Estimation non paramétrique de la fonction de régression pour des données censurées : méthodes locale linéaire et erreur relative

Feriel Bouhadjera

▶ **To cite this version:**

Feriel Bouhadjera. Estimation non paramétrique de la fonction de régression pour des données censurées : méthodes locale linéaire et erreur relative. Statistiques [math.ST]. Université du Littoral Côte d'Opale; Université Badji Mokhtar-Annaba, 2020. Français. NNT : 2020DUNK0561 . tel-03134914

HAL Id: tel-03134914
https://theses.hal.science/tel-03134914v1

Submitted on 8 Feb 2021

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

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

Université du Littoral Côte d'Opale
Université Badji Mokhtar Annaba

École doctorale **ED Régionale SPI 72**
Unité de recherche **LMPA Joseph Liouville**

Thèse présentée par **Feriel Bouhadjera**

Soutenue le **15 décembre 2020**

En vue de l'obtention du grade de docteur de l'Université du Littoral Côte d'Opale et de l'Université Badji Mokhtar Annaba

Discipline **Mathématiques et leurs interactions**
Spécialité **Statistique mathématique**

Estimation non paramétrique de la fonction de régression pour des données censurées : méthodes locale linéaire et erreur relative

Thèse dirigée par	Elias Ould Saïd	directeur
	Mohamed Riad Remita	co-directeur

Composition du jury

Rapporteurs	Zohra Guessoum	Pr. à l'U.S.T.H.B. d'Alger	
	Mustapha Rachdi	Pr. à l'U.G.A. de Grenoble	
Examinateurs	Célestin Kokonendji	Pr. à l'U.B.F.C. de Besançon	président du jury
	Hacène Boutabia	Pr. à l'U.B.M.A. d'Annaba	
Invité	Ouafae Benrabah	M.C.F. à l'U.L.C.O. de Calais	
Directeurs de thèse	Elias Ould Saïd	Pr. à l'U.L.C.O. de Calais	
	Mohamed Riad Remita	Pr. à l'U.B.M.A. d'Annaba	

L'Université du Littoral Côte d'Opale et l'Université Badji Mokhtar Annaba n'entendent donner aucune approbation ni improbation aux opinions émises dans les thèses : ces opinions devront être considérées comme propres à leurs auteurs.

Mots clés : convergence uniforme presque sûre ; données α-mélangeante ; données incomplètes ; fonction de régression ; erreur relative ; normalité asymptotique ; méthode linéaire locale.

Keywords: α-mixing data; asymptotic normality; incomplete data; local linear fit; regression function; relative error; uniform almost sure convergence.

Cette thèse a été préparée dans les laboratoires suivants.

LMPA Joseph Liouville

Maison de la Recherche Blaise Pascal
50, rue Ferdinand Buisson
CS 80699
62228 Calais Cedex
France

☎ (33)(0)3 21 46 55 86
🖷 (33)(0)3 21 46 55 75
✉ secretariat@lmpa.univ-littoral.fr
Site http://www-lmpa.univ-littoral.fr/

LaPS

Bat. Laboratoires de recherche
Badji Mokhtar -Annaba- B.P.12
23000 Annaba
Algérie

☎ 213 (0) 38 87 10 57
🖷 213 (0) 38 87 10 57
✉ laps@univ-annaba.dz
Site http://laps.univ-annaba.dz/

Je dédie cette thèse
à mes parents et à mon frère

Résumé

Dans cette thèse, nous nous intéressons à développer des méthodes robustes et efficaces dans l'estimation non paramétrique de la fonction de régression. Le modèle considéré ici est le modèle censuré aléatoirement à droite qui est le plus utilisé dans différents domaines pratiques.

Dans un premier temps, nous proposons un nouvel estimateur de la fonction de régression en utilisant la méthode linéaire locale. Nous étudions sa convergence uniforme presque sûre avec vitesse. Enfin, nous comparons ses performances avec celles de l'estimateur de la régression à noyau classique à l'aide de simulations.

Dans un second temps, nous considérons l'estimateur de la fonction de régression par erreur relative (RER en anglais), basé sur la minimisation de l'erreur quadratique relative moyenne. Ainsi, nous établissons la convergence uniforme presque sûre (sur un compact) avec vitesse de l'estimateur défini pour des observations indépendantes et identiquement distribuées. En outre, nous prouvons sa normalité asymptotique en explicitant le terme de variance. Enfin, nous conduisons une étude de simulations pour confirmer nos résultats théoriques et nous appliquons notre estimateur sur des données réelles.

Par la suite, nous étudions la convergence uniforme presque sûre (sur un compact) avec vitesse de l'estimateur RER pour des observations soumises à une structure de dépendance du type α-mélange. Une étude de simulation montre le bon comportement de l'estimateur étudié. Des prévisions sur données générées sont réalisées pour illustrer la robustesse de notre estimateur.

Enfin, nous établissons la normalité asymptotique de l'estimateur RER pour des observations α-mélangeantes où nous construisons des intervalles de confiance afin de réaliser une étude de simulations qui valide nos résultats.

Pour conclure, le fil conducteur de cette modeste contribution, hormis l'analyse des données censurées est la proposition de deux méthodes de prévision alternative à la régression classique. La première approche corrige les effets de bord crée par les estimateurs à noyaux classiques et réduit le biais. Tandis que la seconde est plus robuste et moins affectée par la présence de valeurs aberrantes dans l'échantillon.

Mots clés : Convergence uniforme presque sûre ; Données α-mélangeante ; Données incomplétes ; Fonction de régression ; Erreur relative ; Normalité asymptotique ; Méthode linéaire locale.

Abstract

In this thesis, we are interested in developing robust and efficient methods in the nonparametric estimation of the regression function. The model considered here is the right-hand randomly censored model which is the most used in different practical fields.

First, we propose a new estimator of the regression function by the local linear method. We study its almost uniform convergence with rate. We improve the order of the bias term. Finally, we compare its performance with that of the classical kernel regression estimator using simulations.

In the second step, we consider the regression function estimator, based on the minimization of the mean relative square error (called : relative regression estimator). We establish the uniform almost sure consistency with rate of the estimator defined for independent and identically distributed observations. We prove its asymptotic normality and give the explicit expression of the variance term. We conduct a simulation study to confirm our theoretical results. Finally, we have applied our estimator on real data.

Then, we study the almost sure uniform convergence (on a compact set) with rate of the relative regression estimator for observations that are subject to a dependency structure of α-mixing type. A simulation study shows the good behaviour of the studied estimator. Predictions on generated data are carried out to illustrate the robustness of our estimator.

Finally, we establish the asymptotic normality of the relative regression function estimator for α-mixing data. We construct the confidence intervals and perform a simulation study to validate our theoretical results.

In addition to the analysis of the censored data, the common thread of this modest contribution is the proposal of two alternative prediction methods to classical regression. The first approach corrects the border effects created by classical kernel estimators and reduces the bias term. While the second is more robust and less affected by the presence of outliers in the sample.

Key words : Alpha mixing data, Asymptotic normality, Incomplete data, Local linear fit, Regression function, Relative error, Uniform almost sure convergence.

Remerciements

Ce document est le fruit de trois années de recherche, qui n'aurait pas été possible sans l'aide et le soutiens de nombreuses personnes.

Je tiens à remercier particulièrement mes directeurs de thèse le professeur *Elias OULD SAÏD* et le professeur *Riad M. REMITA* pour leurs encadrements tout au long de ce travail. Merci pour m'avoir donnée l'opportunité de vivre cette aventure, de m'avoir fait confiance et pour avoir su être exigeant tout en me laissant libre d'explorer toutes les pistes que je souhaitais entreprendre. Je tiens à leurs exprimer ma plus profonde gratitude pour le soutien moral, l'encouragement et l'extrême patience dont ils ont fait part. Merci aussi de m'avoir insufflé la rigueur mathématique nécessaire à l'exercice délicat de la rédaction de ce manuscrit.

Je suis très reconnaissante aux professeurs *Zohra GUESSOUM* et *Mustapha RACHDI* qui ont accepté la tâche fastidieuse de rapporter cette thèse. Leurs présences à la soutenance de cette thèse m'honore énormément.

Je remercie également les professeurs *Célestin KOKONENDJI* et *Hacène BOUTABIA* d'avoir accepté d'être examinateurs et de bien vouloir participer au jury de cette thèse.

Je suis très heureuse que madame *Ouafae BENRABAH* soit membre invité de ma thèse. J'ai eu la chance de travailler avec elle et je tiens sincèrement à la remercier pour ses encouragements, ses conseils et son aide aussi bien sur le plan personnel que professionnel.

Durant ces trois années de thèse, j'ai eu l'occasion et la chance d'exposer mon travail au sein du laboratoire MSTD de l'USTHB. Je tiens à remercier les Pr. *Zohra GUESSOUM*, *Ourida SADKI* et *Abdelkader TATACHAK* pour leurs disponibilités et accueil durant mes visites à l'USTHB, l'intérêt qu'ils ont porté à mon travail et leurs commentaires constructifs.

Je tiens à remercier de manière générale tous les membres du laboratoire de mathématiques pures et appliquées (LMPA, ULCO) et le laboratoire de probabilités et statistiques (LaPS, UBMA). Je remercie l'UBMA et l'ULCO pour m'avoir attribuée le financement qui m'a permit de travailler dans de bonnes conditions. Je remercie egalement le professeur *Rachid AMARA* qui a été l'instigateur de cette cotutelle.

Je remercie par la même occasion Isabelle, Romuald, Nicky, Ayoub et Arij pour leurs soutiens et aides durant ces années et pour cela je leurs suis reconnaissante.

Quoi que je puisse dire, je ne pourrai jamais exprimer ma gratitude à mes parents et mon frère. Malgré l'éloignement et le confinement suite à la pandémie qui nous séparent, ils ont toujours été à mes cotés pour m'encourager. Merci du fond du cœur pour le soutien et les sacrifices qu'ils ont fait pour moi ainsi que leur amour inconditionnel. Je tiens enfin à remercier tous les membres de ma famille et toutes les personnes proches qui m'ont apportée leur soutien durant ces années.

A tous ceux que je viens de citer, et à tous ceux que j'aurais oublié, je vous dédie cette thèse, chacun de vous a contribué d'une manière ou d'une autre à la façonner.

Articles et communications

Affiliation :

- Actuellement attachée temporaire d'enseignements et de recherches (ATER) au département de mathématiques de l'université de Lille.

Articles inclus dans la thèse :

1. Feriel Bouhadjera, Elias Ould Saïd and Riad M. Remita. Nonparametric relative error estimation of the regression function for censored data. **ArXiv :1901.09555**.

2. Feriel Bouhadjera, Elias Ould Saïd and Riad M. Remita. Strong consistency of the nonparametric local linear regression estimation under censorship model. *Communications in Statistics : Theory and Methods*. **hal-02532639, v1**. En révision mineure.

3. Feriel Bouhadjera and Elias Ould Saïd. On the strong uniform consistency for relative error of the regression function estimator for censoring times series model. Soumis. **ArXiv :1910.01964**.

4. Feriel Bouhadjera and Elias Ould Saïd. Asymptotic normality of the relative error regression function estimator for censored time series data. Soumis.

Articles hors thèse :

1. Ouafae Benrabah, Feriel Bouhadjera and Elias Ould Saïd. Local linear estimation of the regression function for twice censored data. En révision dans *Statistical Papers*.

2. Feriel Bouhadjera and Elias Ould Saïd. Nonparametric local linear estimation of the relative error regression function for censorship model. **hal-02532655, v1**.

3. Feriel Bouhadjera and Elias Ould Saïd. Nonparametric local linear estimation of the relative error regression function for twice censored data.

4. Feriel Bouhadjera, Mohamed Lemdani and Elias Ould Saïd. Strong uniform consistency of the local linear relative error regression estimator under left truncation.

Communications

Fevr. 2020. Exposé au séminaire hebdomadaire du laboratoire de Probabilités, Statistiques et Applications. *Université des sciences et de la technologie Houari Boumédiène, Alger.*

Oct. 2019. Exposé au séminaire des doctorant du laboratoire de Mathématique de l'institut Alexander Grothendieck de Montpellier. *Université de Montpellier.*

Sep. 2019. Exposé à la treizième rencontre doctorale en mathématiques du Nord-Pas-de-Calais. *Faculté des Sciences Jean Perin, Lens.*

Juin 2019. Participation à la 51^{me} journée de Statistiques de la Société Française de Statistiques. *Université de Lorraine, Nancy.*

Avr. 2019. Présentation à la conférence internationale de modélisation mathématique et applications. *Université Mohamed V, Rabat.*

Dec. 2018 Exposé aux journées jeunes chercheurs. *Université Badji Mokhtar Annaba.*

Dec. 2018 Exposé au séminaire hebdomadaire du laboratoire de Probabilités, Statistiques et Applications. *Université des sciences et technologie Houari Boumédiène, Alger.*

Oct. 2018. Rencontres doctorales Lebesgues. *Université de Bretagne Occidentale, Brest.*

Sep 2018. Douzième rencontre doctorale en mathématiques du Nord-Pas-de-Calais, Lille.

Oct. 2017. 1^{ere} édition des doctoriales nationales en mathématiques. *Université des Frères Mentouri, Constantine, Algérie.*

Oct. 2017. Conférence internationale : Evolution des mathématiques contemporaines et leurs impacts dans les sciences et technologies. *Université des Frères Mentouri, Constantine, Algérie.*

Sommaire

Résumé vi

Abstract vii

Remerciements viii

Articles et communications ix

Sommaire xi

Liste des tableaux xiii

Table des figures xiv

1 Introduction générale 1
- 1.1 Données incomplètes . . . 5
- 1.2 Le phénomène de censure . . . 6
- 1.3 Dépendance . . . 12
- 1.4 Contexte et résultats préliminaires . . . 14
- 1.5 Régression non paramétrique à noyau . . . 14
- 1.6 Régression non paramétrique locale linéaire . . . 17
- 1.7 Régression pour un modèle de censure . . . 19
- 1.8 Choix de la fenêtre . . . 20
- 1.9 Contribution de la thèse . . . 21

2 Estimation locale linéaire de la fonction de régression 25
- 2.1 Modèle . . . 25
- 2.2 Hypothèses et principaux résultats . . . 28
- 2.3 Étude numérique . . . 30
- 2.4 Preuves et résultats auxiliaires . . . 32
- 2.5 Conclusion . . . 45

3 Estimation à noyau : cas i.i.d. 46
- 3.1 Introduction . . . 46
- 3.2 Définition de l'estimateur . . . 50
- 3.3 Hypothèses et principaux résultats . . . 52
- 3.4 Étude numérique . . . 54
- 3.5 Preuves et résultats auxiliaires . . . 63

4 Convergence uniforme : cas α-mélangeant **71**
4.1 Introduction 71
4.2 Hypothèses et résultat principal 72
4.3 Étude numérique 73
4.4 Preuves et résultats auxiliaires 80

5 Normalité asymptotique : cas α-mélangeant **87**
5.1 Introduction 87
5.2 Hypothèses et résultat principal 88
5.3 Étude numérique 90
5.4 Preuves et résultats auxiliaires 92

Conclusion **102**

Extensions et perspectives **103**

Annexe **104**
Théorème central limite 104
Inégalités exponentielles 104
Classes de Vapnik-Cervonenkis (V-C classes) 105

Bibliographie **107**

Liste des tableaux

2.1 Tableau comparatif des EQM. 32

3.1 EQM en fonction de α avec C.P. $\approx 30\%$ et $\lambda = 1.5$. 58
3.2 EQM en fonction de α avec C.P. $\approx 30\%$ et $\lambda = 3$. 58
3.3 EAM en fonction des valeurs aberrantes pour C.P. $\approx 33\%$. 60

4.1 Tableau comparatif des EQM pour un C.P. $\approx 12\%$. 79

Table des figures

1.1 Âge en fonction du % de graisse pour 18 adultes. 2
1.2 $T_i = 2X_i + 3$ pour $1 \leq i \leq 100$ avec 9 valeurs aberrantes. 3
1.3 (a) Scatter-plot du temps de survie en fonction de l'âge des patients atteints d'insuffisance cardiaque. (b) Fonctions de survie pour les patients atteints d'insuffisance cardiaque. (c) Fonctions de survie pour les hommes et femmes atteint d'insuffisance cardiaque. 11
1.4 Exemples de noyaux classiques. 15
1.5 $2X_i + 1$ avec $1 \leq i \leq 100$ pour différentes valeurs de h_n allant de la plus petite à la plus grande (de gauche à droite). 17

2.1 $\widehat{m}_{LLR}(\cdot)$ avec C.P.$\approx 30\%$ pour $n = 100, 300$ et 500 respectivement. 31
2.2 $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ pour C.P. $\approx 8, 25$ et 60% respectivement. 31
2.3 $\widehat{m}_{LLR}(\cdot)$, $\widehat{m}_{CR}(\cdot)$ avec $n = 300$ pour C.P. $\approx 10, 35$ et 65% respectivement. . 31

3.1 $\widehat{m}_{RER}(\cdot)$ pour C.P.$\approx 50\%$ et $n = 100, 300$ et 500 respectivement. 55
3.2 $\widehat{m}_{RER}(\cdot)$ pour $n = 300$ et C.P.$\approx 11, 48, 74\%$ respectivement. 56
3.3 $\widehat{m}_{RER}(\cdot)$ pour $n = 300$ et C.P.$\approx 42, 44$ and 33% pour les foncions Exponentielle, Parabolique et Sinusoïdale respectivement. 56
3.4 EQM$(\cdot)$ avec $n = 100, 300$ et 500 (de gauche à droite) et C.P.$\approx 10\%$ pour M.F. $= 10, 50$ et 100 respectivement. 57
3.5 $\widehat{m}_{RER}(\cdot)$ pour $n = 300$, C.P.$\approx 50\%$ et $\alpha = 0.01, 0.05$ et 0.1 respectivement. 57
3.6 $\widehat{m}_{RER}(\cdot)$ et $\widehat{m}_{CR}(\cdot)$ pour $n = 100$ et C.P.$\approx 10, 35, 80\%$ respectivement. . . . 59
3.7 $\widehat{m}_{RER}(\cdot)$, $\widehat{m}_{CR}(\cdot)$, et $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ et C.P.$\approx 50\%$ pour M.F. $= 10, 25$ et 50 respectivement. 59
3.8 $\widehat{m}_{RER}(\cdot)$, $\widehat{m}_{CR}(\cdot)$ e $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ et M.F.$= 25$ pour C.P.$\approx 8, 45, 70\%$ respectivement. 59
3.9 C.P.$\approx 66\%$, $m = 200$ pour $n = 100, 300$ et 500, respectivement. 61
3.10 $\widehat{m}_{RER}(\cdot)$ avec C.P.$\approx 30\%$ pour $n = 50, 150$, et 250 respectivement. 61
3.11 **(a)** Données sur le mélanome malin. o : données censurées et $\diamond$: données non censurées. **(b)** Losange : vraies valeurs et croix : prédiction du RER. 63
3.12 **(c)** et **(d)** Losange : vraies valeurs, croix : prévision du RER, rond : prévision du CR et triangle : prévision du LLR. 63

4.1 $\rho = 0.1$ et C.P. $\approx 35\%$ pour $n = 100, 300$ et 500 respectivement. 74
4.2 $n = 300$ et $\rho = 0.1$ pour C.P. $\approx 7, 40$ et 67% respectivement. 74
4.3 $n = 300$, $\rho = 0.1$ et C.P. $\approx 35\%$ pour M.F.$= 50, 100$ et 150 respectivement. 75
4.4 $\rho = 0.9$ et C.P.$\approx 15\%$ pour $n = 100, 300$ et 500 respectivement. 75
4.5 $n = 300$ et $\rho = 0.9$ pour C.P.$\approx 3, 15$ et 56% respectivement. 76
4.6 $n = 300$, $\rho = 0.9$ et C.P.$\approx 20\%$ pour M.F.$= 50, 100$, et 150 respectivement. 76
4.7 $n = 300$, $\rho = 0.5$ et C.P.$\approx 15\%$. 77

4.8 $\rho = 0.1$ et $n = 300$ pour C.P.$\approx 10, 33$ et 54% respectivement. 77
4.9 $\rho = 0.1$ et $n = 300$ pour C.P.$\approx 5\%$ et M.F.$= 50, 100$ et 150 respectivement. 78
4.10 $\rho = 0.9$ et $n = 300$ pour C.P.$\approx 7, 32$ et 65% respectivement. 78
4.11 $\rho = 0.9$ et $n = 300$ pour C.P.$\approx 5\%$ et M.F.$= 50, 100$ et 150 respectivement. 78
4.12 $n = 300$, $\rho = 0.3$ et C.P.$\approx 30\%$. (a) Nuage de points des données censurées et non censurées. (b) Performance de l'estimateur RER en prévision. (c) Comparaison entre le RER et CR en prévision. 80

5.1 C.P. $\approx 55\%$ et $\rho = 0.3$ pour $n = 50, 200$ et 400 respectivement. 91
5.2 $n = 200$ et C.P. $\approx 55\%$ pour $\rho = 0.9,\ 0.6$ et 0.3 respectivement. 91
5.3 $n = 200$ et $\rho = 0.6$ pour C.P. $\approx 28, 58$ et 80% respectivement. 92
5.4 $\rho = 0.3$ et C.P. $\approx 25\%$ pour $n = 50, 200$ et 400 respectivement. 92
5.5 $\rho = 0.3$ avec C.P. $\approx 85\%$ pour $n = 50, 200$ et 400 respectivement. 92
5.6 $\rho = 0.9$ avec C.P. $\approx 25\%$ pour $n = 50, 200$ et 400 respectivement. 93
5.7 $\rho = 0.9$ avec C.P. $\approx 85\%$ pour $n = 50, 200$ et 400 respectivement. 93

Chapitre 1

Introduction générale

Les fonctions de régression non paramétrique ont été largement utilisées ces dernières décennies, pas seulement en statistiques, mais dans différents domaines tels que la médecine, le traitement de signal, l'économie et la biologie … La fonction de régression est une fonction générale qui caractérise la relation entre deux variables. Par exemple, nous voulons savoir si la réduction de la vitesse permet de diminuer le nombre d'accidents sur la route ; est-ce que l'augmentation des heures d'études permettent d'améliorer la moyenne de l'étudiant, etc. Cette dernière représente l'une des premières quantités qu'un praticien peut étudier lorsqu'il s'intéresse à expliquer une variable à travers une autre. On peut voir ce problème de la manière suivante : Nous avons deux variables aléatoires (v.a.) réelles T (variable d'intérêt / réponse) et X (variable explicative / co-variable) liées par la relation suivante :

$$T = m(X) + \varepsilon \tag{1.1}$$

et on cherche à étudier le lien entre ces deux variables. Ce lien est modélisé par la fonction $m(\cdot)$, dite fonction de régression et ε est une v.a. représentant les erreurs d'observations. Nous supposons que la co-variable X admet une densité marginale notée $f(\cdot)$ et que le couple de v.a. (X, T) admet une densité jointe dans $\mathbb{R}^2$ notée $f(\cdot,\cdot)$. Afin d'estimer la fonction $m(\cdot)$, deux principales approches sont possibles : paramétrique et non-paramétrique. Tout d'abord, l'approche paramétrique stipule l'appartenance de la loi de probabilité réelle des observations à une classe particulière de lois, qui dépendent d'un nombre fini de paramètre à estimer. L'avantage de cette méthode est la facilité d'estimation des paramètres, par contre l'inconvénient majeur de cette approche est l'inadéquation qu'il peut y avoir entre le modèle choisi et le phénomène réel qu'on étudie.

Exemple 1 (L'âge et l'obésité chez l'homme.) *Les données proviennent d'une étude portant sur une nouvelle méthode de mesure de la composition corporelle, et donnent le pourcentage de graisse corporelle (en % de graisse), l'âge et le sexe de* 18 *adultes âges de* 23 *à* 61 *ans (voir Mazess et al. (1984)). On cherche à établir la relation entre l'âge et le pourcentage de graisse.*

L'approche non paramétrique quand a elle, permet d'estimer une fonction sans aucune restriction paramétrique sur cette dernière. Cette approche statistique globale cherche à limiter le nombre d'hypothèses sur la forme / type / nature de la fonction à estimer. Il s'agit donc dès lors d'un problème d'estimation fonctionnelle. Donc, en ayant moins

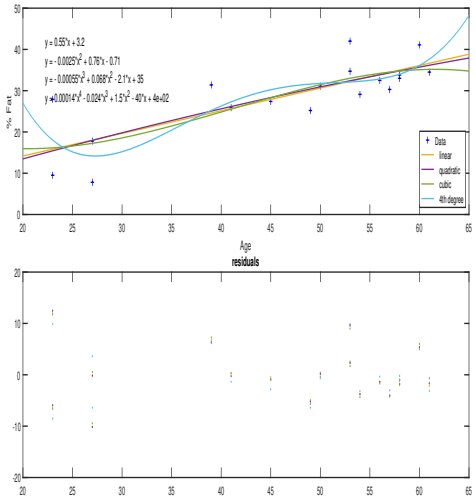


FIGURE 1.1 – Âge en fonction du % de graisse pour 18 adultes.

d'apriori sur les observations on génère un modèle plus générale qui est de ce fait plus robuste. l'inconvénient d'une telle approche est sa nécessité d'un nombre important d'observations et l'appartenance de la fonction à estimer à un espace de dimension infinie. Généralement les estimateurs non paramétriques sont moins efficaces que les estimateurs paramétriques lorsque le modèle paramétrique choisi correspond aux données.

Cependant, comme il est souvent difficile d'avoir une idée sur la loi de T, nous avons privilégié dans cette thèse l'approche non paramétrique pour les estimateurs proposés. Ce thème de recherche fait preuve ces dernières années de grand développement au niveau théorique et pratiques. La modélisation par la régression a fait l'objet d'une littérature abondante, nous renvoyons ici aux revues bibliographiques de Collomb (1981), Prakasa Rao (1983), Bosq and Lecoutre (1987) et Hardle (1990). Dans le contexte non paramétrique, les résultats sur la régression sont très récents comparés à ceux du cas paramétrique.

Étudier le lien entre deux variables a généralement pour but de prédire l'une étant donné l'autre. Il existe plusieurs manières d'aborder ce problème de prévision et l'une des plus utilisées est la régression basée sur l'espérance conditionnelle. Le critère utilisé dans ce cas ci pour estimer la fonction de régression est le critère des moindres carrées donné par :

$$\mathbb{E}\Big[(T - m(X))^2 | X\Big]. \tag{1.2}$$

De ce fait, la fonction de régression $m(x)$ définie par $\mathbb{E}[T|X]$ réalise pour tout x fixé la meilleure approximation de T sachant $X = x$, au sens des moindres carrées. La méthode des moindres carrés est largement utilisée dû à sa simplicité. Cependant, cette dernière atteint ses limites et perd de son efficacité lorsque les données traitées contiennent des valeurs aberrantes. Ainsi, la validité des résultats devient compromise, et la méthode des moindres carrés devient infructueuse et peut être biaisée. Pour des raisons de robustesse, deux approches alternatives ont été proposées dans cette thèse. D'une part, la prédiction au moyen de l'erreur quadratique relative moyenne. Il s'avère que cette dernière est résistante à la présence de valeurs aberrantes dans les données. D'autre part,

l'estimation de la fonction de régression par la méthode linéaire locale qui a l'avantage d'être résistante aux effets de bord et réduit le terme du biais.

En pratique les données contiennent souvent des erreurs de mesures. Imaginons par exemple qu'un appareil de mesure a un dysfonctionnement passager ou une erreur de retranscription des données. Les données collectées contiennent alors ce qu'on appelle des données aberrantes. Une observation est dite aberrante si elle est distante de manière "anormale" des autres observations effectuées sur un phénomène. Ce type de données est généralement observé dans les études de type suivis médicales, les données financières (indices boursiers) mais aussi les études sociologique. La figure 1.2 retranscrit un exemple (que nous avons généré) de données contenant des valeurs aberrantes issues d'un modèle linéaire où 9 valeurs aberrantes ont été crées volontairement pour illustrer ce type de données. Comment avons nous crée ces données aberrantes ! Tout simplement

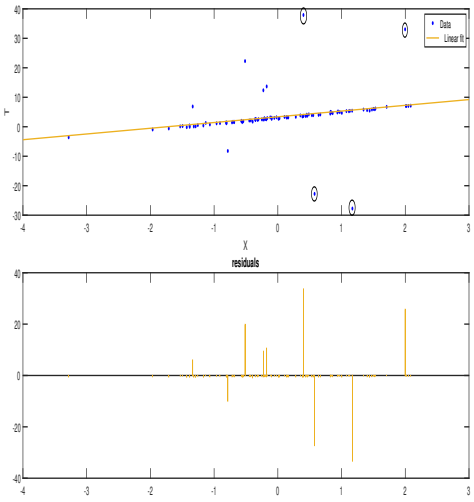


FIGURE 1.2 – $T_i = 2X_i + 3$ pour $1 \leq i \leq 100$ avec 9 valeurs aberrantes.

en choisissant aléatoirement un certain nombre de temps de survie T auxquels on a attribué un poids différent (d'une valeur à une autre) pour se rapprocher le plus possible de la réalité. En observant la figure 1.2, il semble alors essentiel de trouver des techniques de modélisation robuste capable d'étudier la relation entre ces variables.

Une méthode pour détecter ce type de données est la détection graphique soit par une boite à moustaches (Box plot), Histogramme où bien un nuage de points. Une valeur aberrante peut mal conduire les analyses statistiques si elle n'est pas correctement traitée. La recherche sur les traitements de valeurs aberrantes est portée exclusivement sur des méthodes d'estimation résistantes aux valeurs aberrantes.

C'est pour toutes ces raisons que nous nous sommes penchés sur l'estimation non paramétrique de la fonction de régression lorsque les données contiennent des valeurs aberrantes. Comme illustré dans cet exemple, le caractère aberrant de certaines données peut révéler leur appartenance à une population différente du reste des valeurs de la série. Les estimateurs capables de composer avec les données aberrantes sont dits robustes tandis que la moyenne conditionnelle n'en est pas un.

La présence de données aberrantes peut amener à des résultats non pertinents, d'où l'apparition des méthodes d'estimation robuste qui permettent de résoudre ce genre

de problèmes. Le but de ce travail est d'étudier l'estimation robuste de la fonction de régression. Dans la littérature, les méthodes de régression robuste sont les approches les plus utilisées lorsqu'on est en présence de données aberrantes. Par exemple, la régression avec M-estimation, la régression quantile, la médiane et le mode sont des méthodes robuste à ce type de données. La régression robuste s'interprète le plus souvent comme la résistance de l'estimateur aux perturbations dans les données. Cette dernière permet de surmonter les limites des méthodes traditionnelles. C'est pour cette raison que nous nous sommes intéressés à cette fonction de perte quadratique qui s'écrit pour tout T strictement positif par :

$$\mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)^2 \Big| X\right]. \tag{1.3}$$

Cette dernière est appelée erreur quadratique relative moyenne. Les pionniers qui ont considéré cette fonction de perte furent Narula and Wellington (1977). Néanmoins, l'article qui a été l'inspiration (de la problématique) de cette thèse est celui de Park and Stefanski (1998) qui ont considéré (1.3) comme fonction de perte en substitution de (1.2). Le calcul de la solution de (1.3) est détaillé en chapitre 3 et le résultat est donné par :

$$m(X) = \frac{\mathbb{E}\left[T^{-1}|X\right]}{\mathbb{E}\left[T^{-2}|X\right]}, \tag{1.4}$$

en admettant que les deux premiers moments inverse conditionnels de T sachant X existent et soient finis. Ils ont aussi noté que :

$$\frac{\mathbb{E}\left[T^{-1}|X\right]}{\mathbb{E}\left[T^{-2}|X\right]} \leq \mathbb{E}\left[T|X\right]$$

presque sûrement. En raison de la robustesse de l'estimation via l'erreur relative aux valeurs aberrantes, cette dernière est plus adéquate que la méthode basée sur l'EQM classique. Si nous prenons l'exemple de la prévision de la consommation d'électricité d'un foyer, les données peuvent être faibles durant une période (exemple l'hiver) et fortes pour une autre période (pour plus de détails, voir Hirose and Masuda (2018)). Ainsi, la variable d'intérêt peut alors contenir des valeurs aberrantes.

Park and Stefanski (1998) ont considéré une approche paramétrique pour estimer la fonction de régression $m(\cdot)$ qui se concentre sur l'estimation des paramètres moyenne et variance de l'inverse de la variable d'intérêt T (voir Carroll and Ruppert (1988)). Dans ce cadre d'estimation linéaire, nous renvoyons le lecteur à Lin and Chen (2013) et Chen et al. (2010). Narula and Wellington (1977) ont étudié une méthode d'estimation pour la minimisation de la somme des erreurs relatives absolues. Farum (1990) a quant à lui développé une méthode qui permet de réduire l'erreur relative absolue. Khoshgoftaar et al. (1992) ont étudié les propriétés asymptotiques de l'estimateur minimisant la somme des erreurs quadratiques relatives.

Dans notre étude, nous nous sommes concentrés sur les approches non paramétriques. Dans ce cadre, nous rappelons que Jones et al. (2008) ont étudié l'estimateur de la fonction de régression relative par la méthode linéaire locale. Ces derniers ont établi des résultats asymptotiques pour les termes du biais et de la variance. Hu (2019) a examiné le modèle de régression multiplicative à coefficient variable, qui est très utile

pour les modèles à variable d'intérêt positive. Le critère d'erreur relative du produit est étendu au modèle de régression multiplicatif à coefficients variables par des techniques de lissage à noyau. Dans cet article, l'auteur a établi la convergence et la normalité asymptotique de l'estimateur proposé. Dans le cadre fonctionnel, Demongeot et al. (2016) ont établi la convergence uniforme presque sûre avec vitesse et la normalité asymptotique de l'estimateur à noyau de la fonction de régression relative. Parallèlement, Chahad et al. (2017) considèrent la méthode linéaire locale pour estimer la fonction de régression relative et établissent la convergence uniforme presque complète.

La présence de données aberrantes dans l'estimation de la fonction de régression n'est pas l'unique difficulté rencontrée en statistique. Une autre complication souvent rencontré dans le domaine de l'analyse de survie est l'existence de données incomplètes. L'objet de cette thèse est d'estimer la fonction de régression pour des données de survie. L'analyse des données de survie a pour première particularité de ne considérer que les v.a. positives. Une conséquence de cette particularité est que la loi normale ne sera plus ici la référence en matière de distribution. Toute autre loi à support dans $\mathbb{R}_+$ lui sera préférée (par exemple la loi exponentielle). Une seconde particularité de cette analyse est l'incomplétude des données, qui équivaut à une perte d'information.

1.1 Données incomplètes

Ces dernières années, le nombre d'études réalisées en vue de fournir des estimations sur divers sujets dans différents domaine a augmenté considérablement. Que peut on dire de la fiabilité des estimations fondées sur des données incomplètes ? Une question fréquemment soulevée est celle de savoir si, en ignorant les données incomplètes, les estimations basées sur l'information fournie par le phénomène étudié sont ils fiables ? La plupart des statisticiens praticiens ou des analystes des données reconnaissent que l'incomplétude des données a une incidence sur les estimations. C'est pour cette raison que nous nous intéressons dans cette thèse à l'estimation des durées de vie. Qu'entendons nous par durée de vie ? Une durée de vie est une variable aléatoire positive qui représente le temps qui s'écoule jusqu'à la survenue d'un événement. En Médecine, elle désigne le temps de rémission / rechute d'un patient. En industrie, elle identifie le temps entre deux pannes successifs alors qu'en finance elle renvoie au temps d'inflation d'un indice boursier. Bien que cette notion est pluridisciplinaire, on revient à étudier le temps écouler pour passer d'un état à un autre.

La spécificité des données de survie est de comporter des observations incomplètes. Cette incomplétude est essentiellement due à deux phénomènes : la censure et la troncature. La troncature à gauche modélise les cadres d'études expérimentales pour des durées de vie qui doivent dépasser un certain seuil pour être observées. En effet, la variable d'intérêt T doit être supérieur à une certaine variable de troncature Y pour pouvoir être observée (c-à-d il n'y a des observations que si $T \geq Y$). Un exemple de ce type de données est l'étude de la durée de vie après la retraite de sujets qui entrent dans l'enquête à la suite d'un tirage au sort dans une caisse de retraite. Un sujet n'est donc observé que si sa durée de vie après la retraite excède le délai entre sa prise de retraite et l'instant de l'enquête. La durée de vie après la retraite est donc tronquée à gauche par ce délai. Pour ce type de données Ould Saïd and Lemdani (2006) obtiennent un résultat de convergence uniforme presque sûre ainsi que de normalité asymptotique pour l'estimateur de la fonction de régression classique solution du problème de minimisation (1.2). Les auteurs ont utilisé la technique des classes de

Vapnik-Cervonenkis (les VC classes, voir Annexe) afin d'estimer le terme dominant (variance). En utilisant l'approche linéaire locale, Wang et al. (2015) ont construit l'estimateur de la fonction de régression quantile pour des données qui présente une certaine forme de dépendance (appelée α-mélange que nous verrons un peu plus loin). Les auteurs ont établi la normalité asymptotique de l'estimateur proposé. Dans un contexte de données associées, Guessoum and Hamrani (2017) ont étudié la convergence uniforme presque sûre de l'estimateur à noyau de la fonction de régression. Dans un cadre fonctionnel, Altendi et al. (2018) ont proposé un estimateur de la fonction de régression relative défini par la formule (1.4) en utilisant la méthode à noyau et ont étudié sa convergence uniforme presque sure et sa normalité asymptotique.

Nous nous intéressons dans cette thèse à un autre type de données incomplètes qui est celui des données censurées. De ce fait, nous allons commencer par présenter le modèle de censure.

1.2 Le phénomène de censure

Une donnée est dite censurée, si on n'en connaît pas la valeur exacte, mais seulement une estimation inférieure ou supérieure, c-à-d une information grossière du type $T \leq M$ ou $T \geq m$. Une telle information est très pauvre, plus pauvre que de dire que $T \in [m, M]$, puisque une seule des deux bornes est connue. Afin d'illustrer cette notion de censure, nous allons présenter deux situations pratiques :

- Prenons par exemple le cas de données de radioactivité dans l'air. on a mesuré l'activité de certains radionucléides, mais lorsque cette activité est inférieure au seuil réglementaire, on ne publie pas la valeur réelle, mais seulement l'information "inférieure au seuil". Cette situation se rencontre aussi dans bon nombre de mesures liées à l'environnement (niveau de pollution, etc.). La véritable valeur de la mesure est généralement perdue, plus exactement, elle n'est conservée, parce que les responsables n'en voient pas l'intérêt.
- Un second exemple dans un tout autre registre " dans le milieu médical". Il est fréquent de voir un médicament essayé sur des malades/bien-portants, mais, pour une raison ou une autre, certaines personnes quittent l'étude avant la fin de l'expérience et sont perdues de vue : imaginons que les habitants d'une petite ville soient suivies pour un traitement contre l'obésité. Si un des habitant quitte la ville avant la fin de l'observation, les données le concernant ne sont pas complètes. Si l'observation porte sur la survie d'un patient, on pourra dire seulement dans ces conditions que sa durée de vie à été supérieure ou égale à ce qui a été observé. Nous pouvons penser à se prémunir contre l'absence de ce patient en lui demandant par avance, où qu'il aille de bien vouloir donner de ces nouvelles. Or les conditions de vie, d'alimentation, etc. risquent d'être différent et ce patient ne sera donc plus représentatif de ce que l'on cherche à étudier.

Parmi les censures, nous distinguons deux types :

Censure droite : de la forme $T \geq C$;

Censure gauche : de la forme $T \leq C$.

Nous ne ferons pas de différence théorique entre inégalité stricte et inégalité au sens large car dans la pratique, il suffit de déplacer légèrement la borne pour passer de l'un à l'autre. En revanche, pour passer d'une variable censurée à droite à une variable censurée à gauche, ou inversement, on pourrait penser à remplacer T par $-T$. Mais

comme nous travaillons avec des durées de vie, les valeurs mesurées T sont entièrement positives et avec ce changement les données deviennent entièrement négatives, ce qui n'a pas de sens pour une durée de vie. Il existe un autre type de censure appelé censure par intervalle. L'observation dans ce cas ci devient périodique pendant toute la durée de l'étude. Prenons l'exemple d'examen des animaux dans les élevages périodiques avec un passage tout les six mois, la seule information disponible ici est que l'examen se produit entre deux passages t_i et t_{i+1}.

Lorsqu'une information est censurée, la valeur exacte est irrémédiablement perdue. Il est clair qu'aucune méthode mathématique, si sophistiquée soit-elle, ne peut remplacer des données manquantes. On peut néanmoins, par un travail approprié et en faisant certaines hypothèses, reconstituer une loi de probabilité pour le phénomène étudié, qui prend en compte les données censurées ce qui a été fait dans cette thèse. Dans la sous-section ci-dessous, nous présentons trois types de censure à droite.

1.2.1 Types de censure

Type I. (censure fixe) Ce modèle décrit la situation où un test qui se termine à une certaine période, et nous savons que les individus restants n'ont pas échoué à ce moment. Par exemple, nous démarrons l'étude avec 100 ampoules électriques, et terminons l'expérience après un certain temps. Dans ce cas, le temps de censure est souvent fixé, et le nombre d'individus ayant échoué est une variable aléatoire. Soit un échantillon de durées de survie $(T_1, \cdots, T_n)$. La variable de censure C ici est fixe, les observations deviennent alors $(Y_1, \delta_1), \cdots, (Y_n, \delta_n)$ avec :

$$Y_i = \min(T_i, C) \qquad \text{et} \qquad \delta_i = \mathbb{1}_{\{T_i \leq C\}}$$

pour $i = 1, \cdots, n$.

Type II. Pour ce modèle, l'expérience se poursuit jusqu'à ce qu'un nombre fixe d'individus r ait échoué. Nous reprenons l'exemple des ampoules, où dans ce cas nous arrêtons l'expérience une fois exactement 50 ampoules électriques ont échoué. Ici, le nombre d'individus ayant échoué est fixe, et le temps est une v.a. Soit un échantillon de durées de survie $(T_1, \cdots, T_n)$ et $r > 0$ fixé. On dit qu'il y a censure de type II pour cet échantillon si au lieu d'observer directement $(T_1, \cdots, T_n)$ on observe $(Y_1, \delta_1), \cdots, (Y_n, \delta_n)$ avec :

$$Y_i = \min(T_i, T_{(r)}) \qquad \text{et} \qquad \delta_i = \mathbb{1}_{\{T_i \leq T_{(r)}\}}$$

pour $i = 1, \cdots, n$ avec $T_{(r)}$ la r^{ieme} statistique d'ordre de l'échantillon $(T_1, \cdots, T_n)$. Ce type de censure est analogue au type I avec $C = T_{(r)}$.

Type III. (censure aléatoire) Ce type de censure généralise la censure de type I au cas où C est une v.a. Soient un échantillon de durées de survie $(T_1, \cdots, T_n)$ et un second échantillon indépendant composé de variables positives $(C_1, \cdots, C_n)$. On dit qu'il y a censure de type III pour cet échantillon si au lieu d'observer directement $(T_1, \cdots, T_n)$ on observe les couples $(Y_1, \delta_1), \cdots, (Y_n, \delta_n)$ avec :

$$Y_i = \min(T_i, C_i) \qquad \text{et} \qquad \delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$$

pour $i = 1, \cdots, n$.

Lorsque la censure à droite et la troncature à gauche se réunissent dans un même échantillon, nous parlons alors de modèle LTRC (Tronqué à gauche et Censuré à droite).

1.2.2 Le modèle LTRC

Nous notons : Y le temps de survie, T la variable de troncature et C la variable de censure. Dans le cas de données censurées à droite et tronquées à gauche, nous observons le triplet (Z, T, δ) si $Z \geq T$, avec $Z = \min(Y, C)$ et $\delta = \mathbb{1}_{\{Y \leq C\}}$. Dans le cas où $Z < T$ il n'y a pas d'observation. Dans ce cadre, très récemment Benseradj and Guessoum (2020) ont établi la convergence uniforme presque sûre sur un compact avec vitesse de l'estimateur de la régression robuste.

1.2.3 Le modèle mixte

Lorsque la censure à droite et la censure à gauche apparaissent dans un même échantillon simultanément, nous parlons de modèle mixte. Pour cela notons par, T le temps de survie, R la variable de censure droite et L la variable de censure gauche. Ce modèle est noté modèle I dans Patilea and Rolin (2006). Nous observons un échantillon du couple (Z, A) où la variable observée est $Z = \max(Y, L)$ avec $Y = \min(T, C)$ et l'indicateur de censure est :

$$A = \begin{cases} 0, & \text{si} \quad L < T \leq R, \\ 1, & \text{si} \quad L < R < T, \\ 2, & \text{si} \quad Y \leq L. \end{cases}$$

Pour ce type de données Kebabi and Messaci (2012) proposent un estimateur à noyau de la fonction de régression en se basant sur des données synthétiques, ils obtiennent sa convergence uniforme presque complète avec vitesse ainsi que sa normalité asymptotique. Parallèlement, Volgushev and Dette (2013) établissent la convergence faible d'un estimateur à noyau de la fonction de régression quantile. Plus récemment, Khardani (2019) a établi un résultat de convergence uniforme presque sûre avec vitesse ainsi que la normalité asymptotique de l'estimateur à noyau de la fonction de régression lorsque la fonction de perte est l'erreur quadratique relative moyenne.

Nous nous intéressons dans cette thèse au modèle le plus courant en pratique, la censure aléatoire à droite.

1.2.4 La censure aléatoire à droite

Nous notons T le délai jusqu'à l'événement d'intérêt alors T est une variable aléatoire positive. Pour simplifier les notations de base, nous supposerons que la variable d'intérêt T est la durée de survie et donc on peut considérer que l'événement observé est le décès. Dans les deux exemples, nous n'observons à chaque fois que le minimum entre la variable d'intérêt et une autre variable, dite de censure. D'un point de vue mathématique, nous avons :

Définition 1.2.1 *Étant donné une v.a. positive T, on dit qu'il y a censure aléatoire à droite, s'il existe une autre v.a. C telle que, au lieu d'observer T on observe (Y, δ) avec $Y = \min(T, C)$ et*

$$\delta = \begin{cases} 1 & si \quad T \leq C, \\ 0 & sinon. \end{cases}$$

δ est appelé indicateur de censure et permet de connaître la nature de la donnée observée Y (c-à-d si c'est une vraie durée T où si c'est une censure C).

Un problème alors se pose : si l'on connaît la loi des observations, pourrait-on connaître la loi de T et de C ? La réponse n'est pas évidente. Le fait que les données soient incomplètes entraîne une perte d'informations. Pour remédier à ce problème, nous supposerons dans la suite de cette thèse que les v.a. d'intérêt et censure sont indépendantes. Notons que pour toute fonction de répartition Q, nous désignerons par $\tau_Q = \sup\{x,\ Q(x) < 1\}$ la borne supérieure du support de Q.

Proposition 1.2.1 *Si les v.a. T et C sont indépendantes entre elles et de fonctions de répartition (f.d.r.) F et G respectives et si $\tau_F \leq \tau_G$, alors la loi de T est identifiable à partir du couple (Y, δ).*

Dans la sous section qui va suivre nous nous intéressons à la construction de l'estimateur de Kaplan and Meier (1958), pour faciliter la compréhension, nous donnons à la fin un petit exemple.

1.2.5 Estimateur de Kaplan-Meier

En l'absence de censure, la fonction de répartition F s'estiment de manière très simple en utilisant la fonction de répartition empirique usuelle. Par exemple $F(t) = \mathbb{P}(T \leq t)$ est estimé par :

$$\widehat{F}_{emp}(t) := \frac{1}{n} \sum_{1 \leq i \leq n} \mathbb{1}_{\{T_i \leq t\}}.$$

Malheureusement dans le cas ou les données sont censurées, il est impossible d'utiliser la fonction empirique puisqu'elle fait intervenir des quantités non observées, car tous les Y_i censurés ne sont pas observés. On estime alors généralement F en utilisant l'estimateur Kaplan and Meier (1958). Ce dernier est l'outil de bas en statistique pour estimer de manière non paramétrique la distribution d'une v.a. T censurée à droite.

