigma^2(x)$  la variance de  $Y$  sachant que  $X = x$ .

On suppose que la  $(p + 1)^{\text{ème}}$  dérivée de  $m(\cdot)$  au point  $x$  existe. Alors, on doit ajuster localement la fonction de régression inconnue  $m(x)$  par un polynôme de degré  $p$ .

Dans un petit voisinage de  $x$ , nous pouvons approximer  $m(x)$  localement par un terme linéaire (Expression de Taylor).

$$\begin{aligned} m(z) &\approx m(x) + m^{(1)}(x) \frac{(z-x)^1}{1!} + \dots + m^{(p)}(x) \frac{(z-x)^p}{p!} \\ &\equiv \beta_0 + \beta_1 + \dots + \beta_p (z-x)^p \end{aligned}$$

En vue de 5.3, basée sur l'idée de lissage  $LP$ , l'estimateur  $LP$  :

$$\hat{B}_n(x) = \left( \hat{m}_{LP}(x), \frac{\hat{m}_{LP}^{(1)}(x)}{1!}, \dots, \frac{\hat{m}_{LP}^{(p)}(x)}{p!} \right)^t \quad \text{de} \quad B(x) = \left( m(x), \frac{m^{(1)}(x)}{1!}, \dots, \frac{m^{(p)}(x)}{p!} \right)^t$$

est défini par  $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)^t$  Qui minimise :

$$\sum_{i=1}^n \left( Y_i - \sum_{j=0}^p \beta_j (X_i - x)^j \right)^2 K_{h_n}(X_i - x) G_n^{-1}(Y_i) \quad (5.5)$$

Soient

$$X = \begin{pmatrix} 1 & (X_1 - x) & \dots & (X_1 - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (X_n - x) & \dots & (X_n - x)^p \end{pmatrix}, \quad Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix},$$

$$W = \text{diag}(K_{h_n}(X_i - x) G_n^{-1}(Y_i)).$$

On note par :

$$s_{n,j} = \frac{\theta}{n} \sum_{i=1}^n \left( \frac{X_i - x}{h_n} \right)^j K_{h_n}(X_i - x) G_n^{-1}(Y_i), \quad j = 0, \dots, 2p$$

$$t_{n,j} = \frac{\theta}{n} \sum_{i=1}^n \left( \frac{X_i - x}{h_n} \right)^j K_{h_n}(X_i - x) G_n^{-1}(Y_i) Y_i, \quad j = 0, \dots, p$$

$$S_n = \begin{pmatrix} s_{n,0} & \dots & s_{n,p} \\ \vdots & \ddots & \vdots \\ s_{n,p} & \dots & s_{n,2p} \end{pmatrix} \quad \text{et} \quad t_n = \begin{pmatrix} t_{n,0} \\ \vdots \\ t_{n,p} \end{pmatrix}$$

alors

$$\hat{\mathbf{B}}_n(x) = \text{diag}(1, h_n^{-1}, \dots, h_n^{-p}) S_n^{-1} t_n \quad (5.6)$$

### Remarque

A partir de 5.5, lorsque  $p = 0$ , l'estimateur de  $m(x)$  devient l'estimateur de *NW* définie dans (5.4).

## 4 Hypothèses et résultats

Dans la suite, Soient  $C$ ,  $c_0$  et  $c$  des constantes positives finies, dont les valeurs ne sont pas importantes et peuvent changer de ligne en ligne, et soit  $U(x)$  représente un voisinage de  $x$ .  $A_n = O(B_n)$  signifie que  $|A_n| \leq C |B_n|$ .

Soit

$$\begin{aligned} \sigma^2(x) &= E [(Y - m(X))^2 G_n^{-1}(Y) | X = x]. \\ u_j &= \int_{\mathbb{R}} t^j K(t) dt. \\ v_j &= \int_{\mathbb{R}} t^j K^2(t) dt. \\ S &= \begin{pmatrix} u_0 & \cdots & u_p \\ \vdots & \ddots & \vdots \\ u_p & \cdots & u_{2p} \end{pmatrix}, \quad V = \begin{pmatrix} v_0 & \cdots & v_p \\ \vdots & \ddots & \vdots \\ v_p & \cdots & v_{2p} \end{pmatrix}, \quad U = \begin{pmatrix} u_{p+1} \\ \vdots \\ u_{2p+1} \end{pmatrix}. \end{aligned}$$

Afin de formuler les principaux résultats, nous avons besoin des hypothèses suivantes :

(A0)  $a_G < a_F$  et  $b_G < b_F$ .

(A1) Le noyau  $K(\cdot)$  est une fonction bornée à support compact et  $\int_{\mathbb{R}} K(t) dt = 1$ .

(A2) On suppose que  $nh_n \rightarrow \infty$ , et que la suite  $\alpha(n)$  satisfait :

(i) Il existe des entiers positifs  $q_n$  tels que :

$$q_n = o((nh_n)^{\frac{1}{2}}) \text{ et } \lim_{n \rightarrow \infty} (nh_n^{-1})^{\frac{1}{2}} \alpha(q_n) = 0.$$

(ii) Il existe  $r > 2$  et  $\delta > 1 - 2/r$  tel que :  $\sum_{l=1}^{\infty} l^\delta [\alpha(l)]^{1-2/r} < \infty$

(A3) La densité  $l(\cdot)$  de la covariable est continue au point  $x$  avec  $l(x) > 0$ .

(A4) Pour tout entier  $j \geq 1$ , la densité conjointe  $l_j^*(\cdot, \cdot)$  de  $(X_1, X_{j+1})$  par rapport à  $P$  existe sur  $\mathbb{R} \times \mathbb{R}$  et satisfait  $l_j^*(s_1, s_2) \leq C$  pour  $(s_1, s_2) \in U(x) \times U(x)$ .

(A5) La variance  $\sigma^2(t)$  est continue au point  $x$  avec  $\sigma^2(x) > 0$ , et  $m(\cdot)$  a  $(p+1)$  dérivées continues dans  $U(x)$ .

(A6)

- (i) Pour  $j \geq 1$ , on a  $E(Y_1^2 + Y_{1+j}^2 | X_1 = s, X_{1+j} = t) < \infty$  pour  $(s, t) \in U(x) \times U(x)$ ,
- (ii)  $E(|Y|^r | X = s) < \infty$  pour  $s \in U(x)$  et  $r > 2$ .

(a) Les hypothèses de lissage (A1), (A3), (A5) et (A6) (ii) sont utilisées couramment dans la littérature.

L'hypothèse (A2) implique des restrictions au moment de choisir la largeur de la fenêtre, mais les conditions dans (A2) peut être satisfaites facilement. Par exemple, on choisit  $h_n = cn^{-\eta}$  pour  $0 < \eta < 1$ ,  $q_n = (nh_n/\log n)^{1/2}$  et on suppose que  $\alpha(n) = O(n^{-\gamma})$  pour un certain  $\gamma$ , alors l'hypothèse (A2) détient automatiquement si  $\gamma$  est assez grande, précisément  $\gamma > \max((1+\eta)/(1-\eta), r(1+\delta)/(r-2))$  (noter que  $\gamma$  peut être arbitrairement grande si  $\alpha(k) = O(\rho^k)$  pour certains  $0 < \rho < 1$ ).

Hypothèses (A4) et (A6) (i) sont principalement d'ordre technique, et ils sont utilisés pour simplifier les calculs de covariances dans les épreuves, ces hypothèses sont redondantes pour paramètre indépendant.

(b) Lorsque  $Y$  est bornée, l'hypothèse (A6) est vérifié, et la condition (ii) dans (A2) peut être supprimée.

**Théorème 4.0.1** (*Liang, de Uña-Álvarez et Iglesias-Pérez.(2011)*)

Soit  $\alpha(n) = O(n^{-\gamma})$  pour un certain  $\gamma > 3$ . Supposons que (A0) – (A5) sont vérifiés. Si  $h_n = O(n^{-1/(2p+3)})$ , et l'une des conditions suivantes (i) ou (ii) est vérifié, alors

$$\sqrt{nh_n} \left( \text{diag}(1, \dots, h_n^p) \left[ \hat{\mathbf{B}}_n(x) - \mathbf{B}(x) \right] - \frac{h_n^{p+1} m^{(p+1)}(x)}{(p+1)!} S^{-1} U \right) \xrightarrow{P} N(0, \theta \sigma^2(x) l(x)^{-1} \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}).$$

- (i)  $Y$  est bornée ;
- (ii) (A6) détient,  $\gamma \geq [r(r+2)]/[2(r-2)]$  et  $nh_n^{1+4/r} \geq c_0 > 0$ .

Comme une conséquence directe de ce théorème, nous obtenons le résultat suivant.

**Théorème 4.0.2** (*Liang, de Uña-Álvarez et Iglesias-Pérez.(2011)*)

Sous les hypothèses du théorème 4.0.1, pour  $0 \leq \nu \leq p$ , nous avons

$$\sqrt{nh_n^{2\nu+1}} \left( \hat{m}_{LP}^{(\nu)}(x) - m^{(\nu)}(x) - \frac{m^{(p+1)}(x) \nu! B_\nu}{(p+1)!} h_n^{p+1-\nu} \right) \xrightarrow{P} N(0, \theta \sigma^2(x) l(x)^{-1} (\nu!)^2 V_\nu).$$

Où  $B_\nu$  et  $V_\nu$  sont respectivement le  $\nu^{\text{ème}}$  élément de  $S^{-1}U$  et le  $\nu^{\text{ème}}$  élément de la diagonale de  $S^{-1}VS^{-1}$ . (Voir chapitre 2, Section 6.3).

$$\begin{aligned} V_\nu &= \int K_\nu^{*2}(t)dt \\ B_\nu &= \left[ \int t^{p+1}K_\nu^*(t)dt \right]^2 \end{aligned}$$

(a) Le théorème 4.0.2 montre que le biais et la variance asymptotiques de  $\hat{m}_{LP}^{(\nu)}(x)$  sont respectivement donnés par : (voir chapitre 2)

$$\frac{m^{(p+1)}(x)\nu!B_\nu}{(p+1)!}h_n^{p+1-\nu} \quad \text{et} \quad \frac{\theta\sigma^2(x)(\nu!)^2V_\nu}{nh_n^{2\nu+1}l(x)}.$$

Notons que la fenêtre optimale pour estimer la dérivée  $\nu^{\text{th}}$  est définie comme étant celle qui minimise l'erreur quadratique moyenne  $EQM$ . Alors la fenêtre optimale est donnée par :

$$h_{\nu,opt} = \left( \frac{(2\nu+1)[(p+1)!]^2\theta\sigma^2(x)l(x)^{-1}V_\nu}{2(p+1-\nu)[m^{(p+1)}(x)]^2B_\nu^2} \right)^{1/(2p+3)} \frac{1}{n^{1/(2p+3)}}.$$

(b) Lorsque  $p = 0$ , l'estimateur  $\hat{m}_{LP}(x)$  se réduit à l'estimateur NW noté  $\hat{m}_{NW}(x)$ . Ainsi, d'après le théorème 4.0.2, nous avons :

$$\sqrt{nh_n} \left( \hat{m}_{NW}(x) - m(x) - h_n m'(x) \int_{\mathbb{R}} tK(t)dt \right) \xrightarrow{P} N(0, \Delta^2(x)) \quad (5.7)$$

Où  $\Delta^2(x) = \theta\sigma^2(x)l^{-1}(x) \int_{\mathbb{R}} K^2(t)dt$ .

En outre, si  $nh_n^3 \rightarrow 0$ ,  $\int_{\mathbb{R}} tK(t)dt = 0$  alors

$$\sqrt{nh_n} (\hat{m}_{NW}(x) - m(x)) \xrightarrow{P} N(0, \Delta^2(x))$$

La variance asymptotique est :

$$\Delta^2(x) = \theta\nu_0 [\Xi_0(x)l^2(x) + \Xi_2(x)\gamma^2(x) - 2l(x)\gamma(x)\Xi_1(x)] l^{-4}(x).$$

Où

$$\Xi_j(x) = \int y^{2-j} f(x, y) dy \quad \text{et} \quad \gamma(x) = \int y f(x, y) dy.$$

On définit les estimateurs de  $l_n(x)$ ,  $\gamma(x)$  et  $\Xi_j(x)$  respectivement.

$$\begin{aligned}\hat{l}_n(x) &= \frac{\theta_n}{nh_n} \sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{x - X_i}{h_n}\right), \\ \hat{\gamma}_n(x) &= \frac{\theta_n}{nh_n} \sum_{i=1}^n \frac{Y_i}{G_n(Y_i)} K\left(\frac{x - X_i}{h_n}\right), \\ \text{et } \hat{\Xi}_j(x) &= \frac{\theta_n}{nh_n} \sum_{i=1}^n \frac{Y_i^{2-j}}{G_n^2(Y_i)} K\left(\frac{x - X_i}{h_n}\right), \quad j = 0, 1, 2.\end{aligned}$$

Ainsi, nous obtenons l'estimateur de  $\Delta^2(x)$  :

$$\hat{\Delta}_n^2(x) = \theta\nu_0 \left[ \hat{\Xi}_0(x)\hat{l}_n^2(x) + \hat{\Xi}_2(x)\hat{\gamma}_n^2(x) - 2\hat{l}_n(x)\hat{\gamma}_n(x)\hat{\Xi}_1(x) \right] \hat{l}_n^{-4}(x).$$

Alors, l'intervalles de confiance de  $m(X)$  au niveau asymptotique  $1 - \zeta$  :

$$\left[ \hat{m}_{NW}(x) - u_{1-\zeta/2} \hat{\Delta}_n(x)(nh_n)^{-1/2}, \hat{m}_{NW}(x) + u_{1-\zeta/2} \hat{\Delta}_n(x)(nh_n)^{-1/2} \right]$$

où  $u_{1-\zeta/2}$  dénote le quantile d'ordre  $1 - \zeta/2$  de la distribution normale standard.

Pour les données indépendantes, Ould-Saïd et Lemdani (2006)[34] ont prouvé que

$$\sqrt{nh_n}(\hat{m}_{NW}(x) - m(x)) \xrightarrow{P} N(0, \Gamma^2(x))$$

avec :

$$\Delta^2(x) = \theta^{-1}\nu_0 \left[ \Sigma_0(x)l^2(x) + \Sigma_2(x)\psi^2(x) - 2l(x)\psi(x)\Sigma_1(x) \right] l^{-4}(x).$$

Où  $\psi(x) = \int_{\mathbb{R}} yf(x, y)dy$  et  $\Sigma_j(x) = \int \frac{y^{2-j}}{G(y)}f(x, y)dy$  ( $j = 0, 1, 2$ ).

De toute évidence, 5.7 améliore le résultat de Ould-Saïd et Lemdani (2006)[34] et détend l'hypothèse *i.i.d* d'une condition de  $\alpha$ -mélange. En même temps, 5.7 donne aussi le biais asymptotique de l'estimateur dans le cas où  $p = 0$ .

## 5 Preuve des résultats principaux

Soient

$$\begin{aligned} t_{n,j}^* &= \frac{\theta}{n} \sum_{i=1}^n \left( \frac{X_i - x}{h_n} \right)^j K_{h_n}(X_i - x) G_n^{-1}(Y_i) (Y_i - m(X_i)) \\ t_{1n,j}^* &= \frac{\theta}{n} \sum_{i=1}^n \left( \frac{X_i - x}{h_n} \right)^j K_{h_n}(X_i - x) G^{-1}(Y_i) (Y_i - m(X_i)) \\ t_n^* &= (t_{n,0}^*, \dots, t_{n,p}^*)^t, \quad t_{1n}^* = (t_{1n,0}^*, \dots, t_{1n,p}^*)^t. \end{aligned}$$

**Lemme 5.0.1** Soit  $\alpha(n) = O(n^{-\gamma})$  pour  $\gamma > 3$ . Supposons que (A0) – (A4) est vérifié. Si  $nh_n \rightarrow \infty$ , et  $s_{n,j} \xrightarrow{P} l(x)u_j$  pour  $0 \leq j \leq 2p+1$ , de plus  $S_n \xrightarrow{P} l(x)S$ .