L'estimateur de Kaplan-Meier (K-M) est le modèle de durée le plus utilisé en pratique. Il intervient dans toutes les applications qui requièrent la modélisation de durées. Notons par : $\overline{F}(t) := \mathbb{P}(T > t)$, $\overline{G}(t) := \mathbb{P}(C > t)$ et $\overline{H}(t) := \mathbb{P}(Y > t) = \overline{F}(t)\overline{G}(t)$. L'idée de ce modèle est la suivante : survivre après un temps t c'est être en vie juste avant t et ne pas mourir au temps t, c-à-d si $t_2 < t_1 < t$, en utilisant les probabilités composées, nous avons :

$$\begin{aligned} \overline{F}(t) &= \mathbb{P}(T > t_1, T > t) \\ &= \mathbb{P}(T > t | T > t_1) \times \mathbb{P}(T > t_1) \\ &= \mathbb{P}(T > t | T > t_1) \times \mathbb{P}(T > t_1 | T > t_2) \times \mathbb{P}(T > t_2), \end{aligned}$$

et ainsi de suite. En considérant pour $i = 1, \cdots, n$ seulement les dates où l'événement d'intérêt se produit (décès ou censure), nous estimons des quantités du type :

$$p_i := \mathbb{P}(T > Y_{(i)} | T > Y_{(i-1)}),$$

où p_i est la probabilité de survivre dans l'intervalle $]Y_{(i-1)}, Y_{(i)}]$ sachant que l'on était vivant en $Y_{(i-1)}$. Considérons les notations suivantes : r_i le nombre d'individus à risque de subir l'événement juste avant le temps $Y_{(i)}$ et d_i le nombre de décès en $Y_{(i)}$. Nous

notons par $q_i = 1 - p_i$ la probabilité de mourir pendant l'intervalle $]Y_{(i-1)}, Y_{(i)}]$ sachant que l'on était vivant en début de cet intervalle. Alors q_i peut être estimée par :

$$\hat{q}_i = \frac{d_i}{r_i}.$$

Comme les temps d'événements sont supposés distincts (c-à-d qu'il n'y ait pas d'ex-aequo), on a : $d_i = 0$ en cas de censure en $Y_{(i)}$, i.e. quand $\delta_i = 0$ et $d_i = 1$ en cas de décès en $Y_{(i)}$, i.e. quand $\delta_i = 1$. Il est clair que $r_i = n - i + 1$, on obtient alors :

$$\hat{p}_i = \begin{cases} 1 - \dfrac{1}{n-i+1} & \text{si} \quad \delta_i = 1, \\ 1 & \text{si} \quad \delta_i = 0. \end{cases}$$

D'où, nous parvenons enfin à l'estimateur de K-M de la fonction de survie de notre durée d'intérêt T donné par :

$$\widehat{S}(t) := \overline{F}_n(t) \begin{cases} \displaystyle\prod_{Y_{(i)} \le t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{(i)}} & \text{si} \quad t < Y_{(n)}, \\ 0 & \text{si} \quad t \ge Y_{(n)}. \end{cases}$$

En remarquant que la situation de censure de T par C est symétrique à la censure de C par T, on peut définir l'estimateur de K-M de la fonction de survie de la variable de censure $\overline{G}(\cdot)$ en remplaçant $\delta_{(i)}$ par $1 - \delta_{(i)}$ ce qui donne :

$$\overline{G}_n(t) := \begin{cases} \displaystyle\prod_{Y_{(i)} \le t} \left(1 - \frac{1}{n-i+1}\right)^{1-\delta_{(i)}} & \text{si} \quad t < Y_{(n)}, \\ 0 & \text{si} \quad t \ge Y_{(n)}. \end{cases}$$

où $Y_{(1)}, \cdots, Y_{(n)}$ sont les valeurs ordonnées des Y_i et $\delta_{(i)}$ est l'indicatrice de non censure concomitantes aux $Y_{(i)}$.

Remarque 1.2.1 *L'estimateur de K-M peut s'écrire sous la forme suivante :*

$$\overline{G}_n(t) := \begin{cases} \displaystyle\prod_{1 \le i \le n} \left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbb{1}\{Y_{(i)} \le t\}} & si \quad t < Y_{(n)}, \\ 0 & si \quad t \ge Y_{(n)}. \end{cases} \tag{1.5}$$

Pour plus de clarté, notons que tout au long de cette thèse c'est cette dernière forme de l'estimateur K-M que nous allons utilisée dans la construction de nos estimateurs.

Cet estimateur est également appelé produit limite car il s'obtient comme une limite de produits. C'est une fonction en escalier décroissante, constante par morceaux avec des sauts qui se produisent en des observations non censurées. Il n'atteint 0 que si la plus grande des observations $Y_{(n)}$ correspond à une observation non censurée. Les propriétés de cet estimateur ont été très étudiée dans la littérature, voir par exemple Stute and Wang (1993). D'autres contributions importantes concernant la compréhension de l'estimateur K-M ont été apportées par Stute (1995).

Remarque 1.2.2 *Lorsqu'il n'y a pas de censure, l'estimateur de K-M se réduit à la fonction de survie empirique.*

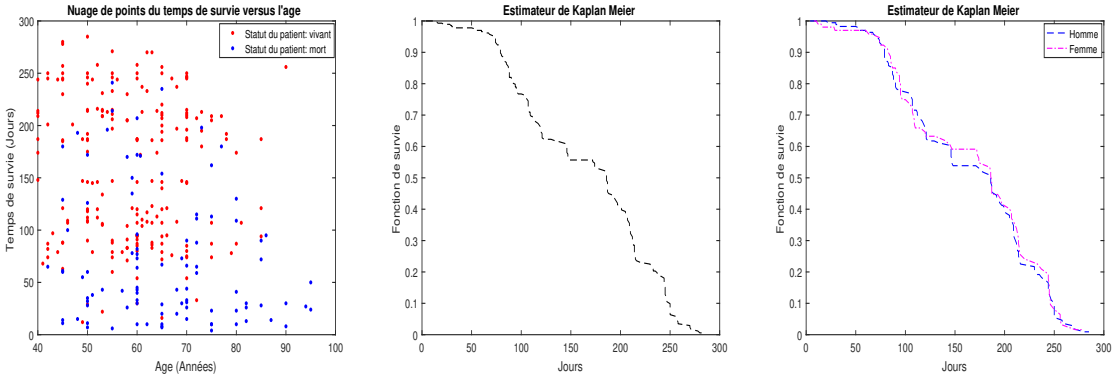
Afin de mieux comprendre les données censurées aléatoirement à droite, nous avons préféré illustrer le traitement de ce genre de données considérées comme incomplètes par un exemple de données réelles (voir : https://plos.figshare.com/articles/Survival_analysis_of_heart_failure_patients_A_case_study/5227684/1).

1.2.6 Exemple (*Insuffisance cardiaque*)

Nous avons analysé un ensemble de données contenant les dossiers médicaux de 299 patients souffrant d'insuffisance cardiaque, recueillis à l'Institut de cardiologie de Faisalabad et à l'hôpital Allied de Faisalabad (Penjab, Pakistan), entre avril et décembre 2015 souffrant de dysfonctionnement systolique ventriculaire gauche, appartenant aux classes III et IV de la *New York Heart Association* (NYHA). Les patients étaient composés de 105 femmes et 194 hommes, et leur âge se situe entre 40 et 95 ans.

Les dossiers médicaux contiennent un certain nombres d'informations sur chaque patient notamment : l'âge, la fraction d'éjection, la créatinine sérique, le sodium sérique, l'anémie, les plaquettes, la créatinine phosphokinase, la pression artérielle, le sexe, le diabète et le tabagisme comme des facteurs potentiels de mortalité. Pour notre part, ce qui nous a intéressé était l'analyse de la survie (temps de survie) des ces patients par rapport à l'âge. Un nuage de point représentant les données où on dissocie les données censurées des non-censurées est donné par la Figure 1.3 (a).

Le graphique de Kaplan Meier (voir les Figures 1.3 (b) et (c)) a été utilisé pour étudier le modèle général de survie qui a montré une forte intensité de la mortalité dans les premiers jours, puis une augmentation progressive jusqu'à la fin de l'étude. Les propriétés de l'estimateur de K-M ont été largement étudiées dans la littérature



(a) (b) (c)

FIGURE 1.3 – (a) Scatter-plot du temps de survie en fonction de l'âge des patients atteints d'insuffisance cardiaque. (b) Fonctions de survie pour les patients atteints d'insuffisance cardiaque. (c) Fonctions de survie pour les hommes et femmes atteint d'insuffisance cardiaque.

(nous renvoyons le lecteur à Andersen et al. (1993)). Soit F_n la f.d.r. empirique basée sur $T_1, T_2, \cdots, T_n$. Nous donnons dans la suite quelques propriétés de convergence :

Théorème 1.2.1 *Étant donnée une variable d'intérêt T de f.d.r. F et une censure C de f.d.r. G indépendantes, alors :*

$$\sup_{0\leq t\leq \tau_F} |F_n(t) - F(t)| \underset{n\to\infty}{\longrightarrow} 0 \qquad p.s.$$

En notant par F (respectivement G) la fonction de répartition de la variable d'intérêt T (respectivement de la variable de censure C), nous pouvons énoncer le résultat suivant :

Théorème 1.2.2 (Földes and Rejtö (1981)) *On suppose que F et G sont continues, et si le réel T est tel que $G(T) < 1$, alors*

$$\mathbb{P}\left(\sup_{-\infty<t\leq T^\star} \left|\overline{F}_n(t) - F(t)\right| = O\left(\frac{\log\log n}{n}\right)\right) = 1.$$

où $T^\star = \min(T, T_F)$.

Dans ce contexte de censure et dans le cadre d'observations i.i.d., Dabrowska (1987) a estimé la fonction de survie conditionnelle en présence de censure à droite. Un résultat de convergence a été établit pour l'estimateur à noyau et l'estimateur des plus proches voisins pour la fonction de hasard cumulative, la fonction de survie et le quantile basés sur l'estimateur de Beran (1981). Quelques années plus tard, Dabrowska (1989) a établi un résultat de convergence uniforme de l'estimateur de Kaplan and Meier (1958) conditionnel à noyau. Stute (1993) a établi la normalité asymptotique de l'estimateur de K-M. Lopez et al. (2013) quand à eux ont étudié les propriétés asymptotiques de l'estimateur de la fonction de répartition jointe pour une co-variable d-dimensionnelle. Lemdani and Ould Saïd (2017) ont défini l'estimateur à noyau de la fonction de régression robuste et ont établit sa convergence uniforme presque sure ainsi que sa normalité asymptotique.

1.3 Dépendance

De nombreux résultats de statistique ont été établi en considérant des échantillons indépendants. Cependant, il est parfois intéressant de considérer et d'étudier des échantillons dépendants afin de pouvoir répondre à des situations pratiques où les données ne sont pas i.i.d. Prenons l'exemple où nous nous intéressons à la contamination par un virus qui se transmet d'une personne à une autre. Effectivement, un individu côtoyant des personnes atteintes par un virus donné, aura plus de chance d'être contaminé que s'il résidait dans une zone relativement épargnée par ce virus. Alors, les données ne sont pas indépendantes. En effet, les données de deux individus géographiquement proches sont dépendantes. Ce type de données est modélisé par des données mélangeantes. Il existe plusieurs types de modélisation de la dépendance au sein d'un échantillon. Nous nous intéresserons dans cette thèse aux phénomènes de dépendance faible. Il existe diverses modélisations de la dépendance faible à l'aide, des notions de variables α-mélangeantes (voir Rosenblatt (1956b)), β-mélangeantes (voir Volkonskii and Rozanov (1959)) ou ϕ-mélangeantes (voir Ibragimov (1962)) et bien d'autres. Pour un point de vue globale sur les différents types de mélange nous renvoyons à Doukhan (1994) et Bradley (2005, 2007). Parmi toutes ces formes de mélanges, l'α-mélange est le plus faible et donc le moins restrictif. Toute suite de v.a. β, ϕ… mélangeantes implique forcement qu'elle est α-mélangeante.

Nous rappelons maintenant la définition d'une suite de v.a. α-mélangeante introduite par Rosenblatt (1956b).

Définition 1.3.1 *Soit $(X_i)_{i\geq 0}$ une suite de variables aléatoires réelles. On note $\mathcal{F}_1^k$ la tribu engendrée par les X_i, $1 \leq i \leq k$ et $\mathcal{F}_{n+k}^{\infty}$ la tribu engendrée par les X_i, $n+k \leq i < \infty$. On définit les coefficients d'α-mélange associés à la suite $(X_i)_{i\geq 0}$ par :*

$$\alpha(n) = \sup_k \sup_{A\in\mathcal{F}_1^k, B\in\mathcal{F}_{n+k}^{\infty}} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)\right|$$

On dit que cette suite est α-mélangeante si le coefficient de mélange vérifie $\alpha(n) \to \infty$ lorsque $n \to \infty$.

On peut définir deux types de mélanges forts :

- Les coefficients d'α-mélange sont dits arithmétiques d'ordre $a > 0$ s'il existe une constante $C \in \mathbb{R}_+^{\star}$ telle que $\alpha(n) \leq Cn^{-a}$.
- Les coefficients d'α-mélange sont dits géométriques s'il existe une constante $C \in \mathbb{R}_+^{\star}$ et $0 \leq \rho < 1$ tels que $\alpha(n) \leq C\rho^n$.

Les processus AR et ARMA sont des exemples de processus géométriquement mélangeants.

Remarque 1.3.1 *Le coefficient de mélange α est tel que $0 \leq \alpha \leq \frac{1}{4}$.*

Ce coefficient est plus faible que d'autres coefficients de mélange β, ϕ, ... (voir Doukhan et al. (1994)), mais comme la notion de variables α-mélangeantes est la plus générale, les résultats impliqueront une classe plus importante de processus. Nous trouvons dans la littérature un grand nombre de travaux consacrés à l'étude de variables α-mélangeantes réelles, nous renvoyons aux travaux Doukhan et al. (1994), Bosq (1998) et Rio (2000).

Ces dernières années, en raison de nombreuses applications dans différents domaines tel que le domaine médical, social où économique, une très grande importance est portée au cas où le temps de survie peut présenter une certaine forme de dépendance. Par exemple, dans le cadre des essais cliniques, il arrive fréquemment que des patients du même hôpital ont des durées de survies corrélés en raison de variables non mesurées comme la qualité des équipements de l'hôpital. Pour plus de détails concernant la dépendance dans les données, nous renvoyons le lecteur à Lipsitz and Ibrahim (2000). Dans le cadre de données censurées, Lecoutre and Ould Saïd (1995) ont établi la convergence uniforme presque complète de l'estimateur de K-M conditionnel. Cai (2001) a développé une méthode d'estimation de la fonction de régression par la méthode locale linéaire. L'auteur a établi la convergence ainsi que la normalité asymptotique de l'estimateur étudié en considérant les deux cas : observations i.i.d. et α-mélangeantes. El Ghouch and Van Keilegom (2008) ont considéré quand à eux le problème d'estimation de la fonction $\mathbb{E}[\phi(T)|X]$ pour toute fonction ϕ connue. Leurs approche consiste à définir une nouvelle variable $T^{\star}$ qui n'est pas sujet à la censure et qui satisfait cette égalité $\mathbb{E}[\phi(T)|X] = \mathbb{E}[T^{\star}|X]$ pour ensuite l'estimer par la méthode linéaire locale. Ces derniers ont établi la convergence uniforme de l'estimateur sous des conditions de mélange. Les mêmes auteurs (El Ghouch and Van Keilegom (2009)) proposent un estimateur de la fonction de régression quantile. Guessoum and Ould Saïd (2010) ont étudié la convergence uniforme presque sûre de l'estimateur à noyau de la fonction de régression défini dans Guessoum and Ould Saïd (2008) pour une co-variable d-dimensionnelle dans le cadre de données α-mélangeantes censurées à droite.

1.4 Contexte et résultats préliminaires

Le but de cette partie n'est bien entendu pas de dresser un panorama complet de tous les résultats obtenus dans l'estimation non paramétrique de la fonction de régression. Nous souhaitons plutôt présenter deux modélisations par deux approches différentes (estimation à noyau et locale linéaire) proposées dans la littérature et des résultats de convergence qui leur sont rattachées.

1.5 Régression non paramétrique à noyau

Soit $X_1, \cdots, X_n$ des v.a. i.i.d. de densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et de fonction de répartition (f.d.r.) $F(x) = \int_{-\infty}^{x} f(t)dt$. Un estimateur intuitif de la f.d.r. F est l'estimateur empirique donné par :

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

où $\mathbb{1}_A$ est la fonction indicatrice de l'événement A. D'après la loi forte des grands nombres (L.F.D.N.), nous avons :

$$F_n(x) \longrightarrow F(x) \qquad \text{p.s.} \qquad \text{quand } n \to \infty.$$

Donc F_n est un estimateur consistant de F. Comment peut-on estimer la fonction de densité f ? Un premier résultat à été proposé par Rosenblatt (1956a) pour un $h_n > 0$ donné par :

$$\widehat{f}_{PR}(x) \approx \frac{F_n(x + h_n) - F_n(x - h_n)}{2h_n}.$$

Ce dernier est appelé un estimateur de *Rosenblatt*. Nous pouvons aussi l'écrire sous cette forme :

$$\widehat{f}_{PR}(x) = \frac{1}{2nh_n} \sum_{1 \leq i \leq n} \mathbb{1}_{\{x-h \leq X_i \leq x+h\}} = \frac{1}{nh_n} \sum_{1 \leq i \leq n} K_0\left(\frac{x - X_i}{h_n}\right).$$

où $K_0(u) = \frac{1}{2}\mathbb{1}_{\{-1 \leq u \leq 1\}}$. Quelques années plus tard Parzen (1962) a suggéré une généralisation de cet estimateur :

$$\widehat{f}_{PR}(x) = \frac{1}{nh_n} \sum_{1 \leq i \leq n} K\left(\frac{x - X_i}{h_n}\right),$$

où K est une densité, telle que $\int K(u)du = 1$. Ce dernier ce dénomme l'estimateur à noyau de *Parzen- Rosenblatt* (P-R). La fonction K est appelée noyau (en anglais "Kernel") et le paramètre h_n est appelé fenêtre (en anglais "bandwidth") de l'estimateur. Nous donnons quelques exemples de noyaux classiques :
Supposons maintenant que la loi du couple (X, T) admet une densité $f(x, t)$ par rapport

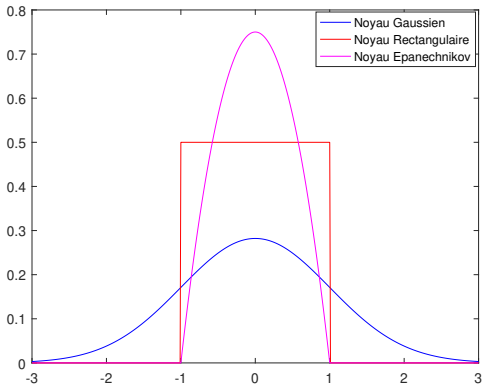


FIGURE 1.4 – Exemples de noyaux classiques.

à la mesure de Lebesgue sur $\mathbb{R}^2$ telle que $f(x) = \int f(x,t)dt > 0$. Alors

$$m(x) = \mathbb{E}[T|X = x] = \frac{\int t f(x,t)dt}{\int f(x,t)dt} = \frac{\int t f(x,t)dt}{f(x)}. \tag{1.6}$$

L'estimateur de Parzen (1962) et Rosenblatt (1956a) (P-R) admet une version multidimensionnelle. Considérons un n-échantillon (X_i, T_i) de couples de variables aléatoires indépendants et de même loi que le couple (X, T), nous définissons, en dimension 2, l'estimateur à noyau de $f(x,t)$ par :

$$\widehat{f}_{PR}(x,t) = \frac{1}{nh_n^2} \sum_{1\leq i\leq n} K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{t - T_i}{h_n}\right) \tag{1.7}$$

où K est un noyau intégrable sur $\mathbb{R}$ tel que $\int K(t)dt = 1$ et $h_n > 0$ est une suite de fenêtres. En remplaçant les densités marginale $f(x)$ et jointe $f(x,t)$ par les estimateurs de P-R $\widehat{f}_{PR}(x)$ et $\widehat{f}_{PR}(x,t)$ respectivement dans (1.6), nous obtenons :

$$\widehat{m}_{NW}(x) = \frac{\int t\widehat{f}_{PR}(x,t)dt}{\widehat{f}_{PR}(x)}.$$

Preuve. D'après (1.7), nous savons que :

$$\int t\widehat{f}_{PR}(x,t)dt = \frac{1}{nh_n^2} \sum_{1\leq i\leq n} K\left(\frac{x - X_i}{h_n}\right) \times \int tK\left(\frac{t - T_i}{h_n}\right)dt. \tag{1.8}$$

Par un changement de variable et le fait que le noyau est symétrique et que c'est une densité de probabilité $\left(\int K(t)dt = 1\right)$, nous avons :

$$\begin{aligned}\frac{1}{h_n}\int tK\left(\frac{t-T_i}{h_n}\right)dt &= \int \frac{t-T_i}{h_n}K\left(\frac{t-T_i}{h_n}\right)dt + \int \frac{T_i}{h_n}K\left(\frac{t-T_i}{h_n}\right)dt \\ &= -h_n\int uK(u)\,du + T_i\int K(u)\,du \\ &= T_i. \end{aligned} \tag{1.9}$$

Donc, à partir d'un noyau K et d'une fenêtre $h_n > 0$, nous pouvons construire des estimateurs à noyau pour la régression non paramétrique analogue à celui construit pour estimer la densité (Rosenblatt (1956a)). En remplaçant (1.9) dans la formule (1.8), nous obtenons l'estimateur de la fonction de régression $m(\cdot)$ défini par :

$$\widehat{m}_{NW}(x) = \frac{\sum_{1\le i\le n} T_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{1\le i\le n} K\left(\frac{x-X_i}{h_n}\right)} \times \mathbb{1}\left\{\sum_{1\le i\le n} K\left(\frac{x-X_i}{h_n}\right) \neq 0\right\}. \tag{1.10}$$

Nous pouvons aussi représenter l'estimateur de N-W comme une somme pondérée des T_i :

$$\widehat{m}_{NW}(x) = \sum_{1\le i\le n} T_i w_i^{NW}(x) \tag{1.11}$$

avec des poids

$$w_i^{NW}(x) = \begin{cases} \dfrac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{1\le i\le n} K\left(\frac{x-X_i}{h_n}\right)} & \text{si} \quad \sum_{1\le i\le n} K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\ 0 & \text{sinon.} \end{cases}$$

Notons que, en restreignant notre étude aux noyaux positifs (c-à-d $K \ge 0$), la fonction indicatrice présentée dans (1.10) disparaît. Ainsi, l'estimateur de N-W est bien linéaire au sens de la définition (1.11) avec pour tout x, la fonction de poids $w_i^{NW}(x)$ satisfont à la relation :

$$\sum_{1\le i\le n} w_i^{NW}(x) = 1.$$

Pour une discussion plus générale sur la fonction de poids dans le cadre de la régression non paramétrique, nous citerons l'article pionnier Stone (1977). Le noyau K détermine la forme du voisinage autour du point x et la fenêtre contrôle la taille de ce voisinage, c-à-d le nombre de d'observations prises pour effectuer la moyenne locale. Intuitivement, la fenêtre h_n a un rôle crucial dans la consistance de l'estimateur N-W (le choix de la fenêtre sera discuté dans la section 1.8).
Si la fenêtre h_n satisfait les conditions suivantes :

$$h_n \longrightarrow 0 \quad \text{et} \quad nh_n \longrightarrow \infty \qquad \text{lorsque } n \to \infty,$$

la variance de l'estimateur de N-W tend vers 0. Cette dernière est une condition nécessaire et suffisante pour obtenir la consistance de l'estimateur de N-W.

Exemple 2 *Pour illustrer l'impact du choix de la fenêtre dans l'estimation non paramétrique, nous étudions un exemple. Supposons que le noyau K est une densité gaussienne, que la fenêtre optimale est de l'ordre de $C(\log n/n)^{1/5}$ pour une constante C convenablement choisie. Nous allons varier la constante C pour changer la valeur de la fenêtre. La figure 1.5, montre le résultat.*

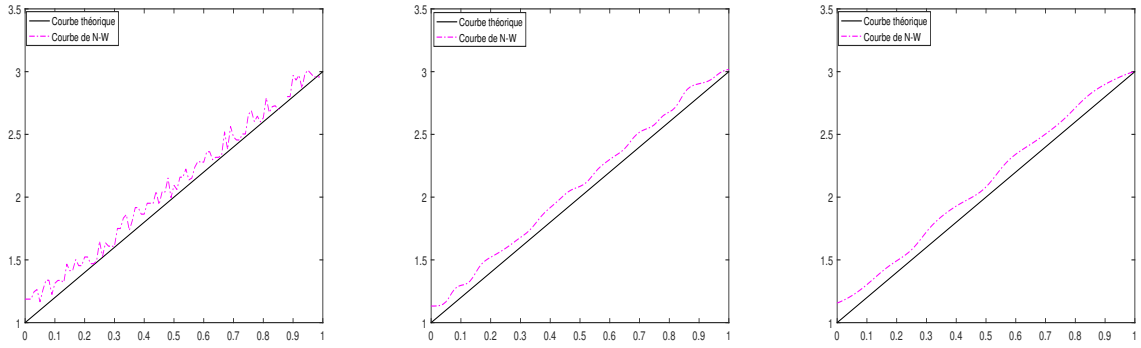


FIGURE 1.5 – $2X_i + 1$ avec $1 \leq i \leq 100$ pour différentes valeurs de h_n allant de la plus petite à la plus grande (de gauche à droite).

1.6 Régression non paramétrique locale linéaire

Dans cette partie, nous étudions la méthode de régression non paramétrique basée sur l'approche locale linéaire pondérée. Cette méthode présente des avantages par rapport aux autres méthodes à noyaux classiques que l'on verra par la suite. Si le noyau K est positif ou nul, l'estimateur de N-W vérifie

$$\widehat{m}_{NW}(\cdot) = \arg\min_{\alpha} \sum_{1\leq i\leq n} (T_i - \alpha)^2 K\left(\frac{x - X_i}{h_n}\right).$$

Ainsi, l'idée maîtresse de l'approche locale linéaire (LL) est une approximation des moindres carrés localement constante des valeurs T_i. Cette démarche est intuitive sachant que la fonction de régression $m(\cdot)$ est elle même solution du problème des moindres carrés. Le but de cette section est de présenter les estimateurs localement linéaires ainsi que leurs propriétés fondamentales.

1.6.1 Construction et définition des estimateurs localement linéaire

Le principe de l'estimation localement linéaire consiste en l'ajustement local d'un polynôme de degré 1 :

$$m(\cdot) \approx \alpha + \beta(\cdot - x) \tag{1.12}$$

aux données $\{(X_i, T_i),\ 1 \leq i \leq n\}$ par la méthode des moindres carrés où α et β sont des paramètres à estimé. De ce fait, nous supposons l'existence de la dérivée seconde de la fonction de régression $m(\cdot)$ au point x. Cette hypothèse est essentielle pour valider

théoriquement la construction de l'estimateur LL. Nous pouvons alors approximer la fonction de régression $m(x)$ par la fonction linéaire (1.12). Par un développement de Taylor au voisinage du point x,

$$m(X) \approx m(x) + m'(x)(X - x) \tag{1.13}$$

pour tout X au voisinage de x. Nous pouvons à présent ajuster notre fonction linéaire aux données $\{(X_i, T_i),\ 1 \leq i \leq n\}$ par la méthode des moindres carrés pondérées par un noyau K. Il faudra alors minimiser par rapport à (α, β) la fonction de perte suivante :

$$\arg\min_{\alpha,\beta} \sum_{1\leq i\leq n} (T_i - \alpha - \beta(X_i - x))^2 K\left(\frac{x - X_i}{h_n}\right). \tag{1.14}$$

Comme pour l'estimateur de N-W, les paramètres K et h_n désigne le noyau et la fenêtre. Notons que, lorsque $\beta = 0$ nous retrouvons l'estimateur de N-W. L'estimateur solution du problème de minimisation (1.14) est appelé *estimateur localement linéaire* et est défini par :

$$\widehat{m}_{LL}(x) = \frac{\displaystyle\sum_{1\leq i,j\leq n} T_i w_{i,j}^{LL}(x)}{\displaystyle\sum_{1\leq i,j\leq n} w_{i,j}^{LL}(x)}$$

où

$$w_{i,j}^{LL}(x) = (X_i - x)\Big((X_i - x) - (X_j - x)\Big) K\left(\frac{X_i - x}{h_n}\right) K\left(\frac{X_j - x}{h_n}\right).$$

Une autre forme de l'estimateur LL est possible et est donnée par :

$$\widehat{m}_{LL}(x) = \frac{\widehat{r_0}(x)\widehat{s_2}(x) - \widehat{r_1}(x)\widehat{s_1}(x)}{\widehat{s_0}(x)\widehat{s_2}(x) - \widehat{s_1^2}(x)}$$

où

$$\begin{aligned} \widehat{r_\ell}(x) &= \frac{1}{nh_n} \sum_{1\leq i\leq n} T_i (X_i - x)^\ell K\left(\frac{X_i - x}{h_n}\right), && \text{pour } \ell = 0, 1, \\ \widehat{s_\ell}(x) &= \frac{1}{nh_n} \sum_{1\leq i\leq n} (X_i - x)^\ell K\left(\frac{X_i - x}{h_n}\right), && \text{pour } \ell = 0, 1, 2. \end{aligned}$$

Nous établirons par la suite que l'estimateur LL est supérieur à l'estimateur à noyau de N-W au sens qu'il atténue les effets de bord. De plus, d'après Fan (1992), l'estimateur LL a un meilleur biais que l'estimateur de N-W et une meilleure variance que l'estimateur de Gasser and Muller (1979). Nous renvoyons le lecteur au livre de Fan and Gijbels (1996) pour plus de détails sur les propriétés de l'estimateur LL en plus de quelques applications sur des données réelles. L'ajustement local linéaire est une méthode attrayante tant du point de vue théorique que pratique.

1.7 Régression pour un modèle de censure

De manière analogue aux données complètes, la fonction de régression $m(\cdot)$, donnée par la relation (1.1) est estimée par :

$$\widehat{m}_{CR}(x) = \frac{\displaystyle\sum_{1\le i\le n} \frac{\delta_i Y_i}{\overline{G}_n(Y_i)} K\left(\frac{x-X_i}{h_n}\right)}{\displaystyle\sum_{1\le i\le n} K\left(\frac{x-X_i}{h_n}\right)}. \tag{1.15}$$

Un résultat de convergence uniforme presque sûre avec vitesse de $\widehat{m}_{CR}(\cdot)$ a été établit par :

Théorème 1.7.1 (Guessoum and Ould Saïd (2008)) *Sous des hypothèses classiques sur la fenêtre h_n, le noyau K et la fonction $m(\cdot)$, on a :*

$$\sup_{x\in\mathcal{C}} |\widehat{m}_{CR}(x) - m(x)| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right) + O(h_n^2) \qquad \text{quand } n \to \infty.$$

La normalité asymptotique de l'estimateur $\widehat{m}_{CR}(\cdot)$ est donnée par :

Théorème 1.7.2 (Guessoum and Ould Saïd (2008)) *Sous des hypothèses sur le noyau K, la fenêtre h_n et d'autres hypothèses de régularité de fonction, on a :*

$$\sqrt{nh_n}\,(\widehat{m}_{CR}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(x)\right).$$

où

$$\sigma^2(x) = \frac{r_2(x)f^2(x) - r_1^2(x)f(x)}{f^4(x)} \int_{\mathbb{R}} K^2(t)dt$$

avec $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi. La fonction $r_\ell(x) = \displaystyle\int \frac{t^\ell}{\overline{G}^\ell(t)} f(x,t)dt$ pour $\ell = 1, 2$ et $f(\cdot)$ est la densité de la co-variable X.

Indication sur la preuve du théorème. La preuve est basée sur l'inégalité exponentielle de Bernstein. Les auteurs ont utilisés une idée de couvrir l'ensemble compact par un nombre finis d'intervalle pour obtenir l'uniformité.

Le même estimateur dans le cadre dépendant (alpha mélangeant) et pour une co-variable d-dimensionnel a été étudié par Guessoum and Ould Saïd (2010, 2012). Deux résultats ont été établi, la convergence uniforme presque sûre sur un compact avec vitesse et la normalité asymptotique de l'estimateur étudié. Les deux résultats sont établis dans ce qui suit :

Théorème 1.7.3 (Guessoum and Ould Saïd (2010)) *Sous des hypothèses sur le noyau K, la fenêtre h_n, des hypothèses de régularité de fonction et d'autres sur le coefficient de mélange, on a :*

$$\sup_{x\in\mathcal{C}} |\widehat{m}_{CR}(x) - m(x)| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right) + O(h_n) \qquad \text{quand } n \to \infty.$$

Théorème 1.7.4 (Guessoum and Ould Saïd (2012)) *Sous des hypothèses sur le noyau K, la fenêtre h_n, des hypothèses de régularité de fonction et d'autres sur le coefficient de mélange, nous avons :*

$$\sqrt{nh_n^d}\,(\widehat{m}_{CR}(x)-m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma^2(x)\right).$$

où

$$\sigma^2(x)=\frac{r_2(x)f(x)-r_1^2(x)}{f^3(x)}\int_{\mathbb{R}^d}K^2(t)dt$$

où la fonction $r_\ell(\cdot)$ pour $\ell=1,2$ a été défini dans le théorème 1.7.2.

1.8 Choix de la fenêtre

Dans cette section, nous supposons le noyau K fixé, et nous nous intéressons qu'au choix de la fenêtre h_n qui est crucial. Nous avons observé que l'efficacité des estimateurs à noyaux et LL est liée au paramètre de lissage h_n. Il faut choisir la fenêtre afin d'équilibrer un terme stochastique (la variance) et un terme déterministe (le biais). Dans le cadre non paramétrique, notons que la vitesse de convergence dans l'estimation de la densité est de l'ordre de $c_1 n^{-1/5}$ qui est plus faible que la vitesse des modèles paramétrique, qui est $c_2 n^{-1/2}$ où c_1 et c_2 sont des constantes qui dépendent de quelques paramètres.

Nous nous intéressons dans notre étude au cadre non paramétrique. Dans ce dernier, du point de vue pratique, un paramètre trop petit fait apparaître des détails insignifiants dans la courbe estimée et au contraire une grande valeur de la fenêtre h_n, la majorité des détails sont effacés ou ne vont pas être considérer. Donc, il est clair que le choix du paramètre h_n est central dans l'estimation non paramétrique. Rappelons que dans la littérature, il y a principalement trois méthodes d'estimation : "rule of thumb", "plug-in" et "cross-validation". Chaque méthode possède des qualités et des inconvénients. Nous soulignons que la méthode la plus populaire (plus utilisée) est la méthode de validation croisée, son idée principale est de minimiser le critère suivant :

$$CV_{CR}=\frac{1}{n-1}\sum_{1\le i\le n}\left(Y_i-\widehat{m}_{CR}^{-i}(X_i)\right)^2, \tag{1.16}$$

où $\widehat{m}_{CR}^{-i}(\cdot)$ est l'estimateur de la fonction de régression classique $m(\cdot)$ par la méthode à noyau obtenu en supprimant le i ème triplet d'observations (X_i,Y_i,δ_i). Même si ce critère a le défaut d'être très variable et peut donner des sous estimations de h_n, il reste la méthode la plus utilisée. Dans toutes nos études numériques, nous avons fait le choix d'utiliser la méthode de la validation croisée.

1.8.1 Implémentation informatique

Tous les algorithmes et les graphiques ont été réalisés sur le logiciel *Matlab*. En ce qui concerne le chapitre 3, même si le résultat est d-dimensionnel, pour faciliter les simulations, nous nous somme placés dans le cas unidimensionnel ($X\in\mathbb{R}$) dans les deux cas : i.i.d. et α-mélangeant. Dans le cas de dépendance, nous avons choisi des co-variables issues d'un processus auto-régressif d'ordre 1 et de paramètre ρ. De plus, différents modèles sont pris en compte dans nos simulations : modèle linéaire,

parabolique, sinusoïdal et exponentiel. Pour chacun des cas, pour montrer la qualité pratique de nos estimateurs, nous avons observé l'effet de la taille d'échantillon, le taux de censure, l'effet des valeurs aberrantes, l'effet de la dépendance (forte et faible) ainsi que la contamination de l'aléa.

1.9 Contribution de la thèse

Cette thèse traite de l'estimation non-paramétrique de la fonction de régression pour des données incomplètes. Le manuscrit est structuré en quatre chapitres en plus de l'introduction et la conclusion.

Le chapitre introductif rappellera les principaux concepts ainsi que les propriétés les plus importantes qui nous permettent de mieux décrire les problèmes traités en situant au fur et à mesure les travaux antérieurs. Ensuite, nous présenterons brièvement les nouveaux résultats obtenus dans cette thèse.

Dans le second chapitre, nous considérons l'estimateur de la fonction de régression par la méthode linéaire locale en présence de censure aléatoire à droite. Nous établissons la convergence uniforme presque sûre de l'estimateur étudié. Nous concluons avec des résultats de simulation et une comparaison avec l'estimateur à noyau.

Le troisième chapitre concerne l'étude de la fonction de régression par la méthode à noyau lorsque la fonction de perte est l'erreur quadratique relative moyenne pour des données censurées aléatoirement à droite. Un estimateur de la fonction de régression relative pour des observations i.i.d. est proposé. La convergence uniforme presque sûre (sur un compact) avec vitesse ainsi que la normalité asymptotique sont établis. Une large étude numérique sur des données générées et un exemple sur un jeu de données réelles sont données.

Le quatrième chapitre est une extension du chapitre précédent au cas où les observations sont issues d'un processus α-mélangeant. La convergence uniforme presque sûre sur un compact avec vitesse a été établi. Des simulation sont faites pour renforcer notre résultat théorique.

Dans le dernier chapitre, nous établissons la normalité asymptotique de l'estimateur de la fonction de régression relative pour des données dépendantes. Les intervalles de confiances sont construite et une étude numérique est établie.

Les différentes partie de cette thèse sont reliées par une problématique commune qui est l'estimation non paramétrique de la fonction de régression. Deux approches sont proposées, la première est basée sur l'estimateur des moindres carrés tandis que la seconde repose sur l'estimateur des erreurs quadratiques relatives moyennes. Les deux approches sont comparées sur une étude sur données simulées. Chaque estimateur a été construit dans le but de trouver une méthode robuste et efficace aux traitement d'un jeu de données contenant de valeurs aberrantes et/ou censurées. L'objectif principal de cette contribution est de réduire la sensibilité de la fonction de régression aux valeurs aberrantes et à la censure.

1.9.1 Brève présentation des estimateurs étudiés dans cette thèse :

Soit $\{(X_i, Y_i, \delta_i), i = 1, \cdots, n\}$ n-réalisation i.i.d. de même loi que le triplet (X, Y, δ). Le premier estimateur que nous avons construit est l'estimateur de la fonction de régression

par la méthode linéaire locale, solution du problème de minimisation suivant :

$$\arg\min_{\alpha,\beta} \sum_{1\le i\le n} \left(\widehat{T}_i - \alpha - \beta(X_i - x)\right)^2 K\left(\frac{X_i - x}{h_n}\right) \tag{1.17}$$

où les données observées $\widehat{T}_i$ appelés données synthétiques sont explicités comme suit :

$$\widehat{T}_i = \frac{\delta_i Y_i}{\overline{G}_n(Y_i)}, \qquad \text{pour } 1 \le i \le n$$

ici δ_i représente l'indicateur de non censure et $\overline{G}_n(\cdot)$ est l'estimateur de Kaplan and Meier (1958) (K-M) défini dans (1.5). La solution au problème de minimisation (1.17) est donnée par :

$$\widehat{m}_{LLR}(x) = \frac{\displaystyle\sum_{1\le i,j\le n} \widehat{T}_j w_{i,j}(x)}{\displaystyle\sum_{1\le i,j\le n} w_{i,j}(x)}$$

avec

$$w_{i,j}(x) = \Big((X_i - x) - (X_j - x)\Big)(X_i - x) K\left(\frac{X_i - x}{h_n}\right) K\left(\frac{X_j - x}{h_n}\right).$$

Le second estimateur que nous avons étudié est l'estimateur de fonction de régression lorsque la fonction de perte est l'erreur quadratique relative moyenne. Cet estimateur est solution de :

$$\arg\min_{\alpha} \sum_{1\le i\le n} \left(\frac{\widehat{T}_i - \alpha}{\widehat{T}_i}\right)^2 K_d\left(\frac{X_i - x}{h_n}\right) \tag{1.18}$$

où cette fois-ci les données observées $\widehat{T}_i^{-\ell}$ pour $\ell = 1, 2$ sont données par :

$$\widehat{T}_i^{-\ell} = \frac{\delta_i Y_i^{-\ell}}{\overline{G}_n(Y_i)}, \qquad \text{pour } 1 \le i \le n.$$

δ_i et $\overline{G}_n(\cdot)$ sont respectivement l'indicateur de non censure et l'estimateur de K-M de la fonction de survie de la variable de censure C donné par la formule (1.5). L'estimateur de $m(\cdot)$ est alors solution du problème de minimisation (1.18) donné par :

$$\widehat{m}_{RER}(x) =: \frac{\widehat{m}_1(x)}{\widehat{m}_2(x)}$$

avec

$$\widehat{m}_\ell(x) = \frac{1}{n h_n^d \widehat{f}(x)} \sum_{1\le i\le n} \widehat{T}_i^{-\ell} K_d\left(\frac{x - X_i}{h_n}\right)$$

pour $\ell = 1, 2$ et $\widehat{f}(\cdot)$ est l'estimateur à noyau de la densité marginale $f(\cdot)$ défini par Parzen (1962) et Rosenblatt (1956a) séparément.

1.9.2 Brève présentation des résultats de convergence obtenus dans cette thèse :

Nous présentons maintenant les principaux résultats obtenus durant cette thèse. Deux résultats ont été établis pour deux estimateurs différents $\widehat{m}_{RER}(x)$ et $\widehat{m}_{LLR}(x)$ pour tout x appartenant à un ensemble compact $\mathcal{C}$.

Méthode locale linéaire

Résultats : Convergence cas i.i.d.

Théorème 1.9.1 *Sous des conditions standard sur le noyau, la fenêtre et des conditions de régularités de fonction, on a*

$$\sup_{x\in\mathcal{C}}|\widehat{m}_{LLR}(x)-m(x)| = O_{p.s.}\left\{\max\left(\sqrt{\frac{\log n}{nh_n^2}}+h_n^{\nu}\right)\right\} \qquad \textit{lorsque} \quad n\to\infty.$$

Pour $\nu>0$ et $\mathcal{C}$ est un ensemble compacte sur $\mathbb{R}$ et la notation p.s. désigne presque sûrement.

Corollaire 1.9.1 *En allégeant les conditions de régularités cités dans le théorème précédent, on obtient un résultat de convergence sans vitesse donné ci-dessous :*

$$\lim_{n\to\infty}\sup_{x\in\mathcal{C}}|\widehat{m}_{LLR}(x)-m(x)| = 0.$$

Méthode à noyau

Résultats : Cas α-mélangeant (d-dimensionnel)

On suppose que le triplet (X,Y,δ) vérifie la condition de mélange dans la définition 1.3.1. Soit $\mathcal{C}$ un compact inclut dans $\mathcal{C}_0=\left\{x\in\mathbb{R}^d \,/\, f(x)>0\right\}$ où $f(\cdot)$ est la densité marginale de la co-variable X.

Théorème 1.9.2 *Sous des hypothèses sur le coefficient de mélange et des conditions standard sur le noyau et la fenêtre, ainsi que d'autres hypothèses de régularité et techniques, nous avons :*

$$\sup_{x\in\mathcal{C}}|\widehat{m}_{RER}(x)-m(x)| = O(h_n)+O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}+\sqrt{h_n^{d(\nu-2)}\log n}\right), \qquad \textit{lorsque} \quad n\to\infty.$$

De plus, définissons $\mathcal{C}^\star=\{\forall x\in\mathcal{C},\ r_\lambda(x)\neq 0 \ \text{ et } \ \mu_\ell(x)\neq 0\}$ pour $\ell=1,2$ et $\lambda=2,3,4$ où

$$r_\lambda(x)=\int\frac{t^{-\lambda}}{\overline{G}(t)}f(x,t)dt \qquad \text{et} \qquad \mu_\ell(x)=\int t^{-\ell}f(x,t)dt.$$

Notons que $f(\cdot,\cdot)$ est la densité jointe du couple (X,T).

Théorème 1.9.3 *Sous les hypothèses sur le coefficient de mélange et des conditions standard sur le noyau et la fenêtre, ainsi que d'autres hypothèses de régularité et techniques, pour n suffisamment grand, nous avons :*

$$\sqrt{nh_n^d}\,(\widehat{m}_{RER}(x)-m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\sigma^2(x)\right), \qquad \textit{pour } x\in\mathcal{C}^\star$$

où

$$\sigma^2(x) = \frac{r_2(x)\mu_2^2(x) - 2\mu_1(x)\mu_2(x)r_3(x) + \mu_1^2(x)r_4(x)}{\mu_2^4(x)} \int_{\mathbb{R}^d} K_d^2(s)ds \tag{1.19}$$

avec $\xrightarrow{\mathcal{L}}$ *qui désigne la convergence en loi.*

Établir la normalité asymptotique de l'estimateur défini, nous permet de construire nos intervalles de confiances. Pour cela, nous devons estimer l'expression de la variance asymptotique (1.19). Ceci est possible en estimant les quantités qui la compose. Par conséquent, nous avons :

$$\widehat{\sigma}^2(x) = \frac{\widehat{r}_2(x)\widehat{\mu}_2^2(x) - 2\widehat{\mu}_1(x)\widehat{\mu}_2(x)\widehat{r}_3(x) + \widehat{\mu}_1^2(x)\widehat{r}_4(x)}{\widehat{\mu}_2^4(x)} \int_{\mathbb{R}^d} K_d^2(s)ds. \tag{1.20}$$

En remplaçant (1.20) dans le théorème 1.9.3, nous avons :

Corollaire 1.9.2 *Sous les hypothèses du théorème 1.9.3, nous avons :*

$$\sqrt{nh_n^d}\,(\widehat{m}_{RER}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \widehat{\sigma}^2(x)\right).$$

Un intervalle de confiance au niveau $1-\beta$ *avec* $0 < \beta < 1$ *est donné par*

$$\left]\widehat{m}_{RER}(x) - t_{1-\frac{\beta}{2}} \frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}};\quad \widehat{m}_{RER}(x) + t_{1-\frac{\beta}{2}} \frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}}\right[$$

où $t_{1-\frac{\beta}{2}}$ *est le quantile de la loi normale standard* $\mathcal{N}(0,1)$.

Les démonstrations de ces résultats et le détail des conditions imposées seront donnés aux chapitre 2–5.

Chapitre 2

Estimation non paramétrique locale linéaire de la fonction de régression pour des données censurées

[1] Dans ce chapitre, nous établissons un résultat de convergence uniforme presque sûre pour un estimateur de la fonction de régression en utilisant la méthode linéaire locale. Le but ici est de construire un estimateur résistant aux effets de bord pour des données censurées aléatoirement à droite. Des simulations ont été faites pour consolider notre résultat théorique et une comparaison avec l'estimateur à noyau classique confirme la supériorité de la nouvelle approche.

2.1 Modèle

Nous considérons le modèle de régression non paramétrique (1.1) dans lequel le couple aléatoire (X, T) est tel que T correspond à une durée de survie (ou v.a.) qui peut être censurée à droite par une autre v.a. C indépendante de T. Notre but est d'estimer l'espérance conditionnelle de T sachant X dans une optique non paramétrique. Dans le cadre non censuré, une grande variété d'estimateurs non paramétriques de la fonction de régression ont été proposés dans la littérature en commençant par Nadaraya (1964), Watson (1964), Gasser and Muller (1979) (G-M) et Fan (1992).