**Lemme 5.0.2** Supposant que (A0) – (A6) est vérifié. Alors  $nh_n \text{Cov}(t_{1n}^*) \rightarrow \theta \sigma^2(x)l(x)V$  où  $\text{Cov}(t_{1n}^*)$  est la matrice de covariance de  $t_{1n}^*$ .

**Lemme 5.0.3** Supposons que (A0) – (A5) est vérifié. Si une des conditions suivantes (i) ou (ii) se tient, donc  $\sqrt{nh_n}t_{1n}^* \xrightarrow{P} N(0, \theta \sigma^2(x)l(x)V)$ .

(i)  $\phi(\cdot)$  est une fonction bornée.

(ii)  $\alpha(n) = O(n^{-\gamma})$  pour certains  $\gamma \geq [r(r+2)]/[2(r-2)]$ ,  $nh_n^{1+4/r} \geq c_0 > 0$  et (A6) est vérifié.

**Lemme 5.0.4** Supposons que  $\alpha(n) = O(n^{-\gamma})$  pour certains  $\gamma > 3$ . Alors, sous (A0) nous avons  $\sup_{y \geq \alpha_F} |G_n(y) - G(y)| = O(n^{-1/2})$ .

### Preuve du théorème 4.0.1

En utilisant le développement de Taylor, nous avons

$$\begin{aligned} (m(X_1), \dots, m(X_n))^t &= \mathbf{X}\mathbf{B}(x) + \frac{m^{(p+1)}(x)}{(p+1)!} ((X_1 - x)^{p+1}, \dots, (X_n - x)^{p+1})^t \quad (5.8) \\ &\quad + (o(X_1 - x)^{p+1}, \dots, o(X_n - x)^{p+1})^t \end{aligned}$$

et

$$\begin{aligned} S_n^{-1}t_n^* &= S_n^{-1}t_n - \text{diag}(1, h_n, \dots, h_n^p) \mathbf{B}(x) \quad (5.9) \\ &\quad - \frac{h_n^{p+1}m^{(p+1)}(x)}{(p+1)!} \mathbf{S}_n^{-1}(s_{n,p+1}, \dots, s_{n,2p+1})^t + o(h_1^{p+1}) \mathbf{S}_n^{-1}(s_{n,p+1}, \dots, s_{n,2p+1})^t \\ &= \text{diag}(1, \dots, h_n^p) \left[ \hat{\mathbf{B}}(x) - \mathbf{B}(x) \right] \\ &\quad - \frac{h_n^{p+1}m^{(p+1)}(x)}{(p+1)!} \mathbf{S}_n^{-1}(s_{n,p+1}, \dots, s_{n,2p+1})^t + o(h_1^{p+1}) \mathbf{S}_n^{-1}(s_{n,p+1}, \dots, s_{n,2p+1})^t \end{aligned}$$

Notez que  $t_n^* - t_{1n}^* = (t_{n,0}^* - t_{1n,0}^*, \dots, t_{n,p}^* - t_{1n,p}^*)^t$ . En utilisant le lemme 5.4 et l'hypothèse (A0), nous avons :

$$\begin{aligned}
 & \sqrt{nh_n} |t_{n,j}^* - t_{1n,j}^*| \tag{5.10} \\
 \leq & \frac{\sqrt{nh_n} \sup_{y \geq a_{\hat{F}}} |G_n(y) - G(y)|}{G(a_{\hat{F}}) - \sup_{y \geq a_{\hat{F}}} |G_n(y) - G(y)|} \frac{\theta}{nh_n^{j+1}} \sum_{i=1}^n |X_i - x|^j G^{-1}(Y_i) \left| K \left( \frac{X_i - x}{h_n} \right) (Y_i - m(X_i)) \right| \\
 = & O_p(h_n^{1/2}) \frac{\theta}{nh_n^{j+1}} \sum_{i=1}^n \left| (X_i - x)^j K \left( \frac{X_i - x}{h_n} \right) (Y_i - m(X_i)) \right| G^{-1}(Y_i)
 \end{aligned}$$

Compte tenu de 5.2, à partir de (A1), (A3) et (A6) (ii), nous avons :

$$\begin{aligned}
 & E \left( \frac{\theta}{nh_n^{j+1}} \sum_{i=1}^n \left| (X_i - x)^j K \left( \frac{X_i - x}{h_n} \right) (Y_i - m(X_i)) \right| G^{-1}(Y_i) \right) \\
 = & \frac{\theta}{h_n^{j+1}} \int_{\mathbb{R}} \int_{\mathbb{R}} \left| (s - x)^j K \left( \frac{s - x}{h_n} \right) (y - m(s)) \right| G^{-1}(y) F^*(ds, dy) \\
 = & \int_{\mathbb{R}} \int_{\mathbb{R}} |t^j K(t) (y - m(x + h_n t))| f(x + h_n t, y) dt dy \\
 = & \int_{\mathbb{R}} |t^j K(t) l(x + h_n t) E(|y - m(x)| / X = x + h_n t) dt \\
 = & O(1).
 \end{aligned}$$

Et qui implique

$$\begin{aligned}
 & \theta (nh_n^{j+1})^{-1} \sum_{i=1}^n \left| (X_i - x)^j K \left( \frac{X_i - x}{h_n} \right) (Y_i - m(X_i)) \right| G^{-1}(Y_i) \\
 = & O(1).
 \end{aligned}$$

Alors

$$\sqrt{nh_n} |t_{n,j}^* - t_{1n,j}^*| = O(h_n^{1/2})$$

En outre et depuis lemme 5.1 nous avons

$$\sqrt{nh_n} S_n^{-1} t_n^* = \sqrt{nh_n} S_n^{-1} t_n^* + O(h_n^{1/2}) \tag{5.11}$$

Par conséquent, compte tenu des lemmes 5.0.1 et 5.0.3, à partir de 5.9, 5.11 et  $nh_n^{2p+3} = O(1)$ , nous obtenons

$$\begin{aligned} & \sqrt{nh_n} \left( \text{diag}(1, \dots, h_n^p) \left[ \hat{\mathbf{B}}_n(x) - \mathbf{B}(x) \right] - \frac{h_n^{p+1} m^{(p+1)}(x)}{(p+1)!} S^{-1} U \right) \\ &= \sqrt{nh_n} S_n^{-1} t_{1n}^* + O(h_n^{1/2}) + o((h_n^{2p+3})^{1/2}) \xrightarrow{D} N(0, \theta \sigma^2(x) l(x)^{-1} \mathbf{S}^{-1} \mathbf{V} \mathbf{S}^{-1}) \end{aligned}$$

### Preuve de théorème 4.0.2

on note que :

$$\begin{aligned} & \text{diag}(1, \dots, h_n^p) \left[ \hat{\mathbf{B}}(x) - \mathbf{B}(x) \right] \\ &= \text{diag}(1, \dots, h_n^p) \left( \hat{m}_{LP}(x) - m(x), \dots, \frac{\hat{m}_{LP}^{(v)}(x) - m^{(v)}(x)}{(v)!}, \dots, \frac{\hat{m}_{LP}^{(p)}(x) - m^{(p)}(x)}{(p)!} \right)^t \\ &= \left( \hat{m}_{LP}(x) - m(x), \dots, \frac{h_n^v}{(v)!} \left( \hat{m}_{LP}^{(v)}(x) - m^{(v)}(x) \right), \dots, \frac{h_n^p}{(p)!} \left( \hat{m}_{LP}^{(p)}(x) - m^{(p)}(x) \right) \right)^t \end{aligned}$$

Par conséquent, la conclusion est obtenue d'après le théorème 4.0.1.

# Chapitre

# 6

## Etude de Simulation

On a programmé à l'aide de langage Matlab **R2012a**.

### 1 L'estimateurs $\hat{m}_{NW}$

Dans cette section, nous calculons l'erreur quadratique moyenne globale de l'estimateur NW, et nous étudions leur ajustement graphique à la vraie courbe sous-jacente.

Nous avons choisi la distribution marginale de  $X_i$  est l'exponentielle avec moyenne (1). Avant la troncature, la taille initiale de l'échantillon était  $N$

Étant donné la valeur de la covariable,  $X_i = x_i (i = 1, \dots, n)$ , la réponse  $Y_i$  a été générée selon une variable exponentielle de moyenne  $1 + \sin(2\pi x_i/x_{max})$ , où  $x_{max}$  est le centile 99% de modèle exponentiel (1). Par conséquent, la vraie fonction de régression est

$$m(x) = E(Y|X = x) = 1 + \sin(2\pi x/x_{max})$$

mais l'observable n'est pas celui-ci et elle doit être estimée en utilisant les corrections données pour la troncature à gauche.

La variable de troncature  $T$  a été générée de façon indépendante selon une distribution exponentielle avec une moyenne  $t_0$ . Le paramètre  $t_0$  a été choisi afin d'assurer des niveaux différents de troncature dans l'échantillon final :

Quand  $Y_i \geq T_i$  le vecteur  $(X_i, Y_i, T_i)$  a été inclus dans l'échantillon final. Cette procédure est répétée jusqu'à ce qu'un échantillon final avec la taille  $n$  a été obtenu, avec  $n = 100, 300$  et  $500$ .

La fenêtre de lissage  $h_n$  est calculé par

$$h_n = \left( \frac{\log n}{n} \right)^{\frac{1}{5}}$$

$n$	$h_n$
100	0.54
300	0.45
500	0.41

Rappelons l'estimateur de *NW* de  $m(x)$  sous le modèle de troncature à gauche :

$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^n K_h(X_i - x) G_n^{-1}(Y_i) Y_i}{\sum_{i=1}^n K_h(X_i - x) G_n^{-1}(Y_i)}$$

Où l'estimateur de maximum de vraisemblance de  $G$  est défini par :

$$G_n(y) = \prod_{\substack{i=1 \\ T_i > y}}^n \left( 1 - \frac{1}{nC_n(T_i)} \right)$$

et l'estimateur empirique de  $C(y)$  est défini par :

$$C_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq y \leq Y_i\}}$$

Le *MSE* est défini comme suit

$$MSE = \frac{1}{n} \sum_{k=1}^n [\hat{m}(X_k, l) - m(X_k, l)]^2$$

Basant sur un noyau gaussien, nous calculons pour  $\hat{m}$  estimateur de  $m$ , l'erreur quadratique moyenne globale (*GMSE*) le long de  $M = 1000$  essais de Monte Carlo. le *GMSE* est défini comme suit

$$GMSE = \frac{1}{Mn} \sum_{l=1}^M \sum_{k=1}^n [\hat{m}(X_k, l) - m(X_k, l)]^2$$

Dans le tableau suivant, nous allons représentés la valeur minimale de *GMSE* et l'écart type *Sd* de l'estimateur de Nadaraya Watson. On remarque que le *GMSE* diminue à mesure que la taille de l'échantillon augmente, mais, le *GMSE* est en croissance quand la proportion de troncature augmente.

$N$	$n$	$\theta$	$h$	$GMSE$	$SD$
100	91	0.91	0.54	0.137	0.063
300	278	0.92	0.45	0.0798	0.038
500	472	0.94	0.41	0.058	0.025

Tableau 1 : Les valeurs minimales de  $GMSE$  de l'estimateur de Nadaraya Watson le long de 1000 essais de Monte-Carlo, pour plusieurs taux de troncature ( $t_0$ ). Les écarts-types ( $Sd$ ) sont également calculés.

Les trois figures suivante montre que lorsque on fixe le taux de troncature, par exemple à 0.9, et on augmente la taille de l'échantillon ( $n = 100, 300, 500$ ), l'erreur quadratique moyenne généralisé  $GMSE$  se diminue.

## 2 Estimation de la densité des données tronquées

Dans cette partie de simulation, on trace les estimateurs à noyau de la densité dans le cas des données complètes et le cas des données tronquées à gauche, et on compare ces deux estimateurs avec la densité normale centrée réduite.

On simule, dans un premier temps,  $N$  valeurs  $(X_i, T_i), i = 1, \dots, N$  du couple de variables aléatoires  $(X, T)$  avec  $T$  indépendante de  $X$ .

Soit  $T$  la variable de troncature, et  $X$  une variables aléatoires de loi normale centrée réduite. Avant la troncature, la taille initiale de l'échantillon était  $N$

Quand  $X_i \geq T_i$  le vecteur  $(X_i, T_i)$  a été inclu dans l'échantillon final. Cette procédure est répétée jusqu'à ce qu'un échantillon final avec la taille  $n$  a été obtenu, avec  $n = 100, 300$  et 500.

Rappelons l'estimateur à noyau de la densité de la variable  $X$  dans le cas des données complètes

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$$

Dans le cas des données tronquées à gauche, il est définit comme suit

$$f_n(x) = \frac{\theta_n}{nh} \sum_{i=1}^n \frac{1}{G_n(X_i)} K\left(\frac{X_i - x}{h}\right)$$

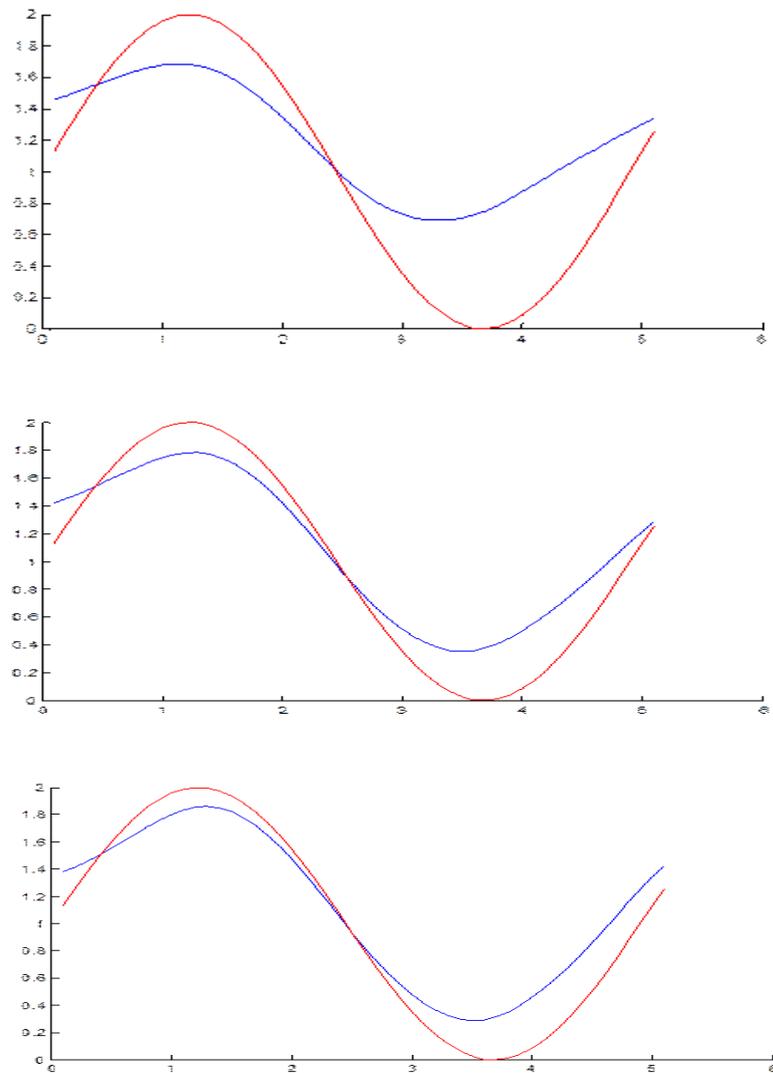


FIG. 9: La fonction de régression (rouge), l'estimateur Nadaraya-Watson (bleu), en moyenne sur 1000 essais de Monte Carlo. Le taux de la troncature est de 0.9, et la taille de l'échantillon est  $n = 100, 300,$  et  $500$  respectivement.