L'idée de base de l'estimation locale linéaire (LL) est d'utiliser la moyenne locale des données proche de x pour construire un estimateur de la fonction de régression en x. En effet, on estime que les observations en X_i proche de x doivent contenir les informations sur la valeur de la fonction de régression en x. Plus formellement, la fonction de régression non paramétrique satisfait :

$$m(x) = \arg\min_{\alpha} \mathbb{E}\Big[(T-\alpha)^2 | X = x\Big].$$

Soit le n-échantillon $(X_1, T_1), \cdots, (X_n, T_n)$ ayant la même loi que le couple (X, T). La

1. En collaboration avec le Pr. E. OULD SAÏD et le Pr. R. M. REMITA. En révision mineure dans le journal Communication in Statistics : Theory and Methods.

fonction de régression $m(x)$ peut être estimée par :

$$\widehat{m}(x) = \arg\min_{\alpha} \sum_{1\leq i\leq n} (T_i - \alpha)^2 K\left(\frac{X_i - x}{h_n}\right) \tag{2.1}$$

où h_n est appelée fenêtre qui est une suite de nombres réelles qui tendent vers 0 à l'infini et K est appelé noyau qui est une densité de probabilité. La solution du problème de minimisation (2.1) conduit à l'estimateur à noyau de N-W donné par :

$$\widehat{m}(x) = \frac{\sum_{1\leq i\leq n} T_i K\left(\frac{X_i - x}{h_n}\right)}{\sum_{1\leq i\leq n} K\left(\frac{X_i - x}{h_n}\right)}.$$

Bien que cet estimateur est connu pour avoir de bonnes caractéristiques tels que la simplicité, la flexibilité et la consistance, son biais asymptotique dépend de la dérivée de la fonction de densité marginale. De plus, il est soumit à des effets de bord. Pour plus de détails concernant les inconvénients de l'estimateur à noyau de N-W et pour une comparaison entre les estimateurs N-W, G-M et LL nous renvoyons le lecteur à l'article de Fan (1992), Fan and Gijbels (1996) et Fan and Yao (2003). Pour surmonter ces inconvénients, Fan (1992) a proposé une nouvelle approche basée sur l'ajustement linéaire locale qui généralise la régression classique à noyau.

Nous supposons maintenant que la fonction de régression $m(\cdot)$ est dérivable et que sa dérivée est finie (c-à-d $m'(x) < \infty$). En se basant sur l'idée que $m(X) \approx m(x)+m'(x)(X-x) \equiv \alpha + \beta(X-x)$ dans un voisinage du point x, l'estimateur de la fonction de régression par la méthode linéaire locale est solution du problème de minimisation suivant :

$$\arg\min_{\alpha,\beta} \sum_{1\leq i\leq n} (T_i - \alpha - \beta(X_i - x))^2 K\left(\frac{X_i - x}{h_n}\right). \tag{2.2}$$

Par un simple calcul algébrique, on a :

$$\widehat{m}(x) = \frac{\sum_{1\leq i,j\leq n} w_{i,j}(x) T_j}{\sum_{1\leq i,j\leq n} w_{i,j}(x)}$$

où

$$w_{i,j}(x) = (X_i - x)\Big((X_i - x) - (X_j - x)\Big) K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_j - x}{h}\right). \tag{2.3}$$

Ainsi $\widehat{m}(\cdot)$ est l'estimateur LL de la fonction de régression lorsque les données sont complètement observées. Cet estimateur possède de nombreuses propriétés statistiques intéressantes (nous renvoyons le lecteur à Fan (1992)).

Dans le contexte des données censurées à droite, parmi les travaux dédiés à l'approche locale linéaire nous citons les travaux de Fan and Gijbels (1994) qui ont établi la convergence en probabilité de l'estimateur de la fonction de régression. Cai (2003) qui a

proposé un estimateur de la fonction de régression. Nous soulignons que ce dernier n'a pas pénalisé les observations par la fonction de survie de la variable de censure. Dans le même contexte, Kim and Truong (1998) ont utilisé la méthode locale linéaire pour définir un estimateur de la fonction de hasard conditionnelle.
Ayant en tête le schéma de censure aléatoire à droite, nous n'observons plus la variable T mais le couple (Y,δ) (voir la sous-section 1.2.4). Sous ce modèle, nous considérons une adaptation spécifique de notre variable observée donnée par :

$$\widetilde{T}_i = \frac{\delta_i Y_i}{\overline{G}(Y_i)}, \qquad \text{pour } 1 \leq i \leq n.$$

où $\overline{G}(\cdot)$ désigne la fonction de survie de la variable de censure C. Ce quotient est appelé "données synthétiques" et permet de prendre en compte l'effet de la censure dans la construction de l'estimateur. Les données synthétiques ont été introduit par Carbonez et al. (1995) et utilisées par Köhler et al. (2002), Guessoum and Ould Saïd (2008), Kebabi and Messaci (2012), … et bien des auteurs. Dans ce qui suit, nous supposerons que :

$$(X_i, T_i)_i \text{ et } (C_i)_i \text{ sont indépendants} \tag{2.4}$$

pour $i = 1,\cdots,n$. En utilisant la propriété de l'espérance conditionnelle et sous la condition d'indépendance , nous avons pour tout x fixé :

$$\begin{aligned}
\mathbb{E}\left[\widetilde{T}_1|X_1 = x\right] &= \mathbb{E}\left[\frac{\delta_1 Y_1}{\overline{G}(Y_1)}\middle|X_1 = x\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}_{\{T_1\leq C_1\}}T_1}{\overline{G}(T_1)}\middle|T_1\right]\middle|X_1 = x\right] \\
&= \mathbb{E}\left[\frac{T_1}{\overline{G}(T_1)}\mathbb{E}\left[\mathbb{1}_{\{T_1\leq C_1\}}|T_1\right]\middle|X_1 = x\right] \\
&= \mathbb{E}[T_1|X_1 = x].
\end{aligned}$$

Nous étendons le problème de minimisation (2.2) au cas censuré en substituant T par $\widetilde{T}$. Le problème d'estimation de la fonction $m(\cdot)$ devient alors :

$$\arg\min_{\alpha,\beta} \sum_{1\leq i\leq n} \left(\widetilde{T}_i - \alpha - \beta(X_i - x)\right)^2 K\left(\frac{X_i - x}{h_n}\right). \tag{2.5}$$

Un calcul élémentaire donne ce qu'on appelle un "pseudo estimateur" défini par :

$$\widetilde{m}_{LLR}(x) = \frac{\sum_{1\leq i,j\leq n} w_{i,j}(x)\widetilde{T}_j}{\sum_{1\leq i,j\leq n} w_{i,j}(x)} =: \frac{\widetilde{\mu}_1(x)}{\widetilde{\mu}_0(x)} \tag{2.6}$$

où $w_{i,j}(x)$ est définie dans (2.3). Ce dernier est un estimateur incalculable du fait que la fonction de survie $\overline{G}(\cdot)$ est inconnue. De la nécessité d'obtenir un estimateur quantifiable, nous remplaçons la fonction de survie par l'estimateur de Kaplan and Meier (1958)

défini dans (1.5), on a :

$$\widehat{T}_i = \frac{\delta_i Y_i}{\overline{G}_n(Y_i)}, \qquad \text{pour} \quad 1 \leq i \leq n, \tag{2.7}$$

En substituant (2.7) dans (2.6), on obtient :

$$\widehat{m}_{LLR}(x) = \frac{\sum\limits_{1\leq i,j\leq n} w_{i,j}(x)\widehat{T}_j}{\sum\limits_{1\leq i,j\leq n} w_{i,j}(x)} =: \frac{\widehat{m}_1(x)}{\widehat{m}_0(x)}, \qquad \left(\frac{0}{0} =: 0\right). \tag{2.8}$$

$\widehat{m}(\cdot)$ est appelé estimateur de la fonction de régression linéaire locale (LLR pour "local linear regression").

Remarque 2.1.1 *En considérant β nul dans (2.5), l'estimateur résultant est celui de la fonction de régression classique (CR pour "classical regression") défini par (1.15) dans Guessoum and Ould Saïd (2008).*

2.2 Hypothèses et principaux résultats

Soit $\mathcal{C}_0 = \{x \in \mathbb{R}/f(x) > 0\}$ et $\mathcal{C}$ un sous ensemble compact de $\mathcal{C}_0$. Nous supposons que pour toute fonction de répartition (f.d.r.) Q, on a $\tau_Q = \sup\{x, Q(x) < 1\}$ l'extrémité supérieure du support de Q. De plus, nous assumons qu'il existe un $\tau > 0$ tel que $\tau < \tau_H$ et $0 < \overline{G}(\tau_H) < \overline{G}(\tau)$.
Pour éviter toute confusion, nous désignons par C toute constante positive. De plus, comme T est une durée de vie, nous supposons qu'elle est bornée.
Les hypothèses nécessaire à la démonstration de notre résultat sont regroupées et listées juste en dessous.

A1. La fenêtre h_n satisfait $\lim\limits_{n\to\infty} h_n = 0$, $\lim\limits_{n\to\infty} \dfrac{nh_n^2}{\log n} = +\infty$ et $\lim\limits_{n\to\infty} \dfrac{nh_n^6}{\log n} = 0$.
A2. Le noyau $K(\cdot)$ est une fonction bornée, positive et symétrique.
A3. La fonction de densité $f(\cdot)$ est continûment différentiable.
A4. La fonction $s_0(\cdot)$ est continûment différentiable.
A5. Il existe $C > 0$ et $\nu > 0$ tel que :

$$\forall (x, y) \in \mathcal{C}^2 \qquad |m(x) - m(y)| \leq C|x - y|^\nu.$$

Quelques remarques sur les hypothèses. L'hypothèse A1 concerne la fenêtre et est standard en estimation non paramétrique. L'hypothèse A2 porte sur le noyau et est nécessaire pour la convergence des termes du biais et de la variance. La majorité des noyaux vérifient cette hypothèse par exemple le noyau d'epanechnicov et gaussien. Les hypothèses A3, A4 et A5 sont des conditions techniques de régularité pour les fonctions $f(\cdot)$, $s_0(\cdot)$ et $m(\cdot)$ respectivement.

Le théorème suivant donne la convergence uniforme presque sûre de $\widehat{m}_{LLR}(\cdot)$ sous le compact $\mathcal{C}$ avec vitesse.

Théorème 2.2.1 *Sous les hypothèses A1–A5, nous avons :*

$$\sup_{x\in\mathcal{C}}|\widehat{m}_{LLR}(x)-m(x)| = O(h_n^v)+O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \qquad \textit{quand} \quad n\to\infty.$$

Remarque 2.2.1 *Un résultat de convergence forte (sans vitesse) pourrait être obtenu en allégeant les hypothèses. Les hypothèses A3 et A4 doivent être remplacées par les conditions respectives suivantes :*

A3 ′. La fonction de densité $f(\cdot)$ est continue au point x.

A4 ′. La fonction $s_0(\cdot)$ est continue au point x.

Théorème 2.2.2 *Sous les mêmes hypothèses que le celle du Théorème 2.2.1 et en remplaçant les hypothèses A3 et A4 par leurs versions analogues A3 ′ et A4 ′, nous avons :*

$$\lim_{n\mapsto+\infty}\sup_{x\in\mathcal{C}}|\widehat{m}_{LLR}(x)-m(x)| = 0 \qquad p.s.$$

La preuve de ce résultat peut être adaptée de celle du Théorème 2.2.1 qui est donnée dans la section 2.4.

La preuve du théorème 2.2.1 est basée sur la décomposition suivante :

$$\widehat{m}_{LLR}(x)-m(x) =: \mathcal{B}_1(x)+\frac{1}{\widehat{m}_0(x)}\{-\mathcal{B}_1(x)\mathcal{B}_2(x)+\mathcal{B}_3(x)+\mathcal{B}_4(x)-m(x)\mathcal{B}_2(x)\}$$

avec

$$\begin{aligned}\mathcal{B}_1(x) &:= \frac{\mathbb{E}[\widetilde{m}_1(x)]}{\mathbb{E}[\widehat{m}_0(x)]}-m(x), \quad \mathcal{B}_2(x):=\widehat{m}_0(x)-\mathbb{E}[\widehat{m}_0(x)],\\ \mathcal{B}_3(x) &:= \widehat{m}_1(x)-\widetilde{m}_1(x) \quad \text{et} \quad \mathcal{B}_4(x):=\widetilde{m}_1(x)-\mathbb{E}[\widetilde{m}_1(x)].\end{aligned}$$

En utilisant l'inégalité triangulaire, nous obtenons :

$$\begin{aligned}\sup_{x\in\mathcal{C}}|\widehat{m}_{LLR}(x)-m(x)| \;\le\; &\inf_{x\in\mathcal{C}}|\mathcal{B}_1(x)|+\frac{1}{\inf_{x\in\mathcal{C}}|\widehat{m}_0(x)|}\Big\{\sup_{x\in\mathcal{C}}|\mathcal{B}_1(x)\mathcal{B}_2(x)|\\ &+\sup_{x\in\mathcal{C}}|\mathcal{B}_3(x)|+\sup_{x\in\mathcal{C}}|\mathcal{B}_4(x)|+\sup_{x\in\mathcal{C}}|m(x)\mathcal{B}_2(x)|\Big\}.\end{aligned}$$

La preuve sera ainsi obtenue par les propositions suivantes :

Proposition 2.2.1 *Sous les hypothèses A1 et A5, nous avons :*

$$\sup_{x\in\mathcal{C}}|\mathcal{B}_1(x)| = O(h_n^v) \quad \textit{quand} \quad n\to\infty.$$

Proposition 2.2.2 *Sous les hypothèses A1–A3, nous avons :*

$$\sup_{x\in\mathcal{C}}|\mathcal{B}_2(x)| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \quad \textit{quand} \quad n\to\infty.$$

Corollaire 2.2.1 *Sous les hypothèses de la proposition 2.2.2, il existe un nombre réel* Γ *strictement positif tel que :*

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{x\in\mathcal{C}} \widehat{m}_0(x) \leq \Gamma\right) < \infty.$$

Proposition 2.2.3 *Sous les hypothèses A1–A4, nous avons :*

$$\sup_{x\in\mathcal{C}} |\mathcal{B}_3(x)| = O_{p.s.}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad quand \quad n \to \infty.$$

Proposition 2.2.4 *Sous les hypothèses A1–A4, nous avons :*

$$\sup_{x\in\mathcal{C}} |\mathcal{B}_4(x)| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \quad quand \quad n \to \infty.$$

Remarque 2.2.2 *Nous rappelons que le choix du noyau K n'affecte pas la qualité de l'estimation contrairement à celui de la fenêtre* h_n*. En effet, l'efficacité de l'estimateur est liée au paramètre de lissage. Nous avons choisi d'utiliser la méthode de validation croisée, où l'idée principale consiste à minimiser, par rapport à* h_n*, le critère suivant :*

$$CV_{LLR} = \frac{1}{n-1} \sum_{1\leq i\leq n} \left(Y_i - \widehat{m}_{LLR}^{i}(X_i)\right)^2 \tag{2.9}$$

où $\widehat{m}_{LLR}^{i}(\cdot)$ *est l'estimateur LLR donné par (2.8) construit avec les* $(n-1)$ *triplets aléatoires* $\{(X_1, Y_1, \delta_1), \cdots, (X_{i-1}, Y_{i-1}, \delta_{i-1}), (X_{i+1}, Y_{i+1}, \delta_{i+1}) \cdots, (X_n, Y_n, \delta_n)\}$.

2.3 Étude numérique

Des simulations sur des données générées sont menées pour évaluer les performances de l'estimateur $\widehat{m}_{LLR}(\cdot)$ donné par (2.8) et comparer son efficacité et sa robustesse par rapport à l'estimateur de la fonction de régression à noyau classique défini dans (1.15). Dans toutes les courbes présentées, nous prenons un noyau gaussien K. En ce qui concerne la fenêtre, cette dernière est sélectionnée par la méthode de validation croisée (voir remarque 2.2.2). On génère n points selon le modèle suivant : $T_i = X_i + 0.2\,\epsilon_i$ ou $X_i \rightsquigarrow \mathcal{N}(0,1)$ et $\epsilon_i \rightsquigarrow \mathcal{N}(0,1)$. Le temps de censure est généré suivant une loi normale $C_i \rightsquigarrow \mathcal{N}(c,1)$ où c est une constante qui ajuste le pourcentage de censure (C.P.). On calcule les données synthétiques selon la formule (2.7) où l'estimateur de K-M est défini en (1.5).

De figure 1 à la figure 3, nous pouvons constaté que l'estimateur LLR est plus efficace lorsque la taille d'échantillon n augmente. De plus, la qualité d'estimation est légèrement affectée par le C.P. mais reste assez résistante et proche de la courbe théorique. En ce qui concerne la comparaison en figure 2.3, nous pouvons voir que les deux estimateurs CR et LLR se comporte de la même manière lorsque la taux de censure est faible. Une remarque importante et qui constitue l'un des avantages majeurs de cette méthode est la résistance de notre estimateur aux effets de bords lorsque l'estimateur est contraint à un fort taux de censure (nous pouvons visualiser cette propriété à travers la figure 2.3). Contrairement à cela, l'estimateur CR est sensible à l'effet de la censure, qui est visible sur les bords.

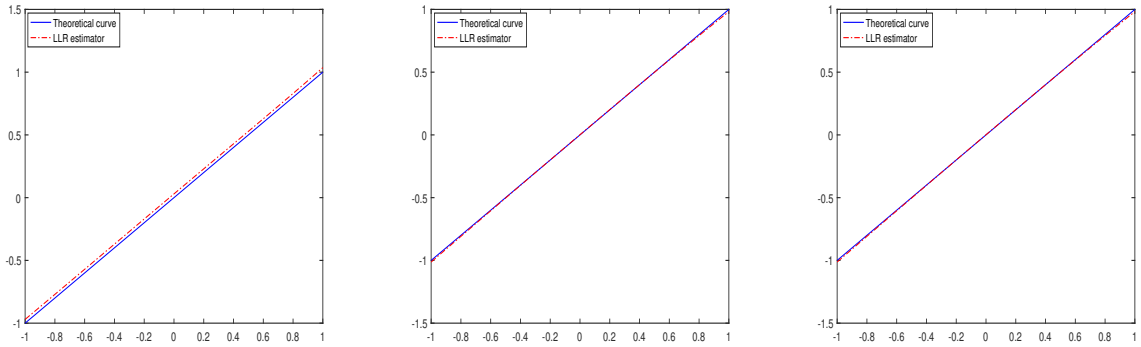


FIGURE 2.1 – $\widehat{m}_{LLR}(\cdot)$ avec C.P.≈ 30% pour $n = 100, 300$ et 500 respectivement.

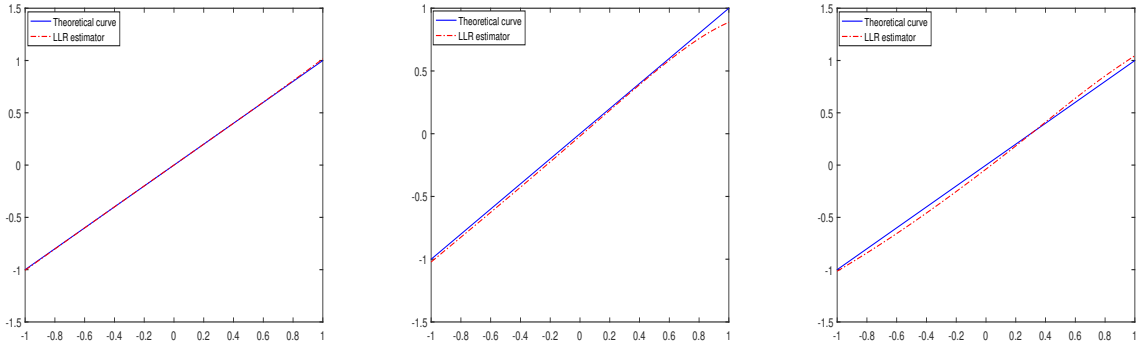


FIGURE 2.2 – $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ pour C.P. $\approx 8, 25$ et 60% respectivement.

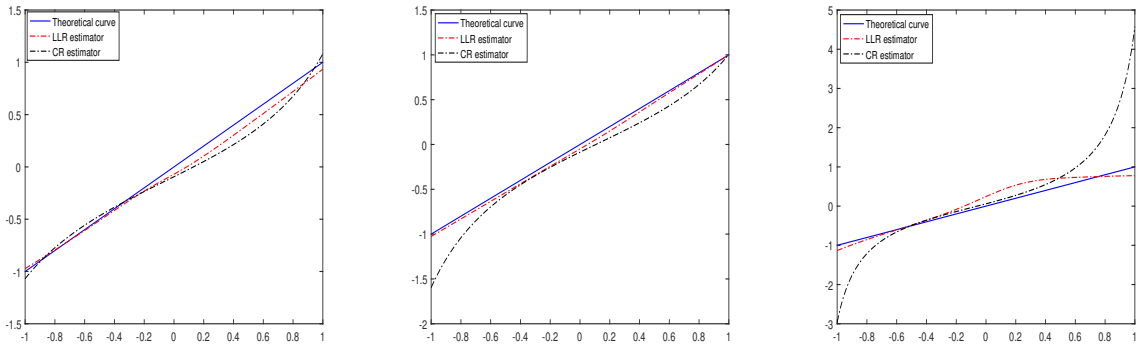


FIGURE 2.3 – $\widehat{m}_{LLR}(\cdot)$, $\widehat{m}_{CR}(\cdot)$ avec $n = 300$ pour C.P. $\approx 10, 35$ et 65% respectivement.

Dans le tableau 2.1, nous prenons différentes valeurs de C.P. et nous reportons l'erreur quadratique moyenne (EQM) des estimateurs LLR et CR. Nous pouvons voir que l'estimateur LLR fonctionne mieux lorsque la taille de l'échantillon augmente et qu'il n'est que légèrement affecté par le pourcentage de données observées.

Tableau 2.1 – Tableau comparatif des EQM.

C.P.	n	CR	LLR
10	100	0.0158	0.0011
	300	0.0024	0.0003
	500	2.48×10^{-4}	2.32×10^{-6}
30	100	0.0836	0.0025
	300	0.0473	0.0020
	500	0.0108	8.10×10^{-4}
50	100	0.0611	0.1181
	300	0.2321	0.0228
	500	0.0258	0.0064

2.4 Preuves et résultats auxiliaires

Avant d'entamer les preuves de propositions 2.2.1–2.2.4, nous avons besoin de définir les quantités suivantes :

$$\widehat{s_\ell}(x) = \frac{1}{nh_n} \sum_{1\le i\le n} \widehat{T_i}(X_i - x)^\ell K\left(\frac{X_i - x}{h}\right), \qquad \text{pour } \ell = 0, 1,$$

$$\widetilde{s_\ell}(x) = \frac{1}{nh_n} \sum_{1\le i\le n} \widetilde{T_i}(X_i - x)^\ell K\left(\frac{X_i - x}{h}\right), \qquad \text{pour } \ell = 0, 1,$$

et

$$\widehat{r_\ell}(x) = \frac{1}{nh_n} \sum_{1\le i\le n} (X_i - x)^\ell K\left(\frac{X_i - x}{h}\right), \qquad \text{pour } \ell = 0, 1, 2.$$

Preuve de la proposition 2.2.3. Nous considérons la décomposition suivante :

$$\begin{aligned} |\widehat{m}_1(x) - \widetilde{m}_1(x)| &= |\widehat{s_0}(x)\widehat{r_2}(x) - \widehat{s_1}(x)\widehat{r_1}(x) - \widetilde{s_0}(x)\widehat{r_2}(x) + \widetilde{s_1}(x)\widehat{r_1}(x)| \\ &\le |\widehat{r_2}(x) - \mathbb{E}[\widehat{r_2}(x)]| \times |\widehat{s_0}(x) - \widetilde{s_0}(x)| + |\mathbb{E}[\widehat{r_2}(x)]\,(\widehat{s_0}(x) - \widetilde{s_0}(x))| \\ &+ |\widehat{r_1}(x) - \mathbb{E}[\widehat{r_1}(x)]| \times |\widehat{s_1}(x) - \widetilde{s_1}(x)| + |\mathbb{E}[\widehat{r_1}(x)]\,(\widehat{s_1}(x) - \widetilde{s_1}(x))|. \end{aligned} \tag{2.10}$$

Nous énonçons et prouvons ensuite le lemme 2.4.1–Lemme 2.4.3 qui sont requis à la preuve de la proposition 2.2.3.

Lemme 2.4.1 *Sous les hypothèses A1, A2 et A3 pour $\ell = 0, 1, 2$ nous avons :*

$$\sup_{x\in\mathcal{C}} |\widehat{r_\ell}(x) - \mathbb{E}[\widehat{r_\ell}(x)]| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}\right) \quad \textit{quand} \quad n \to \infty.$$

Preuve du Lemme 2.4.1. Nous considérons une suite i.i.d. $(X_1, Y_1, \delta_1), \ldots, (X_n, Y_n, \delta_n)$ et nous définissons :

$$\Phi_n = \left\{ \theta_x : \mathbb{R} \to \mathbb{R}^+ / \theta_x(u) = \frac{(u - x)^\ell}{nh_n} K\left(\frac{x - u}{h_n}\right), \quad x \in \mathcal{C} \right\}.$$

D'après la condition (b) du lemme 3 de Giné and Guillou (1999), Φ_n est une Vapnik-Cervonenkis (V-C) classe de fonctions mesurables uniformément bornée pour $\ell = 0, 2$.

Sous la condition (c) du même lemme pour $\ell = 1$, posons $u - x = (u - x + h_n) - h_n$ alors Φ_n est aussi une V-C classe d'enveloppe $\Theta = \dfrac{h_n^{\ell-1}}{n}\|K\|_\infty$. De plus, nous gardons cette écriture pour harmoniser les notations. Sous les hypothèses A2 et A3 nous obtenons :

$$\sup_{x\in\mathcal{C}} \theta_x(X_1) \leq \frac{h_n^{\ell-1}}{n}\|K\|_\infty\|f\|_\infty = \frac{h_n^{\ell-1}}{n}\xi_1 =: U_n.$$

De la même manière, nous avons :

$$\begin{aligned}\sup_{x\in\mathcal{C}} Var\left[\theta_x(X_1)\right] &\leq \sup_{x\in\mathcal{C}} \mathbb{E}\left[\theta_x^2(X_1)\right] \\ &\leq \frac{h_n^{2\ell}\|K\|_2^2\|f\|_\infty}{n^2h_n^2} = \frac{h_n^{2\ell-2}}{n^2}\xi_2 =: \sigma_n^2\end{aligned}$$

avec $\sigma_n \leq U_n$ pour n assez grand.
Maintenant, nous pouvons appliquer l'inégalité de Talagrand (voir proposition 5.4.2). IL existe trois constantes positives C_1, C_2 et C_3 tel que pour $t \geq C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}$, nous obtenons :

$$\mathbb{P}\left[\sup_{\theta_x\in\Phi_n}\left|\sum_{1\leq i\leq n}(\theta_x(X_i) - \mathbb{E}[\theta_x(X_1)])\right| > C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}\right]$$

$$\leq C_2\exp\left(-\frac{C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}}{C_2\frac{h_n^{\ell-1}}{n}\xi_1}\log\left[1 + \frac{C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}\frac{h_n^{\ell-1}}{n}\xi_1}{C_2\left(\sqrt{n}\frac{h_n^{\ell-1}}{n}\sqrt{\xi_2} + \frac{h_n^{\ell-1}}{n}\xi_1\sqrt{\log C_3\frac{\xi_1}{\sqrt{\xi_2}}}\right)^2}\right]\right),$$

En utilisant l'approximation $\log(1+x) \approx x($ pour $x \to 0)$, la partie de droite de la dernière équation devient de l'ordre de

$$C_2\exp\left(-\frac{C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}}{C_2\frac{h_n^{\ell-1}}{n}\xi_1} \times \frac{C_1\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}\frac{h_n^{\ell-1}}{n}\xi_1}{C_2\left(\sqrt{n}\frac{h_n^{\ell-1}}{n}\sqrt{\xi_2}\right)^2}\right) = C_2n^{-\frac{C_1^2}{C_2^2\xi_2}},$$

qui pour des choix appropriés de C_1, C_2 et ξ_2 et sous l'hypothèse A1 nous obtenons un terme général d'une série convergente de l'ordre de $O(n^{-3/2})$. Ensuite, par le lemme de Borel-Cantelli nous obtenons le résultat.

Lemme 2.4.2 *Sous les hypothèses A1, A2 et A4 pour $\ell = 0, 1$, nous avons :*

$$\sup_{x\in\mathcal{C}}|\widehat{s_\ell}(x) - \widetilde{s_\ell}(x)| = O_{p.s.}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad quand \quad n \to \infty.$$

Preuve du Lemme 2.4.2. Pour $\ell = 0, 1$, nous avons

$$\sup_{x\in\mathcal{C}}|\widehat{s_\ell}(x) - \widetilde{s_\ell}(x)| = \sup_{x\in\mathcal{C}}\left|\frac{1}{nh_n}\sum_{1\leq i\leq n}\widehat{T_i}(X_i - x)^\ell K\left(\frac{X_i - x}{h_n}\right) - \frac{1}{nh_n}\sum_{1\leq i\leq n}\widetilde{T_i}(X_i - x)^\ell K\left(\frac{X_i - x}{h_n}\right)\right|$$

$$
\begin{aligned}
&\leq \sup_{x\in\mathcal{C}}\left|\frac{1}{nh_n}\sum_{1\leq i\leq n}T_i(X_i-x)^\ell K\left(\frac{X_i-x}{h_n}\right)\left(\frac{1}{\bar{G}_n(T_i)}-\frac{1}{\bar{G}(T_i)}\right)\right| \\
&\leq \frac{1}{\overline{G}^2(\tau)}\sup_{t\leq\tau}\left|\overline{G}_n(t)-\overline{G}(t)\right|\times\sup_{x\in\mathcal{C}}\left|\frac{1}{nh_n}\sum_{1\leq i\leq n}T_i(X_i-x)^\ell K\left(\frac{X_i-x}{h_n}\right)\right| \\
&=: \sup_{t\leq\tau}\mathcal{L}_1(t)\times\sup_{x\in\mathcal{C}}|\mathcal{L}_2(x)|.
\end{aligned}
$$

Pour $\mathcal{L}_1$, en utilisant le lemme 4.2. in Deheuvels and Einmahl (2000), nous obtenons :

$$
\sup_{t\leq\tau}\mathcal{L}_1(t)=\mathrm{O}_{p.s.}\left(\sqrt{\frac{\log\log n}{n}}\right)\qquad\text{quand } n\longrightarrow\infty.. \tag{2.11}
$$

Pour $\mathcal{L}_2$, nous utilisons la décomposition suivante :

$$
\begin{aligned}
\mathcal{L}_2(x) &= \frac{1}{nh_n}\sum_{1\leq i\leq n}T_i(X_i-x)^\ell K\left(\frac{X_j-x}{h_n}\right)-\mathbb{E}\left[\frac{1}{nh_n}\sum_{1\leq i\leq n}T_i(X_i-x)^\ell K\left(\frac{X_i-x}{h_n}\right)\right] \\
&+ \mathbb{E}\left[\frac{1}{nh_n}\sum_{1\leq i\leq n}T_i(X_i-x)^\ell K\left(\frac{X_i-x}{h_n}\right)\right] \\
&=: \mathcal{L}_{2,1}(x)+\mathcal{L}_{2,2}(x).
\end{aligned} \tag{2.12}
$$

D'une part pour $\mathcal{L}_{2,1}(x)$, sous les hypothèses du Lemme 2.4.1 et pour des données complètes (c-à-d en prenant $C=+\infty$), nous obtenons le résultat suivant :

$$
\sup_{x\in\mathcal{C}}\left|\mathcal{L}_{2,1}(x)\right|=\mathrm{O}_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^2}}\right)\qquad\text{quand } n\longrightarrow\infty. \tag{2.13}
$$

D'autre part, en utilisant la propriété de l'espérance conditionnelle suivis d'un changement de variable pour $\ell=0,1$, nous avons :

$$
\begin{aligned}
\mathcal{L}_{2,2}(x) &= h_n^{-1}\mathbb{E}\left[(X_1-x)^\ell K\left(\frac{X_1-x}{h_n}\right)\mathbb{E}[T_1|X_1]\right] \\
&= \int(u-x)^\ell K\left(\frac{u-x}{h_n}\right)m(u)f(u)dv \\
&= h_n^\ell\int t^\ell K(t)s_0(x+th)dt.
\end{aligned}
$$

En utilisant un développement de Taylor d'ordre 1 et sous les hypothèses A1, A2 et A4 pour $\ell=0,1$, nous obtenons :

$$
\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{L}_{2,2}(x)\right| &\leq h_n^\ell\sup_{x\in\mathcal{C}}|s_0(x)|\int|t|^\ell K(t)dt+h_n^{\ell+1}\sup_{x\in\mathcal{C}}|s_0'(x)|\int|t|^{\ell+1}K(t)dt \\
&= \mathrm{O}(h^\ell).
\end{aligned} \tag{2.14}
$$

Finalement, en combinant les résultats (2.11)–(2.14) nous établissons la preuve du Lemme 2.4.2.

Lemme 2.4.3 *Sous les hypothèses A1, A2 et A3 pour $\ell = 0, 1, 2$, nous avons :*

$$\sup_{x\in\mathcal{C}} |\mathbb{E}[\widehat{r}_\ell(x)]| = O\left(h^\ell\right), \qquad \text{quand } n \to \infty.$$

Preuve du Lemme 2.4.3. En utilisant la propriété de l'espérance conditionnelle, un changement de variable et pour finir un développement de Taylor d'ordre 1, pour $\ell = 0, 1, 2$, nous aurons :

$$\begin{aligned}
\mathbb{E}[\widehat{r}_\ell(x)] &= \mathbb{E}\left[\frac{1}{nh_n}\sum_{1\le i\le n}(X_i - x)^\ell K\left(\frac{X_i - x}{h}\right)\right] \\
&= \frac{1}{h_n}\mathbb{E}\left[(X_1 - x)^\ell K\left(\frac{X_1 - x}{h_n}\right)\right] \\
&= \frac{1}{h_n}\int (u - x)^\ell K\left(\frac{u - x}{h_n}\right) f(u)du \\
&= h_n^\ell \int v^\ell K(v) f(x + hv)dt \\
&= h_n^\ell f(x)\int v^\ell K(v)dv + h_n^{\ell+1}\int v^{\ell+1}K(v)f'(\xi)dv.
\end{aligned}$$

En passant au sup sur le compact $\mathcal{C}$ et sous les hypothèses A1, A2 et A3 nous retrouvons le résultat du lemme 2.4.3.
La combinaison des Lemmes 2.4.1 à 2.4.3 selon la décomposition (2.10) nous donne le résultat de la proposition 2.2.3.

Preuve de la proposition 2.2.4. Par un raisonnement similaire à celui de la preuve de la proposition 2.2.3, nous remarquons :

$$\begin{aligned}
\mathcal{B}_4(x) &= \{\widetilde{s}_0(x)\widehat{r}_2(x) - \mathbb{E}[\widetilde{s}_0(x)\widehat{r}_2(x)]\} - \{\widetilde{s}_1(x)\widehat{r}_1(x) - \mathbb{E}[\widetilde{s}_1(x)\widehat{r}_1(x)]\} \\
&=: \mathcal{B}_{4,1}(x) - \mathcal{B}_{4,2}(x).
\end{aligned}$$

D'une part, nous avons :

$$\begin{aligned}
\mathcal{B}_{4,1}(x) &= (\widetilde{s}_0(x) - \mathbb{E}[\widetilde{s}_0(x)])(\widehat{r}_2(x) - \mathbb{E}[\widehat{r}_2(x)]) + (\widehat{r}_2(x) - \mathbb{E}[\widehat{r}_2(x)])\mathbb{E}[\widetilde{s}_0(x)] \\
&+ (\widetilde{s}_0(x) - \mathbb{E}[\widetilde{s}_0(x)])\mathbb{E}[\widehat{r}_2(x)] + \mathbb{E}[\widetilde{s}_0(x)]\mathbb{E}[\widehat{r}_2(x)] - \mathbb{E}[\widetilde{s}_0(x)\widehat{r}_2(x)]. \qquad (2.15)
\end{aligned}$$

D'autre part, nous avons :

$$\begin{aligned}
\mathcal{B}_{4,2}(x) &= (\widetilde{s}_1(x) - \mathbb{E}[\widetilde{s}_1(x)])(\widehat{r}_1(x) - \mathbb{E}[\widehat{r}_1(x)]) + (\widehat{r}_1(x) - \mathbb{E}[\widehat{r}_1(x)])\mathbb{E}[\widetilde{s}_1(x)] \\
&+ (\widetilde{s}_1(x) - \mathbb{E}[\widetilde{s}_1(x)])\mathbb{E}[\widehat{r}_1(x)] + \mathbb{E}[\widetilde{s}_1(x)]\mathbb{E}[\widehat{r}_1(x)] - \mathbb{E}[\widetilde{s}_1(x)\widehat{r}_1(x)]. \qquad (2.16)
\end{aligned}$$

Cela revient à étudier chaque terme des décompositions (2.15) et (2.16). Pour cela, nous considérons les lemmes suivants :

Lemme 2.4.4 *Sous les hypothèses A1, A2 et A3, pour $\ell = 0, 1$, nous avons :*

$$\sup_{x\in\mathcal{C}} |\widetilde{s}_\ell(x) - \mathbb{E}[\widetilde{s}_\ell(x)]| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^{2-2\ell}}}\right) \qquad \text{quand} \quad n \to \infty.$$

Preuve du Lemme 2.4.4. Nous considérons une suite i.i.d. $(X_1, Y_1, \delta_1), \ldots, (X_n, Y_n, \delta_n)$ et nous définissons :

$$\Xi_n = \left\{ \psi_x : \mathbb{R} \times \mathbb{R}_+^\star \times \{0,1\} \to \mathbb{R}^+ / \psi_x(u,y,\delta) = \frac{\delta y}{n h_n \overline{G}(y)} (u-x)^\ell K\left(\frac{x-u}{h_n}\right), \quad x \in \mathcal{C} \right\}.$$

D'après la condition (b) du lemme 3 de Giné and Guillou (1999), Ξ_n est une Vapnik-Cervonenkis (V-C) classe de fonctions mesurables uniformément bornée pour $\ell = 0, 2$. Sous la condition (c) du même lemme pour $\ell = 1$, posons $u - x = (u - x + h_n) - h_n$ alors Ξ_n est aussi une V-C classe d'enveloppe $\Psi = \dfrac{h_n^{\ell-1}}{n\overline{G}(\tau)} \|K\|_\infty$. De plus, sous les hypothèses A2 nous obtenons :

$$\sup_{x \in \mathcal{C}} \theta_x(X_1, Y_1, \delta_1) \leq \frac{h_n^{\ell-1}}{n\overline{G}(\tau)} \|K\|_\infty = \frac{h_n^{\ell-1}}{n} \xi_3 =: U_n.$$

De la même manière, sous les hypothèses A2 et A3 nous avons :

$$\begin{aligned} \sup_{x \in \mathcal{C}} Var\left[\theta_x(X_1, Y_1, \delta_1)\right] &\leq \sup_{x \in \mathcal{C}} \mathbb{E}\left[\theta_x^2(X_1, Y_1, \delta_1)\right] \\ &\leq \frac{h_n^{2\ell}}{n^2 h_n^2 \overline{G}(\tau)} \|K\|_2^2 \|s_0\|_\infty = \frac{h_n^{2\ell-2}}{n^2} \xi_4 =: \sigma_n^2 \end{aligned}$$

avec $\sigma_n \leq U_n$ pour n assez grand.
Maintenant, nous pouvons appliquer l'inégalité de Talagrand (voir proposition 5.4.2). IL existe trois constantes positives C_1, C_2 et C_3 tel que pour $t \geq C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}}$, nous obtenons :

$$\mathbb{P}\left[\sup_{\psi_x \in \Xi_n} \left| \sum_{1 \leq i \leq n} (\theta_x(X_i, Y_i, \delta_i) - \mathbb{E}[\theta_x(X_1, Y_1, \delta_1)]) \right| > C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}} \right]$$

$$\leq C_2 \exp\left(-\frac{C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}}}{C_2 \frac{h_n^{\ell-1}}{n} \xi_3} \log\left[1 + \frac{C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}} \frac{h_n^{\ell-1}}{n} \xi_3}{C_2 \left(\sqrt{n} \frac{h_n^{\ell-1}}{n} \sqrt{\xi_4} + \frac{h^{\ell-1}}{n} \xi_3 \sqrt{\log C_3 \frac{\xi_3}{\sqrt{\xi_4}}} \right)^2} \right] \right),$$

En utilisant l'approximation $\log(1+x) \approx x$ (pour $x \to 0$), la partie de droite de la dernière équation devient de l'ordre de

$$C_2 \exp\left(-\frac{C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}}}{C_2 \frac{h_n^{\ell-1}}{n} \xi_3} \times \frac{C_1 \sqrt{\frac{\log n}{n h_n^{2-2\ell}}} \frac{h_n^{\ell-1}}{n} \xi_3}{C_2 \left(\sqrt{n} \frac{h_n^{\ell-1}}{n} \sqrt{\xi_4} \right)^2} \right) = C_2 n^{-\frac{C_1^2}{C_2^2 \xi_4}},$$

qui pour des choix appropriés de C_1, C_2 et ξ_4 et sous la dernière partie de l'hypothèse A1, nous obtenons un terme général d'une série convergente de l'ordre de $O(n^{-3/2})$. Ensuite, par le lemme de Borel Cantelli nous obtenons le résultat.

Lemme 2.4.5 *Sous les hypothèses A1, A2 et A4, pour $\ell = 0, 1$, nous avons :*

$$\sup_{x\in\mathcal{C}} |\mathbb{E}[\widetilde{s}_\ell(x)]| = O\left(h_n^\ell\right) \qquad \text{quand} \quad n \to \infty.$$

Preuve du Lemme 2.4.5. Pour $\ell = 0, 1$, en utilisant la propriété de l'espérance conditionnelle, un changement de variable et un développement Taylor, nous avons :

$$\begin{aligned}
\mathbb{E}[\widetilde{s}_\ell(x)] &= \mathbb{E}\left[\frac{1}{nh_n}\sum_{1\le i\le n}\widetilde{T}_i(X_i-x)^\ell K\left(\frac{X_i-x}{h_n}\right)\right] \\
&= \frac{1}{h_n}\mathbb{E}\left[\widetilde{T}_1(X_1-x)^\ell K\left(\frac{X_1-x}{h_n}\right)\right] \\
&= \frac{1}{h_n}\mathbb{E}\left[(X_1-x)^\ell K\left(\frac{X_1-x}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1|X_1\right]\right] \\
&= \frac{1}{h_n}\int (u-x)^\ell K\left(\frac{u-x}{h_n}\right) m(u)f(u)du \\
&= \int (h_n v)^\ell K(v)s_0(x+h_n v)dv \\
&= h_n^\ell s_0(x)\int v^\ell K(v)dv + h_n^{\ell+1}\int v^{\ell+1}K(v)s_0'(\xi)dv.
\end{aligned}$$

En passant au sup sur le compact $\mathcal{C}$ et sous les hypothèses A1, A2 et A4 nous obtenons le résultat.
Désormais, il nous reste qu'à étudié le terme de covariance. A cette fin, nous considérons le lemme suivant :

Lemme 2.4.6 *Sous les hypothèses A1, A2 et A4, nous avons :*

$$\sup_{x\in\mathcal{C}} |Cov(\widetilde{s}_0(x), \widehat{r}_2(x))| = o\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \qquad \text{quand} \quad n \to \infty.$$

Preuve du Lemme 2.4.6. Par définition de la covariance, nous avons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}} |\mathrm{Cov}(\widetilde{s}_0(x), \widehat{r}_2(x))| &\le \sup_{x\in\mathcal{C}} |\mathbb{E}[\widetilde{s}_0(x)\widehat{r}_2(x)]| + \sup_{x\in\mathcal{C}} |\mathbb{E}[\widetilde{s}_0(x)]| \sup_{x\in\mathcal{C}} |\mathbb{E}[\widehat{r}_2(x)]| \\
&=: \sup_{x\in\mathcal{C}} |\mathcal{I}_1(x)| + \sup_{x\in\mathcal{C}} |\mathcal{I}_2(x)| \sup_{x\in\mathcal{C}} |\mathcal{I}_3(x)|.
\end{aligned}$$

D'une part,

$$\begin{aligned}
\mathcal{I}_1(x) &= \mathbb{E}\left[\frac{1}{(nh_n)^2}\sum_{1\le i,j\le n}\widetilde{T}_i(X_j-x)^2 K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right] \\
&= \frac{1}{(nh_n)^2}\left\{\mathbb{E}\left[\sum_{1\le i\le n}\widetilde{T}_i(X_i-x)^2K^2\left(\frac{X_i-x}{h_n}\right)\right] + \mathbb{E}\left[\sum_{\substack{1\le i,j\le n\\ i\ne j}}\widetilde{T}_i(X_j-x)^2 K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right]\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{(nh_n)^2}\left\{n\mathbb{E}\left[\widetilde{T}_1(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]+n(n-1)\mathbb{E}\left[\widetilde{T}_1(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\right\}\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[\widetilde{T}_1(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]+\frac{(n-1)}{nh_n^2}\mathbb{E}\left[\widetilde{T}_1(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\\
&=: \mathcal{I}_{1,1}(x)+\mathcal{I}_{1,2}(x).
\end{aligned}
$$

Pour $\mathcal{I}_{1,1}(\cdot)$, en utilisant la propriété de l'espérance conditionnelle, nous avons :

$$
\begin{aligned}
\mathcal{I}_{1,1}(x) &= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1|X_1\right]\right]\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)m(X_1)\right]\\
&= \frac{1}{nh_n^2}\int(u-x)^2K^2\left(\frac{u-x}{h_n}\right)m(u)f(u)du\\
&= \frac{1}{nh_n^2}\int(u-x)^2K^2\left(\frac{u-x}{h_n}\right)s_0(u)du
\end{aligned}
$$

Par un changement de variable et un développement de Taylor, nous obtenons :

$$
\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{I}_{1,1}(x)\right| &= \frac{1}{nh_n^2}\sup_{x\in\mathcal{C}}\left|\int(th_n)^2K^2(t)s_0(x+th_n)hdt\right|\\
&= \frac{h_n}{n}\sup_{x\in\mathcal{C}}\left|\int t^2K^2(t)s_0(x+th_n)dt\right|\\
&= \frac{h_n}{n}\sup_{x\in\mathcal{C}}\left|\int t^2K^2(t)\{s_0(x)+th_ns'(\xi)\}dt\right|\\
&\leq \frac{h_n}{n}\sup_{x\in\mathcal{C}}|s_0(x)|\int t^2K^2(t)\,dt+\frac{h_n^2}{n}\sup_{x\in\mathcal{C}}\left|s_0'(x)\right|\int|t|^3K^2(t)\,dt.
\end{aligned}
$$

Sous les hypothèses A2 et A4, nous obtenons :

$$
\sup_{x\in\mathcal{C}}\left|\mathcal{I}_{1,1}(x)\right| = \mathrm{O}\left(\frac{h_n}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.17}
$$

Pour $\mathcal{I}_{1,2}(x)$, en utilisant la propriété de l'espérance conditionnelle, nous avons :

$$
\begin{aligned}
\mathcal{I}_{1,2}(x) &= \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1|X_1,X_2\right]\right]\\
&= \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)m(X_1)\right]\\
&= \frac{(n-1)}{nh_n^2}\int\int(v-x)^2K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)m(u)f(u)f(v)dudv\\
&= \frac{(n-1)}{nh_n^2}\int\int(v-x)^2K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)s_0(u)f(v)dudv
\end{aligned}
$$

Par un changement de variable et un développement de Taylor, nous obtenons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{I}_{1,2}(x)\right| &= \frac{(n-1)}{nh_n^2}\sup_{x\in\mathcal{C}}\left|\int\int (sh_n)^2K(t)K(s)s_0(x+th_n)h_n^2dt\,ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s^2K(t)K(s)s_0(x+th_n)dt\,ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s^2K(t)K(s)\{s_0(x)+th_ns'(\xi)\}dtds\right| \\
&\leq \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}|s_0(x)|\left(\int s^2K(s)\,ds\right)\left(\int K(t)dt\right) \\
&+ \frac{(n-1)h_n^3}{n}\sup_{x\in\mathcal{C}}\left|s_0'(x)\right|\left(\int |t|K(t)\,dt\right)\left(\int s^2K(s)ds\right).
\end{aligned}$$

Sous les hypothèses A2 et A4, nous obtenons :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{I}_{1,2}(x)\right| = \mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.18}$$

En combinant (2.17) et (2.18), nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{I}_1(x)| = \mathrm{O}\left(\frac{h_n}{n}\right)+\mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.19}$$

D'autre part, du résultat du lemme 2.4.5 nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{I}_2(x)| = \mathrm{O}(1), \qquad \text{quand } n\to\infty. \tag{2.20}$$

Et du résultat du lemme 2.4.3, nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{I}_3(x)| = \mathrm{O}\left(h_n^2\right), \qquad \text{quand } n\to\infty. \tag{2.21}$$

En associant les résultats (2.19), (2.20) et (2.21), nous obtenons $\mathrm{O}(h_n^2)$. Par la dernière partie de l'hypothèse A1, ce dernier est négligeable devant $\mathrm{o}\left(\sqrt{\frac{\log n}{nh_n^2}}\right)$. D'où nous obtenons le résultat du lemme 2.4.6.