Où l'estimateur de maximum de vraisemblance de  $G$  (Lynden-Bell) est défini par :

$$G_n(y) = \prod_{\substack{i=1 \\ T_i > y}}^n \left(1 - \frac{1}{nC_n(T_i)}\right)$$

et l'estimateur empirique de  $C(y)$  est défini par :

$$C_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{T_i \leq y \leq Y_i\}}$$

Basant sur un noyau gaussien, nous calculons pour  $f_n$  estimateur de  $f$ , l'erreur quadratique moyenne  $MSE$ . Elle est défini comme suit

$$MSE = \frac{1}{n} \sum_{k=1}^n [f_n(X_k, l) - f(X_k, l)]^2$$

Les deux tableaux suivants représentent l'erreur quadratique moyenne  $MSE$  de l'estimateur à noyau de la densité dans le cas des données complètes et le cas des données tronquées à gauche. On à utilisé la fenêtre de lissage  $h = N^{(-1/5)}$  (cas complet) et  $h_1 = \left(\frac{\log n}{n}\right)^{\frac{1}{5}}$  (cas tronqué).

Pour  $\theta \simeq 0.5$ , la distribution de  $Y$  et  $T$  est la  $N(0, 1)$

$N$	$n$	$\theta$	$h$	$h_1$	$MSE(\text{cas complet})$	$MSE(\text{cas tronqué})$
100	49	0.49	0.39	0.6	$3.68 \times 10^{-4}$	0.0023
300	150	0.5	0.31	0.5	$2.46 \times 10^{-4}$	0.0013
500	254	0.5	0.29	0.46	$1.15 \times 10^{-4}$	$6.94 \times 10^{-4}$

Pour  $\theta \simeq 0.75$ , la distribution de  $Y$  est la  $N(1, 1)$  et  $T$  est la  $N(0, 1)$ .

$N$	$n$	$\theta$	$h$	$h_1$	$MSE(\text{cas complet})$	$MSE(\text{cas tronqué})$
100	77	0.77	0.39	0.56	0.0019	0.0024
300	218	0.72	0.31	0.47	$3.54 \times 10^{-4}$	$6.58 \times 10^{-4}$
500	384	0.76	0.28	0.43	$3.22 \times 10^{-4}$	$3.54 \times 10^{-4}$

A travers, les résultats obtenus représentés graphiquement par les figures suivantes, et le calcul de l'  $MSE$  on remarque que notre estimateur à noyau de la densité dans le cas des données tronquées à gauche avec  $\theta \simeq 0.5$  est proche de la densité exacte  $N(0, 1)$ , on dit qu'on a une bonne approximation pour la loi normale.

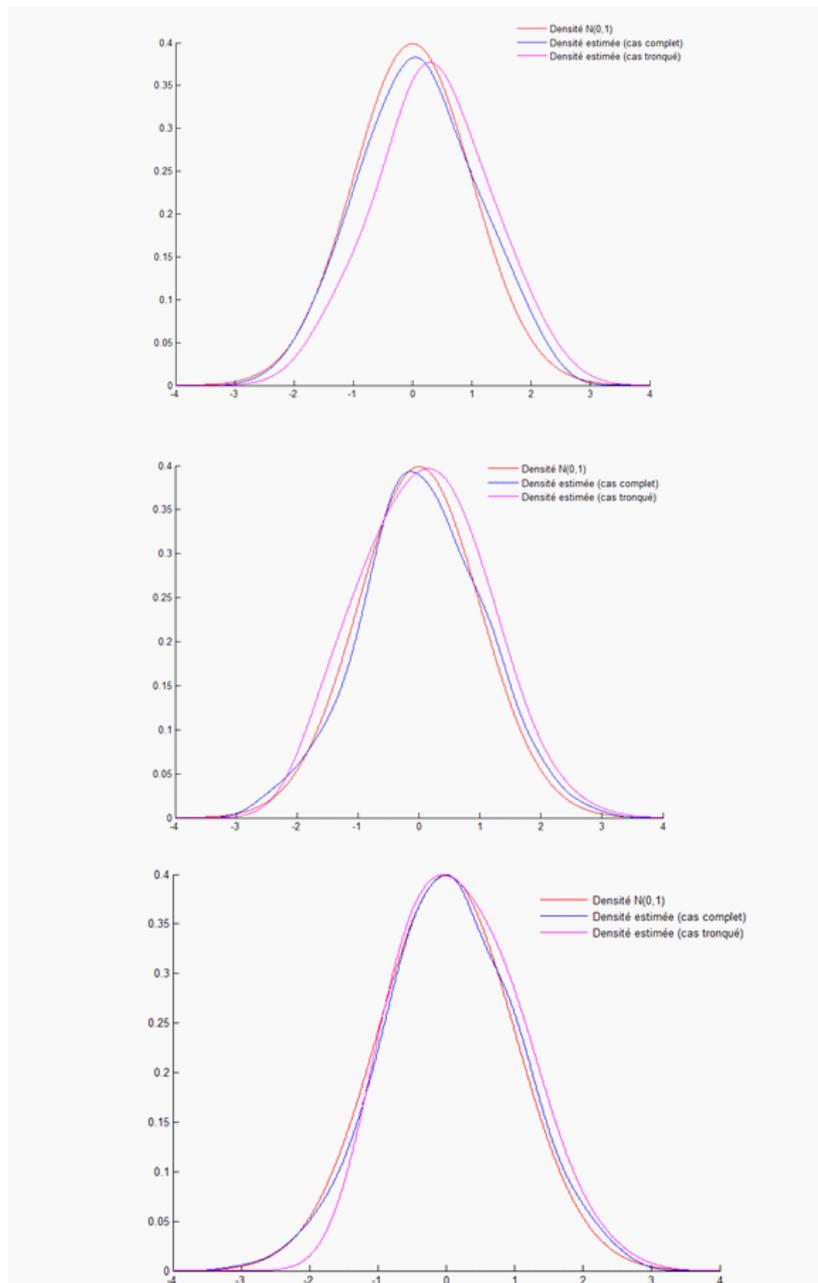


FIG. 10:  $N=100, 300, 500$  et  $\theta=0.5$

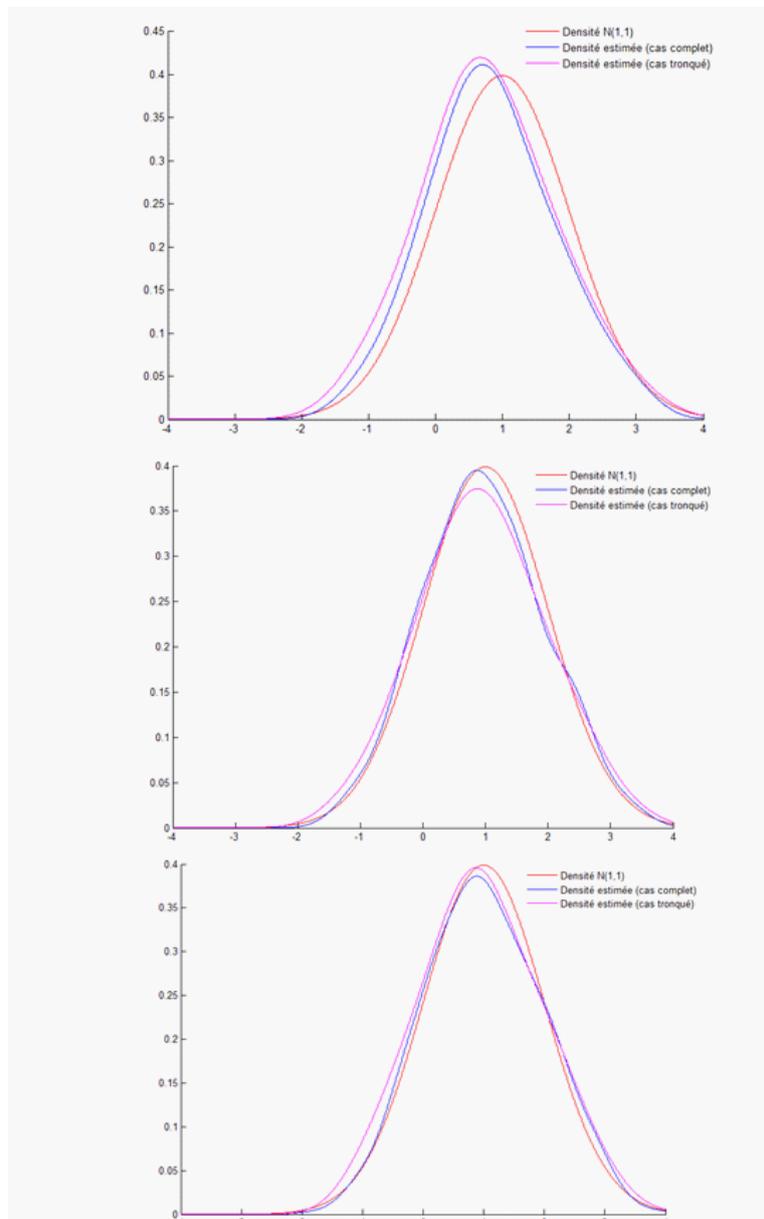


FIG. 11:  $N=100, 300, 500$  et  $\theta = 0.75$ .

# Conclusion

Dans ce mémoire nous avons essayé de faire une synthèse des résultats existant concernant la régression non-paramétrique par la méthode à noyau sur des données complètes et tronquées. Nous avons étudié quelque estimateur à noyau de la fonction de régression, comme l'estimateur de Nadaraya Watson, l'estimateur de Gasser et Muller et la régression par les polynômes locaux.

L'estimation de la fonction de régression des données tronquées dépendantes par la méthode des polynômes locaux est notre but de cette étude, donc a étudié les estimateurs de la fonction de régression et de ses dérivées,

Dans la partie simulation, nous avons calculé les estimateurs à noyau de la densité dans le cas des données complètes et le cas des données tronquées à gauche, et on compare ces deux estimateurs avec la densité normale centrée réduite. De plus, nous avons calculé l'erreur quadratique moyenne globale de l'estimateur NW, et nous étudions leur ajustement graphique à la vraie courbe sous-jacente.

# Appendice

## 1 Loi forte des grands nombres (strong law of large numbers)

**Théorème.** On considère une suite  $(X_n)$  de variables aléatoires (réelles) indépendantes et identiquement distribuées, d'espérance mathématique  $\mu$ . On définit les moyennes :

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Alors la probabilité de l'évènement  $X_n \xrightarrow[n \rightarrow \infty]{} \mu$  est égale à 1.

C'est très exactement la définition de la convergence presque sûre de la suite  $(X_n)$  vers (la v.a certaine)  $\mu$  lorsque  $n$  tend vers l'infini.

## 2 Méthode des moindres carrés

Considérons le modèle général de régression linéaire multiple :

$$Y_i = \beta_0 + \sum_{j=1}^{p-1} X_{ij} \cdot \beta_j + \varepsilon_i, \quad i = 1, \dots, n$$

où

$Y_i$  est la variable dépendante,

$X_{ji}, j = 1, \dots, p - 1$ , sont les variables indépendantes,

$\beta_j, j = 0, \dots, p - 1$ , sont les paramètres à estimer et

$\varepsilon_i$  est le terme d'erreur aléatoire non observable.

L'estimation des paramètres de la fonction par la méthode des moindres carrés consiste à minimiser la somme des carrés des erreurs :

$$\min_{\beta_j} \sum_{i=1}^n \varepsilon_i^2$$

La somme des carrés des erreurs étant égale à :

$$f(\beta_0, \dots, \beta_{p-1}) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left( Y_i - \beta_0 - \sum_{j=1}^{p-1} X_{ij} \cdot \beta_j \right)^2$$

Nous trouvons les paramètres estimés de la fonction en annulant les dérivées partielles de cette fonction par rapport à chaque paramètre :

$$\begin{aligned} \frac{\partial f(\beta_0, \dots, \beta_{p-1})}{\partial \beta_0} &= 0, \\ \frac{\partial f(\beta_0, \dots, \beta_{p-1})}{\partial \beta_1} &= 0, \\ &\vdots \\ \frac{\partial f(\beta_0, \dots, \beta_{p-1})}{\partial \beta_{p-1}} &= 0. \end{aligned}$$

Supposons que  $p = 2$  et que nous disposions de  $n$  observations :  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ . Alors,

$$\begin{aligned} \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \cdot \bar{X} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{aligned}$$

Avec

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

L'estimateur de moindres carrés de  $\hat{Y}_i$  est :

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i, i = 1, \dots, n$$

Nous pouvons également exprimer le modèle de régression linéaire multiple sous forme matricielle :

$$Y = X \cdot \beta + \varepsilon$$

où

$Y$  : est le vecteur ( $n \times 1$ ) des observations relatives à la variable dépendante ( $n$  observations),

$X$  : est la matrice ( $n \times p$ ) des variables indépendantes,

$\varepsilon$  : est le vecteur ( $n \times 1$ ) des erreurs et

$\beta$  : est le vecteur ( $p \times 1$ ) des paramètres à estimer.

L'application de la méthode des moindres carrés consiste à minimiser :

$$\begin{aligned}\varepsilon^t \varepsilon &= (Y - X.\beta)^t (Y - X.\beta) \\ &= Y^t.Y - \beta^t.X^t.Y - Y^t.X.\beta + \beta^t.X^t.X.\beta \\ &= Y^t.Y - 2\beta^t.X^t.Y + \beta^t.X^t.X.\beta.\end{aligned}$$

Les équations sont données matriciellement par :

$$X^t.X.\hat{\beta} = X^t.Y$$

Si  $X^t.X$  est inversible, les paramètres  $\beta$  sont donc estimés par :

$$\hat{\beta} = (X^t.X)^{-1}.X^t.Y.$$

### 3 $O(\cdot)$ et $o(\cdot)$ stochastiques

Pour deux suites de v.a  $\{U_n\}$  et  $\{V_n\}$ , la notation  $U_n = O(V_n)$  signifie que  $\{U_n/V_n\}$  est  $O(1)$ . De plus,  $U_n = o(V_n)$  signifie que  $U_n/V_n \xrightarrow{L} 0$ .  $U_n = o(V_n) \Rightarrow U_n = O(V_n)$ .

### 4 Démonstration de $S_{n,j}$

Soit  $A_n$  v.a, Alors  $A_n = E(A_n) + O\left(\sqrt{\text{Var}(A_n)}\right)$ .

Définition

$T_n = O(R_n)$  signifie que  $T_n = Y_n R_n$  avec la famille  $(Y_n)_n$  uniformément tendue (i.e  $Y_n \xrightarrow{L} y$ ).