Lemme 2.4.7 *Sous les hypothèses A1, A2 et A4, nous avons :*

$$\sup_{x\in\mathcal{C}}|Cov(\widetilde{s_1}(x),\widehat{r_1}(x))| = o\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \qquad \textit{quand} \quad n\to\infty.$$

Preuve du Lemme 2.4.7. Par définition de la covariance, nous avons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}|\mathrm{Cov}(\widetilde{s_1}(x),\widehat{r_1}(x))| &\leq \sup_{x\in\mathcal{C}}|\mathbb{E}[\widetilde{s_1}(x)\widehat{r_1}(x)]|+\sup_{x\in\mathcal{C}}|\mathbb{E}[\widetilde{s_1}(x)]|\sup_{x\in\mathcal{C}}|\mathbb{E}[\widehat{r_1}(x)]| \\
&=: \sup_{x\in\mathcal{C}}|\mathcal{J}_1(x)|+\sup_{x\in\mathcal{C}}|\mathcal{J}_2(x)|\sup_{x\in\mathcal{C}}|\mathcal{J}_3(x)|.
\end{aligned}$$

D'une part, pour $\mathcal{J}_1(x)$ nous avons :

$$\begin{aligned}
\mathcal{J}_1(x) &= \mathbb{E}\left[\frac{1}{(nh_n)^2}\sum_{1\le i,j\le n}\widetilde{T}_i(X_j-x)(X_j-x)K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right]\\
&= \frac{1}{(nh_n)^2}\left\{\mathbb{E}\left[\sum_{1\le i\le n}\widetilde{T}_i(X_i-x)^2K^2\left(\frac{X_i-x}{h_n}\right)\right]\right.\\
&+ \left.\mathbb{E}\left[\sum_{\substack{1\le i,j\le n\\ i\neq j}}\widetilde{T}_i(X_j-x)(X_j-x)K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right]\right\}\\
&= \frac{1}{(nh_n)^2}\left\{n\mathbb{E}\left[\widetilde{T}_1(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]\right.\\
&+ \left.n(n-1)\mathbb{E}\left[\widetilde{T}_1(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\right\}\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[\widetilde{T}_1(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]\\
&+ \frac{(n-1)}{nh_n^2}\mathbb{E}\left[\widetilde{T}_1(X_1-x)(X_2-x)K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\\
&=: \mathcal{J}_{1,1}(x)+\mathcal{J}_{1,2}(x).
\end{aligned}$$

Pour $\mathcal{J}_{1,1}(\cdot)$, du résultat (2.17), nous avons :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{J}_{1,1}(x)\right| = \sup_{x\in\mathcal{C}}\left|\mathcal{I}_{1,1}(x)\right| = \mathrm{O}\left(\frac{h_n}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.22}$$

Pour $\mathcal{J}_{1,2}(x)$, en utilisant la propriété de l'espérance conditionnelle, nous avons :

$$\begin{aligned}
\mathcal{J}_{1,2}(x) &= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1-x)(X_2-x)K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1|X_1,X_2\right]\right]\\
&= \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_1-x)(X_2-x)K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)m(X_1)\right]\\
&= \frac{(n-1)}{nh_n^2}\int\int(u-x)(v-x)K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)m(u)f(u)dudv\\
&= \frac{(n-1)}{nh_n^2}\int\int(u-x)(v-x)K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)s_0(u)dudv.
\end{aligned}$$

Par un changement de variable et un développement de Taylor, nous obtenons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{J}_{1,2}(x)\right| &= \frac{(n-1)}{nh_n^2}\sup_{x\in\mathcal{C}}\left|\int\int h_n^2 s\, tK(t)K(s)s_0(x+th)h_n^2 dt\, ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s\, tK(t)K(s)s_0(x+th_n)dt\, ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s\, tK(t)K(s)\{s_0(x)+th_n s'(\xi)\}dt\, ds\right| \\
&\leq \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}|s_0(x)|\int|s|K(s)ds\int|t|K(t)dt \\
&+ \frac{(n-1)h_n^3}{n}\sup_{x\in\mathcal{C}}\left|s_0'(x)\right|\int t^2K(t)dt\int sK(s)ds.
\end{aligned}$$

Sous les hypothèses A2 et A4, nous obtenons :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{J}_{1,1}(x)\right| = \mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.23}$$

En combinant (2.22) et (2.23), nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{J}_1(x)| = \mathrm{O}\left(\frac{h_n}{n}\right)+\mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.24}$$

D'autre part, du résultat du lemme 2.4.5 nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{J}_2(x)| = \mathrm{O}(1), \qquad \text{quand } n\to\infty. \tag{2.25}$$

Et du résultat du lemme 2.4.3, nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{J}_3(x)| = \mathrm{O}\left(h_n^2\right), \qquad \text{quand } n\to\infty. \tag{2.26}$$

En associant les résultats (2.24), (2.25) et (2.26), nous obtenons $\mathrm{O}(h_n^2)$. Par la dernière partie de l'hypothèse A1, ce dernier est négligeable devant $\mathrm{o}\left(\sqrt{\frac{\log n}{nh_n^2}}\right)$. D'où nous obtenons le résultat du lemme 2.4.7.
Enfin les résultats des Lemmes 2.4.1 et 2.4.3–2.4.7 selon les décompositions (2.15) et (2.16) conclues le résultat de la proposition 2.2.4.
Preuve de la proposition 2.2.2. Soit la décomposition suivante :

$$\begin{aligned}
\mathcal{B}_2(x) &= \widehat{r}_0(x)\widehat{r}_2(x)-\widehat{r}_1^2(x)-\mathbb{E}[\widehat{r}_0(x)\widehat{r}_2(x)-\widehat{r}_1^2(x)] \\
&= \{\widehat{r}_0(x)\widehat{r}_2(x)-\mathbb{E}[\widehat{r}_0(x)\widehat{r}_2(x)]\}-\left\{\widehat{r}_1^2(x)-\mathbb{E}[\widehat{r}_1^2(x)]\right\} \\
&=: \mathcal{B}_{2,1}(x)-\mathcal{B}_{2,2}(x).
\end{aligned}$$

D'une part, nous avons :

$$\begin{aligned}
\mathcal{B}_{2,1}(x) &= (\widehat{r}_0(x)-\mathbb{E}[\widehat{r}_0(x)])(\widehat{r}_2(x)-\mathbb{E}[\widehat{r}_2(x)])+\mathbb{E}[\widehat{r}_0(x)](\widehat{r}_2(x)-\mathbb{E}[\widehat{r}_2(x)]) \\
&+ \mathbb{E}[\widehat{r}_2(x)](\widehat{r}_0(x)-\mathbb{E}[\widehat{r}_0(x)])+\mathbb{E}[\widehat{r}_2(x)]\mathbb{E}[\widehat{r}_0(x)]-\mathbb{E}[\widehat{r}_0(x)\widehat{r}_2(x)].
\end{aligned} \tag{2.27}$$

D'autre part, pour étudier $\mathcal{B}_{2,2}(x)$ il suffit d'étudier $\mathrm{Var}(\widehat{r_\ell}(x))$. Pour les termes qui n'ont pas encore été étudié, nous considérons les lemmes suivants :

Lemme 2.4.8 *Sous les hypothèses A1–A3 pour $\ell = 0, 1, 2$, nous avons :*

$$\sup_{x\in\mathcal{C}} |Var(\widehat{r_\ell}(x))| = O_{p.s.}\left(\frac{h_n^{2\ell-1}}{n}\right) \qquad quand \quad n \to \infty.$$

Preuve du Lemme 2.4.8. En utilisant la propriété de l'espérance conditionnelle, un changement de variable et pour finir un développement de Taylor pour $\ell = 0, 1, 2$, nous obtenons :

$$\begin{aligned}
\mathbb{E}[\widehat{r}_\ell^2(x)] &= \mathbb{E}\left[\frac{1}{(nh_n)^2}\sum_{i\le i\le n}(X_i - x)^{2\ell}K^2\left(\frac{X_i - x}{h_n}\right)\right] \\
&= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1 - x)^{2\ell}K^2\left(\frac{X_1 - x}{h_n}\right)\right] \\
&= \frac{1}{nh_n^2}\int (u - x)^{2\ell}K^2\left(\frac{u - x}{h_n}\right) f(u)du \\
&= \frac{1}{nh_n}\int (sh_n)^{2\ell}K^2(s)f(x + h_n s)ds \\
&= \frac{1}{nh_n}\left\{h_n^{2\ell}f(x)\int s^{2\ell}K^2(s)ds + h_n^{2\ell+1}\int s^{2\ell+1}K^2(s)f'(\xi)ds\right\}.
\end{aligned}$$

En passant au sup sur le compact $\mathcal{C}$ et sous les hypothèses A1, A2 et A3 on obtient le résultat. De plus, le résultat est négligeable devant $o\left(\sqrt{\frac{\log n}{nh_n^2}}\right)$.

Lemme 2.4.9 *Sous les hypothèses A1–A3, nous avons :*

$$\sup_{x\in\mathcal{C}} |Cov(\widehat{r_0}(x), \widehat{r_2}(x))| = o\left(\sqrt{\frac{\log n}{nh_n^2}}\right) \qquad quand \quad n \to \infty.$$

Preuve du Lemme 2.4.9. Par définition de la covariance, nous avons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}} |\mathrm{Cov}(\widehat{r_0}(x), \widehat{r_2}(x))| &= \sup_{x\in\mathcal{C}} |\mathbb{E}[\widehat{r_0}(x)\widehat{r_2}(x)]| + \sup_{x\in\mathcal{C}} |\mathbb{E}[\widehat{r_0}(x)]| \sup_{x\in\mathcal{C}} |\mathbb{E}[\widehat{r_2}(x)]| \\
&=: \sup_{x\in\mathcal{C}} |\mathcal{K}_1(x)| + \sup_{x\in\mathcal{C}} |\mathcal{K}_2(x)| \sup_{x\in\mathcal{C}} |\mathcal{K}_3(x)|.
\end{aligned}$$

D'une part,

$$\begin{aligned}
\mathcal{K}_1(x) &= \mathbb{E}\left[\frac{1}{(nh_n)^2}\sum_{1\le i,j\le n}(X_j-x)^2K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right]\\
&= \frac{1}{(nh_n)^2}\left\{\mathbb{E}\left[\sum_{1\le i\le n}(X_i-x)^2K^2\left(\frac{X_i-x}{h_n}\right)\right]\right.\\
&+ \left.\mathbb{E}\left[\sum_{\substack{1\le i,j\le n\\ i\neq j}}(X_j-x)^2K\left(\frac{X_i-x}{h_n}\right)K\left(\frac{X_j-x}{h_n}\right)\right]\right\}\\
&= \frac{1}{(nh_n)^2}\left\{n\mathbb{E}\left[(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]\right.\\
&+ \left.n(n-1)\mathbb{E}\left[(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\right\}\\
&= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]\\
&+ \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right]\\
&=: \mathcal{K}_{1,1}(x)+\mathcal{K}_{1,2}(x).
\end{aligned}$$

Pour $\mathcal{K}_{1,1}(\cdot)$, en utilisant la propriété de l'espérance conditionnelle, nous avons :

$$\begin{aligned}
\mathcal{K}_{1,1}(x) &= \frac{1}{nh_n^2}\mathbb{E}\left[(X_1-x)^2K^2\left(\frac{X_1-x}{h_n}\right)\right]\\
&= \frac{1}{nh_n^2}\int(u-x)^2K^2\left(\frac{u-x}{h_n}\right)f(u)du
\end{aligned}$$

Par un changement de variable et un développement de Taylor, nous obtenons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{K}_{1,1}(x)\right| &= \frac{1}{nh_n^2}\sup_{x\in\mathcal{C}}\left|\int(th_n)^2K^2(t)f(x+th_n)hdt\right|\\
&= \frac{h_n}{n}\sup_{x\in\mathcal{C}}\left|\int t^2K^2(t)f(x+th_n)dt\right|\\
&= \frac{h_n}{n}\sup_{x\in\mathcal{C}}\left|\int t^2K^2(t)\{f(x)+th_nf'(\xi)\}dt\right|\\
&\le \frac{h_n}{n}\sup_{x\in\mathcal{C}}|f_0(x)|\int t^2K^2(t)\,dt+\frac{h_n^2}{n}\sup_{x\in\mathcal{C}}\left|f_0'(x)\right|\int|t|^3K^2(t)\,dt.
\end{aligned}$$

Sous les hypothèses A2 et A3, nous obtenons :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{K}_{1,1}(x)\right| = \mathrm{O}\left(\frac{h_n}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.28}$$

Pour $\mathcal{K}_{1,2}(x)$, nous avons : en utilisant la propriété de l'espérance conditionnelle, nous

avons :

$$\begin{aligned}
\mathcal{K}_{1,2}(x) &= \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_1-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right] \\
&= \frac{(n-1)}{nh_n^2}\mathbb{E}\left[(X_2-x)^2K\left(\frac{X_1-x}{h_n}\right)K\left(\frac{X_2-x}{h_n}\right)\right] \\
&= \frac{(n-1)}{nh_n^2}\int\int(v-x)^2K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)f(u)f(v)dudv \\
&= \frac{(n-1)}{nh_n^2}\int\int(v-x)^2K\left(\frac{u-x}{h_n}\right)K\left(\frac{v-x}{h_n}\right)f(u)f(v)dudv.
\end{aligned}$$

Par un changement de variable et un développement de Taylor, nous obtenons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathcal{K}_{1,2}(x)\right| &= \frac{(n-1)}{nh_n^2}\sup_{x\in\mathcal{C}}\left|\int\int(sh_n)^2K(t)K(s)f(x+th_n)f(x+sh_n)h_n^2dt\,ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s^2K(t)K(s)f(x+th_n)f(x+sh_n)dt\,ds\right| \\
&= \frac{(n-1)h_n^2}{n}\sup_{x\in\mathcal{C}}\left|\int\int s^2K(t)K(s)(f(x)+th_nf'(\xi_1))(f(x)+sh_nf'(\xi_2))dt\,ds\right|
\end{aligned}$$

Sous les hypothèses A2 et A3, nous obtenons :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{K}_{1,2}(x)\right| = \mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.29}$$

En combinant (2.28) et (2.29), nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{K}_1(x)| = \mathrm{O}\left(\frac{h_n}{n}\right)+\mathrm{O}\left(\frac{(n-1)h_n^2}{n}\right), \qquad \text{quand } n\to\infty. \tag{2.30}$$

D'autre part, du résultat du lemme 2.4.3 nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{K}_2(x)| = \mathrm{O}(1), \qquad \text{quand } n\to\infty. \tag{2.31}$$

Et du résultat du lemme 2.4.3, nous avons :

$$\sup_{x\in\mathcal{C}}|\mathcal{K}_3(x)| = \mathrm{O}\left(h_n^2\right), \qquad \text{quand } n\to\infty. \tag{2.32}$$

En associant les résultats (2.30), (2.31) et (2.32), nous obtenons le résultat du lemme 2.4.9.

Enfin, les résultats du Lemme 2.4.1, Lemme 2.4.3, Lemme 2.4.8 et Lemme 2.4.9 selon la décomposition (2.27) nous établissons la preuve de la proposition 2.2.2.

Preuve du corollaire 1. Il existe un Γ strictement positif tel que pour tout $x\in\mathcal{C}$, $\mathbb{E}[\widehat{m}_0(x)]\geq\Gamma$. Ainsi $\inf_{x\in\mathcal{C}}\widehat{m}_0(x)\leq\frac{\Gamma}{2}$ ce qui implique qu'il existe un $x\in\mathcal{C}$ tel que $|\mathbb{E}[\widehat{m}_0(x)]-\widehat{m}_0(x)|\geq\frac{\Gamma}{2}$

ce qui donne

$$\sup_{x\in\mathcal{C}}\left|\mathbb{E}[\widehat{m}_0(x)]-\widehat{m}_0(x)\right|\geq\frac{\Gamma}{2}.$$

D'ou, le résultat de la proposition 2.2.2 permet d'écrire pour $\frac{\Gamma}{2}=\Gamma'$:

$$\sum_n\mathbb{P}\left(\inf_{x\in\mathcal{C}}\widehat{m}_0(x)\leq\Gamma'\right)\leq\sum_n\mathbb{P}\left(\sup_{x\in\mathcal{C}}\left|\mathbb{E}[\widehat{m}_0(x)]-\widehat{m}_0(x)\right|\leq\Gamma'\right)<\infty.$$

Preuve de la proposition 2.2.1. Nous considérons :

$$\begin{aligned}
|\mathcal{B}_1(x)| &= \left|\frac{\mathbb{E}\left[\widetilde{m}_1(x)\right]-m(x)\mathbb{E}\left[\widehat{m}_0(x)\right]}{\mathbb{E}\left[\widehat{m}_0(x)\right]}\right| \\
&= \left|\frac{h_n^{-2}\left\{\mathbb{E}[w_{1,2}(x)\widetilde{T}_2]-m(x)\mathbb{E}\left[w_{1,2}(x)\right]\right\}}{h_n^{-2}\mathbb{E}\left[w_{1,2}(x)\right]}\right| \\
&= \left|\frac{\mathbb{E}\left[w_{1,2}(x)\left\{\mathbb{E}[\widetilde{T}_2|X_2]-m(x)\right\}\right]}{\mathbb{E}[w_{1,2}(x)]}\right| \\
&= \left|\frac{\mathbb{E}\left[w_{1,2}(x)\left(m(X_2)-m(x)\right)\right]}{\mathbb{E}\left[w_{1,2}(x)\right]}\right| \\
&= |m(X_2)-m(x)|.
\end{aligned}$$

Ainsi, sous les hypothèses A1 et A5, nous obtenons

$$\sup_{x\in\mathcal{C}}|\mathcal{B}_1(x)|\leq C|X_2-x|^{\nu}\leq Ch_n^{\nu}.$$

En conclusion, en additionnant tous les résultats des propositions 2.2.1–2.2.4 selon la décomposition principale, nous obtenons la preuve du théorème 2.2.1.

2.5 Conclusion

Dans ce chapitre, nous avons étudié l'estimateur de la fonction de régression par la méthode linéaire locale pour des données censurées aléatoirement à droite. Nous avons montré sa convergence uniforme presque sûre et sa supériorité théorique et numérique par rapport à l'estimateur de la fonction de régression classique à noyau. À l'aide d'une étude de simulation nous avons montré que le LLR est plus efficace que le CR. D'une part, il réduit les effets de bords et d'autres part il reste résistant quand le taux de censure augmente substantiellement. Nous soulignons que la méthode proposée par Cai (2003) où il utilise deux poids (deux noyaux) le premier noyau est un noyau standard et le second est l'estimateur de Kaplan and Meier (1958). L'estimateur résultant est intéressant toutefois l'auteur n'a pas utilisé de poids au dénominateur avec la loi de survie de la variable de censure. Rappelons que El Ghouch and Van Keilegom (2008) ont estimé la fonction de régression par la méthode linéaire locale en utilisant l'estimateur de Beran (1981). Leurs conditions doivent avoir un résultat sur la loi conditionnelle sur la variable aléatoire censurée que, dans notre cas, nous n'utilisons pas. En outre, leur résultat uniforme n'est donné qu'en probabilité. Nous soulignons qu'à notre connaissance, le type de notre résultat n'a jamais été obtenu.

Chapitre 3

Estimation non paramétrique de la fonction de régression relative pour un modèle de censure

[1] Dans ce chapitre, la problématique abordée est l'estimation non paramétrique de la fonction de régression. Nous considérons un modèle de censure aléatoire à droite et nous construisons un estimateur à noyau de l'erreur relative quadratique moyenne pour la fonction de régression. Nous étudions sa convergence uniforme presque sûre sur un compact avec vitesse ainsi que sa normalité asymptotique. L'expression de la variance asymptotique est explicitement donnée. Une large étude numérique est conduite pour appuyer nos résultats théoriques et montrer les avantages de la nouvelle approche comparé aux autres méthodes. Finalement, nous appliquons la nouvelle approche à un exemple de données réelles.

3.1 Introduction

La fonction de régression classique $m(\cdot) = \mathbb{E}[T|X = \cdot]$ est connue pour être le minimum de l'erreur quadratique moyenne $\text{EQM}(\cdot) := \mathbb{E}[(T - m(X))^2 \big| X = \cdot]$, sauf que cette dernière est très sensible aux valeurs aberrantes. Pour surmonter cet inconvénient, plusieurs méthode ont été utilisée, nous mentionnons la méthode des M-estimateurs (voir Huber (1981), Collomb and Härdle (1986), Boente and Fraiman (1990)) et la méthode polynomiale locale (nous renvoyons le lecteur à Fan and Gijbels (1996)). Dans ce chapitre, nous considérons une autre approche qui permet de construire un estimateur efficace même si les données sont affectés par la présence de forte valeurs aberrantes. Notre fonction de régression est solution de l'erreur quadratique relative moyenne suivante :

$$\mathbb{E}\left[\left(\frac{T - m(X)}{T}\right)^2 \Big| X\right]. \tag{3.1}$$

Cette dernière est une mesure de performance plus significative que l'EQM usuelle en présence de valeurs aberrantes. Park and Stefanski (1998) ont montré que la solution

1. En collaboration avec E. OULD SAÏD et R. M. REMITA. Ce chapitre a fait l'objet d'un article soumis pour publication. ArXiv :1901.09555

du problème (3.1) est :

$$m(X) = \frac{\mathbb{E}\left[T^{-1}|X\right]}{\mathbb{E}\left[T^{-2}|X\right]} =: \frac{m_1(X)}{m_2(X)}. \tag{3.2}$$

Pour prouver ce résultat , d'une part, nous avons que :

$$\begin{aligned}
\mathbb{E}\left[\frac{T-m(X)}{T^2}\middle|X\right] &= \mathbb{E}\left[\left\{T^{-1}-m(X)T^{-2}\right\}\middle|X\right] \\
&= \mathbb{E}\left[T^{-1}\middle|X\right] - m(X)\mathbb{E}\left[T^{-2}\middle|X\right] \\
&= m_1(X) - m(X)m_2(X) \\
&= m_1(X) - m_1(X) \\
&= 0 \qquad \text{p.s.}
\end{aligned} \tag{3.3}$$

D'autre part, nous posons $r(X)$ n'importe quel estimateur de T sachant X. Nous montrons grâce de la linéarité de l'espérance conditionnelle :

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{T-r(X)}{T}\right)^2\middle|X\right] &= \mathbb{E}\left[\left(\frac{T-m(X)+m(X)-r(X)}{T}\right)^2\middle|X\right] \\
&= \mathbb{E}\left[\left(\frac{T-m(X)}{T}+\frac{m(X)-r(X)}{T}\right)^2\middle|X\right] \\
&= \mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)^2+2\left(\frac{T-m(X)}{T}\right)\left(\frac{m(X)-r(X)}{T}\right)+\left(\frac{m(X)-r(X)}{T}\right)^2\middle|X\right] \\
&= \mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)^2\middle|X\right]+\mathbb{E}\left[\left(\frac{m(X)-r(X)}{T}\right)^2\middle|X\right] \\
&+ 2\mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)\left(\frac{m(X)-r(X)}{T}\right)\middle|X\right] \\
&=: \mathcal{T}_1+\mathcal{T}_2+\mathcal{T}_3.
\end{aligned}$$

Pour $\mathcal{T}_3$, de (3.3), nous obtenons :

$$\begin{aligned}
\mathcal{T}_3 &= 2\mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)\left(\frac{m(X)-r(X)}{T}\right)\middle|X\right] \\
&= 2(m(X)-r(X))\mathbb{E}\left[\left(\frac{T-m(X)}{T^2}\right)\middle|X\right] \\
&= 0
\end{aligned} \tag{3.4}$$

Pour $\mathcal{T}_1$, en utilisant la propriété de l'espérance conditionnelle, nous avons :

$$\begin{aligned}
\mathcal{T}_1 &= \mathbb{E}\left[\left(\frac{T-m(X)}{T}\right)^2\middle|X\right] \\
&= \mathbb{E}\left[\left(\frac{T^2-2Tm(X)+m^2(X)}{T^2}\right)\middle|X\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[1 - 2m(X)T^{-1} + m^2(X)T^{-2}\middle|X\right] \\
&= 1 - 2m(X)\mathbb{E}\left[T^{-1}\middle|X\right] + m^2(X)\mathbb{E}\left[T^{-2}\middle|X\right] \\
&= 1 - 2m(X)m_1(X) + m^2(X)m_2(X) \\
&= 1 - 2\frac{m_1^2(X)}{m_2(X)} + \frac{m_1^2(X)}{m_2(X)} \\
&= 1 - \frac{m_1^2(X)}{m_2(X)}. \qquad (3.5)
\end{aligned}
$$

Pour $\mathcal{T}_2$, nous avons :

$$
\begin{aligned}
\mathcal{T}_2 &= (m(X) - r(X))\,\mathbb{E}\left[T^{-2}\middle|X\right] \\
&= (m(X) - r(X))\,m_2(X). \qquad (3.6)
\end{aligned}
$$

En combinant (3.3), (3.5) et (3.6), nous obtenons :

$$
\mathbb{E}\left[\left(\frac{T - r(X)}{T}\right)^2\middle|X\right] = \left\{1 - \frac{m_1^2(X)}{m_2(X)}\right\} + \{m(X) - r(X)\}\,m_2(X) \qquad (3.7)
$$

Le premier terme de (3.7) ne dépend pas de $r(X)$ et le second terme est minimum lorsque $r(x) = m(x)$ presque sûrement, alors $m(X)$ est la solution qui réalise le minimum de (3.7) comme nous l'avons supposé dès le départ. L'équation (3.7) montre aussi que $m(x)$ s'écrit aussi sous deux formes :

$$
m(X) = 1 - \frac{m_1^2(X)}{m_2(X)} = \frac{m_2(X) - m_1^2(X)}{m_2(X)} = \frac{\mathrm{V}ar(T^{-1}|X)}{m_2(X)}
$$

ou

$$
m(X) = \frac{m_1(X)}{m_2(X)} = \frac{m_1(X)}{\mathrm{V}ar\,(T^{-1}|X) + m_1^2(X)}.
$$

Park and Stefanski (1998) ont aussi montré que cette solution satisfait (pour toute fonction décroissante t^{-2}) que

$$
\frac{\mathbb{E}\left[T^{-1}|X\right]}{\mathbb{E}\left[T^{-2}|X\right]} \leq \mathbb{E}\left[T\middle|X\right] \qquad (3.8)
$$

p.s. à condition que les deux premiers moments inverses conditionnels existent et soient finis. En raison de sa robustesse face aux valeurs aberrantes, l'erreur relative est plus adéquate que la méthode basée sur l'EQM classique. Si nous prenons l'exemple de la prévision de la consommation d'électricité d'un foyer, les données peuvent être faibles pendant une période (exemple l'hiver) et fortes pour une autre période (pour plus de détails, voir Hirose and Masuda (2018)). Ainsi, la variable d'intérêt peut alors contenir des valeurs aberrantes.

Park and Stefanski (1998) ont considéré une approche paramétrique pour estimer la fonction de régression $m(\cdot)$ qui se concentre sur l'estimation des fonctions moyenne et variance de l'inverse de la variable d'intérêt T (voir Carroll and Ruppert (1988)). Dans

ce cadre d'estimation linéaire, nous renvoyons le lecteur à Lin and Chen (2013) et Chen et al. (2010). Narula and Wellington (1977) ont étudié une méthode d'estimation pour la minimisation de la somme des erreurs relatives absolues. Farum (1990) a développé une méthode qui permet de réduire l'erreur relative absolue. Khoshgoftaar et al. (1992) ont étudié les propriétés asymptotiques de l'estimateur minimisant la somme des erreurs quadratiques relatives.

Dans notre étude, nous nous sommes concentrés sur les approches non paramétriques. Dans ce cadre, nous rappelons que Jones et al. (2008) ont étudié l'estimateur de la fonction de régression relative par la méthode linéaire locale. Ces derniers ont établi des résultats asymptotiques pour les termes du biais et variance. Hu (2019) a examiné le modèle de régression multiplicative à coefficient variable, qui est très utile pour les modèles à variable d'intérêt positive. Le critère d'erreur relative du produit est étendu au modèle de régression multiplicative à coefficients variables par des techniques de lissage du noyau. Dans cet article, l'auteur a établi la convergence et la normalité asymptotique de l'estimateur proposé. Dans le cadre fonctionnel, nous mentionnons les travaux de Demongeot et al. (2016) et Altendi et al. (2018) qui ont établit la convergence uniforme presque sûre avec vitesse et la normalité asymptotique de l'estimateur à noyau de la fonction de régression relative pour des données complètes et tronquées respectivement.

Dans de nombreux problèmes pratiques, les données étudiées ne sont pas toujours complètement disponibles pour le praticien. Par exemple, dans les études de suivi médical, il arrive souvent, pour diverses raisons, que la durée d'intérêt ne puisse être observée. Cela peut être dû à la perte de vue des patients au début ou à la fin de la période d'étude. Ces valeurs sont censurées. Bien qu'inconnues, ces valeurs doivent être prises en compte (par le biais des dates de censure) pour obtenir une estimation correcte et des prévisions précises. Pour ces données pratiques, les procédures statistiques conventionnelles ne sont plus valables et des techniques plus élaborées sont utilisées pour modéliser ces données.

L'un des cas classiques de données incomplètes est le modèle de censure à droite. Un exemple standard dans le traitement des données censurées à droite est celui des données sur la leucémie aiguë (voir Freireich et al. (1963)). Dans cette étude, la durée de rémission (en semaines) des patients atteints de leucémie aiguë traités soit par placebo soit par 6-mercaptopurine (6-MP) a été observée (pour plus de détails sur ce jeu de données, vous pouvez consulter Klein and Moeschberger (2004)). Les patients ont été suivis jusqu'au retour de leur leucémie (rechute) ou jusqu'à la fin de l'étude. Les données sont censurées et nécessitent un traitement spécial. La suppression des données censurées entraîne une perte d'informations. Si les observations censurées sont supprimées, les durées de rémission les plus longues ne sont pas prises en compte et l'effet du traitement par 6-MP est sous-estimé. Ainsi, pour éviter un tel inconvénient, il est plus pratique de prendre en compte les informations partielles dues au mécanisme de la censure.
La littérature sur les données censurées est très vaste et sans prétendre à l'exhaustivité, nous citons les travaux de Cox (1972), Beran (1981), Koul et al. (1981),Dabrowska (1987, 1989), Stute (1993), Stute (1995), Lecoutre and Ould Saïd (1995), Stute (1996), Carbonez et al. (1995), Köhler et al. (2002), Delecroix et al. (2008), Guessoum and Ould Saïd (2008), Lopez (2011), Guessoum and Ould Saïd (2010, 2012), Lopez et al. (2013) et Lemdani and Ould Saïd (2017). Dans ce cadre de données censurées, nous allons définir une transformation des données dites "données synthétiques" dont la modélisation

permet de prendre en compte l'effet de censure. Pour cela, nous considérons le couple observable (Y,δ) et nous définissons :

$$\widetilde{T}^{-\ell} = \frac{\delta Y^{-\ell}}{\overline{G}(Y)}, \qquad \ell = 1,2 \tag{3.9}$$

où $\overline{G} := 1 - G$ désigne la fonction de survie de la variable de censure. Les données synthétiques est une idée de Carbonez et al. (1995) pour étudier les estimateurs a partitions qui a été reprise par Köhler et al. (2002) et bien d'autres autres. Citons par exemple, Guessoum and Ould Saïd (2008, 2010, 2012), Lemdani and Ould Saïd (2017). Tout au long de ce chapitre, nous allons supposer que :

$$(T,X) \text{ et } C \text{ sont indépendants.} \tag{3.10}$$

Cette condition est plausible lorsque la censure est indépendante des caractéristiques du patient dans l'étude. Ainsi, de l'équation (3.9) et la condition (3.10), nous avons :

$$\begin{aligned}
\mathbb{E}\left[\widetilde{T}^{-\ell}|X\right] &= \mathbb{E}\left[\frac{\delta Y^{-\ell}}{\overline{G}(Y)}|X\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{\delta Y^{-\ell}}{\overline{G}(Y)}|T,X\right]|X\right] \\
&= \mathbb{E}\left[\frac{T^{-\ell}}{\overline{G}(T)}\mathbb{E}\left[\mathbb{1}_{\{T\le C\}}|T\right]|X\right] \\
&= \mathbb{E}[T^{-\ell}|X] =: m_\ell(X) \qquad \text{pour } \ell = 1,2.
\end{aligned} \tag{3.11}$$

Cette égalité signifie qu'estimer la quantité $\mathbb{E}\left[\widetilde{T}^{-\ell}|X\right]$ revient à estimer notre fonction objectif $m(\cdot) = m_1(\cdot)/m_2(\cdot)$.

3.2 Définition de l'estimateur

Nous considérons le modèle générale de régression (1.1), avec $(T_1, T_2, \cdots, T_n)$ un n échantillon i.i.d. de n variables aléatoires réelles issue d'une variable T et $(X_1, X_2, \cdots, X_n)^d$ un n échantillon de co-variables X de dimension d. L'étude que nous réalisons ci-dessous est d'estimer :

$$m(x) = \frac{\displaystyle\int t^{-1} f(t|x)dt}{\displaystyle\int t^{-2} f(t|x)dt} = \frac{m_1(x)}{m_2(x)} =: \frac{\mu_1(x)}{\mu_2(x)} \tag{3.12}$$

où $\mu_\ell(\cdot) = m_\ell(\cdot)f(\cdot)$ pour $\ell = 1,2$. Ainsi, l'estimateur analogue direct au célèbre estimateur de la fonction de régression de Nadaraya (1964) et Watson (1964) de (3.12) est donné par :

$$\widehat{m}_{RER}(x) = \frac{\displaystyle\sum_{1\le i\le n} T_i^{-1} K_d\left(\frac{x-X_i}{h_n}\right)}{\displaystyle\sum_{1\le i\le n} T_i^{-2} K_d\left(\frac{x-X_i}{h_n}\right)},$$

où la fenêtre h_n est une suite de nombres réelles positives qui tends vers 0 lorsque $n \to +\infty$ et le noyau $K(\cdot)$ est une densité de probabilité définie dans $\mathbb{R}^d$. Ce dernier à été définit par Jones et al. (2008) pour une co-variable unidimensionnelle.

Dans le contexte censuré, nous utilisant la transformation (3.9) et nous définissons $\widetilde{m}(\cdot)$ comme un pseudo estimateur de $m(\cdot)$. Ce dernier est donné pour tout x par :

$$\widetilde{m}_{RER}(x) =: \frac{\widetilde{m}_1(x)}{\widetilde{m}_2(x)}$$

où

$$\widetilde{m}_\ell(x) = \frac{\widetilde{\mu}_\ell(x)}{\widehat{f}(x)} = \frac{\displaystyle\sum_{1\le i\le n} \widetilde{T}_i^{-\ell} K_d\left(\frac{x-X_i}{h_n}\right)}{\displaystyle\sum_{1\le i\le n} K_d\left(\frac{x-X_i}{h_n}\right)}.$$

Un estimateur calculable de $m_{RER}(\cdot)$ est donné par :

$$\widehat{m}_{RER}(x) =: \frac{\widehat{m}_1(x)}{\widehat{m}_2(x)} \tag{3.13}$$

où

$$\widehat{m}_\ell(x) = \frac{\widehat{\mu}_\ell(x)}{\widehat{f}(x)} = \frac{\displaystyle\sum_{1\le i\le n} \widehat{T}_i^{-\ell} K_d\left(\frac{x-X_i}{h}\right)}{\displaystyle\sum_{1\le i\le n} K_d\left(\frac{x-X_i}{h}\right)} \tag{3.14}$$

pour $\ell = 1,\ 2$ et $\widehat{f}(\cdot)$ est l'estimateur à noyau de la densité marginale $f(\cdot)$ définie par Parzen (1962) et Rosenblatt (1956a).

Remarque 3.2.1 *L'hypothèse d'indépendance (3.10) entre $(C_i)_i$ et $(T_i, X_i)_i$ peut sembler forte et nous pouvons penser à la remplacer par une hypothèse d'indépendance conditionnelle classique entre $(C_i)_i$ et $(T_i)_i$ étant donné $(X_i)_i$ (voir par exemple Beran (1981), Dabrowska (1987, 1989)). Toutefois, sous cette hypothèse, nous proposons d'écrire les données synthétiques (notre nouvelle variable d'intérêt censurée estimée) comme ceci :*

$$\widehat{T}_i^{-\ell} \frac{\delta_i Y_i^{-\ell}}{\overline{G}_n(Y_i|X_i)} \qquad \text{pour } 1 \le i \le n \tag{3.15}$$

pour $\ell = 1, 2$, avec $\overline{G}_n(Y_i|X_i)$ est l'estimateur de Beran (1981) de la fonction de survie conditionnelle de la variable de censure C sachant X basé sur l'estimateur de K-M conditionnel. Alors, nous pouvons construire un estimateur analogue à celui proposé dans l'équation (3.13) en utilisant (3.15). Pour autant que nous sachions, la convergence uniforme de cet estimateur n'a pas été démontré comme dans le cas inconditionnel (voir Deheuvels and Einmahl (2000)). Nous pensons que cette question doit être abordée si nous voulons obtenir des vitesses de convergence.

Remarque 3.2.2 *El Ghouch and Van Keilegom (2008) ont considérés une nouvelle approche d'estimation par la méthode linéaire locale de la fonction de régression $m(\cdot)$. L'idée consiste à transformer le triplet observé (X, Y, δ) par un nouveau vecteur $(X, T^\star)$ qui lui n'est pas sujet à la censure comme le sont les données au départ.*

L'objet principal de cette section est d'étudier l'estimateur à noyau de la fonction de régression relative $\widehat{m}_{RER}(\cdot)$ donné par la formule (3.13), dans le cas d'observations i.i.d. Nous prouvons la convergence presque sûre et la normalité asymptotique de notre estimateur dans la sections 3.5 et nous établissons une étude numérique sur données générées et réelles dans la section 3.4.

3.3 Hypothèses et principaux résultats

Soit $\mathcal{C}$ un ensemble compact de $\mathbb{R}^d$. Pour établir nos résultats, nous avons besoin de définir ces deux fonctions :

$$\mu_\ell(x) = \int t^{-\ell} f(x,t)dt, \qquad \text{pour } \ell = 1,2 \tag{3.16}$$

et

$$r_\lambda(x) = \int \frac{t^{-\lambda}}{G(t)} f(x,t)dt, \qquad \text{pour } \lambda = 2,3,4. \tag{3.17}$$

où $f(\cdot,\cdot)$ est la densité jointe du couple (X,T). En outre, pour toute f.d.r. H nous définissons par $\tau_H = \sup\{x, \overline{H}(x) > 0\}$ l'extrémité supérieur du support. Nous supposons que $\tau < \tau_H < \infty$ et $\overline{H}(\tau) > \overline{H}(\tau_H) > 0$. De plus, nous utiliserons les hypothèses suivantes, rassemblées pour faciliter les références.

H1. La fenêtre h_n satisfait : $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n^d = +\infty$ et $\lim_{n\to\infty} \frac{\log n}{nh_n^d} = 0$.

H2. $h_n^d \log\log n = o(1)$.

H3. $\lim_{n\to\infty} nh_n^{d+4} = 0$.

K1. Le noyau $K(\cdot)$ est une fonction de densité bornée et à support compact.

K2. $\int \|t\| K_d(t)dt < \infty$ pour $\|t\| = \sum_{1\le i\le d} |t_i|$

K3. $\int \|t\| K_d^2(t)dt < \infty$ et $\int K_d^2(t)dt < \infty$.

D1. La fonction $\mu_\ell(\cdot)$, pour $\ell = 1,2$, est deux fois continûment différentiable.

D2. La fonction $r_\lambda(\cdot)$, pour $\lambda = 2,3,4$, est continûment différentiable.

D3. Il existe un $\Gamma > 0$ tel que $\mu_2(x) > \Gamma$ pour tout $x \in \mathcal{C}$.

M1. Pour toute constante positive C, nous supposons que la variable d'intérêt est bornée tel que :

$$T^{-\ell} \le C, \qquad \text{pour } \ell = 1,2.$$

Commentaires

1. Les hypothèses H1–H3 concerne le paramètre de lissage h_n et sont standards en estimation non paramétrique de la fonction de régression pour des données complètes et incomplètes (voir proposition 1.1 et 1.2 dans Tsybakov (2009)).
2. Les hypothèses K1–K3 concerne le noyau $K(\cdot)$. De plus, les hypothèses K2 et K3 interviennent dans les terme du biais et variance.
3. Les hypothèses D1 et D2 sont des hypothèses de régularité des fonction $\{\mu_\ell(\cdot),\ \ell = 1,2\}$ et $\{r_\lambda(\cdot),\ \lambda = 2,3,4\}$. De plus, D1 intervient dans l'étude du terme du biais et D2 intervient dans le terme de variance.

4. L'hypothèse D3 est une hypothèse technique nécessaire à l'obtention de nos résultats.
5. L'hypothèse M1 déclame la bornitude de l'inverse de la variable d'intérêt T. Cette dernière nous est utile pour la démonstration du terme de variance à l'aide des V-C classes. On note que cette hypothèse de bornitude est récurrente en régression non paramétrique et n'est en aucun cas gênante puisqu'on travaille avec des durées de vie.

Après avoir donné les hypothèses nous sommes en mesure de présenter nos principaux résultats. Le premier résultat concerne la convergence uniforme presque sûre avec vitesse. Les preuves sont établit dans la section 3.5.

Théorème 3.3.1 *Sous les hypothèses H1, K1–K3, D1–D3 et M1, nous avons :*

$$\sup_{x\in\mathcal{C}}|\widehat{m}_{RER}(x)-m(x)| = O_{p.s.}\left\{\max\left(\sqrt{\frac{\log n}{nh_n^d}},h_n\right)\right\} \quad \textit{quand} \quad n\longrightarrow\infty.$$

Remarque 3.3.1 *Dans le cas unidimensionnel (c-à-d pour $d=1$), si nous choisissons $h_n = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{5}}\right)$ alors le théorème 3.3.1 devient :*

$$\sup_{x\in\mathcal{C}}|\widehat{m}_{RER}(x)-m(x)| = O\left(\left(\frac{\log n}{n}\right)^{\frac{2}{5}}\right).$$

Le second théorème traite de la normalité asymptotique de l'estimateur $\widehat{m}_{RER}(\cdot)$. Notons que :

$$\Sigma(x)=\begin{pmatrix} r_2(x) & r_3(x)\\ r_3(x) & r_4(x)\end{pmatrix}$$

est la matrice de variance covariance où la fonction $r_\lambda(\cdot)$ pour $\lambda=2,3,4$ est donnée par la formule (3.17). Maintenant, nous somme en position de donner notre résultat de normalité asymptotique. Soit $\mathcal{C}^\star = \Big\{x\in\mathcal{C}$ tel que $\mu_\ell(x)\neq 0,\ \ell=1,2$ et $r_\lambda(x)\neq 0,\ \lambda=2,3,4\Big\}$.

Théorème 3.3.2 *Sous les hypothèses H1–H3, K1–K3, D1–D3 et M1, pour $x\in\mathcal{C}^\star$, nous avons :*

$$\sqrt{nh_n^d}\,(\widehat{m}_{RER}(x)-m(x))\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\sigma^2(x)\right) \qquad \textit{quand} \quad n\longrightarrow\infty$$

où

$$\sigma^2(x)=\kappa\frac{r_2(x)\mu_2^2(x)-2r_3(x)\mu_1(x)\mu_2(x)+r_4(x)\mu_1^2(x)}{\mu_2^4(x)} \tag{3.18}$$

pour $\kappa=\int K_d^2(t)dt$ et $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi.

3.3.1 Intervalles de confiances

Dans l'estimation non paramétrique, la variance asymptotique dépend de certaines fonction inconnues. Dans notre cas, pour déterminer les intervalles de confiance, nous

devons estimer la quantité inconnue $r_\lambda(\cdot)$ qui apparaît dans l'expression de la variance asymptotique. Définissons un estimateur consistant de $r_\lambda(\cdot)$ pour $\lambda = 2, 3, 4$ par :

$$\widehat{r_\lambda}(x) = \frac{1}{nh_n^d} \sum_{1\leq i\leq n} \frac{\delta_i Y_i^{-\lambda}}{\overline{G}_n^2(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right). \tag{3.19}$$

Nous remplaçons (5.2) dans (5.1) pour obtenir un estimateur calculable :

$$\widehat{\sigma^2}(x) = \kappa \frac{\widehat{r_2}(x)\widehat{\mu_2^2}(x) - 2\widehat{\mu_1}(x)\widehat{\mu_2}(x)\widehat{r_3}(x) + \widehat{\mu_1^2}(x)\widehat{r_4}(x)}{\widehat{\mu_2^4}(x)}. \tag{3.20}$$

Finalement, le théorème 3.3.2 devient :

Corollaire 3.3.1 *Sous les hypothèses du théorème 3.3.2, nous avons :*

$$\frac{\sqrt{nh_n^d}}{\widehat{\sigma}(x)} \left(\widehat{m}_{RER}(x) - m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Les intervalles de confiance de niveau $1 - \beta$ *avec* $0 < \beta < 1$ *sont données par :*

$$\left]\widehat{m}_{RER}(x) - t_{1-\frac{\beta}{2}} \frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}}; \widehat{m}_{RER}(x) + t_{1-\frac{\beta}{2}} \frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}}\right[$$

où $t_{1-\frac{\beta}{2}}$ *désigne le quantile de la loi normale standard.*

Remarque 3.3.2 *Le choix de la fenêtre a un impact capital sur le comportement de l'estimateur. Grace à sa simplicité et à sa consistance, la méthode de validation croisée a été largement utilisé dans la littérature. Nous adaptons cette dernière au problème étudié dans ce chapitre. L'idée principale est de minimiser l'erreur de prédiction, où dans notre cas, l'erreur quadratique relative moyenne donnée par :*

$$CV_{RER} = \frac{1}{n-1} \sum_{1\leq i\leq n} Y_i^{-2} \left(Y_i - \widehat{m}_{RER}^{-i}(X_i)\right)^2 \tag{3.21}$$

où $\widehat{m}_{RER}^{-i}(\cdot)$ *est l'estimateur de la fonction de régression relative obtenu dans (3.13) en supprimant la* i^{me} *observation* (X_i, Y_i, δ_i).

3.4 Étude numérique

Dans cette section, nous mettons en œuvre notre méthodologie pour évaluer la précision de l'estimateur pour les données d'échantillons finis. Les objectifs de cette étude sont les suivants :

— Montrer l'efficacité de la méthode proposée en présence de valeurs aberrantes pour différents taux de censure et tailles d'échantillon.
— Comparer la performance de l'estimateur développé à d'autres modèles de régression tels que le CR et le LLR.
— Illustration de l'efficacité du RER pour des données réelles (mélanome malin).