Théorème central limite

$Y_n = \frac{A_n - E(A_n)}{\sqrt{\text{Var}(A_n)}} \xrightarrow{L} y$  ça implique que

$$\begin{aligned}A_n - E(A_n) &= Y_n \sqrt{\text{Var}(A_n)} \\ &= O\left(\sqrt{\text{Var}(A_n)}\right)\end{aligned}$$

Rappelant la définition 2.13, on note

$$\begin{aligned}
 S_{n,j} &= E(S_{n,j}) + O\left(\sqrt{\text{Var}(S_{n,j})}\right) \\
 &= nh^j \int u^j K(u) f(x+uh) du + O\left(\sqrt{n \left[ E(X_1 - x)^{2j} K_h^2(X_1 - x) \right]}\right) \\
 &= nh^j \left[ f(x)\mu_j + o(1) + O\left(1/\sqrt{nh}\right) \right] \\
 &= nh^j \left[ f(x)\mu_j + o(1) + o(1) \right] \\
 &= nh^j f(x)\mu_j + nh^j o(1) \\
 &= nh^j f(x)\mu_j \left[ 1 + \frac{1}{f(x)\mu_j} o(1) \right] \\
 &= nh^j f(x)\mu_j \{1 + o(1)\}
 \end{aligned}$$

Défini dans 2.17. On utilise le même argument on trouve

$$S_{n,j}^* = nh^{j-1} f(x)\sigma^2(x)\nu_j \{1 + o(1)\}$$

donc

$$S_n^* = nh^{-1} f(x)\sigma^2(x)HS^*H \{1 + o(1)\} \quad (**)$$

## 5 Construction des noyaux d'ordre élevés

Le noyau  $w_\nu$  est appelé noyau d'ordre  $(\nu, p)$  s'il vérifie :

$$\int_{-\tau}^{\tau} x^j w_\nu(x) dx = \begin{cases} 0 & j \in \{0, \dots, p-1, p+1, \dots, k-1\} \\ (-1)^\nu \cdot \nu! & j = \nu \\ \beta_p \neq 0 & j = p. \end{cases}$$

Le support de  $w_\nu$  est  $[-\tau, \tau]$ .

Nous rappelons que le noyau d'ordre  $(\nu, p)$  est utilisé pour l'estimation de la  $\nu^{\text{ème}}$  dérivée de la fonction de régression.

## 6 Démonstration du théorème de Ruppert et Wand (1994)

On commence par 2.9, et en utilisant 2.17 et (\*)

$$\begin{aligned}
\text{Var}(\hat{\beta}|X) &= S_n^{-1} S_n^* S_n^{-1} \\
&= \{nf(x)HSH [1 + o(1)]\}^{-1} \{nh^{-1}f(x)\sigma^2(x)HS^*H [1 + o(1)]\} \\
&\quad \times \{nf(x)HSH [1 + o(1)]\}^{-1} \\
&= \frac{nf(x)\sigma^2(x)}{nhf(x)nf(x)} (HSH)^{-1} (HS^*H) (HSH)^{-1} \frac{[1 + o(1)]}{[1 + o(1)][1 + o(1)]} \\
&= \frac{\sigma^2(x)}{nhf(x)} H^{-1} S^{-1} S^* S^{-1} H^{-1} [1 + o(1)]
\end{aligned}$$

on a  $\hat{m}_\nu(x) = \nu! e_{\nu+1}^t \hat{\beta}$  et  $H^{-1}H^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & h^{-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h^{-2p} \end{pmatrix}_{(p+1) \times (p+1)}$

alors

$$\text{Var}(\hat{m}_\nu(x)|X) = \frac{\nu!^2 \sigma^2(x)}{nh^{2\nu+1} f(x)} e_{\nu+1}^t S^{-1} S^* S^{-1} e_{\nu+1} [1 + o(1)]$$

Concernant le biais, Lorsque  $p - \nu$  est impair, le biais conditionnel peut s'écrire :

$$\begin{aligned}
\mathbb{E}(\hat{\beta}/X) &= (X^t W X)^{-1} X^t W \mathbb{E}(Y|X) \\
&= (X^t W X)^{-1} X^t W m \quad \text{car } \mathbb{E}(Y|X) = X\beta + \mathbb{E}(\varepsilon|X) \\
&= (X^t W X)^{-1} X^t W (X\beta + r) \quad \text{car } m = X\beta + r \\
&= (X^t W X)^{-1} X^t W X\beta + (X^t W X)^{-1} X^t W r \\
&= \beta + S_n^{-1} X^t W r.
\end{aligned}$$

où

$$\begin{aligned}
r &= m - X\beta \\
&= m - \sum_{j=0}^p \beta_j (X_i - x)^j
\end{aligned}$$

et

$$m = \{m(X_1), \dots, m(X_n)\}$$

et

$$\begin{aligned}\beta &= \{\beta_0, \dots, \beta_p\} \\ &= \left\{ m(x), \dots, \frac{m^{(p)}(x)}{p!} \right\}\end{aligned}$$

Pour pouvoir faire le développement de Taylor, on suppose l'existence de la  $(p+2)^{\text{éme}}$  dérivée de la fonction de régression  $m(\cdot)$  au point  $x$ , alors pour  $X_i$  appartenant à un voisinage de  $x$  on a

$$\begin{aligned}m(X_i) &\approx m(x) + m'(x)(X_i - x) + \frac{m''(x)}{2}(X_i - x)^2 + \dots + \frac{m^{(p)}(x)}{p!}(X_i - x)^p \\ &\quad + \frac{m^{(p+1)}(x)}{(p+1)!}(X_i - x)^{p+1} + o(X_i - x)^{p+1} \\ &\approx \sum_{j=0}^{p+1} \frac{m^{(j)}(x)}{j!}(u - x)^j + o(X_i - x)^{p+1} \\ &= \sum_{j=0}^{p+1} \beta_j(u - x)^j + o(X_i - x)^{p+1}\end{aligned}$$

$$\begin{aligned}r_i &= m(x_i) - \sum_{j=0}^p \beta_j(X_i - x)^j \\ &= \beta_{p+1}(X_i - x)^{p+1} + o(X_i - x)^{p+1}, i = 1, \dots, n\end{aligned}$$

alors,

$$\begin{aligned}\mathbb{E}(\widehat{\beta}|X) &= \beta + S_n^{-1} X^t W r. \\ &= \beta + S_n^{-1} X^t W [\beta_{p+1}(X_i - x)^{p+1} + o(X_i - x)^{p+1}]_{i=1, \dots, n} \\ &= \beta + S_n^{-1} [\beta_{p+1} \cdot c_n + o(nh^{p+1})]\end{aligned}$$

où

$$\begin{aligned}
 c_n &= X^t W [(X_i - x)^{p+1}]_{i=1, \dots, n} \\
 &= \begin{pmatrix} 1 & \cdots & 1 \\ (X_1 - x) & \cdots & (X_n - x) \\ \vdots & \ddots & \vdots \\ (X_1 - x)^p & \cdots & (X_n - x)^p \end{pmatrix}_{(p+1) \times n} \times \begin{pmatrix} K_h(X_1 - x) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & K_h(X_n - x) \end{pmatrix}_{n \times n} \\
 &\quad \times \begin{pmatrix} (X_1 - x)^{p+1} \\ \vdots \\ (X_n - x)^{p+1} \end{pmatrix}_{n \times 1} \\
 &= \begin{pmatrix} \sum_{i=1}^n (X_i - x)^{p+1} K_h(X_i - x) \\ \sum_{i=1}^n (X_i - x)^{p+2} K_h(X_i - x) \\ \vdots \\ \sum_{i=1}^n (X_i - x)^{2p+1} K_h(X_i - x) \end{pmatrix}_{(p+1) \times 1} \\
 &= \begin{pmatrix} S_{n,p+1} \\ S_{n,p+2} \\ \vdots \\ S_{n,2p+1} \end{pmatrix}_{(p+1) \times 1}
 \end{aligned}$$

en effet,

$$\begin{aligned}
 \text{Biais}(\widehat{\beta}|X) &= \mathbb{E}(\widehat{\beta}|X) - \beta \\
 &= S_n^{-1} [\beta_{p+1} \cdot c_n + o(nh^{p+1})]
 \end{aligned}$$

on remplace 2.16 et 2.17 dans l'expression de biais, alors le biais asymptotique est

$$\begin{aligned}
 \text{Biais}(\widehat{\beta}|X) &= H^{-1} S^{-1} c_p \beta_{p+1} h^{p+1} [1 + o(1)] \\
 &= H^{-1} S^{-1} c_p \frac{m^{(p+1)}(x)}{(p+1)!} h^{p+1} [1 + o(1)]
 \end{aligned}$$

avec  $c_p = (\mu_{p+1}, \dots, \mu_{2p+1})^t$ .

$$\begin{aligned}
E(m^{(\nu)}(x)|X) &= E(e_{\nu+1}^t \widehat{\beta} \nu! | X) \\
&= e_{\nu+1}^t \nu! E(\widehat{\beta} | X) \\
&= e_{\nu+1}^t H^{-1} S^{-1} c_p \frac{\nu!}{(p+1)!} m^{(p+1)}(x) h^{p+1} [1 + o(1)] \\
&= e_{\nu+1}^t S^{-1} c_p \frac{\nu!}{(p+1)!} m^{(p+1)}(x) h^{p+1-\nu} [1 + o(1)]
\end{aligned}$$

## Références

- [1] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. Springer, New York.
- [2] Bradley, R. C. (2005). *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions. Probability Surveys*. Vol. 2 107-144. ISSN : 1549-5787. DOI : 10.1214/154957805100000104.
- [3] Cai, J. and Kim, J. (2003). *Nonparametric quantile estimation with correlated failure time data*. *Lifetime Data Analysis* 9, 357-371.
- [4] Chu, C.K., and Marron, J.S. (1991). *Choosing a kernel regression estimator (with discussions)*. *Statist. Sci.*, 6, 404-436.
- [5] Cleveland, W.S. (1979). *Robust locally weighted regression and smoothing scatterplots*. *J. Amer. Statist. Assoc.*, 74, 829-836.
- [6] de Uña-Álvarez J, Liang HY, Rodríguez-Casal A (2010) *Nonlinear wavelet estimator of the regression function under left truncated dependent data*. *J Nonparametr Stat* 22(3) :319–344
- [7] Fan, J. (1992). *Design-adaptive nonparametric regression*. *J. Amer. Statist. Assoc.*, 87, 998-1004.
- [8] Fan, J. (1993a). *Local linear regression smoothers and their minimax efficiency*. *Ann. Statist.*, 21, 196-216.
- [9] Fan, J. and Gijbels, I. (1992). *Variable bandwidth and local linear regression smoothers*. *Ann. Statist.*, 20, 2008-2036.
- [10] Fan, J. and Gijbels, I. (1995b). *Adaptive order polynomial fitting : bandwidth robustification and bias reduction*. *J. Compo Graph. Statist.*, 4, 213-227.
- [11] Fan, F. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statist. and Appl. Probab., 66, Chapman and Hill, London.
- [12] Fan, F. and Gijbels, I. (1997). *Local polynomial regression : Optimal kernels and asymptotic minimax efficiency*. *Ann.Inst.Statist. Math*. Vol. 49, N°.1, 79-99.
- [13] Fan J, Yao Q (2003) *Nonlinear time series*. Springer, New York
- [14] Feigelson ED, Babu GJ (1992) *Statistical challenges in modern astronomy*. Springer, Berlin

- 
- [15] Gasser, T. , Müller, H.-G. and Mammitzsch, V. (1985). Kernels for non parametric curve estimation. *J. Royal Statist. Soc. B*, 47, 238-252.
- [16] Gasser, T. and Müller, H.-G. (1979). Kernel estimation of regression functions. In Smoothing Techniques for Curve Estimation, *Lecture Notes in Mathematics*, 757, 23-68. Springer-Verlag, New York.
- [17] Glad, I.K., Hjort, N.L., and Ushakov, N.G. (2007), Mean-Squared Error of Kernel Estimators for Finite Values of the Sample Size. *Journal of Mathematical Sciences*, 146(4), 5977–5983.
- [18] Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. *Theory Probab Appl* 22, 411– 413.
- [19] Guessoum, Z. (2009). Régression non paramétrique dans les modèles censurés. Thèse d'Etat N° .01 -D/M.T.
- [20] Györfi L, Härdle W, Sarda P, Vieu P (1989) Nonparametric curve estimation from time series. In : Lecture notes in statistics, vol 60. Springer, New York
- [21] Hastie, T.J. and Loader, C. (1993). Local regression : automatic kernel carpentry (with discussion). *Statist. Sci.*, 8, 120-143.
- [22] He S, Yang G (1994) Estimating a lifetime distribution under different sampling plan. In : Gupta SS, Berger JO (eds) Statistical decision theory and related topics., vol 5. Springer, Berlin, pp 73–85
- [23] He, S. and Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, 26, 1011–1027.
- [24] Keiding, N. and Gill, R. D. (1990). Random truncation models and Markov processes. *Ann.Statist.* 18:582–602.
- [25] Klein, J. P. and Moeschberger, M. L. (1997). *Survival analysis : techniques for censored and truncated data*. Springer-Verlag, New York.
- [26] Liang, H.Y. , Li, D. , Qi, YC. (2009). Strong convergence in nonparametric regression with truncated dependent data. *J Multivar Anal* 100 :162–174.
- [27] Liang, H.Y. de Uña-Álvarez, J. Iglesias-Pérez, M.C.(2011) Local polynomial estimation of a conditional mean function with dependent truncated data. *Test* 20 653–77. DOI 10.1007/s 11749-011-0234-6
- [28] Lynden-Bell. D (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Mon Not R Astron Soc Lett* 155 :95–118.
- [29] Lejeune, M.(1985) *Statistiques, la théorie et ses applications*. France 2004 – 2005.
- [30] Mack, Y.P. and Müller, H.-G. (1989). Convolution type estimators for nonparametric regression estimation. *Statist. Probe Lett.* 7, 229-239.

- 
- [31] Marron, J.S., and Wand, M.P. (1992). Exact Mean Integrated Squared Error. *Annals of Statistics*, 20, 712–736.
- [32] Masry, E. , Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand J Stat* 24 :165–179
- [33] Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *J. Amer. Statist. Assoc.*, 82, 231-238.  
Müller, H.-G. (1988). Nonparametric Regression Analysis of Longitudinal Data. *Lecture Notes in Statistics*, 46. Springer-Verlag, Berlin.
- [34] Ould-Saïd ,E. Lemdani, M. (2006) Asymptotic properties of a non parametric regression function estimator with randomly truncated data, *Ann. Statist. Mat.* Vol 58, 357-378.
- [35] Priestley, M.B. and Chao, M.T. (1972). Nonparametric function fitting. *J. Royal Statist. Soc. B*, 34, 384-392.
- [36] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, *Proc. Natl. Acad. Sci. USA* 42, 43-47.
- [37] Ruppert, D. and Wand, M.P. (1994). Multivariate weighted least squares regression. *Ann. Statist.*
- [38] Stone, C.J. (1977). Consistent nonparametric regression. *Ann. Statist.*, 5, 595-645.
- [39] Stone, C.J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8, 1348-1360.
- [40] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10, 1040-1053.
- [41] Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Ann Stat* 21 :146-156.
- [42] Tatachak. A. Préviation non paramétrique dans les modèles de troncature via l'estimation du mode. Thèse d'Etat N° d'ordre : 05/2008 - E/MT.
- [43] Tsybakov, A. B. (1986). Robust reconstruction of functions by the local approximation method. *Problems of Information Transmission*, 22, 133- 146.
- [44] Tsybakov, A. B. (2003). *Introduction à l'estimation non-paramétrique*. Mathématiques & Applications,41.Springer.
- [45] Tsai WY, Jewell NP, Wang MC (1987) A note on the product-limit estimator under right censoring and left truncation. *Biometrika* 74 :883–886
- [46] Ushakov, N. G. and Ushakov, V. G. (2012). On bandwidth selection in kernel density estimation. *Journal of Nonparametric Statistics*, DOI :10.1080/10485252.2012.655734.

- [47] Wang MC, Jemell NP, Tsai WY (1986) Asymptotic properties of the product-limit estimate under random truncation. *Ann Stat* 14 :1597–1605
- [48] Withers, C.S. (1981). Conditions for linear processes to be strong mixing. *Z Wahrscheinlichkeitstheor Verw Geb* 57 :477–480
- [49] Woodroffe.M (1985). Estimating a distribution function with truncated data. *Ann Stat* Vol. 13, No. 1, 163-177.