3.4.1 Convergence

Algorithme

Entrées Générer n i.i.d. v.a. $\{\varepsilon_i \rightsquigarrow \mathcal{N}(0.1), X_i \rightsquigarrow \mathcal{N}(3,1)$ et $C_i \rightsquigarrow \mathcal{N}(3+a,1)$, pour $i = 1,\ldots,n,\}$ et a est une constante qui permet de varier le pourcentage de censure.

Étape 1. Calculer $T_i = 2X_i + 1 + 0.2\varepsilon_i$.

Étape 2. Déterminer $Y_i = T_i \wedge C_i$ et $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$ ce qui donne l'ensemble observé $\{(X_i, Y_i, \delta_i), 1 \leq i \leq n\}$.

Étape 3. L'estimateur de K-M de $\overline{G}(\cdot)$ est calculé de (1.5).

Étape 4. Le noyau K est une densité normale standard.

Étape 5. La fonction théorique est :

$$m(x) = 2x + 1. \tag{3.22}$$

Étape 6. La fenêtre est sélectionnée sur l'intervalle $h_n \in [0.01, 2]$.

Étape 7. Calculer l'estimateur RER de (3.13) pour différentes valeurs de h_n et dans l'ensemble compact $\mathcal{C} = [1,4]$.

Étape 8. Le choix de la fenêtre optimale est obtenue en utilisant le critère de la validation croisée (voir la remarque (3.3.2)).

Sorties Compiler l'estimateur RER de (1.8) pour la fenêtre optimale.

1.a Effet de la taille d'échantillon : La figure 3.1 nous permet de vérifier les performances de l'estimateur RER pour un taux de censure fixe. Nous pouvons voir que la qualité d'ajustement à la droite théorique s'améliore lorsque n augmente.

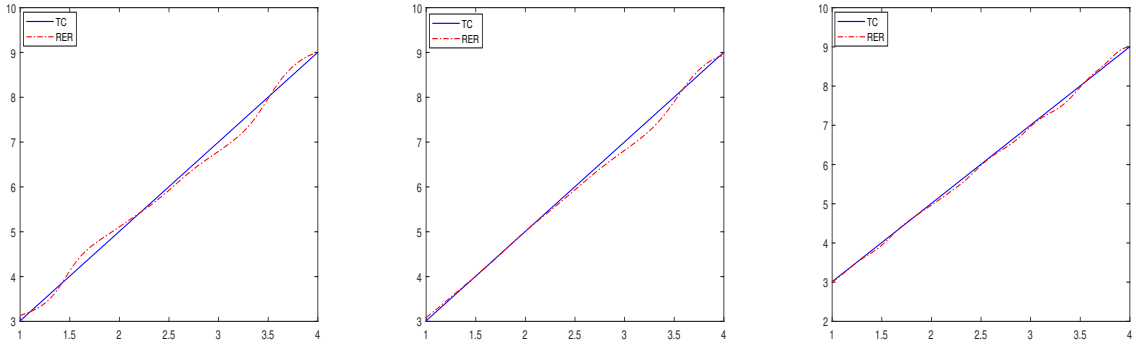


FIGURE 3.1 – $\widehat{m}_{RER}(\cdot)$ pour C.P.$\approx 50\%$ et $n = 100, 300$ et 500 respectivement.

1.b Effet du taux de censure : Figure 3.2 est le fruit d'une simulation de l'estimateur RER pour une taille d'échantillon fixe et une variation du taux de censure. Nous pouvons constater que la qualité d'ajustement se dégrade légèrement quand le taux de censure augmente ce qui est tout à fait normal parce qu'on observe de moins en moins de vraie valeurs (variable d'intérêt).

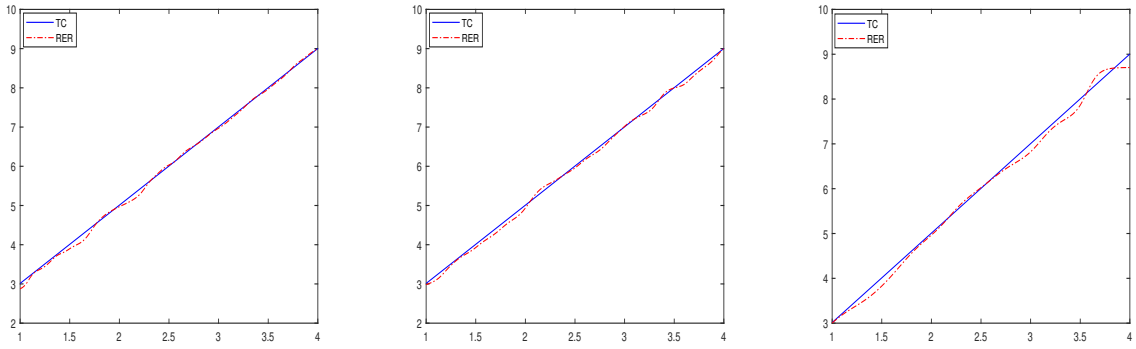


FIGURE 3.2 – $\widehat{m}_{RER}(\cdot)$ pour $n = 300$ et C.P.$\approx 11, 48, 74\%$ respectivement.

1.c. Fonctions non-linéaires : Pour montrer que l'estimateur RER s'adapte aux fonctions non linéaires, nous avons choisis trois modèles :

- Parabolique : $T = X^2 + 1 + 0.2\varepsilon$,
- Sinusoïdale : $T = \sin(\frac{X}{2})^2 + 1 + 0.2\varepsilon$,
- Exponentielle : $T = \exp(\frac{X}{2}) + 0.2\varepsilon$.

Les courbes sont illustrées dans la figure 3.3 Nous pouvons observé que l'estimateur suit bien la courbe théorique.

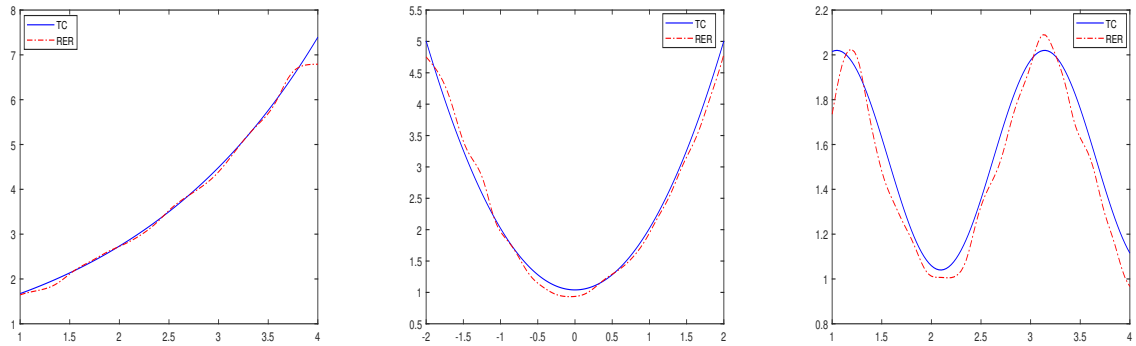


FIGURE 3.3 – $\widehat{m}_{RER}(\cdot)$ pour $n = 300$ et C.P.$\approx 42, 44$ and 33% pour les foncions Exponentielle, Parabolique et Sinusoïdale respectivement.

1.d EQM en fonction des valeurs aberrantes : Nous comparons les performances de notre estimateur par le moyen du calcul des courbes d'erreurs quadratiques moyennes définie par :

$$EQM(x_i) = \frac{1}{B}\sum_{j=1}^{B}\left(\widehat{m}_{RER,j}(x_i) - m(x_i)\right)^2, \tag{3.23}$$

pour un $B = 200$, C.P. fixés et plusieurs valeurs de facteur multiplicateur (M.F.) et n. Nous reportons les résultats dans la figure 3.4 Nous pouvons voir à travers les courbes de l'erreur quadratique moyenne ponctuelle de notre estimateur, que l'E.Q.M. est plus petite lorsque le M.F. est faible et qu'elle diminue encore plus lorsque le

n augmente. Nous pouvons conclure que notre estimateur résiste même lorsque le coefficient multiplicateur est très élevé.

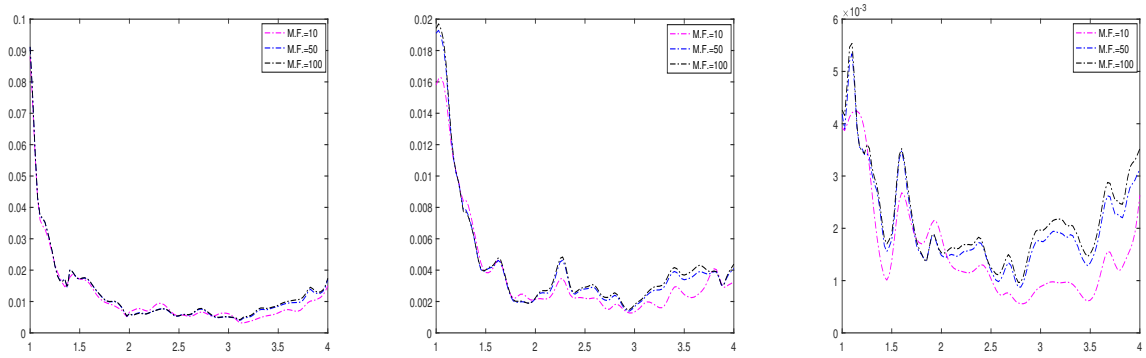


FIGURE 3.4 – EQM(·) avec $n = 100, 300$ et 500 (de gauche à droite) et C.P.$\approx 10\%$ pour M.F. $= 10, 50$ et 100 respectivement.

1.e Effet de la contamination de l'erreur aléatoire ε : Nous prenons le même algorithme qu'avant en changeant l'étape 1 qui devient :

Étape 1'. Calculer $T_i = 2X_i + 1 + \varepsilon_i$, où $\varepsilon_i \rightsquigarrow (1-\alpha)W_1 + \alpha W_2$, $W_1 \rightsquigarrow \mathcal{N}(0,1)$ et $W_2 \rightsquigarrow \mathcal{N}(0,\lambda)$. Nous prenons un niveau $\alpha \in (0,\ 0.1)$ et λ est souvent pris $\lambda = 3$.

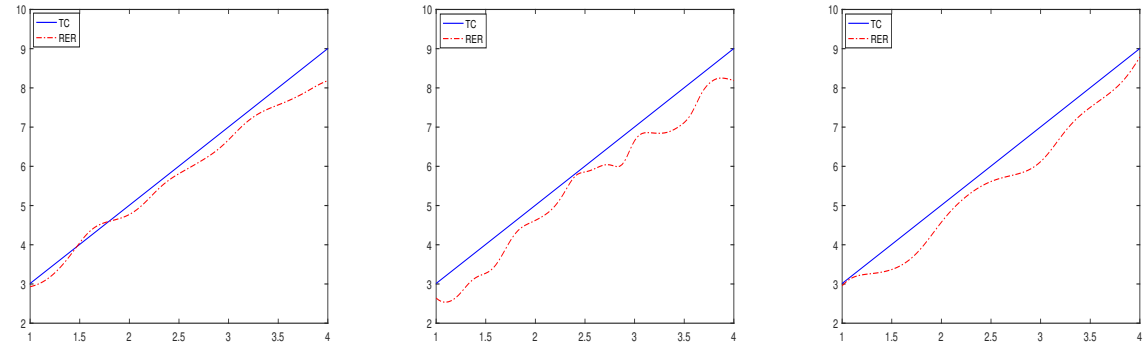


FIGURE 3.5 – $\widehat{m}_{RER}(\cdot)$ pour $n = 300$, C.P.$\approx 50\%$ et $\alpha = 0.01, 0.05$ et 0.1 respectivement.

Nous observons dans la figure 3.5 que plus le niveau alpha est élevé, moins il y a d'approximation, ce qui est prévisible.

2. Comparaison de l'EQM pour une erreur aléatoire contaminée ε : Dans cette seconde partie, nous étudions le comportement de trois estimateur (RER, CR et LLR) en comparant leurs EQM Nous prenons le même modèle comme en partie i et nous étudions les cas suivants : Le niveau $\alpha = 0,01,\ 0,05$ et $0,1$ et la magnitude $\lambda = 1,5$ et 3. Nous pouvons voir clairement dans le tableau 3.1 et 3.2, que l'estimateur RER est meilleur dans pratiquement tout les cas. Le seul cas où l'estimateur RER est moins bon que l'estimateur LLR est le cas où $n = 300$, avec $\alpha = 0.01$ ou 0.05 pour $\lambda = 3$.

3. Étude de comparaison : Afin de mettre en évidence l'efficacité de l'estimation des erreurs relatives qui est l'idée principale de cette partie, nous établissons une étude comparative entre CR et RER.

n	α	RER	CR	LLR
100	0.01	0.7842	1.2388	1.1314
	0.05	0.2809	1.9404	1.4796
	0.1	0.4646	1.2919	1.3820
300	0.01	0.4883	1.1906	0.8857
	0.05	0.2594	2.2350	1.5706
	0.1	0.3641	0.4566	1.4198
500	0.01	0.3663	0.6420	0.5632
	0.05	0.3631	1.3042	1.2343
	0.1	0.3212	0.3569	0.8328

Tableau 3.1 – EQM en fonction de α avec C.P. $\approx 30\%$ et $\lambda = 1.5$.

n	α	RER	CR	LLR
100	0.01	0.3925	1.2175	1.4218
	0.05	0.3984	1.3758	1.8018
	0.1	0.3926	1.6613	0.7891
300	0.01	0.4188	0.8304	0.1393
	0.05	0.3762	0.6885	0.3299
	0.1	0.1892	0.9223	0.6857
500	0.01	0.3784	1.5956	1.6327
	0.05	0.3113	1.1179	0.8262
	0.1	0.3082	0.7595	0.3735

Tableau 3.2 – EQM en fonction de α avec C.P. $\approx 30\%$ et $\lambda = 3$.

3.a CR versus RER : Dans cette partie, nous avons comparé notre estimateur à l'estimateur CR (défini par Guessoum and Ould Saïd (2008)) pour différents taux de censure. Il est clair de la figure 3.6 que la courbe RER est collée à la courbe théorique contrairement à la courbe CR qui s'éloigne de plus en plus quand le taux de censure C.P. augmente.

3.b Effet des valeurs aberrantes : Pour renforcer notre théorie qui repose sur l'efficacité du RER en présence de valeurs aberrantes, nous réalisons une étude comparative dans laquelle deux estimateurs à noyau sont également simulés pour des données censurées à droite de façon aléatoire. Le premier est l'estimateur CR (voir la formule (1.15)) et le second est l'estimateur LLR défini dans (2.8). Pour créer cet effet de valeurs aberrantes, chaque valeur de rang 20 de l'échantillon est multipliée par un facteur (M.F.) que nous augmentons progressivement. Ensuite, à partir de figure 3.7, nous pouvons observer très clairement que la courbe RER est très proche de la courbe théorique, contrairement à la courbe CR et LLR. Ainsi, notre estimateur est plus efficace en présence de valeurs aberrantes, ce qui est l'objectif principal de cet article.

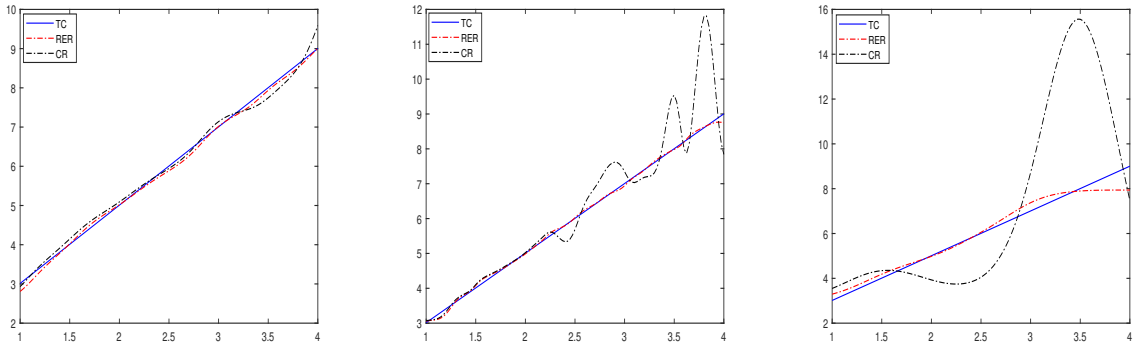


FIGURE 3.6 – $\widehat{m}_{RER}(\cdot)$ et $\widehat{m}_{CR}(\cdot)$ pour $n = 100$ et C.P.$\approx 10, 35, 80\%$ respectivement.

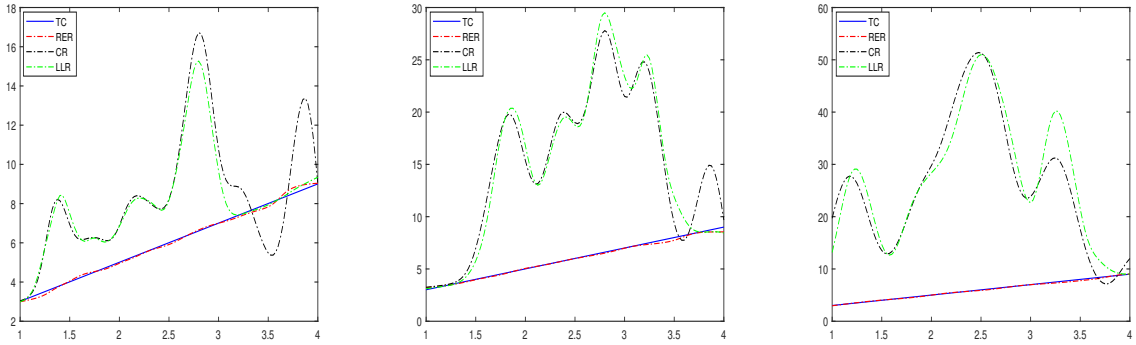


FIGURE 3.7 – $\widehat{m}_{RER}(\cdot)$, $\widehat{m}_{CR}(\cdot)$, et $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ et C.P.$\approx 50\%$ pour M.F. $= 10, 25$ et 50 respectivement.

3.c Effet de la censure : Pour prouver la solidité de notre approche, nous changeons de rôle. Nous fixons la taille de l'échantillon, la M.F. et faisons varier le C.P. Nous pouvons observer à partir de figure 3.8 que l'estimateur RER est résistant en présence de censure, contrairement aux deux autres (CR et LLR) qui s'écartent considérablement de la courbe théorique.

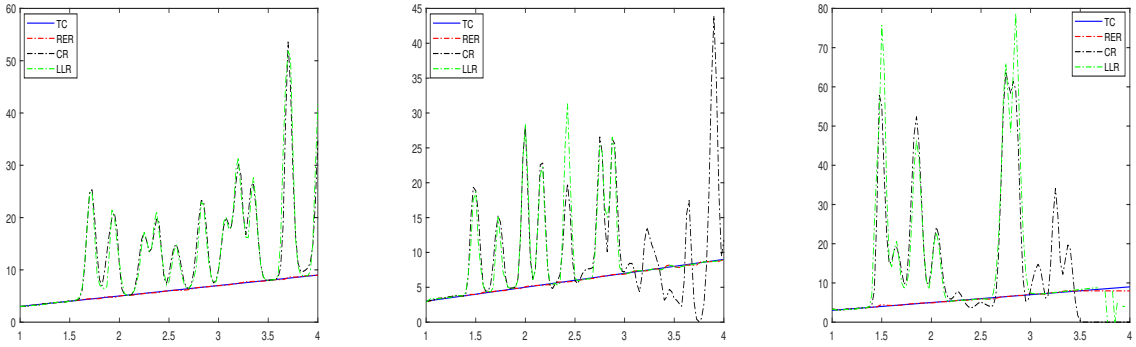


FIGURE 3.8 – $\widehat{m}_{RER}(\cdot)$, $\widehat{m}_{CR}(\cdot)$ e $\widehat{m}_{LLR}(\cdot)$ avec $n = 300$ et M.F.$= 25$ pour C.P.$\approx 8, 45, 70\%$ respectivement.

4. Tableau des EAM : Pour finaliser cette partie, nous comparons la moyenne des erreurs absolues pour les méthodes RER (EAM^1), CR (EAM^2) et LLR (EAM^3) sont représentées par :

$$\text{EAM}^1 = \frac{1}{n}\sum_{1\leq i\leq n} |\widehat{m}_{RER}(x_i) - m(x_i)|,$$

$$\text{EAM}^2 = \frac{1}{n}\sum_{1\leq i\leq n} |\widehat{m}_{CR}(x_i) - m(x_i)|,$$

et

$$\text{EAM}^3 = \frac{1}{n}\sum_{1\leq i\leq n} |\widehat{m}_{LLR}(x_i) - m(x_i)|,$$

pour différentes n, C.P. et M.F. et pour le même modèle linéaire donné par (3.22). Nous observons du tableau 3.3 que EAM^1 est significativement petites que EAM^2 et EAM^3 en présence de valeurs aberrantes. Ceci prouve que l'estimation de la fonction de régression via l'erreur relative est robuste dans le cas de valeurs aberrantes.

n	M.F.	EAM^1	EAM^2	EAM^3
100	10	0.1217	2.0864	3.1224
	25	0.1237	8.4469	9.9929
	50	0.1275	11.0175	11.2813
300	10	0.0413	2.6293	2.2035
	25	0.0468	8.2153	11.1731
	50	0.0575	12.3789	12.0252
500	10	0.0381	2.8884	2.2822
	25	0.0409	5.7154	6.1769
	50	0.0418	8.8666	8.1872

Tableau 3.3 – EAM en fonction des valeurs aberrantes pour C.P. $\approx 33\%$.

3.4.2 Normalité asymptotique

Dans cette sous section, nous illustrons la normalité asymptotique. Pour ce faire, nous comparons la forme de l'estimateur à noyau de la densité d'une suite générer selon notre modèle avec la densité de la loi normale standard (centrée et réduite).

Algorithme

Pour $j = 1 : m$ faire

Étape 1. Générer n i.i.d. v.a. $\{\varepsilon_i \rightsquigarrow \mathcal{N}(0,1), X_i \rightsquigarrow \exp(1.5)$ et $C_i \rightsquigarrow \exp(3+a)$, pour $i = 1,\ldots,n,\}$ et a est une constante qui permet d'adapter le C.P.

Étape 2. La variable d'intérêt est calculer selon le modèle linéaire suivant : $T_i = 2X_i + 1 + 0.2\varepsilon_i$.

Étape 3. Nous fixons $x = 0$ et nous calculons $\{\widehat{\mu_\ell}(x),\ \ell = 1,2\}$ de (3.16) et $\{\widehat{r_\lambda}(x),\ \lambda = 2,3,4\}$ de (3.17).

Étape 4. La fenêtre optimale de l'estimateur à noyau de la densité est $h_m = Cm^{-\frac{1}{5}}$ (voir e.g. Silverman (1986)). Une densité gaussienne standard est prise comme noyau K.

Étape 5. Calculer la variance asymptotique :

$$\widehat{\sigma}^2(0) = \kappa \frac{\widehat{r}_2(0)\widehat{\mu}_2^2(0) - 2\widehat{r}_3(0)\widehat{\mu}_1(0)\widehat{\mu}_2(0) + \widehat{r}_4(0)\widehat{\mu}_1^2(0)}{\widehat{\mu}_2^4(0)},$$

Étape 6. Calculer la suite $A_j = \left(\frac{nh_n}{\widehat{\sigma}_j^2(0)}\right)^{1/2} (\widehat{m}_{RER,j}(0) - 1),\ j = 1,\dots,m$ pour lequel nous construisons un estimateur de la densité du type N-W et nous le comparons à la courbe de la loi normale standard.

Nous pouvons voir de la figure 3.9 la bonne qualité de l'estimateur étudié.

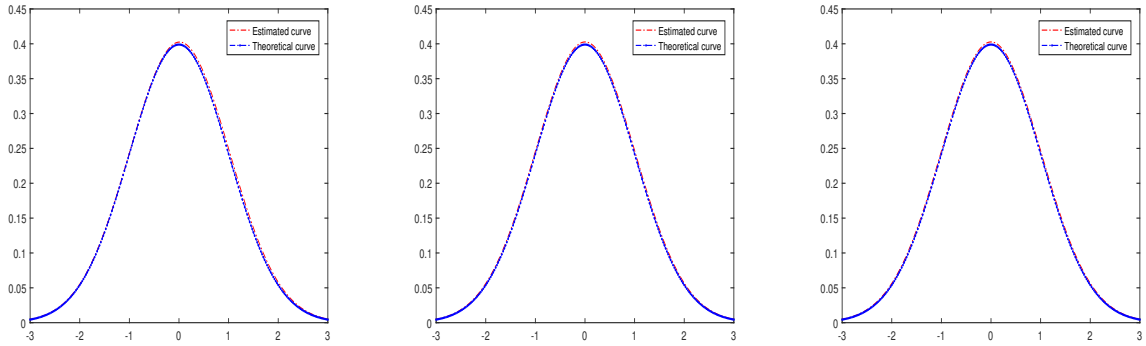


FIGURE 3.9 – C.P.≈ 66%, $m = 200$ pour $n = 100, 300$ et 500, respectivement.

3.4.3 Intervalles de confiances

Nous avons construit les courbes de confiance (à 95%) pour différentes valeurs de n. Nous avons illustré sur le même graphe les deux courbes correspondant à la fonction théorique $m(\cdot)$ et l'estimateur $\widehat{m}_{RER}(\cdot)$ pour un $x \in [1, 4]$. Le paramètre de lissage optimal h_{opt} nous utilisons la méthode le validation croisée (voir la remarque 3.3.2).

La qualité des intervalles de confiance s'améliore avec l'augmentation de la taille de l'échantillon, comme le montre la figure 3.10.

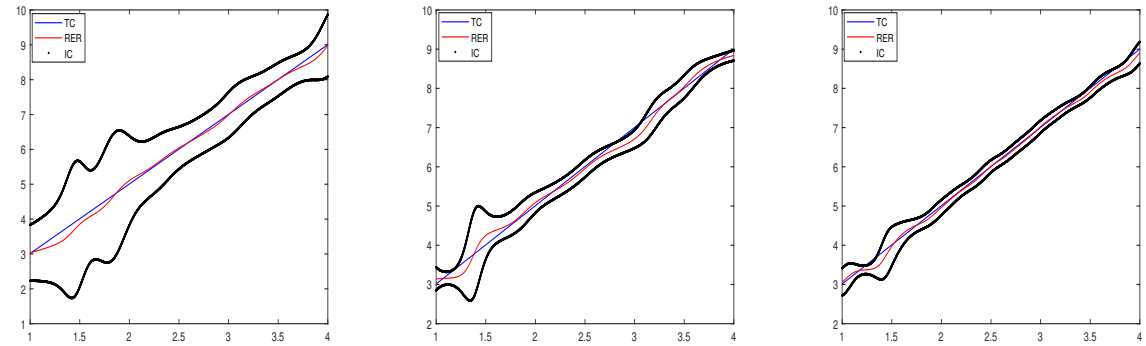


FIGURE 3.10 – $\widehat{m}_{RER}(\cdot)$ avec C.P.≈ 30% pour $n = 50, 150$, et 250 respectivement.

3.4.4 Application sur des données réelles

Dans cette partie, nous analysons un ensembles de données réelles pour illustrer l'efficacité de l'estimateur de régression des erreurs relatives (RER) en présence des données de censure. De plus, nous le comparons aux l'estimateurs CR étudié Guessoum and Ould Saïd (2008) et LLR définit dans le chapitre 2.
Les données consistent en des mesures effectuées sur des patients atteints de mélanome malin (cancer de la peau). Chaque patient a subi une opération radicale au département de chirurgie plastique de l'hôpital universitaire d'Odense, au Danemark. C'est-à-dire que la tumeur a été complètement retirée avec la peau à une distance d'environ 2.5 cm autour d'elle. Tous les patients ont été suivis jusqu'à la fin de l'année 1977.
Les données couvrent 205 temps de survie (en jours) des patients depuis l'opération "mélanome malin", voir Andersen et al. (1993). L'ensemble de données fournit des informations sur :

- Le sexe (homme/femme), l'âge (en années).
- Statut (1 indique qu'ils sont morts d'un mélanome, 2 indique qu'ils étaient encore en vie, 3 indique qu'ils sont morts de causes non liées à leur mélanome (pour nous 2=3=censuré)).
- Âge en années au moment de l'opération (notez que bien que les patients entrent dans l'étude à des moments différents du calendrier, tous les patients entrent au moment 0 dans l'échelle de temps (temps depuis l'opération) utilisée comme variable de réponse) de chaque patient.
- L'épaisseur de la tumeur (en mm) et si elle était ulcéreuse ou non (1 présente et 0 absente).

La variable temporelle considérée comme la plus importante est le temps écoulé depuis l'opération. Nous considérons le lien entre le temps de survie (la variable d'intérêt) et l'épaisseur de la tumeur (la variable explicative ou co-variable). Pour comparer la performance de prévision, 190 données ont été sélectionnés au hasard comme échantillon d'apprentissage, désignées par (X_i, Y_i, δ_i), $i = 1, \ldots, 190$, et les points de données restants ont été traités comme échantillon de test, désignées par (X_i, Y_i, δ_i), $i = 191, \ldots, 205$. Comme nous ne pouvons pas prédire des données qui sont censurées, nous éliminons les données censurées des valeurs prédites. Alors, au lieu de prédire 15 valeurs, nous ne faisons la prévision que de 6 valeurs sur les 15 qui ne sont pas censurées. Nous avons pris un noyau gaussien standard K et la largeur de fenêtre h_n est choisie par la méthode de validation croisée (voir la remarque 3.3.2).

Afin d'obtenir plus de précision sur notre estimateur, nous évaluons les erreurs quadratiques moyennes (EQM), avec

$$\text{EQM} := \frac{1}{6}\sum_{i=1}^{6}(Y_i - \widehat{Y_i})^2$$

où Y_i (resp. $\widehat{Y_i}$) est la variable réelle (resp. estimée) du i-ème patient. Nos prévisions sont telles que les valeurs réelles et les valeurs prédites sont si proches qu'il nous est difficile de les différencier.

En conclusion, la méthode RER semble améliorer la qualité de la prévision par rapport aux méthodes CR et LLR. Nous soulignons que le taux de censure des données réelles est de l'ordre de 72, 2% et le fait que les prédictions de notre estimateur correspondent à 99% des données de l'échantillon contrairement aux deux autres estimateurs

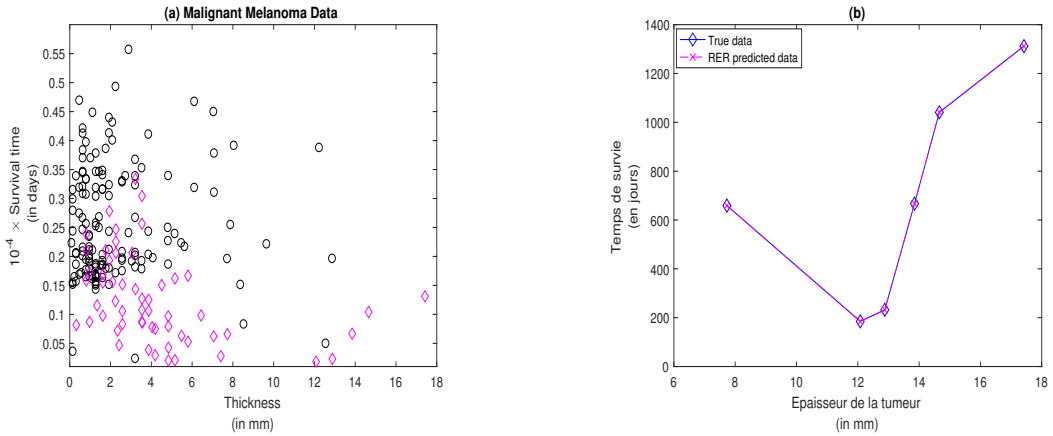


FIGURE 3.11 – **(a)** Données sur le mélanome malin. o : données censurées et ⋄ : données non censurées. **(b)** Losange : vraies valeurs et croix : prédiction du RER.

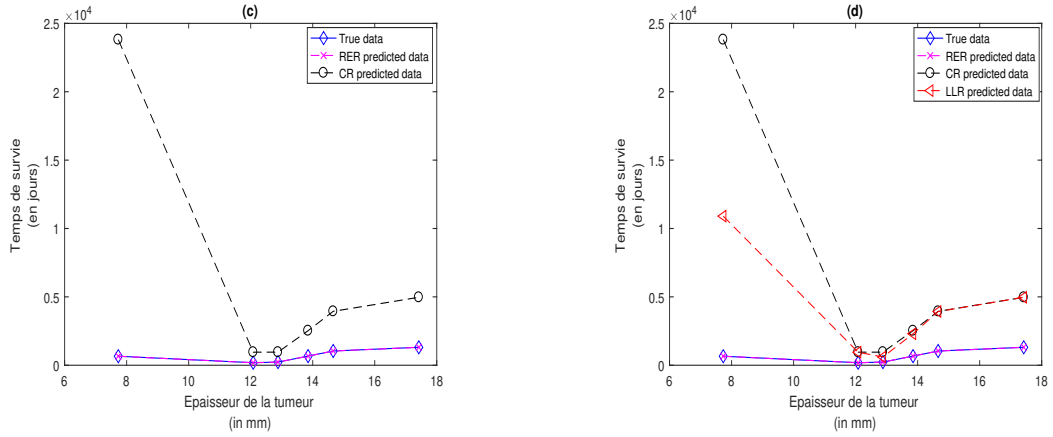


FIGURE 3.12 – **(c)** et **(d)** Losange : vraies valeurs, croix : prévision du RER, rond : prévision du CR et triangle : prévision du LLR.

qui correspondent de manière globale à 30%.

3.5 Preuves et résultats auxiliaires

Rappelons que nous disposons d'un n-triplets aléatoires i.i.d. $(X_1, Y_1, \delta_1), \cdots, (X_n, Y_n, \delta_n)$ à valeurs dans $\mathbb{R}^d \times \mathbb{R}_+^\star \times \{0,1\}$. Dans ce qui va suivre, nous allons donner une suite de lemmes nécessaires à la preuve du théorème 3.3.1. Le premier lemme aborde le terme "dominant" celui de variance, pour lequel on a opté pour un traitement avec la méthodes des V-C classes (voir la sous-section 5.4.2 du chapitre 1).
La preuve du théorème est basée sur la décomposition suivante :

$$\widehat{m}_{RER}(x) - m(x) = \frac{1}{\widehat{\mu}_2(x)} \left\{ \left[\left(\widehat{\mu}_1(x) - \widetilde{\mu}_1(x) \right) + (\widetilde{\mu}_1(x) - \mathbb{E}[\widetilde{\mu}_1(x)]) + (\mathbb{E}[\widetilde{\mu}_1(x)] - \mu_1(x)) \right] \right.$$
$$\left. + m(x) \left[(\widetilde{\mu}_2(x) - \widehat{\mu}_2(x)) + (\mathbb{E}[\widetilde{\mu}_2(x)] - \widetilde{\mu}_2(x)) + (\mu_2(x) - \mathbb{E}[\widetilde{\mu}_2(x)]) \right] \right\}.$$

En utilisant l'inégalité triangulaire, nous avons :

$$\begin{aligned}
&\sup_{x\in\mathcal{C}}|\widehat{m}_{RER}(x)-m(x)| \\
&\quad\leq \frac{1}{\inf\limits_{x\in\mathcal{C}}|\widehat{\mu}_2(x)|}\left\{\sup_{x\in\mathcal{C}}\left[|\widehat{\mu}_1(x)-\widetilde{\mu}_1(x)|+|\widetilde{\mu}_1(x)-\mathbb{E}[\widetilde{\mu}_1(x)]|+|\mathbb{E}[\widetilde{\mu}_1(x)]-\mu_1(x)|\right]\right. \\
&\qquad\left.+\sup_{x\in\mathcal{C}}|m(x)|\left[|\widetilde{\mu}_2(x)-\widehat{\mu}_2(x)|+|\mathbb{E}[\widetilde{\mu}_2(x)]-\widetilde{\mu}_2(x)|+|\mu_2(x)-\mathbb{E}[\widetilde{\mu}_2(x)]|\right]\right\}. \quad (3.24)
\end{aligned}$$

Lemme 3.5.1 *Sous les hypothèses H1, K1, K3 et D2, pour $\ell = 1,2$, nous avons :*

$$\sup_{x\in\mathcal{C}}\left|\widetilde{\mu}_\ell(x)-\mathbb{E}[\widetilde{\mu}_\ell(x)]\right| = O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \quad quand \quad n\longrightarrow\infty.$$

Preuve du lemme 3.5.1. Nous considérons la suite i.i.d. $(X_1,Y_1,\delta_1),\ldots,(X_n,Y_n,\delta_n)$ et définissons :

$$\Phi_n=\left\{\theta_x:\mathbb{R}^d\times\mathbb{R}_+^\star\times\{0,1\}\to\mathbb{R}^+/\theta_x(u,y,\delta)=\frac{\delta y^{-\ell}}{nh_n^d\overline{G}(y)}K_d\left(\frac{x-u}{h_n}\right),\quad x\in\mathcal{C}\right\}.$$

Du lemme (3b) dans Giné and Guillou (1999), Φ_n est une Vapnik-Cervonenkis (V-C) classe de fonctions mesurables positives. Cette dernière est uniformément bornée par l'enveloppe $\Theta=\dfrac{C\|K\|_\infty}{nh_n^d\,\overline{G}(\tau)}$. De plus, sous l'hypothèse K1, nous avons :

$$\sup_{x\in\mathcal{C}}\theta_x(X_1,Y_1,\delta_1)\leq\frac{C\|K\|_\infty}{nh_n^d\,\overline{G}(\tau)}=:U_n.$$

Par ailleurs, en utilisant le propriété de l'espérance conditionnelle, un changement de variable et sous les hypothèses K3 et D2, pour $\gamma=2\ell$, avec $\ell=1,2$ nous avons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\mathbb{V}ar\left[\theta_x(X_1,Y_1,\delta_1)\right] &\leq \sup_{x\in\mathcal{C}}\mathbb{E}\left[\theta_x^2(X_1,Y_1,\delta_1)\right] \\
&= \frac{1}{n^2h_n^{2d}}\sup_{x\in\mathcal{C}}\left|\mathbb{E}\left[K_d^2\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1^{-2\ell}|X_1\right]\right]\right| \\
&= \frac{1}{n^2h_n^{2d}}\sup_{x\in\mathcal{C}}\left|\int K_d^2\left(\frac{x-u}{h_n}\right)\int\frac{t^{-\lambda}}{\overline{G}(t)}f(t|u)dtf(u)du\right| \\
&= \frac{1}{n^2h_n^{2d}}\sup_{x\in\mathcal{C}}\left|\int K_d^2\left(\frac{x-u}{h_n}\right)\int\frac{t^{-\lambda}}{\overline{G}(t)}f(u,t)dtdu\right| \\
&= \frac{1}{n^2h_n^{d}}\sup_{x\in\mathcal{C}}\left|\int K_d^2(t)\,r_\lambda(x-th_n)dt\right| \\
&\leq \frac{\|K\|_\infty^2\|r_\lambda\|_\infty}{n^2h_n^d}=:\sigma_n^2
\end{aligned}$$

avec $\sigma_n\leq U_n$ pour n assez grand. Ainsi, nous pouvons appliqué l'inégalité de Talagrand (voir la sous-section 5.4.2 du chapitre 2). Il existe trois constantes positives C_1, C_2 et C_3

avec $t \geq C_1\sqrt{\frac{\log n}{nh_n^d}}$. Nous avons :

$$\mathbb{P}\left[\sup_{\theta_x\in\Phi_n}\left|\sum_{i=1}^{n}(\theta_x(X_i,Y_i,\delta_i)-\mathbb{E}[\theta_x(X_1,Y_1,\delta_1)])\right| > C_1\sqrt{\frac{\log n}{nh_n^d}}\right]$$

$$\leq C_2\exp\left(-\frac{C_1\sqrt{\frac{\log n}{nh_n^d}}}{C_2\frac{C\|K\|_\infty}{nh_n^d\overline{G}(\tau)}}\log\left[1+\frac{C_1\sqrt{\frac{\log n}{nh_n^d}}\times\frac{C\|K\|_\infty}{nh_n^d\overline{G}(\tau)}}{C_2\left(\sqrt{n}\frac{\|K\|_\infty\sqrt{\|r_\lambda\|_\infty}}{nh_n^{d/2}}+\frac{C\|K\|_\infty}{nh_n^d\overline{G}(\tau)}\sqrt{\log C_3\frac{\|\mu_\ell\|_\infty}{\sqrt{\|r_\lambda\|_\infty}}}\right)^2}\right]\right),$$

en utilisant $\log(1+x)\approx x($ pour $x\to 0)$, la partie de droite de la dernière inégalité devient :

$$C_2\exp\left(-\frac{C_1\sqrt{\frac{\log n}{nh_n^d}}}{C_2\frac{C\|K\|_\infty}{nh_n^d\overline{G}(\tau)}}\times\frac{C_1\sqrt{\frac{\log n}{nh_n^d}}\times\frac{C\|K\|_\infty}{nh_n^d\overline{G}(\tau)}}{C_2\left(\sqrt{n}\frac{\|K\|_\infty\sqrt{\|r_\lambda\|_\infty}}{nh_n^{d/2}}\right)^2}\right)=C_2 n^{-\left(\frac{C_1}{C_2}\right)^2\frac{1}{\|K\|_\infty^2\|r_\lambda\|_\infty}},$$

où par un choix approprié des constantes C_1 et C_2, le majorant est de l'ordre de $n^{-3/2}$. Ce dernier est le terme général d'une série convergente qui par le lemme de Borel-Cantelli conclut la preuve.

Lemme 3.5.2 *Sous les hypothèses H1, D1, D2, K1–K3, pour $\ell=1,2$, nous avons :*

$$\sup_{x\in\mathcal{C}}|\widehat{\mu}_\ell(x)-\widetilde{\mu}_\ell(x)| = O_{p.s.}\left\{\sqrt{\frac{\log\log n}{n}}\right\} \quad \textit{quand} \quad n\longrightarrow\infty. \tag{3.25}$$

Preuve du Lemme 3.5.2. Pour $\ell=1,2$, nous avons :

$$\begin{aligned}
\widehat{\mu}_\ell(x)-\widetilde{\mu}_\ell(x) &= \frac{1}{nh_n^d}\sum_{1\leq i\leq n}\left\{\widehat{T}_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)-\widetilde{T}_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\right\}\\
&= \frac{1}{nh_n^d}\sum_{1\leq i\leq n}\left\{\frac{T_i^{-\ell}}{\overline{G}_n(T_i)}\mathbb{1}_{\{T_i\leq C_i\}}K_d\left(\frac{x-X_i}{h_n}\right)-\frac{T_i^{-\ell}}{\overline{G}(T_i)}\mathbb{1}_{\{T_i\leq C_i\}}K_d\left(\frac{x-X_i}{h_n}\right)\right\}\\
&= \frac{1}{nh_n^d}\sum_{1\leq i\leq n}T_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\left(\frac{1}{\overline{G}_n(T_i)}-\frac{1}{\overline{G}(T_i)}\right)\\
&= \frac{1}{nh_n^d}\sum_{1\leq i\leq n}T_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\left(\frac{\overline{G}(T_i)-\overline{G}_n(T_i)}{\overline{G}_n(T_i)\overline{G}(T_i)}\right).
\end{aligned}$$

Alors

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\widehat{\mu}_\ell(x)-\widetilde{\mu}_\ell(x)\right| &\leq \frac{1}{\overline{G}_n(\tau)\overline{G}(\tau)}\sup_{t\leq\tau}\left|\overline{G}_n(t)-\overline{G}(t)\right|\times\sup_{x\in\mathcal{C}}\left|\frac{1}{nh_n^d}\sum_{1\leq i\leq n}T_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\right|\\
&=: \frac{1}{\overline{G}^2(\tau)}\sup_{t\leq\tau}|\mathcal{L}_1(t)|\times\sup_{x\in\mathcal{C}}|\mathcal{L}_2(x)|.
\end{aligned} \tag{3.26}$$

D'une part, pour $\mathcal{L}_1$ d'après le résultat de Deheuvels and Einmahl (2000), nous avons :

$$\sup_{t\leq\tau}|\mathcal{L}_1(t)| = O_{p.s.}\left\{\sqrt{\frac{\log\log n}{n}}\right\} \quad \text{quand} \quad n\longrightarrow\infty. \tag{3.27}$$

Pour $\mathcal{L}_2$, nous avons :

$$\begin{aligned}\mathcal{L}_2(x) &= \left\{\frac{1}{nh_n^d}\sum_{1\leq i\leq n} T_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right) - \frac{1}{h_n^d}\mathbb{E}\left[T_1^{-\ell}K_d\left(\frac{x-X_1}{h_n}\right)\right]\right\} + \frac{1}{h_n^d}\mathbb{E}\left[T_1^{-\ell}K_d\left(\frac{x-X_1}{h_n}\right)\right]\\ &=: \mathcal{L}_{2,1}(x)+\mathcal{L}_{2,2}(x)\end{aligned}$$

Pour $\mathcal{L}_{2,2}$, en utilisant la propriété de l'espérance conditionnelle et un changement de variable, nous avons :

$$\begin{aligned}\mathcal{L}_{2,2}(x) &= \frac{1}{h_n^d}\mathbb{E}\left[K_d\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[T_1^{-\ell}|X_1\right]\right]\\ &= \frac{1}{h_n^d}\mathbb{E}\left[K_d\left(\frac{x-X_1}{h_n}\right)m_\ell(X_1)\right]\\ &= \frac{1}{h_n^d}\int K_d\left(\frac{x-u}{h_n}\right)m_\ell(u)f(u)du\\ &= \frac{1}{h_n^d}\int K_d\left(\frac{x-u}{h_n}\right)\mu_\ell(u)du\\ &= \frac{1}{h_n^d}\int K_d(t)\,\mu_\ell(x-h_nt)h_n^d dt\\ &= \int K_d(t)\,\mu_\ell(x-h_nt)dt.\end{aligned}$$

À ce stade, nous allons introduire le développement de Taylor à l'ordre 2 de la fonction $\mu_\ell(\cdot)$ pour $\ell=1,2$ qui nous sera utile dans nos calculs par :

$$\mu_\ell(x-h_nt) = \mu_\ell(x) - h_n\sum_{1\leq i\leq d} t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}. \tag{3.28}$$

D'où, en utilisant le développement de Taylor (3.28) à l'ordre 1 et sous les hypothèse K1, K2 et D1, nous avons :

$$\begin{aligned}\sup_{x\in\mathcal{C}}\left|\mathcal{L}_{2,2}(x)\right| &= \sup_{x\in\mathcal{C}}\left|\int K_d(t)\left(\mu_\ell(x) - h_n\sum_{1\leq i\leq d} t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right)dt\right|\\ &\leq \sup_{x\in\mathcal{C}}\left|\mu_\ell(x)\right| + h_n\sum_{1\leq i\leq d}\sup_{x\in\mathcal{C}}\left|\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right|\int|t_i|K_d(t)\,dt\\ &= o(1).\end{aligned} \tag{3.29}$$

Pour $\mathcal{L}_{2,1}$, un preuve analogue au lemme 3.5.1 pour $C = +\infty$, donne :

$$\sup_{x\in\mathcal{C}}\left|\mathcal{L}_{2,1}(x)\right| = O_{p.s.}\left\{\sqrt{\frac{\log n}{nh_n^d}}\right\} \quad \text{quand} \quad n \longrightarrow \infty. \tag{3.30}$$

En combinant (3.27)–(3.30) nous obtenons notre résultat.

Lemme 3.5.3 *Sous les hypothèses H1, K1, K2 et D1, pour $\ell = 1, 2$, nous avons :*

$$\sup_{x\in\mathcal{C}}\left|\mathbb{E}[\widetilde{\mu}_\ell(x)] - \mu_\ell(x)\right| = O(h_n) \qquad \textit{quand} \quad n \longrightarrow \infty.$$

Preuve du lemme 3.5.3. En utilisant la propriété de l'espérance conditionnelle, nous avons dans un premier temps :

$$\begin{aligned}
\mathbb{E}[\widetilde{\mu}_\ell(x)] &= h_n^{-d}\mathbb{E}\left[\widetilde{T}_1^{-\ell}K_d\left(\frac{x-X_1}{h_n}\right)\right] \\
&= h_n^{-d}\mathbb{E}\left[K_d\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1^{-\ell}|X_1\right]\right] \\
&= h_n^{-d}\int K_d\left(\frac{x-u}{h_n}\right)m_\ell(u)f(u)du \\
&= h_n^{-d}\int K_d\left(\frac{x-X_1}{h_n}\right)\mu_\ell(u)du.
\end{aligned}$$

Dans un second temps, par un changement de variable et le développement de Taylor (3.28), nous obtenons :

$$\begin{aligned}
\mathbb{E}[\widetilde{\mu}_\ell(x)] - \mu_\ell(x) &= h_n^{-d}\int K_d\left(\frac{x-u}{h_n}\right)\mu_\ell(u)du - \mu_\ell(x) \\
&= \int K_d(t)[\mu_\ell(x-h_nt) - \mu_\ell(x)]dt \\
&= \int K_d(t)\left[-h_n\sum_{1\le i\le d} t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right]dt.
\end{aligned}$$

Pour finir, sous les hypothèses D1, K1, K2, nous obtenons :

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\left|\mathbb{E}[\widetilde{\mu}_\ell(x)] - \mu_\ell(x)\right| &= \sup_{x\in\mathcal{C}}\left|\int K_d(t)\left[-h_n\sum_{1\le i\le d} t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right]\right| \\
&\le h_n\sum_{1\le i\le d}\sup_{x\in\mathcal{C}}\left|\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right|\int |t_i|K_d(t)dt \\
&= \mathrm{O}(h_n).
\end{aligned}$$

Ce qui conclut la preuve du lemme 3.5.3.
Finalement, les résultats du lemme 3.5.1 au lemme 3.5.3 permettent de conclure la preuve du théorème 3.3.1.

Preuve du théorème 3.3.2. Notre but est de montrer :

$$\sqrt{nh_n^d}(\widehat{m}_{RER}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2(x)) \quad \text{quand} \quad n \longrightarrow \infty.$$

Notons que, pour $\ell = 1, 2$,

$$\begin{aligned}\sqrt{nh_n^d}(\widehat{\mu}_\ell(x) - \mu_\ell(x)) &= \sqrt{nh_n^d}\,(\widehat{\mu}_\ell(x) - \widetilde{\mu}_\ell(x)) + \sqrt{nh_n^d}\,(\widetilde{\mu}_\ell(x) - \mathbb{E}[\widetilde{\mu}_\ell(x)]) \\ &+ \sqrt{nh^d}\,(\mathbb{E}[\widetilde{\mu}_\ell(x)] - \mu_\ell(x)) \\ &=: \Lambda_\ell(x) + \Gamma_\ell(x) + \Xi_\ell(x). \end{aligned} \tag{3.31}$$

D'une part, nous considérons les termes Λ_ℓ et Ξ_ℓ que nous montrons négligeable.

Lemme 3.5.4 *Sous H2 et H3 , par le lemme 3.5.2 et lemme 3.5.3 les deux quantités $\Lambda_\ell(x)$ et $\Xi_\ell(x)$ sont $o(1)$ lorsque $n \to \infty$.*

Preuve du lemme 3.5.4. Du lemme 3.5.2 et sous l'hypothèse H2, nous obtenons :

$$\Lambda_\ell(x) = \sqrt{nh_n^d}\,(\widehat{\mu}_\ell(x) - \widetilde{\mu}_\ell(x)) = O_{p.s.}\left(\sqrt{h_n^d \log\log n}\right) = o_{p.s.}(1). \tag{3.32}$$

De la même manière, du lemme 3.5.3 et sous l'hypothèse H3, nous obtenons :

$$\Xi_\ell(x) = \sqrt{nh_n^d}\,(\mathbb{E}[\widetilde{\mu}_\ell(x)] - \mu_\ell(x)) = O_{p.s.}\left(\sqrt{nh_n^{d+4}}\right) = o_{p.s.}(1). \tag{3.33}$$

Nous considérons maintenant le terme dominant $\Gamma_\ell(x)$ pour $\ell \in \{1, 2\}$ et prouvons :

Lemme 3.5.5 *Sous les hypothèses H1, K1, D1 et D2, nous avons :*

$$(\Gamma_1(x), \Gamma_2(x))^t \xrightarrow{\mathcal{L}} \mathcal{N}(0, \kappa\Sigma(x)) \qquad \textit{quand} \quad n \to \infty.$$

Preuve du lemme 3.5.5. Nous commençons par calculer la variance asymptotique. Par définition de la variance, pour $\ell = 1, 2$, nous avons :

$$\begin{aligned}Var(\Gamma_\ell(x)) &= nh_n^d\left\{\mathbb{E}\left[\widetilde{\mu}_\ell^2(x)\right] - \mathbb{E}^2[\widetilde{\mu}_\ell(x)]\right\} \\ &= h_n^{-d}\mathbb{E}\left[\frac{\delta_1 Y_1^{-2\ell}}{\overline{G}^2(Y_1)} K_d^2\left(\frac{x - X_1}{h_n}\right)\right] - h_n^{-d}\mathbb{E}^2\left[\widetilde{T}_1^{-\ell} K_d\left(\frac{x - X_1}{h_n}\right)\right] \\ &= \mathcal{V}_1 - \mathcal{V}_2.\end{aligned}$$

D'une part, pour $\mathcal{V}_2$ en procèdent comme dans le lemme 3.5.3 et sous les hypothèses

H1, K1, K2 et D1, nous avons :

$$\begin{aligned}
\sqrt{\mathcal{V}_2} &= h_n^{-d/2}\mathbb{E}\left[K_d\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1^{-\ell}|X_1\right]\right] \\
&= h_n^{-d/2}\int K_d\left(\frac{x-u}{h_n}\right)\mu_\ell(u)du \\
&= h_n^{d/2}\int K_d(t)\mu_\ell(x-h_nt)dt \\
&= h_n^{d/2}\int K_d(t)\left\{\mu_\ell(x)-h_n\sum_{1\le i\le d}t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}\right\}dt \\
&= h_n^{d/2}\mu_\ell(x)\int K_d(t)dt - h_n^{d/2+1}\int\sum_{1\le i\le d}t_i\frac{\partial\mu_\ell(\xi)}{\partial x_i}K_d(t)dt \\
&= \mathrm{O}\left(h_n^{d/2}\right). \qquad (3.34)
\end{aligned}$$

D'autre part, pour $\mathcal{V}_1$, en utilisant la propriété de l'espérance conditionnelle, pour $\gamma = 2\ell = 2,4$, alors par un changement de variable et un développement de Taylor et sous les hypothèses K3 et D2, nous avons :

$$\begin{aligned}
\mathcal{V}_1 &= h_n^{-d}\mathbb{E}\left[K_d^2\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[\frac{T_1^{-2\ell}}{\overline{G}(T_1)}|X_1\right]\right] \\
&= h_n^{-d}\int K_d^2\left(\frac{x-u}{h_n}\right)\int\frac{t^{-2\ell}}{\overline{G}(t)}f(t|u)dtf(u)du \\
&= h_n^{-d}\int K_d^2\left(\frac{x-u}{h_n}\right)\int\frac{t^{-\lambda}}{\overline{G}(t)}f(u,t)dtdu \\
&= h_n^{-d}\int K_d^2\left(\frac{x-u}{h_n}\right)r_\lambda(u)du \\
&= h_n^{-d}\int K_d^2(t)r_\lambda(x-ht)\,h_n^d\,dt \\
&= r_\lambda(x)\int K_d^2(t)dt + o(h_n) \\
&= r_\lambda(x)\kappa + o(h_n). \qquad (3.35)
\end{aligned}$$

En combinant (3.34) et (3.35) nous obtenons pour $\ell = 1,2$ et $\lambda = 2,4$:

$$V(\Gamma_\ell(x)) \longrightarrow r_\lambda(x)\kappa \qquad \text{quand} \quad n\to\infty.$$

D'autre part, sous les hypothèses D2 et K3, nous obtenons facilement :

$$\begin{aligned}
Cov(\Gamma_1(x),\Gamma_2(x)) &= \mathbb{E}[\Gamma_1(x)\Gamma_2(x)] - \mathbb{E}[\Gamma_1(x)]\mathbb{E}[\Gamma_2(x)] \\
&= h^{-d}\left\{\mathbb{E}\left[\frac{\delta_1 Y_1^{-3}}{\overline{G}^2(Y_1)}K_d^2\left(\frac{x-X_1}{h_n}\right)\right] - \mathbb{E}\left[\widetilde{T}_1^{-1}K_d\left(\frac{x-X_1}{h_n}\right)\right]\mathbb{E}\left[\widetilde{T}_1^{-2}K_d\left(\frac{x-X_1}{h_n}\right)\right]\right\} \\
&= r_3(x)\kappa + o(1).
\end{aligned}$$

Ensuite, nous allons montrons que toute combinaison linéaire est asymptotiquement

gaussienne. Pour tout nombres réels $(c_1, c_2)^t$ nous posons :

$$\Delta(x) = \sum_{\ell=1}^{2} c_\ell \Gamma_\ell(x) =: \sum_{i=1}^{n} \left(c_1 \Delta_i^1(x) + c_2 \Delta_i^2(x) \right) \tag{3.36}$$

où pour $\ell = 1, 2$

$$\Delta_i^\ell(x) := (nh_n^d)^{-1/2} \left\{ \widetilde{T}_i^{-\ell} K_d \left(\frac{x - X_i}{h_n} \right) - \mathbb{E}\left[\widetilde{T}_i^{-\ell} K_d \left(\frac{x - X_i}{h_n} \right) \right] \right\}.$$

Maintenant, pour montrer que (3.36) est asymptotiquement normale nous vérifions la condition de Berry-Essèen (Chow and Teicher (1997), p. 322). Pour cela, nous devons prouver :

$$\rho_n^3 := \sum_{i=1}^{n} \mathbb{E}\left[\left| \Delta_i^\ell(x) \right|^3 \right] \longrightarrow 0 \tag{3.37}$$

avec

$$\mathbb{E}\left[\left| \Delta_i^\ell(x) \right|^3 \right] = \left(\frac{h_n^d}{n} \right)^{3/2} \mathbb{E}\left[\left| \frac{\widetilde{T}_i^{-\ell}}{h_n^d} K_d \left(\frac{x - X_i}{h_n} \right) - \mathbb{E}\left[\frac{\widetilde{T}_i^{-\ell}}{h_n^d} K_d \left(\frac{x - X_i}{h_n} \right) \right] \right|^3 \right].$$

En appliquant l'inégalité C_r (voir Loève (1963), p. 155), nous obtenons :

$$\begin{aligned} \mathbb{E}\left[\left| \Delta_i^\ell(x) \right|^3 \right] &\leq 4 \left(\frac{h_n^d}{n} \right)^{3/2} \left\{ \mathbb{E}\left[\frac{\delta_i |Y_i|^{-3\ell}}{h_n^{3d} \overline{G}^3(Y_i)} K_d^3 \left(\frac{x - X_i}{h_n} \right) \right] + \left| \mathbb{E}\left[\frac{\widetilde{T}_i^{-\ell}}{h_n^d} K_d \left(\frac{x - X_i}{h_n} \right) \right] \right|^3 \right\} \\ &\leq 4 \left(\frac{h_n^d}{n} \right)^{3/2} \{ \mathcal{M}_1 + \mathcal{M}_2 \}. \end{aligned}$$

et comme $\mathcal{M}_1$ et $\mathcal{M}_2$ sont bornées sous l'hypothèse K1 ce qui nous donne $\rho_n^3 = O\left(\left(\frac{h_n^{3d}}{n} \right)^{1/2} \right) = o(1)$. Maintenant, sous l'hypothèse H1 la propriété (3.37) est satisfaite ce qui prouve que la normalité asymptotique de $\Gamma_\ell(x)$. Alors, de (3.32) et (3.33) nous pouvons conclure la preuve du lemme 3.5.5.

Maintenant pour compléter la preuve du théorème 3.3.2, nous considérons θ de $\mathbb{R} \times \mathbb{R}_+^*$ vers $\mathbb{R}$ définie par $\theta(x, y) = x/y$. Nous déduisons du théorème de Mann-Wald (voir Rao (1965), p. 321) que

$$\sqrt{nh_n^d}(\widehat{m}_{RER}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \kappa \nabla\theta^T \Sigma(x) \nabla\theta\right)$$

où le gradient $\nabla\theta^T = \left(\frac{\partial\theta}{\partial x}, \frac{\partial\theta}{\partial y} \right)$ est évalué au point $(\mu_1(x), \mu_2(x))$. Un calcul algébrique simple donne alors le terme de variance suivante :

$$\sigma^2(x) = \kappa \frac{r_2(x)\mu_2^2(x) - 2r_3(x)\mu_1(x)\mu_2(x) + r_4(x)\mu_1^2(x)}{\mu_2^4(x)},$$

complète la preuve du théorème 3.3.2.

Chapitre 4

Convergence uniforme de l'estimateur de la fonction de régression relative pour des données censurées dépendantes

[1] Dans ce chapitre, nous considérons un vecteur (X, T) où X est une co-variable d-dimensionnelle et T est une variable d'intérêt sujet à une censure aléatoire à droite. Le vecteur (X, T) satisfait la propriété d'α-mélange aussi appelée mélange fort. Le but de ce chapitre est d'étudier le comportement de l'estimateur à noyau de la fonction de régression relative pour des des observations mélangeantes. Notons que cette étude est très utiles dans la prévision des série temporelle. Nous avons établi la convergence uniforme presque sûre sur un compact avec vitesse de l'estimateur étudié dans le chapitre 3 où nous avons mis en évidence le terme de covariance. Des simulations montrent que l'estimateur proposé est performant pour une taille d'échantillon fini et dans différents cas. Une prévision sur données générées a été réalisée pour montrer la qualité du modèle.

4.1 Introduction

De nombreux résultats de statistique ont été établi en considérant des échantillons indépendants. Cependant, il est parfois intéressant de considérer et d'étudier des échantillons dépendants afin de pouvoir répondre à des situations où les données ne sont pas indépendantes. C'est par exemple le cas lorsqu'on s'intéresse à l'étude des séries temporelles. Il existe plusieurs types de modélisation de la dépendance au sein d'un échantillon. Nous nous intéressons dans ce chapitre au phénomène de dépendance fort ou α-mélange introduit par Rosenblatt (1956a). Pour plus de détails concernant les différentes formes de dépendance, nous référons à Doukhan (1994). Il apparaît notamment que la notion de variable α mélangeantes est la plus générale, d'où notre choix.

Plusieurs processus satisfont la propriété d'α-mélange. Nous citerons les processus ARMA qui sont fortement mélangeant c-à-d qu'il existe $\rho \in (0,1)$ et un $a > 0$ tel que, pour tout $n \geq 1$, $\alpha(n) \leq a\rho^n$ (voir par exemple, Jones (1978)). les modèles EXPAR (voir Ozaki

1. En collaboration avec le Pr. E. OULD SAÏD et le Pr. R. M. REMITA. Ce chapitre a fait l'objet d'un article soumis pour publication.

(1979), les modèles ARCH simples (voir Engle (1982), leur extension GARCH (voir Bollerslev (1986)) et les modèles bilinéaires markoviens sont des modèles fortement mélangeant sous certaines conditions générales d'ergodicité.
Soit $(X_i)_{1\leq i\leq n}$ une suite de vecteurs aléatoires de la même loi que $X \in \mathbb{R}^d$ où nous désignons par $X_1,\ldots,X_d$ les composantes de X. Nous supposons que les suites $\{T_i, i \geq 1\}$ et $\{C_i, i \geq 1\}$ sont α-mélangeantes avec des coefficients $\alpha_1(n)$ et $\alpha_2(n)$, respectivement. Dans son lemme 2, Cai (2001) a montré que $\{Y_i, i \geq 1\}$ est alors fortement mélangeant avec un coefficient :

$$\alpha(n) = 4 \max(\alpha_1(n), \alpha_2(n))$$

ou la preuve est principalement basé sur le résultat de Dhompongsa (1984). Désormais, nous supposons que $\{(Y_i, \delta_i, X_i),\ \ i = 1,...,n\}$ est fortement mélangeant avec un coefficient $\alpha(n)$ tel que $\alpha(n) = O(n^{-\nu})$ pour un $\nu > 3$.

Dans un contexte de censure aléatoire à droite (voir le modèle sous-section 1.2.4 dans le chapitre 2), certains auteurs abordent l'estimation la fonction de régression pour des données dépendantes. Nous pouvons citer Cai (1998) qui a étudié les propriétés asymptotiques de l'estimateur de Kaplan-Meier pour des données dépendantes censurées et Cai (2001) qui a abordé l'estimation de la f.d.r pour des séries temporelles. Les auteurs El Ghouch and Van Keilegom (2008, 2009) ont estimé les fonctions de régression classique et régression quantile respectivement en appliquant la méthode linéaire locale tandis que dans Guessoum and Ould Saïd (2010, 2012) ont établi la convergence uniforme presque sûre sur un compact avec vitesse (où ils ont mit en évidence le terme de covariance) et la normalité asymptotique de l'estimateur à noyau de la fonction de régression pour des données α-mélangeante. Récemment, Khardani and Slaoui (2019) ont établi la convergence uniforme ainsi que la normalité asymptotique de l'estimateur de la fonction de régression relative dans le cas i.i.d. Toutefois, aucun point de vue pratique n'a été avancé pour démontrer la faisabilité et la qualité de l'estimateur en présence de valeurs aberrantes. Dans ce chapitre, nous établissons un résultat de convergence uniforme presque sûre avec vitesse sur un ensemble compact dans le cas de données fortement mélangeantes. Comme la dépendance est mesurée par la covariance, nous avons mis en évidence le terme de covariance dans la vitesse de convergence de notre estimateur. Nous montrons par une étude de simulation, que l'estimateur est robuste en présence de valeurs aberrantes et est significativement meilleur que l'estimateur de la fonction de régression à noyau classique solution du problème de minimisation de l'erreur quadratique moyenne.

4.2 Hypothèses et résultat principal

Afin de présenter notre résultat de convergence uniforme presque sûre pour des données censurées α-mélangeantes, nous considérons les notations et hypothèses H1, K1–K3, D1–D3 et M1 énoncées dans le chapitres 3 et nous introduisons les hypothèses suivantes :

H4. La fenêtre h_n satisfait $\exists\, \psi > 0, \exists\, c > 0$, tel que $cn^{\frac{\gamma(3-\nu)}{\gamma(\nu+1)+2\gamma+1}+\psi d} \leq h_n^d$, pour tout $\nu > 3$ et $\gamma > 0$.

H5. $\lim\limits_{n\to\infty} h_n^{d(\nu-2)} \log n = 0$.

K4. $\forall\, (t_1,t_2) \in \mathcal{C}^2, |K_d(t_1) - K_d(t_2)| \leq \|t_1 - t_2\|^\gamma$, pour $\gamma > 0$,

D4. La densité jointe $f_{i,j}(\cdot,\cdot)$ de (X_i, X_j) existe et satisfait :

$$\sup_{\mathbb{R}^d\times\mathbb{R}^d} |f_{i,j}(\cdot,\cdot) - f_i(\cdot)f_j(\cdot)| \leq C < \infty, \qquad \text{for any } i,j \geq 1.$$

Commentaires :

1. Les hypothèses H4 et H5 concernent la fenêtre. La première est une hypothèse technique nécessaire pour obtenir notre résultat et la seconde est la vitesse de convergence du terme de covariance.
2. L'hypothèse K4 est technique et désigne que la fonction noyau est holderienne.
3. L'hypothèse D4 est une hypothèse technique qui intervient dans le calcul du terme de covariance.

Théorème 4.2.1 *Sous les hypothèses H1, H4, H5, K1–K4, D1–D4 et M1, nous avons :*

$$\sup_{x\in\mathcal{C}} |\widehat{m}_{RER}(x) - m(x)| = O(h_n) + O_{p.s.}\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right), \quad \textit{quand} \quad n \longrightarrow \infty.$$

4.3 Étude numérique

L'objectif de cette partie est d'examiner la performance de notre estimateur $\widehat{m}_{RER}(x)$ en considérant certains cas particuliers de taille fixe. Nous varions le taux de dépendance et le pourcentage de censure (C.P.). Nous comparons l'efficacité de l'approche RER à celle de l'approche CR définie dans Guessoum and Ould Saïd (2010).

Algorithme

Require: $0 < \rho < 1$, $X_0 \rightsquigarrow \mathcal{N}(0,1)$ et $\varepsilon \rightsquigarrow \mathcal{N}(0,1)$.

Etape 1. Nous considérons un processus auto-régressif bi-dimensionnel AR(1) généré par

$$\begin{cases} X_i = 3 + \rho X_{i-1} + \sqrt{1-\rho^2}\varepsilon_i, \\ T_i = X_{i+1}, \qquad i = 1,\ldots,n. \end{cases}$$

Etape 2. Étant donné $X_1 = x$, nous avons $T_1 = 3 + \rho x + \sqrt{1-\rho^2}\varepsilon_2$. Il est clair que la distribution conditionnelle de T_1 sachant x est une v.a. gaussienne de loi $\mathcal{N}(3+\rho x, 1-\rho^2)$. En ce qui concerne la variable de censure, nous avons généré $C_i = X_{i+1} + \lambda$ où λ est un paramètre qui permet d'adapter le pourcentage de censure (C.P.).

Etape 3. Pour chaque individus i nous calculons la variable observée $Y_i = T_i \wedge C_i$ et son indicateur de censure $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$ correspondant. Ainsi, l'échantillon observé est $\{(X_i, Y_i, \delta_i), 1 \leq i \leq n\}$.

Etape 3'. On multiplie chaque valeur de Y (notre variable observée) au rang $\{20 \times ii$ où $1 \leq ii \leq \frac{n}{20}\}$ par un facteur multiplicateur (M.F.) qu'on fait varier pour créer des valeurs aberrantes dans l'échantillon de taille n.

Etape 4. L'estimateur de K-M de $\overline{G}(\cdot)$ est calculé par son expression (1.5).

Etape 5. Nous avons choisi d'utiliser un noyau gaussien, pour le choix de la fenêtre optimale, on utilise la méthode de validation croisée (voir la remarque 1.8) de $[0.01, 2]$ avec un pas de 0.01.

Output : Calculer l'estimateur RER donné par (3.14) pour un intervalle $x \in [1, 4]$ et la fenêtre optimale h_{opt}.

Cas linéaire :

Dans cette sous section, nous observons les performances de l'estimateur RER pour une faible et forte dépendance quand la fonction théorique est du type linéaire.

1. Faible dépendance

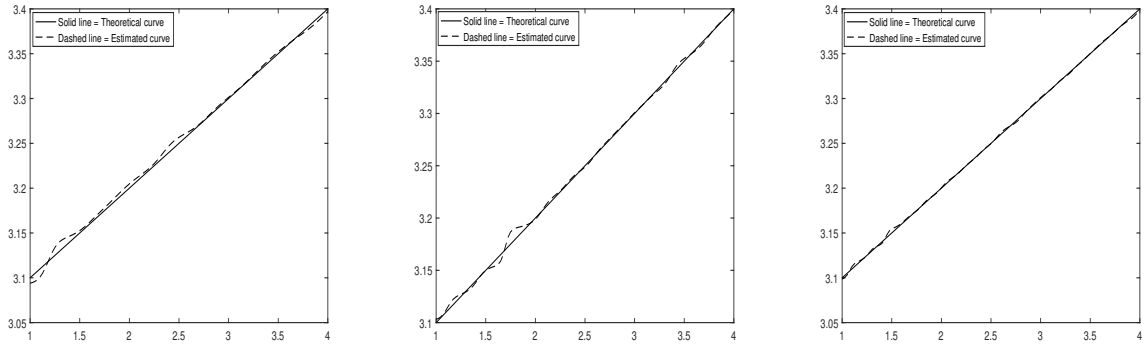


FIGURE 4.1 – $\rho = 0.1$ et C.P. $\approx 35\%$ pour $n = 100, 300$ et 500 respectivement.

1.a Effet de la taille d'échantillon : Il est facile à voir de la figure 4.1 que la qualité d'ajustement est meilleur quand la taille d'échantillon n augmente pour un taux de censure C.P. et un taux de dépendance ρ fixe.

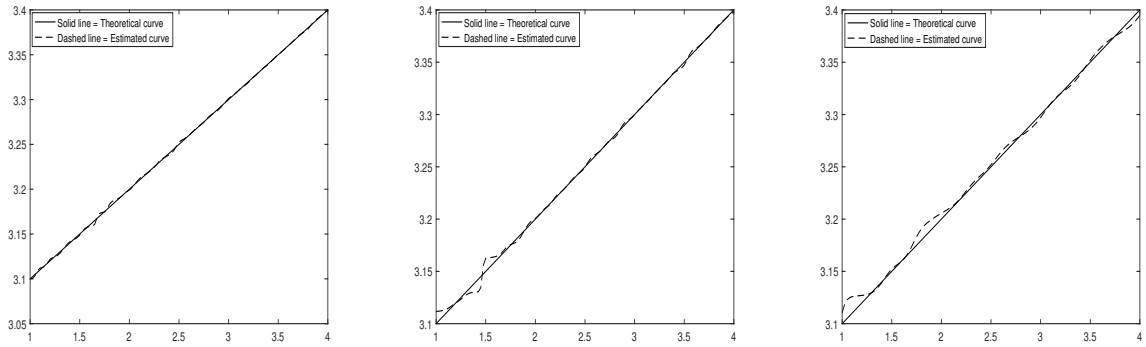


FIGURE 4.2 – $n = 300$ et $\rho = 0.1$ pour C.P. $\approx 7, 40$ et 67% respectivement.

1.b. Effet du taux de censure : On peut dire que quand le pourcentage de censure augmente, il entraîne une plus grande variation de l'estimateur résultant, mais reste

généralement proche de la courbe théorique même pour un C.P. élevé (la figure 4.2 affiche les résultats). Ainsi, la qualité de l'ajustement est affectée par la censure qui est prévisible.

1.c. Effet des valeurs aberrantes : Pour montrer la robustesse de notre approche, nous avons générer la variable observée selon l'étape 3'. De la figure 4.3, nous pouvons constater que notre estimateur est très proche de la courbe théorique sachant que l'on observe que 65% des vraies valeurs. Donc, nous pouvons conclure que notre estimateur est résistant aux valeurs aberrantes.

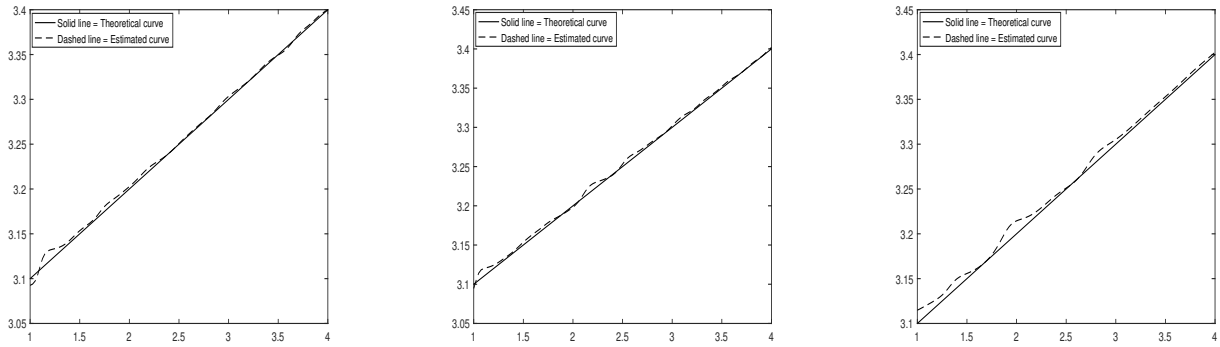


FIGURE 4.3 – $n = 300$, $\rho = 0.1$ et C.P. $\approx 35\%$ pour M.F.= 50, 100 et 150 respectivement.

2. Forte dépendance

2.a. Effet de la taille d'échantillon : Dans le cas où les données sont fortement dépendantes ($\rho = 0.9$) et pour un taux de censure fixé, nous pouvons observer au travers la figure 4.4 que l'ajustement de l'estimateur RER à la droite théorique s'améliore quand la taille d'échantillon augmente.

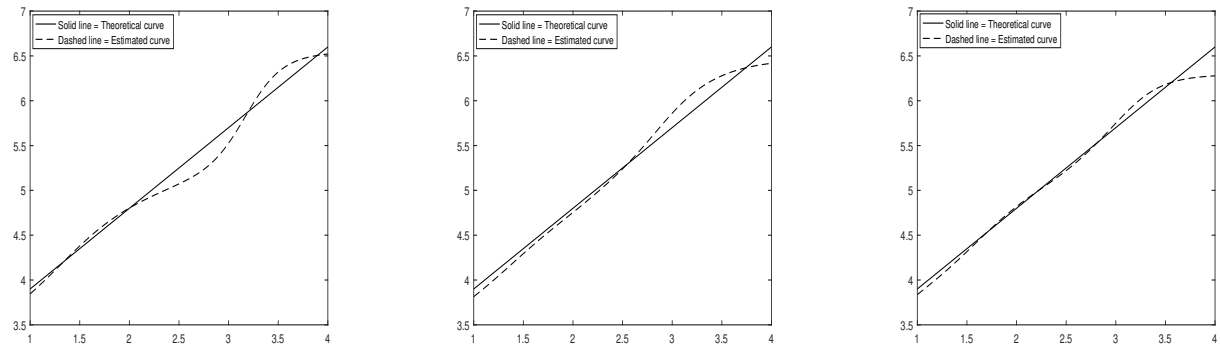


FIGURE 4.4 – $\rho = 0.9$ et C.P.$\approx 15\%$ pour $n = 100, 300$ et 500 respectivement.

2.b. Effet du C.P. : Nous pouvons voir que la qualité d'ajustement est meilleure pour une taille d'échantillon grande et un faible taux de censure (voir figure 4.5).

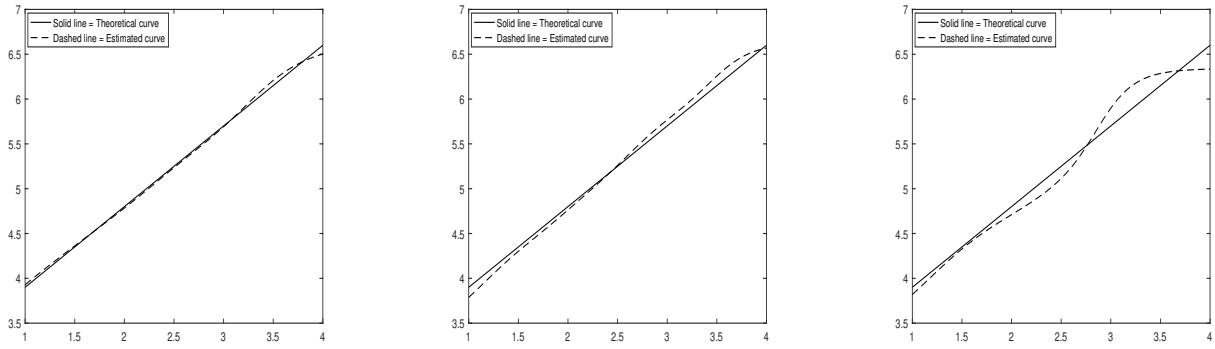


FIGURE 4.5 – $n = 300$ et $\rho = 0.9$ pour C.P.$\approx 3,15$ et 56% respectivement.

1.c. Effet des valeurs aberrantes : Il est clair de la figure 4.6 que notre estimateur résiste en présence de valeurs aberrantes même si le M.F. est très grand.

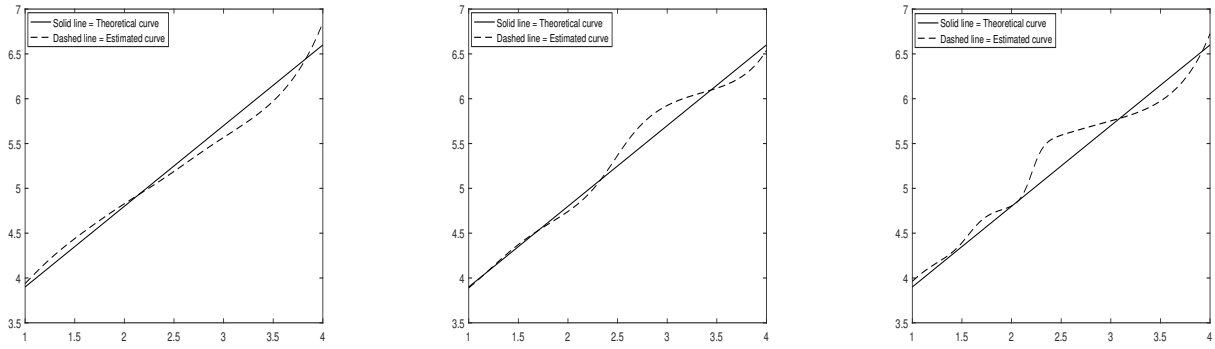


FIGURE 4.6 – $n = 300$, $\rho = 0.9$ et C.P.$\approx 20\%$ pour M.F.$= 50, 100$, et 150 respectivement.

4.3.1 Cas non-linéaire

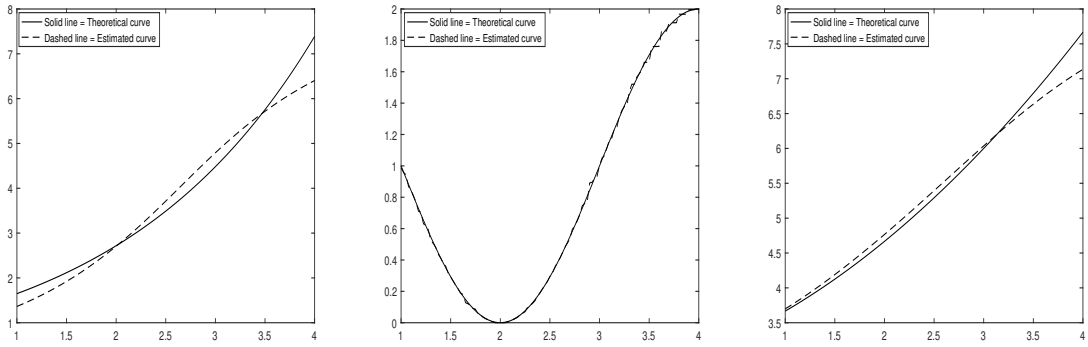
Nous considérons maintenant trois fonctions non-linéaires :

$$\begin{aligned} T_i &= 1 + \cos\left(\frac{\pi}{2} X_i\right), \quad \text{Modèle cosinus,} \\ T_i &= \exp(X_i), \quad \text{Modèle exponentiel,} \\ T_i &= \frac{2}{3} X_i^2 + X_i + 3, \quad \text{Modèle quadratique.} \end{aligned}$$

La figure 4.7 montre que la qualité d'ajustement est aussi bonne que le cas linéaire. La qualité d'estimation est meilleure quand la taille d'échantillon augmente.

4.3.2 Étude comparative

Pour montrer l'efficacité de l'estimateur RER, nous établissons une étude comparative dans laquelle nous considérons l'estimateur de la régression classique de Guessoum and Ould Saïd (2010) défini dans (1.15) pour une faible et forte dépendance.



(a) Exponentielle (b) Cosinus (c) Quadratique

FIGURE 4.7 – $n = 300$, $\rho = 0.5$ et C.P.$\approx 15\%$.

1. faible dépendance

1.a. Effet du C.P. : Nous fixons la taille d'échantillon et nous varions le taux de censure. Nous pouvons constater clairement de la figure 4.8 que l'estimateur RER est proche de la courbe théorique contrairement à celui du CR qui est distant de la courbe quand le C.P. augmente.

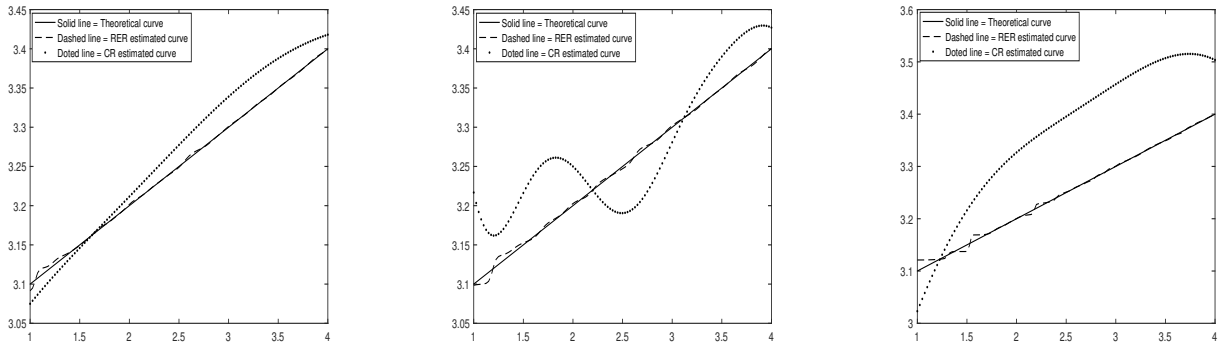


FIGURE 4.8 – $\rho = 0.1$ et $n = 300$ pour C.P.$\approx 10, 33$ et 54% respectivement.

1.b. Effet des valeurs aberrantes : Nous fixons la taille d'échantillon, le pourcentage de censure et nous varions le M.F. Il ressort de la figure 4.9 que l'estimateur RER se superpose à la courbe réelle, contrairement à l'estimateur CR qui est significativement affecté par la M.F. lorsque la dépendance est faible.

2. Forte dépendance

2.a Effet du C.P. : Nous fixons ρ, n et nous faisons varier le C.P. pour examiner l'effet de la censure sur les estimateurs RER et CR lorsque la dépendance est forte. Nous pouvons observer sur la figure 4.10 que l'estimateur du RER reste proche de la courbe théorique par rapport à l'estimateur du CR qui est le but de notre étude.

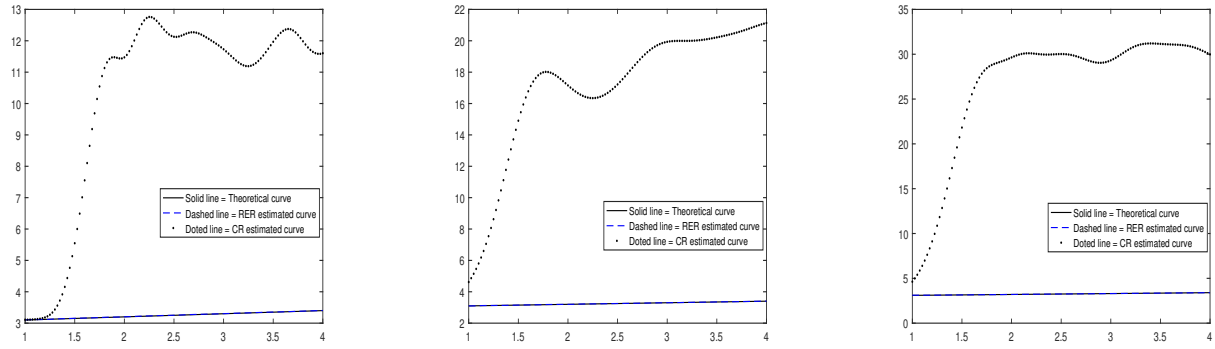


FIGURE 4.9 – $\rho = 0.1$ et $n = 300$ pour C.P.$\approx 5\%$ et M.F.$= 50, 100$ et 150 respectivement.

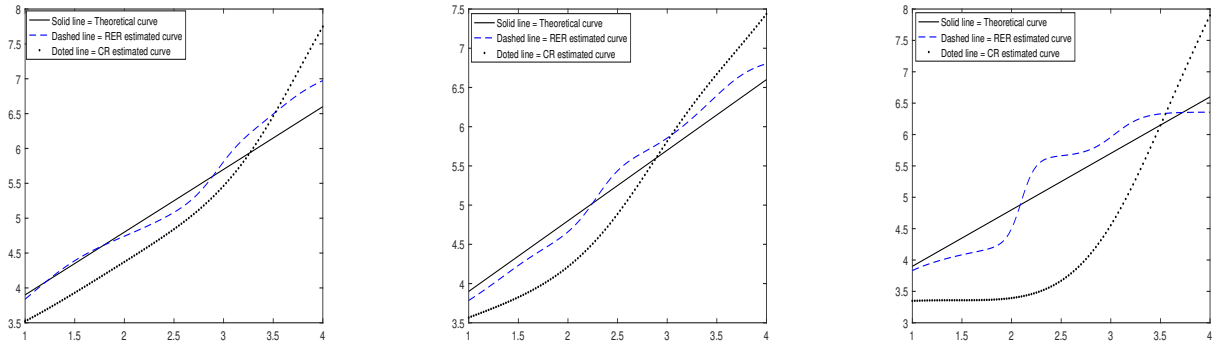


FIGURE 4.10 – $\rho = 0.9$ et $n = 300$ pour C.P.$\approx 7,32$ et 65% respectivement.

2.b. Effet des valeurs aberrantes : Nous fixons ρ, n, C.P. et nous faisons varier le M.F. (voir : **étape 3'**) pour évaluer l'effet des valeurs aberrantes sur les deux estimateurs (CR et RER) lorsque la dépendance est élevée. Comme prévu, notre estimateur reste résistant aux valeurs aberrantes sous une forte dépendance, contrairement à celui de CR qui est plus éloigné lorsque la M.F. devient importante.

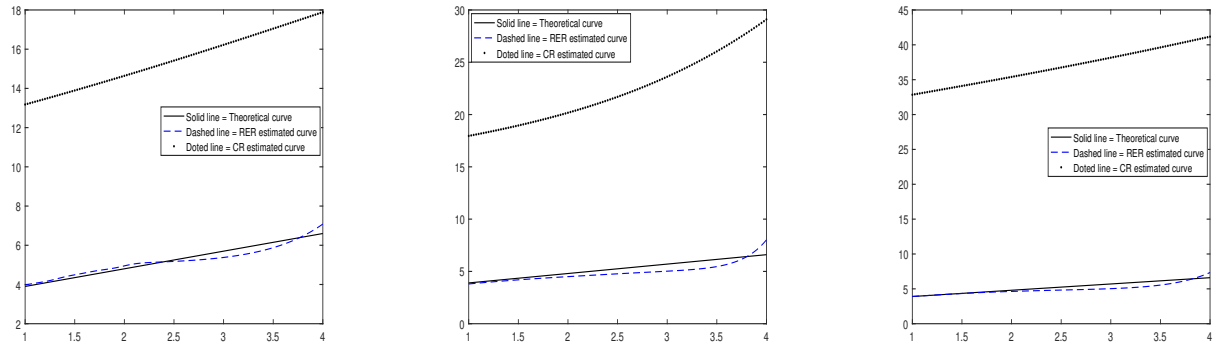


FIGURE 4.11 – $\rho = 0.9$ et $n = 300$ pour C.P.$\approx 5\%$ et M.F.$= 50, 100$ et 150 respectivement.

3. Tableau comparatif des écarts quadratiques moyens : Les performances numériques des différents modèles considérés, le tableau 4.1 indique l'erreur quadratique moyenne (EQM) correspondante aux estimateurs CR et RER. On voit clairement que

l'EQM du RER diminue fortement par rapport à celle du CR à mesure que la taille augmente.

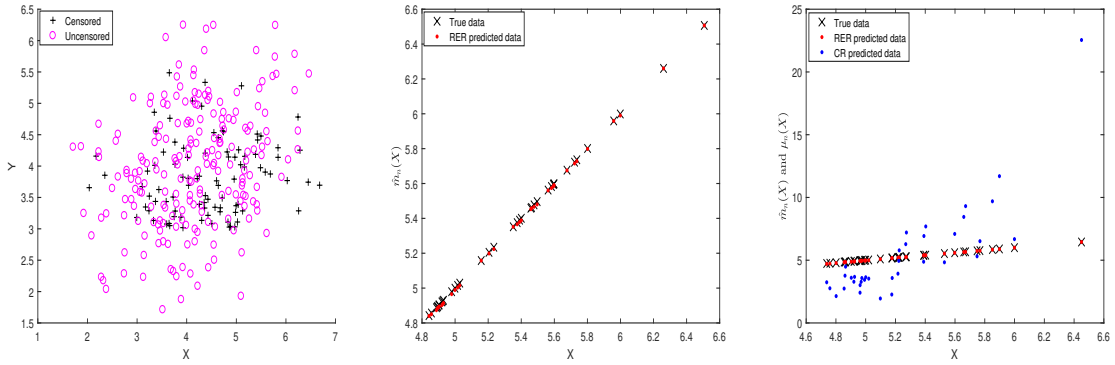
Tableau 4.1 – Tableau comparatif des EQM pour un C.P. ≈ 12%.

n	ρ	M.F.	CR	RER
		50	3.7481×10^2	0.0265
	0.1	100	1.7085×10^3	0.0390
		150	8.5971×10^3	0.0434
		50	3.5218×10^2	0.0394
100	0.5	100	9.1558×10^2	0.0077
		150	6.1379×10^3	0.0025
		50	3.0441×10^2	0.5410
	0.9	100	1.4437×10^3	0.0307
		150	8.5500×10^3	0.8906
		50	2.1442×10^2	0.0564
	0.1	100	7.9874×10^2	0.0366
		150	3.9463×10^3	0.0451
		50	98.7668	0.0109
300	0.5	100	5.4635×10^2	0.0148
		150	1.1780×10^3	0.0065
		50	1.1642×10^2	0.0328
	0.9	100	6.4634×10^2	0.6603
		150	1.8572×10^3	0.1870
		50	42.2378	0.1219
	0.1	100	5.0131×10^2	0.3718
		150	1.7212×10^3	0.0378
		50	6.2931	0.0021
500	0.5	100	4.2397×10^2	0.0017
		150	7.8780×10^2	0.0060
		50	89.3259	0.0293
	0.9	100	5.0478×10^2	0.3342
		150	1.0438×10^3	0.0202

4.3.3 Prévision expérimentale

Dans cette partie, nous évaluons la performance des prédicteurs RER et CR pour le même jeu de données générées. Pour cela, nous considérons un échantillon de taille $N = 300$ du processus AR(1) (voir : Étape 1). Dans la figure 4.12 (a), nous montrons un nuage de points où nous distinguons les données non censurées (o) et censurées (+). Le taux de censure global est d'environ 35%. Pour notre étude, nous avons divisé au hasard notre échantillon de N en deux sous-ensembles. L'échantillon d'apprentissage, de taille $n = 250$, sera notre échantillon statistique c'est-à-dire (X_i, Y_i, δ_i) avec $i = 1, \cdots, n$ pour lequel les estimateurs sont calculés. Ensuite, pour chaque nouvelle co-variable X_i, $i = n+1, \cdots, N$, nous calculons les valeurs prédites, par estimateur RER, correspondant à $\widehat{Y}_i, i = n+1, \cdots, N$ considéré comme l'échantillon test. Notez que, nous éliminons

les données censurées du point prédit (c'est-à-dire que pour tout $n+1 \leq i \leq N$, si la variable observée $Y_i = C_i$, nous retirons l'observation Y_i des valeurs prédites. Il est inapproprié de prédire une valeur censurée, ce qui est évident). Dans la figure 4.12 (b) nous traçons les valeurs prévues par rapport aux valeurs réelles.
Pour le dernier cas et pour chaque estimateur, nous utilisons la largeur de bande optimale h_n pour $\widehat{m}_{RER}(\cdot)$ et $\widehat{m}_{CR}(\cdot)$ en minimisant le critère de validation croisée (voir : la remarque 1.8). Par conséquent, par souci de concision, nous nous limitons à montrer les résultats pour $\rho = 0,3$ et C.P.$\approx 35\%$.
En conclusion, la méthode RER semble améliorer la qualité des prévisions par rapport à la méthode CR. En outre, un fait intéressant peut être observé dans la figure 4.12 (c) qui est la plus grande différence entre les prédictions obtenues par les deux méthodes. Comme prévu, la performance de l'échantillon fini se détériore lorsque la dépendance augmente. Globalement, de meilleurs résultats sont obtenus lorsque la dépendance et le niveau de censure sont faibles. Enfin, nous pouvons conclure que l'estimateur du RER est plus efficace et plus précis.



(a) (b) (c)

FIGURE 4.12 – $n = 300$, $\rho = 0.3$ et C.P.$\approx 30\%$. (a) Nuage de points des données censurées et non censurées. (b) Performance de l'estimateur RER en prévision. (c) Comparaison entre le RER et CR en prévision.

4.4 Preuves et résultats auxiliaires

Preuve du théorème 4.2.1. La démonstration est essentiellement basée sur les mêmes arguments analytiques utilisés dans la section 3.5 sous la même décomposition (3.24) du chapitre 4, on peut dire que la propriété de l'indépendance des observations n'a aucune influence sur les termes : du biais $(\mathbb{E}[\widehat{\mu}_\ell(\cdot)] - \mu_\ell(\cdot))$ et de l'erreur $(\widehat{\mu}_\ell(\cdot) - \widetilde{\mu}_\ell(\cdot))$ pour un $\ell = 1, 2$. Autrement dit, la vitesse de convergence des deux termes sera la même dans le cas de mélange. Cependant, le terme de variance ou de dispersion est basé sur le lemme suivant :

Lemme 4.4.1 *Sous les hypothèses H1, H4, H5, K1–K4 et D1, pour $\ell = 1, 2$, nous avons :*

$$\sup_{x \in \mathcal{C}} |\widetilde{\mu}_\ell(x) - \mathbb{E}[\widetilde{\mu}_\ell(x)]| = O_{p.s.}\left(\sqrt{\frac{\log n}{n h_n^d}} + \sqrt{h_n^{d(\nu-2)} \log n} \right), \quad \textit{quand} \quad n \longrightarrow \infty.$$

Preuve du Lemme 4.4.1. $\mathcal{C}$ est un ensemble compact, alors il admet une couverture $\mathcal{S}$ par un nombre fini s_n de boules $\mathrm{B}_k(x_k^*, a_n^d)$ de centre $x_k^* = (x_{1,k}^*, \ldots, x_{d,k}^*), 1 \leq k \leq s_n$. Ainsi, pour tout $x \in \mathcal{C}$ il existe k tel que $\|x - x_k^*\| \leq a_n^d$ où a_n vérifie $a_n^{d\gamma} = h_n^{d\left(\gamma+\frac{1}{2}\right)} n^{-\frac{1}{2}}$ avec γ est la constante de la condition de Lipshitz définie dans l'hypothèse K5. Puisque $\mathcal{C}$ est borné alors il existe une constante $M > 0$ tel que $s_n \leq M a_n^{-d}$.
Soit pour tout $x \in \mathcal{C}$ et $\ell = 1, 2$, l'ensemble donné :

$$\mathcal{A}_{\ell,i}(x) = (nh_n^d)^{-1}\left(\widetilde{T}_i^{-\ell} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left[\widetilde{T}_i^{-\ell} K_d\left(\frac{x - X_i}{h_n}\right)\right]\right),$$

alors

$$\sum_{1\leq i\leq n} \mathcal{A}_{\ell,i}(x) = \widetilde{\mu}_\ell(x) - \mathbb{E}[\widetilde{\mu}_\ell(x)],$$

qui se décompose comme suit :

$$\begin{aligned}\sum_{1\leq i\leq n} \mathcal{A}_{\ell,i}(x) &= \left\{\left(\widetilde{\mu}_\ell(x) - \widetilde{\mu}_\ell(x_k^*)\right) - \left(\mathbb{E}\left[\widetilde{\mu}_\ell(x)\right] - \mathbb{E}\left[\widetilde{\mu}_\ell(x_k^*)\right]\right)\right\} + \left(\widetilde{\mu}_\ell(x_k^*) - \mathbb{E}\left[\widetilde{\mu}_\ell(x_k^*)\right]\right) \\ &=: \sum_{1\leq i\leq n} \widetilde{\mathcal{A}}_{\ell,i}(x) + \sum_{1\leq i\leq n} \mathcal{A}_{\ell,i}(x_k^*),\end{aligned}$$

d'où

$$\begin{aligned}\sup_{x\in\mathcal{C}}\left|\sum_{1\leq i\leq n} \mathcal{A}_{\ell,i}(x)\right| &\leq \max_{1\leq k\leq s_n}\sup_{x\in B_k}\left|\sum_{1\leq i\leq n} \widetilde{\mathcal{A}}_{\ell,i}(x)\right| + \max_{1\leq k\leq s_n}\left|\sum_{1\leq i\leq n} \mathcal{A}_{\ell,i}(x_k^*)\right| \\ &=: \mathcal{B}_1 + \mathcal{B}_2.\end{aligned}$$

D'une part, nous nous occupons du premier terme $\mathcal{B}_1$. Pour cela, nous avons :

$$\begin{aligned}\left|\sum_{1\leq i\leq n} \widetilde{\mathcal{A}}_{\ell,i}(x)\right| &= \left|\left[\widetilde{\mu}_\ell(x) - \widetilde{\mu}_\ell(x_k^*)\right] - \mathbb{E}\left[\widetilde{\mu}_\ell(x) - \widetilde{\mu}_\ell(x_k^*)\right]\right| \\ &= \left|\frac{1}{nh_n^d}\sum_{1\leq i\leq n} \widetilde{T}_i^{-\ell}\left(K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right)\right. \\ &+ \left.\frac{1}{h_n^d}\mathbb{E}\left[\widetilde{T}_1^{-\ell}\left(K_d\left(\frac{x - X_1}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right)\right]\right| \\ &\leq \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{|T_i|^{-\ell}}{\overline{G}(T_i)}\left|K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right| \\ &+ \frac{1}{h_n^d}\mathbb{E}\left[\frac{|T_1|^{-\ell}}{\overline{G}(T_1)}\left|K_d\left(\frac{x - X_1}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right|\right] \\ &=: \mathcal{D}_{1,\ell}(x) + \mathcal{D}_{2,\ell}(x).\end{aligned}$$

Des hypothèses M1 et K4, nous avons :

$$\begin{aligned}\sup_{x\in B_k}\mathcal{D}_{1,\ell}(x) &\leq \frac{C}{\bar{G}(\tau)}\frac{1}{h_n^d}\sup_{x\in\mathcal{C}}\left|K_d\left(\frac{x-X_i}{h_n}\right)-K_d\left(\frac{x_k^*-X_i}{h_n}\right)\right| \\ &\leq \frac{C}{h_n^d\overline{G}(\tau)}\frac{\|x-x_k^*\|^\gamma}{h_n^\gamma} \\ &\leq C\frac{a_n^{d\gamma}}{h_n^{d+\gamma}}.\end{aligned}$$

De la même manière, sous les hypothèses M1 et K4, nous avons :

$$\sup_{x\in B_k}\mathcal{D}_{2,\ell}(x)\leq C\frac{a_n^{d\gamma}}{h_n^{d+\gamma}},$$

alors

$$\sup_{x\in B_k}\left|\sum_{1\leq i\leq n}\widetilde{\mathcal{A}}_{\ell,i}(x)\right| = \sup_{x\in B_k}\mathcal{D}_{1,\ell}(x)+\sup_{x\in B_k}\mathcal{D}_{2,\ell}(x)\leq\frac{2Ca_n^{d\gamma}}{h_n^{d+\gamma}}\leq\frac{Ch_n^{d(\gamma+\frac{1}{2})}n^{-\frac{1}{2}}}{h_n^{d+\gamma}}=\frac{C}{\sqrt{nh_n^d}}h_n^{\gamma(d-1)},$$

ce qui nous conduit à :

$$\mathcal{B}_1=\max_{1\leq k\leq s_n}\sup_{x\in B_k}\left|\sum_{i=1}^n\widetilde{\mathcal{A}}_{\ell,i}(x)\right|=\mathrm{O}\left(\frac{1}{\sqrt{nh_n^d}}\right) \tag{4.1}$$

D'autre part, nous procédons au calcul du second terme $\mathcal{B}_2$. Soit

$$U_i=U_{i,k}=nh_n^d\mathcal{A}_{\ell,i}(x_k^*)=\widetilde{T}_i^{-\ell}K_d\left(\frac{x_k^*-X_i}{h_n}\right)-\mathbb{E}\left[\widetilde{T}_i^{-\ell}K_d\left(\frac{x_k^*-X_i}{h_n}\right)\right].$$

Pour appliquer le Lemme 5.4.1, nous devons dans un premier temps calculer la quantité :

$$\begin{aligned}S_n^2 &= \sum_i\sum_j\left|Cov(U_i,U_j)\right|=\sum_{i\neq j}\left|Cov(U_i,U_j)\right|+n\mathrm{V}ar(U_1) \\ &=: \mathcal{V}+n\mathrm{V}ar(U_1).\end{aligned} \tag{4.2}$$

D'un coté, nous considérons :

$$\begin{aligned}\mathrm{V}ar(U_1) &= Var\left[\widehat{T}_i^{-\ell}K_d\left(\frac{x_k^*-X_1}{h_n}\right)\right] \\ &= \mathbb{E}\left[\frac{\delta_1Y_1^{-2\ell}}{\bar{G}^2(Y_1)}K_d^2\left(\frac{x_k^*-X_1}{h_n}\right)\right]-\mathbb{E}^2\left[\widehat{T}_1^{-\ell}K_d\left(\frac{x_k^*-X_1}{h_n}\right)\right] \\ &=: \mathcal{R}_1-\mathcal{R}_2.\end{aligned}$$

Pour $\mathcal{R}_1$, en utilisant le propriété de l'espérance conditionnelle et un changement de variable, nous obtenons :

$$\begin{aligned}
\mathcal{R}_1 &= \mathbb{E}\left[\frac{\delta_1 Y_1^{-2\ell}}{\bar{G}^2(Y_1)} K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\right] \\
&= \mathbb{E}\left[K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\mathbb{E}\left[\frac{\delta_1 Y_1^{-2\ell}}{\bar{G}^2(Y_1)}|X_1\right]\right] \\
&= \mathbb{E}\left[K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\mathbb{E}\left[\frac{T_1^{-2\ell}}{\overline{G}^2(T_1)}\mathbb{E}\left[\mathbb{1}_{\{T_1\leq C_1\}}|T_1\right]|X_1\right]\right] \\
&= \mathbb{E}\left[K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\mathbb{E}\left[\frac{T_1^{-2\ell}}{\overline{G}(T_1)}|X_1\right]\right] \\
&= \int K_d^2\left(\frac{x_k^* - u}{h_n}\right)\int \frac{t^{-2\ell}}{\overline{G}(t)} f(t|u)dt f(u)du \\
&= \int K_d^2\left(\frac{x_k^* - u}{h_n}\right)\int \frac{t^{-\lambda}}{\overline{G}(t)} f(x,u)dtdu \\
&= h_n^d \int K_d^2(t) r_\lambda(x_k^* - h_n t)dt,
\end{aligned}$$

par un développement de Taylor de la fonction $r_\lambda(\cdot)$ autour de x_k^* et sous les hypothèses K2, K3 et D2, nous obtenons :

$$\mathcal{R}_1 = \mathrm{O}\left(h_n^d\right). \tag{4.3}$$

Pour $\mathcal{R}_2$, nous avons :

$$\begin{aligned}
\sqrt{\mathcal{R}_2} &= \mathbb{E}\left[K_d\left(\frac{x_k^* - X_1}{h_n}\right)\mathbb{E}\left[\widetilde{T}_1^{-\ell}|X_1\right]\right] \\
&= \int K_d\left(\frac{x_k^* - u}{h_n}\right) m_\ell(u) f(u)du \\
&= \int K_d\left(\frac{x_k^* - u}{h_n}\right)\mu_\ell(u)du \\
&= h_n^d \int K_d(t)\mu_\ell(x_k^* - h_n t)dt,
\end{aligned}$$

par un développement de Taylor cette fois ci pour la fonction $\mu_\ell(\cdot)$ autour de x_k^* et sous les hypothèses D1 et K2, nous avons :

$$\mathcal{R}_2 = \mathrm{O}\left(h_n^{2d}\right). \tag{4.4}$$

Alors de (4.3) et (4.4), nous obtenons :

$$n\mathrm{V}ar(U_1) = n(\mathcal{R}_1 - \mathcal{R}_2) = \mathrm{O}\left(nh_n^{2d}\right) + \mathrm{O}\left(nh_n^d\right) = \mathrm{O}\left(nh_n^d\right). \tag{4.5}$$

D'une part,

$$\begin{aligned}
|Cov(U_i, U_j)| &= |\mathbb{E}[U_i U_j] - \mathbb{E}[U_i]\mathbb{E}[U_j]| \\
&= \left| \mathbb{E}\left[\widetilde{T}_i^{-\ell} \widetilde{T}_j^{-\ell} K_d\left(\frac{x_k^* - X_i}{h_n}\right) K_d\left(\frac{x_k^* - X_j}{h_n}\right)\right] \right. \\
&- \left. \mathbb{E}\left[\widetilde{T}_i^{-\ell} K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right] \mathbb{E}\left[\widetilde{T}_j^{-\ell} K_d\left(\frac{x_k^* - X_j}{h_n}\right)\right] \right| \\
&\leq h_n^{2d} \int\int K_d(t) K_d(s) |f_{i,j}(x_k^* - h_n t, x_k^* - h_n s) \\
&- f_i(x_k^* - h_n t) f_j(x_k^* - h_n s)| dt ds,
\end{aligned}$$

qui sous l'hypothèse D4 conduit à

$$|Cov(U_i, U_j)| = O\left(h_n^{2d}\right), \tag{4.6}$$

uniformément sur i et j.
Maintenant, pour évaluer le comportement asymptotique de $\mathcal{V}$ selon la décomposition de Masry (1986), nous définissons les deux ensembles :

$$E_1 = \{(i, j) \text{ tel que } 0 < |i - j| \leq \beta_n\}$$

et

$$E_2 = \{(i, j) \text{ tel que } \beta_n < |i - j| \leq n\}$$

où $\beta_n \to \infty$ quand $n \to \infty$ a une vitesse lente de l'ordre de $\beta_n = o(n)$. Soit les sommes de covariances $\mathcal{V}_1$ et $\mathcal{V}_2$ sous les ensembles E_1 et E_2 respectivement.

$$\mathcal{V} = \sum_{E_1} |Cov(U_i, U_j| + \sum_{E_2} |Cov(U_i, U_j| =: \mathcal{V}_1 + \mathcal{V}_2.$$

Nous obtenons alors de (4.6)

$$\mathcal{V}_1 = \sum_{E_1} |Cov(U_i, U_j| = \sum_{0<|i-j|\leq\beta_n} h_n^{2d} = O\left(n h_n^{2d} \beta_n\right).$$

Pour $\mathcal{V}_2$, nous utilisons l'inégalité modifiée de Davydov (1970) pour des processus mélangeants (voir aussi le livre Rio (2000) p. 87, formule 6.19b). Ce qui nous mène, pour tout $i \neq j$, à

$$|Cov(U_i, U_j)| \leq C\alpha(|i - j|),$$

nous obtenons alors

$$\begin{aligned}
\mathcal{V}_2 &\leq C \sum \sum_{\beta_n < |i-j| \leq n} \alpha(|i - j|) \\
&= C n^2 \alpha(\beta_n).
\end{aligned}$$

Nous posons $\beta_n = [h_n^{-d}]$ (où $[\cdot]$ désigne la partie entière) ce qui nous permet d'obtenir un terme de covariance non négligeable devant le terme de variance d'ordre :

$$\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2 = O\left(n h_n^d\right) + O\left(n^2 h_n^{dv}\right). \tag{4.7}$$

Finalement, de (4.2), (4.5) et (4.7) nous obtenons

$$S_n^2 = \mathcal{V} + nVar(U_1) = \mathrm{O}\left(nh_n^d\right) + \mathrm{O}\left(nh_n^{dv}\right).$$

Maintenant, nous somme prêt à appliquer l'inégalité du Lemme 5.4.1, où pour $\varepsilon > 0$ nous avons :

$$\begin{aligned}
\mathbb{P}\left[\left|\sum_{1\le i\le n} \mathcal{A}_{\ell,i}(x_k^*)\right| > \varepsilon\right] &= \mathbb{P}\left[\left|\sum_{1\le i\le n} U_i\right| > nh_n^d\varepsilon\right] \\
&\le C\left(1 + \frac{nh_n^d\varepsilon^2}{r}\right)^{-\frac{r}{2}} + nCr^{-1}\left(\frac{r}{nh_n^d\varepsilon}\right)^{v+1} \\
&=: C(\mathcal{E}_1 + \mathcal{E}_2).
\end{aligned}$$

En prenant $\varepsilon = \varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(v-2)}\log n}\right)$ avec $\varepsilon_0 > 0$, nous avons pour la première partie :

$$\mathcal{E}_1 = \left(1 + \frac{\varepsilon_0^2 \log n}{r}\right)^{-\frac{r}{2}}. \tag{4.8}$$

En choisissant $r = (\log n)^{1+b}$ avec $b > 0$, (4.8) devient

$$\mathcal{E}_1 = \left(1 + \varepsilon_0^2(\log n)^{-b}\right)^{-\frac{(\log n)^{1+b}}{2}}$$

En passant au logarithme et en utilisant un développement de Taylor $\log(1+x)$

$$\log \mathcal{E}_1 \simeq \log n^{-\frac{\varepsilon_0^2}{2}}$$

d'où nous déduisons

$$\mathcal{E}_1 = n^{-\frac{\varepsilon_0^2}{2}}. \tag{4.9}$$

Pour un même choix de ε et r, nous avons :

$$\mathcal{E}_2 \simeq n(\log n)^{v(1+b)}\varepsilon_0^{-(v+1)}(nh_n^d \log n)^{-\frac{v+1}{2}}.$$

En prenant une fois de plus l'inégalité de Fuk-Nagaev et en utilisant $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)$, nous pouvons écrire :

$$\begin{aligned}
\mathbb{P}\left[\max_{1\le k\le S_n}\left|\sum_{1\le i\le n} \mathcal{A}_{\ell,i}(x_k^*)\right| > \varepsilon_n\right] &\le Ma_n^{-d}C\left(n^{-\frac{C\varepsilon_0^2}{2}} + n(\log n)^{v(1+b)}\varepsilon_0^{-(v+1)}(nh_n^d\log n)^{-\frac{v+1}{2}}\right) \\
&\le Mh_n^{-d(1+\frac{1}{2\gamma})}n^{\frac{1}{2\gamma}}C\left(n^{-\frac{c\varepsilon_0^2}{2}} + n(\log n)^{v(1+b)}\varepsilon_0^{-(v+1)}(nh_n^d\log n)^{-\frac{v+1}{2}}\right) \\
&\le MCn^{\frac{1}{2\gamma}-\frac{C\varepsilon_0^2}{2}}h_n^{-d(1+\frac{1}{2\gamma})} \\
&+ MC\varepsilon_0^{-(v+1)}n^{1+\frac{1}{2\gamma}}h_n^{-d(1+\frac{1}{2\gamma})}(\log n)^{v(1+b)}(nh_n^d\log n)^{-\frac{v+1}{2}} \\
&=: MC(\mathcal{Z}_1 + \varepsilon_0^{-(v+1)}\mathcal{Z}_2).
\end{aligned} \tag{4.10}$$

Sous l'hypothèse H4, nous avons :

$$\begin{aligned}\mathcal{Z}_2 &\leq Cn^{1+\frac{1}{2\gamma}-\frac{\nu+1}{2}} h_n^{-d(1+\frac{1}{2\gamma}+\frac{\nu+1}{2})} (\log n)^{\nu(1+b)-\frac{\nu+1}{2}} \\ &\leq Cn^{1+\frac{1}{2\gamma}-\frac{\nu+1}{2}} n^{-\frac{(3-\nu)}{2}-\psi d\left[\frac{\gamma(\nu+1)+2\gamma+1}{2\gamma}\right]} (\log n)^{\nu(1+b)-\frac{\nu+1}{2}} \\ &\leq Cn^{-1+\frac{1-\psi d[\gamma(\nu+3)+1]}{2\gamma}} (\log n)^{\nu(1+b)-\frac{\nu+1}{2}}.\end{aligned}$$

Alors, par un choix approprié de ψ, $\mathcal{Z}_2$ est un terme général d'une série convergente. De la mème manière, nous pouvons choisir ε_0 tel que $\mathcal{Z}_1$ est un terme général d'une série convergente. Finalement, en appliquant le lemme de Borel-Cantelli à (4.10) nous obtenons le résultat.

Remarque 4.4.1 *Le paramètre ψ de l'hypothèse H4 peut être choisi tel que :*

$$\psi > \frac{1}{\gamma(\nu+3)+1}.$$

Cette condition assure la convergence des séries du lemme 4.4.1.
Finalement, en considérant les résultats du lemme 3.5.2 et du lemme 3.5.3 selon l'inégalité (3.24) nous concluons la preuve du théorème.

4.4.1 Remarques

Dans ce chapitre, un estimateur solution d'un problème de minimisation de l'erreur quadratique relative moyenne pour une co-variable multidimensionnelle à été proposé, lorsque les données sont dépendantes et sont sujet à une censure aléatoire à droite. Nous comparons l'estimateur RER à celui de la CR, nous avons constaté quelques remarques. Comme nous l'avons soupçonné, le comportement asymptotique de l'estimateur RER est meilleur pour une faible dépendance (une petite valeur de ρ) et un pourcentage faible de censure. Nous avons aussi montré que la qualité d'estimation est influencée par plusieurs paramètres (C.P., ρ, M.F., n).

Nous pouvons aussi dire que le comportement de l'estimateur RER reste résistant dans tout nos résultats en comparaison avec l'estimateur CR qui est significativement affecté par la présence de valeurs aberrantes, le taux de censure mais aussi la taille d'échantillon. Une autre remarque intéressante reliée à l'aspect dépendance est le fait que pour un petit ρ l'estimateur reste résistant.

Chapitre 5

Normalité asymptotique de l'estimateur de la fonction de régression relative pour des données censurées dépendantes

[1] Dans ce chapitre, pour un modèle de censure aléatoire à droite, nous étudions la normalité asymptotique de l'estimateur à noyau de la fonction de régression relative (dont la convergence a été étudiée dans le chapitre 4) lorsque les données présentent une forme de dépendance forte appelée α-mélange. La variance asymptotique est explicitement donnée. Des simulations sont réalisées pour consolider notre résultat théorique et illustrent la bonne qualité de la méthode étudiée. Une application aux intervalles de confiance est donnée.

5.1 Introduction

La propriété asymptotique abordée dans ce chapitre est la normalité asymptotique, il s'agit d'un sujet très important en statistique. En effet, la normalité asymptotique nous permet de construire les intervalles de confiance et de faire les tests. Nous rappelons le cadre de travail dans lequel nous sommes : Soit $T \in \mathbb{R}_+$ une v.a. réelle d'intérêt et $X_1, \cdots, X_d$ sa co-variable associée à valeurs dans $\mathbb{R}^d$. On introduit de même les variables aléatoire de censure $C_1, \cdots, C_n$ de même loi que $C \in \mathbb{R}_+$. Notre étude porte sur le modèle de censure aléatoire à droite, nos observations sont donc :

$$\begin{cases} Y_i = T_i \wedge C_i & 1 \leq i \leq n, \\ \delta_i = \mathbb{1}_{\{T_i \leq C_i\}} & 1 \leq i \leq n, \\ X_i \in \mathbb{R}^d & 1 \leq i \leq n. \end{cases}$$

L'importance d'étudier ce type de données est mis en valeur par l'application de ces derniers aux données réelles (pour des exemples réels sur les données censurées, nous renvoyons le lecteur aux livres de Andersen et al. (1993) et Klein and Moeschberger (2004)).

Dans plusieurs situations pratiques, les données observées ne sont pas toujours in-

1. En collaboration avec le Pr. E. OULD SAÏD. Ce chapitre a fait l'objet d'un article soumis pour publication.

dépendantes. Nous nous intéressons dans ce chapitre à une forme de dépendance forte appelée α-mélange (nous renvoyons à la définition 1.3.1). Dans ce qui va suivre, nous allons supposer que $(T_i)_{i\geq 1}$ et $(C_i)_{i\geq 1}$ sont deux suites indépendantes de v.a. fortement mélangeantes avec des coefficients de mélanges $\alpha_1(n)$ et $\alpha_2(n)$ respectivement. Un résultat de Cai (1998) montre que les observations $(Y_i)_{i\geq 1}$ sont fortement mélangeants avec un coefficient de mélange $\alpha(n) = 4\max(\alpha_1(n), \alpha_2(n))$. Par conséquent, nous supposons que $(Y_i, \delta_i, X_i)_{i\geq 1}$ est α-mélangeant avec un coefficient de mélange égal à α.

Dans le contexte de données dépendantes, il existe une large littérature sur l'estimation non paramétrique. Nous citons, El Ghouch and Van Keilegom (2008) qui ont considéré l'estimation de la fonction de régression pour des données α-mélangeantes censurées à droite basé sur une transformation de données. Une année plus tard, El Ghouch and Van Keilegom (2009) établissent un nouveau résultat cette fois ci sur le fonction de régression quantile dans le même contexte. Guessoum and Ould Saïd (2010, 2012) qui ont montré respectivement la convergence uniforme presque sûre sur un compact avec vitesse et la normalité asymptotique de l'estimateur de la fonction de régression. Pour plus de références concernant l'estimation non paramétrique pour des données dépendantes, nous renvoyons au livre de Fan and Yao (2003).

Le but principal de ce chapitre est d'établir la normalité asymptotique de l'estimateur de la fonction de régression relative définit par (3.13) dans le chapitre 4 sous des conditions de α-mélange. Cette étude étend le résultat de Khardani (2019) au cadre dépendant et constitue une continuité du chapitre 4 dans lequel nous avons établi la convergence uniforme presque sûre de l'estimateur de la fonction (3.12).

5.2 Hypothèses et résultat principal

Afin d'établir la normalité asymptotique de l'estimateur $\widehat{m}_{RER}(\cdot)$, nous gardons les même hypothèses et notations que celles du chapitre 4 auxquelles nous rajoutons les hypothèses suivantes :

D5. La fonction $\mu_\ell(\cdot,\cdot)$, pour $\ell = 1, 2$ existe et satisfait pour tout $(u, v) \in \mathbb{R}^{2d}$

$$\sup_{i,j}\sup_{u,v}\left|\mu_{i,j,\ell}(u,v) - \mu_{i,\ell}(u)\mu_{j,\ell}(v)\right| \leq C < \infty.$$

D6. $f(\cdot)$ est continûment différentiable.

M2. $h^{-\frac{d}{a_1}}\sum_{s>v_n}\alpha^{\frac{1}{a_1}}(s) \leq \infty$ pour $0 < a_1 < 1$ et $d > 1$.

B1. Il existe deux suites d'entiers p_n et q_n tendant vers ∞ en même temps que n tel que :

$$\frac{p_n}{q_n} \to \infty, \quad \frac{k(p_n+q_n)}{n} \to 1, \quad \frac{p_n}{\sqrt{nh_n^d}} \to 0 \quad \text{and} \quad k\alpha(q_n) \to 0.$$

De plus, pour $k := k_n = \left[\frac{n}{p_n+q_n}\right]$ (où $[\cdot]$ désigne la fonction partie entière), nous avons :

$$kq_n^{-\nu} \longrightarrow 0.$$

Commentaires :

1. D5 est une hypothèse technique nécessaire dans le calcul du terme de covariance. Dans le cas où nous majorant le terme de covariance , nous pouvons remplacer D5 par l'hypothèse D4.
2. L'hypothèse D6 est une condition de régularité sur la fonction de densité $f(\cdot)$.
3. Les hypothèses M2 et B1 sont des hypothèses techniques. La première hypothèse concerne le coefficient de mélange et la seconde hypothèse est nécessaire pour utiliser la technique de Doob (1953).

Ce théorème traite de la normalité asymptotique de l'estimateur $\widehat{m}_{RER}(\cdot)$. Notons que :

$$\Sigma(x) = \begin{pmatrix} r_2(x) & r_3(x) \\ r_3(x) & r_4(x) \end{pmatrix}$$

est la matrice de variance covariance où les fonctions $r_\lambda(\cdot)$ sont données par (3.17). Soit $\mathcal{C}^\star = \Big\{ x \in \mathcal{C} \text{ tel que } \mu_\ell(x) \neq 0,\ \ell = 1,\ 2 \text{ et } r_\lambda(x) \neq 0,\ \lambda = 2,3,4 \Big\}$. Nous établissons le résultat suivant :

Théorème 5.2.1 *Sous les hypothèses H1–H3, K1–K4, D1–D3, D5, D6, M2 et B1, pour $x \in \mathcal{C}^\star$, nous avons :*

$$\sqrt{nh_n^d}\,(\widehat{m}_{RER}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}\Big(0, \sigma^2(x)\Big), \qquad \textit{quand} \quad n \longrightarrow \infty$$

où

$$\sigma^2(x) = \kappa \frac{r_2(x)\mu_2^2(x) - 2r_3(x)\mu_1(x)\mu_2(x) + r_4(x)\mu_1^2(x)}{\mu_2^4(x)} \tag{5.1}$$

pour $\kappa = \displaystyle\int K_d^2(t)dt$ *et* $\xrightarrow{\mathcal{L}}$ *désigne la convergence en loi.*

5.2.1 Intervalles de confiance

Dans l'estimation non paramétrique, la variance asymptotique dépend de certaines fonction inconnues. Dans notre cas, pour déterminer les intervalles de confiance, nous devons estimer la quantité inconnue $r_\lambda(\cdot)$ qui apparaît dans l'expression de la variance asymptotique. Définissons un estimateur consistant de $r_\lambda(\cdot)$ pour $\lambda = 2,3,4$ par :

$$\widehat{r}_\lambda(x) = \frac{1}{nh_n^d} \sum_{1 \leq i \leq n} \frac{\delta_i Y_i^{-\lambda}}{\overline{G}_n^2(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right). \tag{5.2}$$

Nous remplaçons (5.2) dans (5.1) pour obtenir un estimateur calculable :

$$\widehat{\sigma}^2(x) = \kappa \frac{\widehat{r}_2(x)\widehat{\mu}_2^2(x) - 2\widehat{\mu}_1(x)\widehat{\mu}_2(x)\widehat{r}_3(x) + \widehat{\mu}_1^2(x)\widehat{r}_4(x)}{\widehat{\mu}_2^4(x)}. \tag{5.3}$$

Corollaire 5.2.1 *Sous les hypothèses du théorème 5.2.1, nous avons :*

$$\frac{\sqrt{nh_n^d}}{\widehat{\sigma}(x)}\,(\widehat{m}_{RER}(x) - m(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1).$$

Les intervalles de confiance de niveau $0 < \beta < 1$ *sont données par :*

$$\left]\widehat{m}_{RER}(x) - t_{1-\frac{\beta}{2}}\frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}}; \widehat{m}_{RER}(x) + t_{1-\frac{\beta}{2}}\frac{\widehat{\sigma}(x)}{\sqrt{nh_n^d}}\right[$$

où $t_{1-\frac{\beta}{2}}$ *désigne le quantile de la loi normale standard.*

Remarque 5.2.1 *Notons que ce résultat est analogue à celui obtenu dans le cas i.i.d. De plus, le choix de la fenêtre optimale est obtenu en essayons d'équilibrer les termes du biais et de la variance, ce qui donne* $h_{opt} = O\left(\frac{\log n}{n}\right)^{\frac{1}{d+2}}$. *Pour* $d = 1$, *nous trouvons le résultat (très connu)* $h_{opt} = O\left(\frac{\log n}{n}\right)^{\frac{1}{3}}$. *Ceci est due à nos hypothèses sur le noyau* K. *Si nous supposons que notre noyau est symétrique, nous obtenons le résultat suivant* $h_{opt} = O\left(\frac{\log n}{n}\right)^{\frac{2}{d+4}}$ *et alors pour* $d = 1$, *nous aurons* $h_{opt} = O\left(\frac{\log n}{n}\right)^{\frac{2}{5}}$.

5.3 Étude numérique

Le but de cette partie est d'examiner le comportement asymptotique de $\widehat{m}_{RER}(\cdot)$. Une illustration des données générées de la normalité asymptotique de notre estimateur de la RER est réalisée. Nous comparons ensuite la forme de leur densité estimée avec celui de densité normale standard. À cette fin, nous considérons le processus de mélange fort unidimensionnel ($d = 1$) généré par :

$$\begin{cases} X_i &= \rho X_{i-1} + \sqrt{1-\rho^2}\epsilon_i \\ T_i &= \frac{1}{2}X_{i+1} + 2 \qquad i = 1, \cdots, n. \end{cases}$$

où $0 \leq \rho \leq 1$ et ϵ_i est une suite de bruit blanc et $X_0 \rightsquigarrow \mathcal{N}(0,1)$. Ce dernier est un un processus auto-régressif d'ordre 1 AR(1), étant donné $X_i = x$ nous obtenons $T_i = \frac{1}{2}\rho x + 2 + \sqrt{1-\rho^2}\epsilon_i$ pour tout i. D'où,

$$m(x) = \mathbb{E}[T_i | X_i = x] = \frac{\rho}{2}x + 2$$

pour tout $x \in [1,4]$. Nous générons $C_i = X_{i+1} + \lambda$ selon un processus AR(1) où le paramètre λ représente une constante qui nous permet d'adapter le pourcentage de censure. Les simulations sont effectuées pour différentes tailles d'échantillons $n = 100$, $n = 200$ et $n = 400$ et les résultat sont obtenues pour $B = 200$ répétitions. Nous calculons la quantité suivante :

$$\sqrt{\frac{nh_n}{\widehat{\sigma}^2(x)}}\left(\widehat{m}_{RER,j}(0) - 2\right), \qquad \text{pour } 1 \leq j \leq B.$$

où $\widehat{\sigma}_n^2(x)$ est déterminée à partir de la formule (5.3). Pour les poids qui apparaissent dans notre estimateur $\widehat{m}_{RER}(x)$, nous utilisons la loi normale standard (i.e. de densité $\left(K(u) = \exp(-u^2/2)/\sqrt{2\pi}\right)$). Dans cette étude, cinq valeurs du paramètre ρ font objet $\rho = 0.3, 0.6$ et 0.9 où $\rho \rightsquigarrow 1$ est une forte dépendance. Pour chaque figure, nous utilisons une valeur de h_n différente que nous sélectionnons sur une grille de 200 valeurs équidistantes sur l'intervalle $[0.01, 2]$. Nous prenons la valeur qui minimise le critère de la validation croisée (voir la sous-section 1.8). En ce qui concerne la fenêtre optimale dans l'estimation de la densité (voir Silverman (1986)), nous avons choisi $h^\star = CB^{-0.2}$ où la constante C est choisi de manière appropriée.

De la Figure 5.1 nous pouvons constater que la qualité d'estimation s'améliore

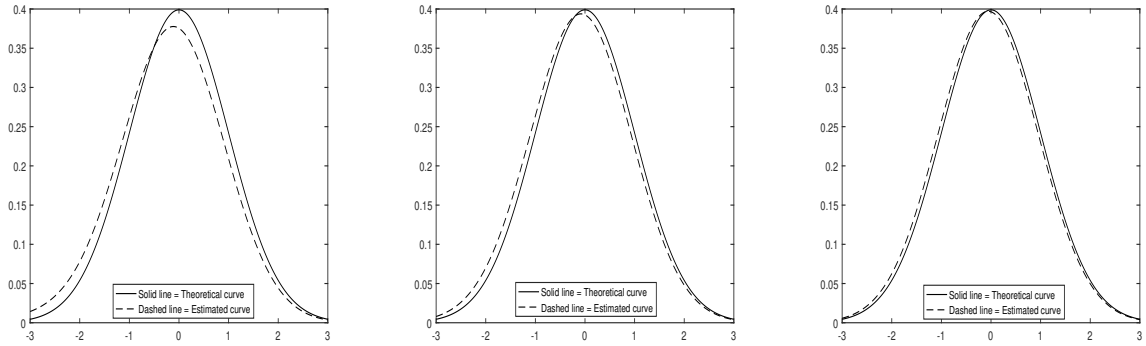


FIGURE 5.1 – C.P. $\approx 55\%$ et $\rho = 0.3$ pour $n = 50, 200$ et 400 respectivement.

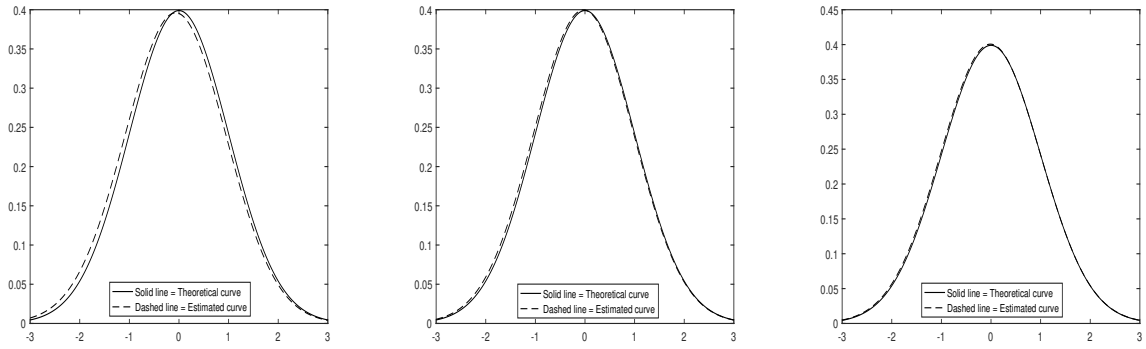


FIGURE 5.2 – $n = 200$ et C.P. $\approx 55\%$ pour $\rho = 0.9,\ 0.6$ et 0.3 respectivement.

lorsque n augmente. L'estimateur et la courbe de la loi normale standard sont tellement proches qu'il est difficile de les distinguer. Nous pouvons voir aussi que lorsque le taux de dépendance devient élevé ($\rho = 0.9$) l'estimateur $\widehat{m}_{RER}(\cdot)$ reste résistant.

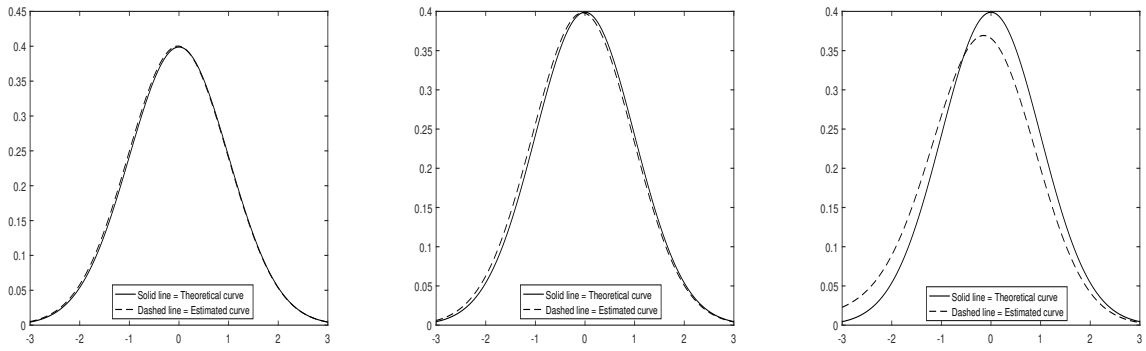


FIGURE 5.3 – $n = 200$ et $\rho = 0.6$ pour C.P. $\approx 28, 58$ et 80% respectivement.

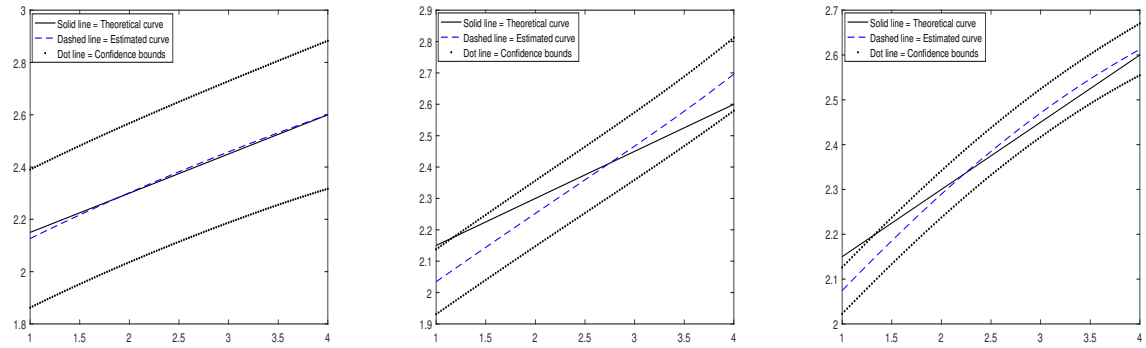


FIGURE 5.4 – $\rho = 0.3$ et C.P. $\approx 25\%$ pour $n = 50, 200$ et 400 respectivement.

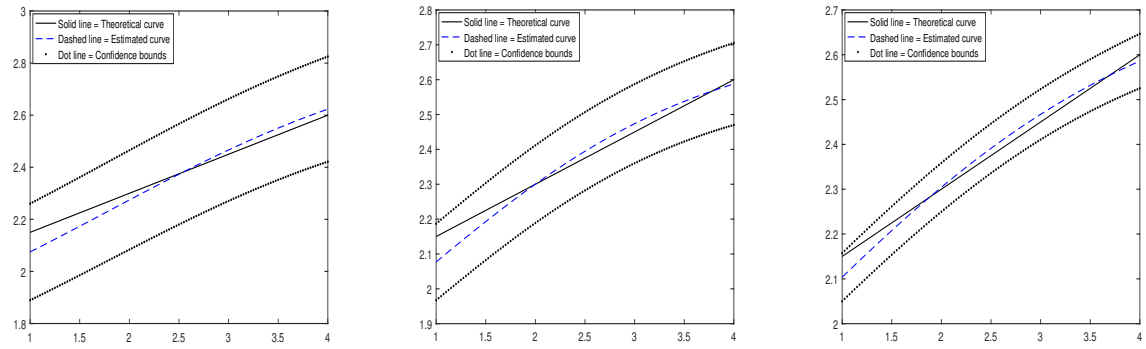


FIGURE 5.5 – $\rho = 0.3$ avec C.P. $\approx 85\%$ pour $n = 50, 200$ et 400 respectivement.

5.4 Preuves et résultats auxiliaires

Preuve du théorème 5.2.1. Nous rappelons que l'objectif ici est de prouver :

$$\sqrt{nh_n^d}\left(\widehat{m}_{RER}(x) - m(x)\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma^2(x)\right) \qquad \text{quand } n \to \infty.$$

Sous la décomposition (3.31) nous avons trois termes $\Gamma_\ell(x)$, $\Lambda_\ell(x)$ et $\Xi_\ell(x)$. Du lemme 3.5.4 les deux termes derniers termes sont négligeables sous les hypothèses H2 et H3 et valent $o_{p.s.}(1)$. Nous allons démontrer que le terme $\Gamma_\ell(x)$ est asymptotiquement normal. Nous commençons par évaluer la variance asymptotique à travers le lemme suivant :

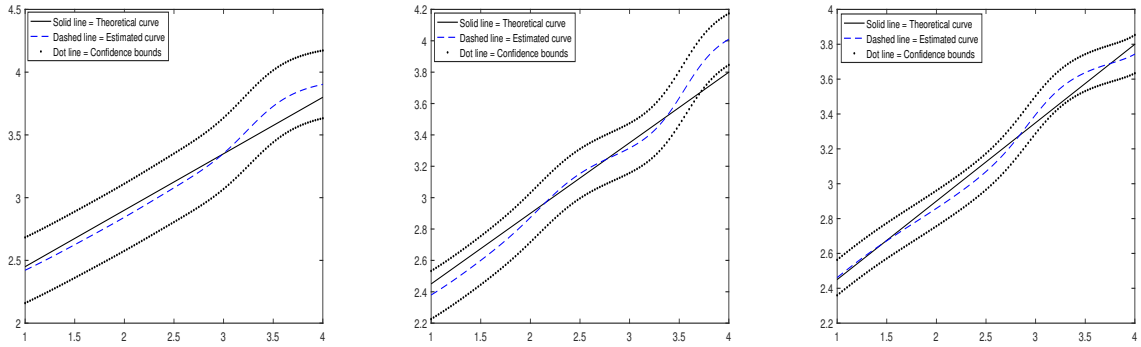


FIGURE 5.6 – $\rho = 0.9$ avec C.P. $\approx 25\%$ pour $n = 50, 200$ et 400 respectivement.

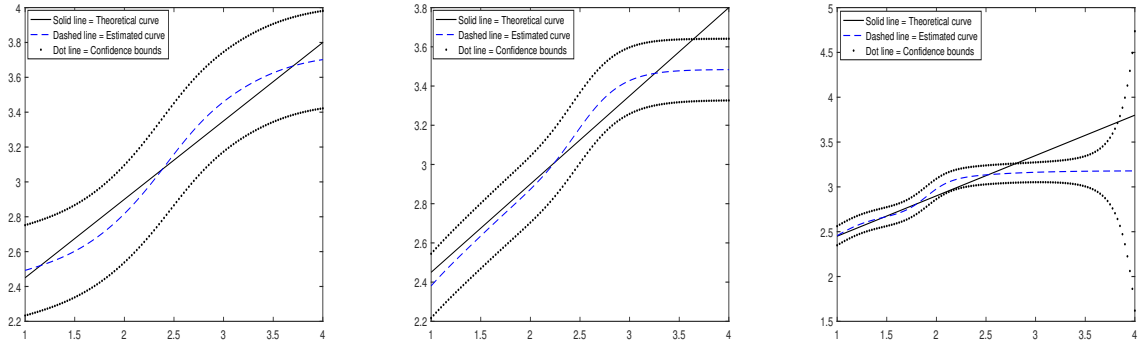


FIGURE 5.7 – $\rho = 0.9$ avec C.P. $\approx 85\%$ pour $n = 50, 200$ et 400 respectivement.

Lemme 5.4.1 *Sous les hypothèses K1, K2, D1, D2, D5 et M2 pour $\ell = 1, 2$ et $\lambda = 2\ell$, nous avons :*

$$Var\left(\Gamma_\ell(x)\right) \longrightarrow \sigma_\lambda(x) := r_\lambda(x)\kappa.$$

Preuve du lemme 5.4.1. Soit pour $1 \leq i \leq n$

$$U_{i,\ell}(x) := \widetilde{T}_i^{-\ell} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left[\widetilde{T}_i^{-\ell} K_d\left(\frac{x - X_i}{h_n}\right)\right], \qquad \text{for } \ell = 1, 2.$$

Nous avons :

$$\widetilde{\mu}_\ell(x) - \mathbb{E}\left[\widetilde{\mu}_\ell(x)\right] = \frac{1}{nh_n^d} \sum_{1 \leq i \leq n} U_{i,\ell}(x).$$

Pour $\ell = 1, 2$

$$\begin{aligned} Var\left(\Gamma_\ell(x)\right) &= nh_n^d Var\left(\frac{1}{nh_n^d} \sum_{1 \leq i \leq n} U_{i,\ell}(x)\right) \\ &= \frac{1}{nh_n^d} Var\left(\sum_{1 \leq i \leq n} U_{i,\ell}(x)\right) \\ &= \frac{1}{h_n^d} \mathbb{E}\left[U_{1,\ell}^2(x)\right] + \frac{1}{nh_n^d} \sum_{\substack{1 \leq i,j \leq n \\ i \neq j}} \mathbb{E}\left[U_{i,\ell}(x) U_{j,\ell}(x)\right] \end{aligned}$$

$$=: \quad \mathcal{A}_1 + \mathcal{A}_2.$$

Nous avons :

$$\begin{aligned}
\mathcal{A}_1 &= \frac{1}{h_n^d}\mathbb{E}\left[U_{1,\ell}^2(x)\right] \\
&= \frac{1}{h_n^d}\mathbb{E}\left[\left(\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)-\mathbb{E}\left[\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right]\right)^2\right] \\
&= \frac{1}{h_n^d}\mathbb{E}\left[\left(\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right)^2-2\left(\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right)\mathbb{E}\left[\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right]+\left(\mathbb{E}\left[\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right]\right)^2\right] \\
&= \frac{1}{h_n^d}\left\{\mathbb{E}\left[\frac{\delta_1 Y^{-2\ell}}{\overline{G}^2(Y_1)}K^2\left(\frac{x-X_1}{h_n}\right)\right]-2\mathbb{E}^2\left[\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right]+\mathbb{E}^2\left[\widetilde{T}_1^{-\ell}K\left(\frac{x-X_1}{h_n}\right)\right]\right\} \\
&= \frac{1}{h_n^d}\left\{\mathbb{E}\left[\frac{\delta_1 Y^{-2\ell}}{\overline{G}^2(Y_1)}K_d^2\left(\frac{x-X_1}{h_n}\right)\right]-\mathbb{E}^2\left[\widetilde{T}_1^{-\ell}K_d\left(\frac{x-X_1}{h_n}\right)\right]\right\} \\
&=: \frac{1}{h_n^d}\{\mathcal{A}_{1,1}-\mathcal{A}_{1,2}\}.
\end{aligned}$$

D'une part, en utilisant la propriété de l'espérance conditionnelle, un changement de Taylor et un développement de Taylor pour la fonction $r_\lambda(\cdot)$ pour $\lambda = 2, 4$, nous obtenons sous les hypothèse K1, K2 et D2 :

$$\begin{aligned}
\mathcal{A}_{1,1} &= h_n^d \int K_d^2(t) r_\lambda(x - h_n t)dt \\
&= h_n^d r_\lambda(x)\kappa,
\end{aligned}$$

D'autre part, de manière analogue et sous les hypothèses D1, K1 et K2, nous avons :

$$\mathcal{A}_{1,2} \leq \mathrm{O}\left(h_n^{2d}\right).$$

Ensuite, nous combinons les résultats obtenue ci-dessous par :

$$\begin{aligned}
\mathcal{A}_1 &\leq h_n^{-d}\{\mathcal{A}_{1,1}-\mathcal{A}_{1,2}\} \\
&\leq h_n^{-d}\left\{h_n^d r_\lambda(x)\kappa - h_n^{2d}\right\} \\
&= r_\lambda(x)\kappa. \qquad (5.4)
\end{aligned}$$

Ensuite, en utilisant le propriété de l'espérance conditionnelle, un changement de variable, nous avons :

$$\begin{aligned}
|E[U_{i,\ell}U_{j,\ell}]| &= \left|\mathbb{E}\left[\widetilde{T}_i^{-\ell}\widetilde{T}_j^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)K_d\left(\frac{x-X_j}{h_n}\right)\right]-\mathbb{E}\left[\widetilde{T}_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\right]\mathbb{E}\left[\widetilde{T}_j^{-\ell}K_d\left(\frac{x-X_j}{h_n}\right)\right]\right| \\
&= \left|\int\int K_d\left(\frac{x-u}{h_n}\right)K_d\left(\frac{x-v}{h_n}\right)\mu_{i,j,\ell}(u,v)dudv\right. \\
&- \left.\int K_d\left(\frac{x-u}{h_n}\right)\mu_{i,\ell}(u)du\int K_d\left(\frac{x-v}{h_n}\right)\mu_{j,\ell}(v)dv\right|
\end{aligned}$$

$$\begin{aligned}
&\leq \int\int K_d\left(\frac{x-u}{h_n}\right)K_d\left(\frac{x-v}{h_n}\right)\left|\mu_{i,j,\ell}(u,v)-\mu_{i,\ell}(u)\mu_{j,\ell}(v)\right|dudv\\
&\leq h_n^{2d}\int\int K_d(t)K_d(s)\left|\mu_{i,j,\ell}(x-h_nt,x-h_ns)-\mu_{i,\ell}(x-h_nt)\mu_{j,\ell}(x-h_ns)\right|dtds
\end{aligned}$$

où sous l'hypothèse D5, nous obtenons :

$$\left|\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]\right| \leq Ch_n^{2d}. \tag{5.5}$$

Nous traitons maintenant le terme $\mathcal{A}_2$ et pour cela nous considérons une suite d'entiers $v_n \to \infty$ lorsque $n \to \infty$, de sorte que :

$$\mathcal{A}_2 = \frac{1}{nh_n^d}\left\{\sum_{1\leq|i-j|\leq v_n}\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]+\sum_{v_n<|i-j|\leq n}\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]\right\}.$$

D'une part, de (5.5) nous avons :

$$\sum_{1\leq|i-j|\leq v_n}\left|\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]\right| \leq Ch_n^{2d}nv_n.$$

En choisissant pour un $b>1$, $v_n=[(h_n^{-d})^{1/b}]$, nous avons :

$$\frac{1}{nh_n^d}\sum_{1\leq|i-j|\leq v_n}\left|\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]\right| \leq Ch_n^{(b-1)d/b}\longrightarrow 0. \tag{5.6}$$

D'autre part, nous utilisons une inégalité des moments Rio (2000) (voir la proposition 5.4.3). Soit a_1, a_2 et a_3 prenant leurs valeurs dans $\mathbb{R}^\star$ plus grands ou égales à 1 tel que : $\frac{1}{a_1}+\frac{1}{a_2}+\frac{1}{a_3}=1$ et $b<a_1<\frac{\nu}{2}$. Nous avons :

$$\begin{aligned}
\frac{1}{nh_n^d}\sum_{v_n<|i-j|\leq n}\left|\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big]\right| &\leq (nh_n^d)^{-1}\sum_{v_n<|i-j|\leq n}C\,\alpha^{\frac{1}{a_1}}(|i-j|)\\
&\times\left(\mathbb{E}\left|K_d\left(\frac{x-X_i}{h_n}\right)\right|^{a_2}\right)^{\frac{1}{a_2}}\left(\mathbb{E}\left|K_d\left(\frac{x-X_j}{h_n}\right)\right|^{a_3}\right)^{\frac{1}{a_3}}.
\end{aligned}$$

Nous avons :

$$\mathbb{E}\left|K_d\left(\frac{x-X_i}{h_n}\right)\right|^{a_2} \leq Ch_n^d$$

Alors, sous l'hypothèse M2 et pour $b<a<\frac{\nu}{2}$ nous avons :

$$\begin{aligned}
\frac{1}{nh_n^d}\sum_{v_n<|i-j|\leq n}\mathbb{E}\Big[U_{i,\ell}(x)U_{j,\ell}(x)\Big] &\leq \frac{C}{nh_n^d}\sum_{v_n<|i-j|\leq n}\alpha^{\frac{1}{a_1}}(|i-j|)(h_n^d)^{\frac{1}{a_2}}(h_n^d)^{\frac{1}{a_3}}\\
&\leq \frac{C}{nh_n^d}\sum_{v_n<|i-j|\leq n}\alpha^{\frac{1}{a_1}}(|i-j|)h_n^{\frac{d}{a_1}(a_1-1)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{C}{nh_n^{\frac{d}{a_1}}} \sum_{v_n<|i-j|\leq n} \alpha^{\frac{1}{a_1}}(|i-j|) \\
&\leq \frac{C}{nh_n^{\frac{d}{a_1}}} \sum_{v_n<|i-j|\leq n} \frac{|i-j|}{v_n} \alpha^{\frac{1}{a_1}}(|i-j|) \\
&\leq \frac{C}{nh_n^{\frac{d}{a_1}}} + \sum_{s>v_n} \frac{s}{h_n^{-\frac{d}{b}}} \alpha^{\frac{1}{a_1}}(s) \\
&\leq \frac{C}{nh_n^{d\left(\frac{1}{a_1}-\frac{1}{b}\right)}} \sum_{s>v_n} s\,\alpha^{\frac{1}{a_1}}(s) \longrightarrow 0. \qquad (5.7)
\end{aligned}
$$

Finalement, en combinant (5.4), (5.6) et (5.7) nous obtenons le résultat du Lemme 5.4.1.

Lemme 5.4.2 *Sous les hypothèses K1–K3 et D1–D3, nous avons :*

$$Cov\left(\Gamma_1(x),\Gamma_2(x)\right) \longrightarrow \sigma_3(x) := r_3(x)\kappa.$$

Preuve du lemme 5.4.2. Par définition, nous avons :

$$
\begin{aligned}
Cov\left(\Gamma_1(x),\Gamma_2(x)\right) &= nh_n^d\{\mathbb{E}[\Gamma_1(x)\Gamma_2(x)] - \mathbb{E}[\Gamma_1(x)]\mathbb{E}[\Gamma_2(x)]\} \\
&=: nh_n^d\{\mathcal{S}_3 - \mathcal{S}_1\mathcal{S}_2\}.
\end{aligned}
$$

D'une part, nous avons :

$$
\begin{aligned}
\mathcal{S}_3 &= \mathbb{E}[\widetilde{\mu}_1(x)\widetilde{\mu}_2(x)] - \mathbb{E}[\widetilde{\mu}_1(x)]\mathbb{E}[\widetilde{\mu}_2(x)] \\
&=: \mathcal{S}_{3,1} - \mathcal{S}_{3,2}.
\end{aligned}
$$

Pour $\mathcal{S}_{3,1}$, nous avons :

$$
\begin{aligned}
\mathcal{S}_{3,1} &= \frac{1}{(nh_n^d)^2} \sum_{1\leq i,j\leq n} \mathbb{E}\left[\widetilde{T}_i^{-1}\widetilde{T}_j^{-2} K_d\left(\frac{x-X_i}{h_n}\right) K_d\left(\frac{x-X_j}{h_n}\right)\right] \\
&= \frac{1}{(nh_n^d)^2}\left\{\sum_{1\leq i\leq n} \mathbb{E}\left[\frac{\delta_i Y_i^{-3}}{\overline{G}^2(Y_i)} K_d^2\left(\frac{x-X_i}{h_n}\right)\right] + \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \mathbb{E}\left[\widetilde{T}_i^{-1}\widetilde{T}_j^{-2} K_d\left(\frac{x-X_i}{h_n}\right) K_d\left(\frac{x-X_j}{h_n}\right)\right]\right\} \\
&= \frac{1}{(nh_n^d)^2}\left\{\sum_{1\leq i,j\leq n} \mathcal{S}_{3,1,1} + \sum_{\substack{1\leq i,j\leq n \\ i\neq j}} \mathcal{S}_{3,1,2}\right\}
\end{aligned}
$$

Pour $\mathcal{S}_{3,1,1}$, en utilisant la propriété de l'espérance conditionnelle, un changement de variable et un développement de Taylor. Sous les hypothèses K3 et D2, nous avons :

$$
\begin{aligned}
\mathcal{S}_{3,1,1} &= \mathbb{E}\left[\frac{\delta_1 Y_1^{-3}}{\overline{G}^2(Y_1)} K_d^2\left(\frac{x-X_1}{h_n}\right)\right] \\
&= \mathbb{E}\left[K_d^2\left(\frac{x-X_1}{h_n}\right)\mathbb{E}\left[\frac{T_1^{-3}}{\overline{G}(T_1)}\Big|X_1\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \int K_d^2\left(\frac{x-u}{h_n}\right)\int \frac{t^{-3}}{\overline{G}(t)} f(t|u)dt f(u)du \\
&= \int K_d^2\left(\frac{x-u}{h_n}\right)\int \frac{t^{-3}}{\overline{G}(t)} f(u,t)dtdu \\
&= \int K_d^2\left(\frac{x-u}{h_n}\right) r_3(u)du \\
&= h_n^d \int K_d^2(t) r_3(x-h_n t)dt \\
&\leq h_n^d r_3(x)\int_{\mathbb{R}^d} K_d^2(t)dt =: h_n^d r_3(x)\kappa.
\end{aligned}
$$

De (5.5), nous avons $\left|\mathbb{E}[U_{i,\ell}(x)U_{j,\ell}(x)]\right| \leq Ch_n^{2d}$ alors

$$
\begin{aligned}
\mathcal{S}_{3,1} &\leq (nh_n^d)^{-2}\left\{nh_n^d r_3(x)\kappa + O\left((nh_n^d)^2\right)\right\} \\
&= (nh_n^d)^{-1} r_3(x)\kappa + o(1)
\end{aligned}
$$

nous concluons la preuve du lemme 5.4.2.
Pour prouver la normalité asymptotique de $\Gamma_\ell(x)$ pour $\ell = 1,2$ et pour toute combinaison linéaire. Nous utilisons la technique de Doob (1953). Pour un couple de nombre réel (c_1, c_2), définissons :

$$
\Gamma_n(x) := \sqrt{nh_n^d}\sum_{\ell=1}^{2} c_\ell \Gamma_\ell(x) = c_1\Gamma_1(x) + c_2\Gamma_2(x). \tag{5.8}
$$

Lemme 5.4.3 *Sous les hypothèses K1, D2 et M2 nous avons :*

$$
\Gamma_n(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_4^2(x))
$$

où $\sigma^2(x) = \kappa c^t \Sigma_x c$ *pour* $c = (c_1, c_2)^t$ *et*

$$
\Sigma_x = \begin{pmatrix} r_2(x) & r_3(x) \\ r_3(x) & r_4(x) \end{pmatrix}.
$$

Preuve du lemme 5.4.3. Nous avons de (5.8) que :

$$
\begin{aligned}
Var(\Gamma_n(x)) &= Var\left(c_1\Gamma_1(x) + c_2\Gamma_2(x)\right) \\
&= c_1^2 Var\left(\Gamma_1(x)\right) + c_2^2 Var\left(\Gamma_2(x)\right) + 2c_1c_2 Cov\left(\Gamma_1(x), \Gamma_2(x)\right).
\end{aligned}
$$

Les deux termes de variances ont été traité dans le lemme 5.4.1, quand au terme de covariance il a été traité dans le lemme 5.4.2. Maintenant, soit $W_i(x) := c_1 U_{i,1}(x) + c_2 U_{i,2}(x)$, alors

$$
\Gamma_n(x) = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}\frac{W_i(x)}{\sqrt{h_n^d}} =: \frac{\widetilde{W}_n(x)}{\sqrt{n}}. \tag{5.9}
$$

Notre but ici est de montrer :

$$\frac{\widetilde{W}_n(x)}{\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_4^2(x)).$$

En utilisant la technique de Doob (1953) p. 228–232, nous divisons l'ensemble en un nombre k larges bloques p et un nombre k petits bloques q. Soit :

$$\begin{aligned} w_{1,j} &= \sum_{i=(j-1)(p_n+q_n)+1}^{(j-1)(p_n+q_n)+p_n} \frac{W_i(x)}{\sqrt{h_n^d}}, \\ w_{2,j} &= \sum_{i=(j-1)(p_n+q_n)+p_n+1}^{j(p_n+q_n)} \frac{W_i(x)}{\sqrt{h_n^d}}, \\ w_{3,k} &= \sum_{i=k(p_n+q_n)+1}^{n} \frac{W_i(x)}{\sqrt{h_n^d}} \end{aligned}$$

et

$$W_{1,n} = \sum_{j=1}^{k} w_{1,j}, \quad W_{2,n} = \sum_{j=1}^{k} w_{2,j} \quad \text{and} \quad W_{3,n} = \sum_{j=1}^{k} w_{3,j}$$

où $W_n(x) = W_{1,n} + W_{2,n} + W_{3,n}$.

$$\frac{1}{\sqrt{n}} W_{1,n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_4^2(x)), \tag{5.10}$$

et

$$\frac{1}{n}\mathbb{E}\left[W_{2,n}^2\right] + \frac{1}{n}\mathbb{E}\left[W_{3,n}^2\right] \longrightarrow 0. \tag{5.11}$$

Nous commençons par prouver dans un premier temps :

$$\begin{aligned} \mathbb{E}\left[W_{2,n}^2\right] &= \mathbb{E}\left[\left(\sum_{j=1}^{k} w_{2,j}\right)^2\right] \\ &= \sum_{j=1}^{k} Var\left(w_{2,j}\right) + 2 \sum_{1\le j,l\le k} Cov(w_{2,j}, w_{2,\ell}). \end{aligned} \tag{5.12}$$

D'une part, nous avons :

$$\begin{aligned} Var\left(w_{2,j}\right) &= Var\left(\sum_{i=(j-1)(p_n+q_n)+p_n+1}^{j(p_n+q_n)} \frac{W_i(x)}{\sqrt{h_n^d}}\right) \\ &= q_n Var\left(\frac{W_1(x)}{h_n^d}\right) + \sum_{0\le|i-j|\le q_n} Cov\left(\frac{W_i(x)}{h_n^d}, \frac{W_l(x)}{h_n^d}\right) \\ &\le q_n Var\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right) + o(q_n). \end{aligned} \tag{5.13}$$

Alors,

$$\sum_{j=1}^{k} Var\left(w_{2,j}\right) = kqVar\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right) + o(kq_n). \tag{5.14}$$

D'autre part, nous avons :

$$Var\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right) = c_1^2\sigma_1^2(x) + c_2^2\sigma_2^2(x) + 2c_1c_2\sigma_3(x) + o(1). \tag{5.15}$$

De B1 nous avons que $kq_n = \dfrac{nq_n}{(p_n+q_n)} = o(n)$ alors

$$\sum_{j=1}^{k} Var\left(w_{2,j}\right) = o(n). \tag{5.16}$$

Maintenant, par stationnarité de l'hypothèse M2, we have

$$\begin{aligned} 2\sum_{j,\ell=1}^{k} Cov\left(w_{2,j}, w_{2,\ell}\right) &\leq C\frac{kq_n}{h_n^d}\sum_{l=1}^{k-1}\sum_{m=1}^{q_n} h^{d\left(1-\frac{1}{a_1}\right)}\alpha^{\frac{1}{a_1}}(l(p_n+q_n)+m) \\ &\leq C\frac{kq_n}{h^{\frac{d}{a}}}\sum_{j\geq p_n+q_n+1}\alpha^{\frac{1}{a_1}}(j) = o(n.) \end{aligned} \tag{5.17}$$

Alors, de (5.12), (5.16) et (5.17), nous déduisons que :

$$\frac{1}{n}\mathbb{E}\left[W_{2,n}^2\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{5.18}$$

À présent, nous traitons $W_{3,n}$ dans le même esprit de $W_{2,n}$, nous avons :

$$\begin{aligned} \frac{1}{n}\mathbb{E}\left[W_{3,n}^2\right] &= \frac{1}{n}\left\{(n-k(p_n+q_n))Var\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right)\right. \\ &+ \left.\sum\sum_{0\leq|i-j|\leq n-k(p_n+q_n)} Cov\left(\frac{W_i(x)}{\sqrt{h_n^d}}, \frac{W_j(x)}{\sqrt{h_n^d}}\right)\right\} \\ &\leq \left(1-\frac{k(p_n+q_n)}{n}\right)Var\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right) + o(1). \end{aligned}$$

Sous B1 nous savons que $\frac{k(p_n+q_n)}{n} \to 1$, alors nous obtenons :

$$\frac{1}{n}\mathbb{E}\left[W_{3,n}^2\right] \longrightarrow 0. \tag{5.19}$$

Alors, (5.18) et (5.19) termine la preuve de (5.11).
Maintenant concernant (5.10), soient $\psi(s)$ et $\psi(j)$ les fonctions caractéristiques de

$\frac{1}{\sqrt{n}}W_{1,n}$ et $w_{1,j}$ respectivement. Nous montrons que :

$$\lim\left|\psi(x)-\prod_{j=1}^{k}\psi_j(s)\right|=0$$

ce qui prouve que les v.a. $w_{1,j}$ sont asymptotiquement indépendant, et nous prouvons que

$$\prod_{j=1}^{k}\psi_j(s)\xrightarrow{\mathcal{D}}\mathcal{N}(0,\sigma_4^2(x))$$

Pour la première partie, en utilisant l'inégalité de Volkonskii and Rozanov (1959) (voir la proposition 5.4.4), nous avons :

$$\begin{aligned}\left|\psi(s)-\prod_{j=1}^{k}\psi_j(s)\right| &= \mathbb{E}\left[e^{it\frac{W_{1,n}}{\sqrt{n}}}\right]-\prod_{j=1}^{k}\mathbb{E}\left[e^{itw_{1,j}}\right]\\ &\leq 16k\alpha(q_n+1)\leq C\,k\alpha(q_n)\longrightarrow 0.\end{aligned}$$

Pour la seconde partie, nous avons :

$$\frac{1}{n}\sum_{j=1}^{k}\mathbb{E}\left[w_{1,j}^2\right]\longrightarrow\sigma_4^2(x) \tag{5.20}$$

et pour tout $\epsilon>0$

$$\frac{1}{n}\sum_{j=1}^{k}\mathbb{E}\left[w_{1,j}^2\mathbb{1}_{\{w_{1,j}>\epsilon\sqrt{n}\sigma_{4,n}(x)\}}\right]\longrightarrow 0. \tag{5.21}$$

De plus,

$$\begin{aligned}Var\left(w_{1,j}\right) &= \left\{p_nVar\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right)+\sum\sum_{0<|i-\ell|\leq p_n}Cov\left(\frac{W_i(x)}{\sqrt{h_n^d}},\frac{W_\ell(x)}{\sqrt{h_n^d}}\right)\right\}\\ &\leq p_nVar\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right)+o(p_n).\end{aligned}$$

alors

$$\frac{1}{n}\sum_{j=1}^{k}\mathbb{E}\left[w_{1,j}^2\right]=\frac{kp_n}{n}Var\left(\frac{W_1(x)}{\sqrt{h_n^d}}\right)+o\left(\frac{kp_n}{n}\right).$$

De (5.15) et sachant que $\frac{p_nk}{n}\to 1$ (5.20) est démontrée.
Maintenant, pour établir (5.21) nous avons que $\frac{W_i(x)}{\sqrt{h_n^d}}\leq\frac{C}{\sqrt{h_n^d}}$. Alors, sous l'hypothèse B1, nous obtenons :

$$\frac{1}{\sqrt{h_n^d}}|w_{1,j}|\leq\frac{Cp_n}{\sqrt{nh_n^d}}\to 0.$$

D'où, l'ensemble $\{|w_{1,j}|>\epsilon\sqrt{n}\sigma_{4,n}(x)\}$ est vide pour n assez grand et termine la preuve

de (5.21).

Corollaire 5.4.1 *Sous les hypothèses K3 et D2, pour $\ell = 1, 2$, nous avons :*

$$\Gamma_\ell(x) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\ell^2(x)).$$

Preuve du corollaire 5.4.1. Si $\ell = 1$ et nous fixons $c_1 = 1$ et $c_2 = 0$ (Si $\ell = 2$ et nous fixons $c_1 = 0$ et $c_2 = 1$ respectivement) dans (5.8), en utilisant les résultats du lemme 5.4.1 et Lemme 5.4.3 nous terminons la preuve.

Corollaire 5.4.2 *Sous les hypothèses K1 et D2, nous avons :*

$$(\Gamma_1(x), \Gamma_2(x)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(x)\kappa)$$

où $\Sigma(x)$ est défini dans le lemme 5.4.3.

Preuve du corollaire 5.4.2. En utilisant le résultat du lemme 5.4.3 et le corollaire 5.4.1 nous obtenons le résultat.

Conclusion

Dans cette thèse, nous nous sommes intéressés à l'étude des modèles non paramétriques dans un but d'estimer la fonction de régression. Les résultats que nous énonçons sont liés aux propriétés asymptotiques des estimateurs à noyau et local linéaire pour un modèle de censure.

Notre intérêt est dans un premier temps, l'estimation par la méthode linéaire locale de la fonction de régression. Nous supposons que l'échantillon que nous étudions est constitué de variables i.i.d. et que la variable d'intérêt est censurée aléatoirement à droite. Nous établissons la convergence uniforme presque sûre avec vitesse de notre estimateur. Une étude de simulation montre les performances de la méthode étudiée.

Nous considérons dans un second temps le problème d'estimation de la fonction de régression par la méthode à noyaux pour un échantillon constitué de variable i.i.d. La fonction à minimiser ici est l'erreur quadratique relative moyenne qui est robuste aux valeurs aberrantes. Nous proposons un estimateur à noyau pour la fonction de régression lorsque la variable d'intérêt est sujet à une censure aléatoire à droite. Nous établissons la convergence uniforme presque sûre sur un compact de notre estimateur avec vitesse et sa normalité asymptotique et nous donnons l'expression explicite des termes asymptotiquement dominants du biais et de la variance. Une conséquence directe de ce dernier résultat est la construction d'intervalles de confiance asymptotiques ponctuels dont nous étudions les propriétés aux travers des simulations. Une large étude numérique sur données générées à été entrepris dans le but de renforcer notre résultat théorique. Nous appliquons la nouvelle approche sur un exemple de données réelles liées à un cancer de la peau.

Nous supposons par la suite que l'échantillon que nous étudions est constitué de variables α-mélangeantes et que le modèle de régression est solution du problème de minimisation de l'erreur quadratique relative moyenne. Nous établissons un résultat de convergence uniforme presque sûre avec vitesse de l'estimateur sous des conditions générales. Une étude de simulation est conduite afin de tester le comportement de cet estimateur pour un échantillon de taille finie, différents taux de censures et dépendances.

Enfin, nous étendons le résultat de convergence et prouvons la normalité asymptotique de l'estimateur de la fonction de régression sous des conditions de forte dépendance. Nous établissons la normalité asymptotique et nous donnons l'expression explicite de sa variance asymptotique. Nous conduisons une étude de simulation pour confirmer notre résultat théorique.

Extensions et perspectives

Il reste beaucoup de questions sans réponses :

- Dans le chapitre 2, nous n'avons obtenu qu'un résultat de convergence. Il serait alors intéressant d'étudier la normalité asymptotiques de l'estimateur de la fonction de régression par la méthode linéaire locale. On pourra aussi penser à étudier ce dernier dans le cadre dépendant (pour des données α-mélangeantes).

- Un autre axe de recherche auquel je m'intéresse tout particulièrement et dans lequel je souhaiterais m'investir à l'avenir, est l'étude des données manquantes (MAR : missing at random) dans le cadre de données censurées pour l'estimation de la fonction de régression dans un contexte non-paramétrique.

- Des travaux en collaboration avec le Pr. *E. Ould Saïd* et le Pr. *M. Lemdani* ont déjà été fait dans le cadre de l'estimation locale linéaire de la fonction de régression pour des données tronquées à gauche. On pourra penser à étendre l'étude aux données dépendantes. De plus, pourquoi pas étudier la normalité asymptotique de l'estimateur étudié.

- Une autre piste consiste à étendre nos travaux au cas d'une co-variable censurée.

- Il serait aussi intéressant d'étudier le comportement de l'estimation de la fonction de régression pour des données censurées lorsque la dépendance entre le temps de censure C et le temps de survie T est décrite par une copule.

Annexe

Théorème central limite

Théorème 5.4.1 (Loève (1963)) *Soit $(X_n)_{n\in\mathbb{Z}}$ une suite de v.a. i.i.d. centrée. Notons par s_n l'écart type de la somme partielle S_n. Si*

$$\frac{1}{s_n^{2+\gamma}}\sum_{k=1}^{n}\mathbb{E}\left[|X_k|^{2+\gamma}\right]\longrightarrow 0,$$

alors pour n assez grand

$$\frac{S_n}{s_n}\xrightarrow{\mathcal{L}}\mathcal{N}(0,1).$$

Théorème 5.4.2 (Doukhan (1994)) *Soit $(X_n)_{n\in\mathbb{Z}}$ une suite stationnaire et fortement mélangeante de variables aléatoires réelles centrées, de suite de coefficients de mélange $(\alpha_n)_n \geq 0$. Notons la fonction quantile de la variable X définie par $Q_x(u) = \inf\{t : \mathbb{P}(|X| > t) \leq u\}$, $S_n = \sum_{i=1}^{n} X_i$ la somme partielle et α^{-1} la fonction définie par $\alpha^{-1}(u) = \sum_{i\in\mathbb{N}} \mathbb{1}_{u<\alpha_i}$. Si*

$$\int_0^1 \alpha^{-1}(u)Q_{X_0}^2(u)du \leq +\infty$$

alors la série $\sum_{n\in\mathbb{Z}}\mathbb{E}[X_0X_n]$ converge vers $\sigma^2 \geq 0$ et $n^{-1}\mathrm{V}ar(S_n)$ converge aussi vers σ^2. De plus, si $\sigma^2 > 0$, alors

$$\frac{S_n}{\sqrt{n}}\xrightarrow{\mathcal{L}}\mathcal{N}(0,\sigma^2).$$

Inégalités exponentielles

La première inégalité est celle de Fuk-Nagaev. Cette dernière est une extension de l'inégalité de Bernstein au cadre de variables fortement mélangeantes.

Proposition 5.4.1 (A.11 ii, p.237. dans Ferraty and Vieu (2006)) *Soit $\{U_i, i \geq 1\}$ une suite de v.a. réelles avec un coefficient de mélange $\alpha(n) = O(n^{-v})$, $v > 1$ tel que $\forall n \in \mathbb{N}$, $\forall i \in \mathbb{N}$, $1 \leq i \leq n$ $|U_i| < +\infty$. Alors, pour tout $\varepsilon > 0$ et pour chaque $r > 1$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} U_i\right| > \varepsilon\right) \leq C\left(1+\frac{\varepsilon^2}{rS_n^2}\right)^{-r/2} + \frac{nC}{r}\left(\frac{2r}{\varepsilon}\right)^{v+1}$$

où $S_n^2 = \sum_{i,j} |\mathrm{C}ov(U_i, U_j)|$.

Classes de Vapnik-Cervonenkis (V-C classes)

On se donne un espace métrique (E, d) et un $\varepsilon > 0$. Le nombre de ε-recouvrement de l'espace métrique (E, d) noté $\mathcal{N}(E, d, \varepsilon)$ est défini comme le nombre minimal de boules ouvertes d de centres dans E et de rayon ε, requis pour couvrir l'ensemble E.
Une classe de fonctions mesurables $\mathcal{M}$ est une V-C classe de fonctions par rapport à l'enveloppe M s'il existe une fonction mesurable M presque partout finie avec $|\theta| \leq M$ pour toute fonction $\theta \in \mathcal{M}$, et des nombres réels C_1 et C_2 tels que :

$$\mathcal{N}\left(\mathcal{M}, \|\cdot\|_2, \varepsilon \|\mathcal{M}\|_2\right) \leq \left(\frac{C_1}{\varepsilon}\right)^{C_2},$$

pour tout $\varepsilon \in (0,1)$ et toute mesure de probabilité $\mathbb{P}$ pour laquelle :

$$\int M^2 d\mathbb{P} < \infty.$$

Pour approfondir cette notion de V-C classe, vous pouvez consulter le livre de Pollard (1984).

Lemme 5.4.4 (Giné and Guillou (1999)) —

(a). Si $\mathcal{M}$ est finie alors $\mathcal{M}$ est une V-C classe par rapport à l'enveloppe $\max\{|\theta| / \theta \in \mathcal{M}\}$.

(b). Si $\mathcal{M} = \{\theta_x, x \in E\}$ où E est une partie de $\mathbb{R}$ et $0 \leq h_x(s) \leq h_y(s)$ pour tout $x, y \in E$, $x < y$ et $s \in S$, alors $\mathcal{M}$ est une V-C classe par rapport à $M = \sup\{|\theta| / \theta \in \mathcal{M}\}$.

(c). Si $\mathcal{M}_1$ et $\mathcal{M}_2$ sont deux V-C classe par rapport à M_1 et M_2 respectivement, alors $\{\theta_1 + \theta_2 / \theta_1 \in \mathcal{M}_1, \theta_2 \in \mathcal{M}_2\}$ et $\{\theta_1 - \theta_2 / \theta_1 \in \mathcal{M}_1, \theta_2 \in \mathcal{M}_2\}$ sont des V-C classes par rapport à $\sqrt{M_1^2 + M_2^2}$.

Pour la preuve de ce lemme, veuillez consulter l'article de Giné and Guillou (1999).
L'inégalité qui va suivre est celle de Talagrand.

Proposition 5.4.2 (Giné and Guillou (2002))) *Si $\{\theta_i,\ i = 1, \cdots, n\}$ sont n v.a.r. i.i.d. et si $\mathcal{M}$ est une V-C classe mesurable et uniformément bornée de fonctions telles que* $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq U_n$ *et* $\sup_{f \in \mathcal{F}} \mathrm{V}ar(f(\theta_i)) \leq \sigma_n^2$ *où σ_n et U_n sont des nombres réels vérifiant $0 \leq \sigma_n \leq U_n$, alors il existe des constantes C_1 et C_2 ne dépendant que des caractéristiques A et v de la V-C classe telle que :*

$$\mathbb{P}\left\{\sup_{f \in \mathcal{F}} \left| \sum_{1 \leq i \leq n} (f(\theta_i) - \mathbb{E}[f(\theta_1)]) \right| > t \right\}$$

$$\leq C_1 \exp\left\{ -\frac{1}{C_2} \frac{t}{U_n} \log\left(1 + \frac{t U_n}{C_1 \left(\sqrt{n}\sigma_n + U_n \sqrt{\log\left(A \frac{U_n}{\sigma_n}\right)} \right)^2} \right) \right\}$$

pour tout $t \geq C_3 \sqrt{\log\left(A \frac{U_n}{\sigma_n}\right)}$.

La deux dernières inégalités sont l'inégalité de covariance de Davydov et l'inégalité de Volkonskii et Rozanov respectivement.

Proposition 5.4.3 (Ferraty and Vieu (2006), proposition A10, p236) *Soit $(X_n)_{n\in\mathbb{Z}}$ une suite de variables aléatoires α-mélangeante stationnaire. Pour $k \in \mathbb{Z}$, considérons les variables réelles $\mathcal{X}$ (resp. $\mathcal{X}'$) qui sont $\mathcal{F}_{-\infty}^{k}$ mesurable (resp. $\mathcal{F}_{n+k}^{\infty}$).*
Si pour $p,q,r > 1$ satisfaisant $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$ nous avons $\mathbb{E}(\mathcal{X})^p < \infty$ et $\mathbb{E}(\mathcal{X}')^p < \infty$, alors il existe $0 < C < \infty$ tel que

$$Cov(\mathcal{X},\mathcal{X}') \leq C\left(\mathbb{E}(\mathcal{X})^p\right)^{1/p}\left(\mathbb{E}(\mathcal{X})^q\right)^{1/q}\left(\alpha(n)\right)^{1/r},$$

où la tribu $\mathcal{F}_a^b = \sigma\{X_i, a \leq i \leq b\}$.

Proposition 5.4.4 (Volkonskii and Rozanov (1959), Lemme 1.1) *Soient $X_1,\cdots,X_m$ des variables aléatoires α mélangeantes mesurable par rapport aux σ-algèbres $\mathcal{F}_{i_1}^{j_1},\cdots,\mathcal{F}_{i_m}^{j_m}$ respectivement, avec $1 \leq i_1 < j_1 < \cdots < i_m < j_m \leq n$, $i_{l+1} - j_l \geq w \geq 1$ et $|X_j| \leq 1$ pour $l, j = 1,\cdots,m$ alors*

$$\left|\left(\mathbb{E}\prod_{j=1}^{m} X_j\right) - \prod_{j=1}^{m}\mathbb{E}\left[X_j\right]\right| \leq 16(m-1)\alpha(w).$$

Bibliographie

B. Altendi, J. Demongeot, A. Laksaci, and M. Rachdi. Functional data analysis : estimation of the relative error in functional regression under left truncated model. *J. Nonparametric Statist.*, 30 :472–490, 2018.

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer-Verlag, 1993.

H. Benseradj and Z. Guessoum. Strong uniform consistency rate of an m-estimator of regression function for incomplete data under α-mixing condition. *Communications in Statist. Theory and Methods*, page DOI : 10.1080/03610926.2020.1764037, 2020.

R. Beran. Nonparametric regression with randomly censored survival data. Technical report, Department of Statistics, University of California, Berkeley., 1981.

G. Boente and R. Fraiman. Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *Ann. Statist.*, 18 :891–906, 1990.

T. Bollerslev. General autoregressive conditional heteroskedasticity. *J. Economt.*, 31 : 307–327, 1986.

D. Bosq. *Nonparametric statistics for stochastics processes. estimation and prediction*, volume 110. Springer-Verlag. New-York, 1998.

D. Bosq and J. P. Lecoutre. *Théorie de l'estimation fonctionnelle*. Economica., 1987.

R. C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surveys.*, 2 :107–144, 2005.

R. C. Bradley. *Introduction to strong mixing conditions.* Kendrick press, 3 :1-3, 2007.

Z. Cai. Asymptotic propreties of kaplan-meier estimator for censored dependent data. *Stat. Probab. Lett.*, 37 :381–389, 1998.

Z. Cai. Estimating a distribution function for censored time series data. *J. Multiv. Analysis*, 78 :299–318, 2001.

Z. Cai. *Weignted local linear approach to censored nonparametric regression.* Recent Advances and Trends in Nonparametric Statist., 2003.

A. Carbonez, L. Gyorfi, and E. C. Van Der Meulen. Partitioning estimates of a regression function under random censoring. *Statist. and Decisions.*, 76 :1335–1344, 1995.

R. J. Carroll and D. Ruppert. *Transformation and weighting in regression*. Chapman and Hall, London, 1988.

A. Chahad, L. Ait-Hennani, and A. Laksaci. Functional local linear estimate for functional relative-error regression. *Journal of Statistical Theory and Practice*, 2017.

K. Chen, S. Guo, Y. Lin, and Z. Ying. Least absolute relative error estimation. *J. Amer. Statist. Assoc.*, 105 :1104–1112, 2010.

Y. S. Chow and H. Teicher. *Probability theory. independence, interchangeability, martingales*. Springer, New York, 1997.

G. Collomb. Estimation non paramétrique de la régression : revue bibliographique. *Int. Statist. Rev.*, 49 :75–93, 1981.

G. Collomb and W. Härdle. Strong uniform convergence rates in robust nonparametric time series analysis and prediction : kernel regression estimation from dependent observations. *Stochastic Process. Appl.*, 23 :77–89, 1986.

D. R. Cox. Regression models and life-tables (with discussion). *J. R. Statist. Soc. Ser. B.*, 34 :187–202, 1972.

D. Dabrowska. Nonparametric regression with censored survival data. *Scand. J. Statist.*, 14 :181–197, 1987.

D. Dabrowska. Uniform consistency of the kernel conditional kaplan-meier estimate. *Ann. of Statist.*, 17 :1157–1167, 1989.

Y. A. Davydov. The invariance principale for stationary processes. *Theory. Probab. Application*, 14 :487–498, 1970.

P. Deheuvels and J. H. Einmahl. Functional limit laws for the increments of kaplan-meier product limit processes and applications. *Ann Probab.*, 28 :1301–1335, 2000.

M. Delecroix, O. Lopez, and V. Patilea. Nonlinear censored regression using synthetic data. *Scand. J. Statist.*, 35 :248–256, 2008.

J. Demongeot, A. Hamie, A. Laksaci, and M. Rachdi. Relative-error prediction in nonparametric functional statistics : theory and practice. *J. of Multivariate Anal.*, 146 : 261–268, 2016.

S. Dhompongsa. A note on the almost sure approximation of the empirical process of weakly dependent random vectors. *Yokohama Math.*, 32 :113–121, 1984.

J. L. Doob. *Stochastic processes*. New York, NY. John Wiley & Sons., 1953.

P. Doukhan. *Mixing : Properties and examples*. Lecture Notes in Statistics, 85, Springer-Verlag, New York, 1994.

P. Doukhan, P. Massart, and E. Rio. The functional central limit theorem for strongly mixing processes. *Ann. Inst. H. Poincaré. Probab. Statist.*, 30(1) :63–82, 1994.

A. El Ghouch and I. Van Keilegom. Nonparametric regression with dependent censored data. *Scandinavian J. of Statist.*, 35(2) :228–247, 2008.

A. El Ghouch and I. Van Keilegom. Local linear quantile regression with dependent censored data. *Statist. Sinica*, 19 :1621–1640, 2009.

R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50 :987–1007, 1982.

J. Fan. Design adaptative nonparametric regression. *J. of the American Statist. Association*, 87 :998–1004, 1992.

J. Fan and I. Gijbels. *Censored regression : local linear approximations and their applications.*, volume 89. 1994.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications.*, volume 66. Chapman & Hall/CRC., 1996.

J. Fan and Q. Yao. *Nonlinear time series : nonparametric and parametric methods*. Springer, New York, 2003.

N. R. Farum. Improving the relative error of estimation. *The Amer. Stat.*, 44 :288–289, 1990.

F. Ferraty and P. Vieu. *Nonparametric functional data analysis : theory and practice.* Springer, New York., 2006.

A. Földes and L. Rejtö. A l.i.l. type result for the product limit estimator. *Probability Theory and Related Fields*, 56(1) :75–86, 1981.

E. J. Freireich, E. Gehan, E. Frei, L. R. Schroeder, I. J. Wolman, R. Anbari, E. O. Burgert, S. D. Mills, D. Pinkel, O. S. Selawry, J. H. Moon, B. R. Gendel, C. L. Spurr, R. Storrs, F. Haurani, B. Hoogstraten, and S. Lee. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukaemia : a model for evaluation of other potentially useful therapy. *Blood.*, 21 :699–716, 1963.

T. Gasser and H. G. Muller. Kernel estimation of regression function, in smoothing techniques for curve estimation. *Lecture Notes in Mathematics.*, pages 23–68, 1979.

E. Giné and A. Guillou. Law of the iterated logarithm for censored data. *Ann. of Probab.*, 27 :2042–2067, 1999.

E. Giné and A. Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. I. H. Poincaré*, 38 :907–921, 2002.

Z. Guessoum and F. Hamrani. Convergence rate of the kernel regression estimator for associated and truncated data. *J. of Nonparametric Statist.*, 2 :425–446, 2017.

Z. Guessoum and E. Ould Saïd. On nonparametric estimation of the regression function under random censorship model. *Statist. and Decisions*, 26 :1001–1020, 2008.

Z. Guessoum and E. Ould Saïd. Kernel regression uniform rate estimation for censored data under alpha-mixing condition. *Elect. J. of statist.*, 4 :117–132, 2010.

Z. Guessoum and E. Ould Saïd. Central limit theorem for the kernel estimator of the regression function for censored time series. *J. of Nonparametric Statist.*, 24 :379–397, 2012.

W. Hardle. *Applied nonparametric regression*. Cambridge Univ. Press. London., 1990.

K. Hirose and H. Masuda. Robust relative error estimation. *Entropy*, 20(632) :24, 2018.

D. H. Hu. Local least product relative error estimation for varying coefficient multiplicative regression model. *Acta. Math. Appl. Sinica.*, 35 :274–286, 2019.

P. J. Huber. *Robust statistics*. Wiley Series in Probability and Statistics, 1981.

L. A. Ibragimov. Some limit theorems for stationary processes. *Theory. Probab. Application*, 7 :349–382, 1962.

D. A. Jones. Nonlinear autoregressive processes. *Proc. Roy. Soc. London A*, 360 :71–95, 1978.

M. C. Jones, H. Park, K. I. Shin, S. K. Vines, and S. O. Jeong. Relative error prediction via kernel regression smoothers. *Journal of Statist. Plann. and Infer.*, 138 :2887–2898, 2008.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *J. Amer. Stat. Assoc.*, 53 :458–481, 1958.

K. Kebabi and F. Messaci. Rate of the almost complete convergence of a kernel regression estimate with twice censored data. *Statist. and Pobab. Letters*, 82 :1908–1913, 2012.

S. Khardani. Relative error prediction for twice censored data. *Math. Methods of Statist.*, 28(4) :291–306, 2019.

S. Khardani and Y. Slaoui. Nonparametric relative regression under random censorship model. *Stat. and Probab. Letters*, 151 :116–122, 2019.

T. M. Khoshgoftaar, B. B. Bhattacharyya, and G. D. Richardson. Prediction software errors, during development, using nonlinear regression models : comparative study. *IEEE Trans. Reliab.*, 41 :390–395, 1992.

H. T. Kim and Y. K. Truong. Nonparametric regression estimates with censored data : local linear smoothers and their applications. *Biometrics.*, 54 :1434–1444, 1998.

J. P. Klein and M. L. Moeschberger. *Survival analysis : techniques for censored and truncated data*. Springer-Verlag New York, 2004.

M. Köhler, K. Máthè, and M. Pintër. Prediction from randomly right censored data. *J. Multivar. Anal.*, 80 :73–100, 2002.

H. Koul, V. Susarla, and T. Van Ryzin. Regression analysis with randomly right-censored data. *Ann. Statist.*, 9 :1276–1288, 1981.

J. P. Lecoutre and E. Ould Saïd. Convergence of the conditional kaplan-meier estimate under strong mixing. *J. of Statist. Planning and Inference*, 44(3) :359–369, 1995.

M. Lemdani and E. Ould Saïd. Nonparametric robust regression estimation for censored data. *Statist. Papers*, 58 :505–525, 2017.

Y. Lin and K. Chen. Efficient estimation of the censored linear regression model. *Biometrika.*, 100 :525–530, 2013.

S. R. Lipsitz and J. G. Ibrahim. Estimation with correlated censored survival data with missing covariates. *Biostatistics*, 1 :315–327, 2000.

M. Loève. *Probability theory*. Springer-Verlag. New York, 1963.

O. Lopez. Nonparametric estimation of the multivariate distribution function in a censored regression model with applications. *Comm. Statist. Theory and Methods*, 40 : 2639–2660, 2011.

O. Lopez, V. Patilea, and I. Van Keilegom. Single index regression models in the presence of censoring depending on the covariates. *Bernoulli*, 19 :721–747, 2013.

E. Masry. Recursive probability density estimation for weakly dependent stationary processes. *IEEE Trans. Inform. theory*, 32 :254–267, 1986.

R. B. Mazess, W. W. Peppler, and M. Gibbons. Total body composition by dualphoton (153gd) absorptiometry. *American J. of clinical Nutrition.*, 40 :834–839, 1984.

E. A. Nadaraya. On estimating regression. *Theor. Probab. Appl.*, 9 :141–142, 1964.

S. C. Narula and J. F. Wellington. Prediction, linear regression and the minimum sum of relative errors. *Technometrics,*, 19 :185–190, 1977.

E. Ould Saïd and M. Lemdani. Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. *Ann. Inst. Statist. Math.*, 58 : 357–378, 2006.

T. Ozaki. Nonlinear time series models for nonlinear random vibrations. Technical report, Univ. of Manchester, 1979.

H. Park and L. A. Stefanski. Relative error prediction. *Statist. & Probab. Lett.*, 40 : 227–236, 1998.

E. Parzen. On estimating of the probability density function and mode. *Ann. Math. Statist.*, 33 :1065–1076, 1962.

V. Patilea and J. M. Rolin. Product limit estimators of the survival function with twice censored data. *Ann. Statist.*, 34(2) :925–938, 2006.

D. Pollard. *Convergence of stochastic processes.* Springer Verlag. Berlin., 1984.

B. L. S. Prakasa Rao. *Nonparametric Functional Estimation*. Z. W. Birnbaum E. Lukacs, 1983.

C. R. Rao. *A linear statistical inference and its applications*. Wiley. New-York, 1965.

E. Rio. Theorie asymptotique des processus aléatoires faiblement dépendants. *Math.*, 42 :43–47, 2000.

M. Rosenblatt. Remark on some nonparametric estimates of density function. *Ann. Math. Statist.*, 27 :832–837, 1956a.

M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U. S. A.*, 42 :43–47, 1956b.

B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

J. C. Stone. Consistence nonparametric regression. *Ann. Statist.*, 5.4 :595–645, 1977.

W. Stute. Censorship when covariables are present. *J. Multivariate Anal.*, 45 :89–103, 1993.

W. Stute. The central limit theorem under random censorship. *Ann. Statist.*, 23 :422–439, 1995.

W. Stute. Distributional convergence under random censorship when covariables are present. *Scand. J. Statist.*, pages 461–471, 1996.

W. Stute and J. L. Wang. The strong law under random censorship. *Ann. Statist.*, 21 : 1591–1607, 1993.

A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009.

S. Volgushev and H. Dette. Nonparametric quantile regression for twice censored data. *Bernoulli*, 19(3) :748–779, 2013.

V. A. Volkonskii and Y. A. Rozanov. Some limit theorem for random functions. *I. Theory Probab. Appl.*, 4 :178–197, 1959.

J. F. Wang, W. M. Ma, G. F. Fan, and L. M. Wen. Local linear quantile regression with truncated and dependent data. *Statist. and Probab. Letters*, 96 :332–340, 2015.

G. S. Watson. Smooth regression analysis. *Sankhyà*, 26 :359–372, 1964.