



Estimateurs fonctionnels récurrents et leurs applications à la prévision

Aboubacar Amiri

► To cite this version:

Aboubacar Amiri. Estimateurs fonctionnels récurrents et leurs applications à la prévision. Mathématiques générales [math.GM]. Université d'Avignon, 2010. Français. NNT : 2010AVIG0407 . tel-00565221v2

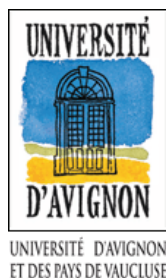
HAL Id: tel-00565221

<https://theses.hal.science/tel-00565221v2>

Submitted on 22 Apr 2011

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.



ACADÉMIE D'AIX-MARSEILLE
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

THÈSE DE DOCTORAT

présentée à l'Université d'Avignon et des Pays de Vaucluse
pour l'obtention du grade de Docteur

Spécialité : **Mathématiques**

Option : **Statistiques**

*Estimateurs fonctionnels récurrents et leurs
applications à la prévision*

par

Aboubacar AMIRI

Soutenue publiquement le 6 décembre 2010 devant un jury composé de :

M. A. BERLINET	Professeur, Université de Montpellier 2	Rapporteur
M ^{lle} . D. BLANKE	Professeur, Université d'Avignon	Directrice
M. D. BOSQ	Professeur, Université Paris 6	Examineur
M ^{me} . V. MAUME-DESCHAMPS	Professeur, Université Lyon 1	Examinatrice
M. P. SARDA	Professeur, Université Toulouse 3	Rapporteur
M. R. SENOSSI	Directeur de recherche, INRA Avignon	Examineur

École Doctorale ED 536 Sciences et Agrosiences
Laboratoire d'Analyse Non Linéaire et Géométrie (EA 2151)

Remerciements

Tout au long de mon parcours universitaire, et particulièrement ces trois dernières années, j'ai eu la chance de rencontrer des personnes formidables, sur qui j'ai pu compter et qui m'ont aidé à réaliser ce travail. Je tiens à les saluer et les remercier.

Je voudrais, en tout premier lieu, exprimer ma profonde reconnaissance à ma directrice de thèse, le professeur Delphine Blanke, d'avoir accepté de m'initier à la recherche et diriger ce travail. Comme tous ceux qui ont eu la chance d'être sous sa direction, j'ai pu constater à quel point elle est attachante, attentionnée et profondément gentille. Elle restera pour moi un modèle, pour ses grandes compétences scientifiques, pour son exigence de la clarté, sa franchise et ses qualités humaines. Je lui adresse mes remerciements les plus chaleureux, d'avoir accepté de me prendre sous son aile, pour tout ce qu'elle m'a appris et pour ses encouragements. J'espère dans l'avenir pouvoir transmettre ne serait-ce qu'une partie de tout ce qu'elle m'a appris, la meilleure façon selon moi de lui rendre hommage.

Je suis très honoré que les professeurs Alain Berlinet et Pascal Sarda aient accepté de rapporter cette thèse. Je les remercie vivement pour le travail que cela leur a demandé. Leurs remarques pertinentes et leurs commentaires précieux m'ont permis de préciser et d'améliorer un certain nombre de points. Les échanges que nous avons pu avoir avec Pascal Sarda, lors de mon séjour en Allemagne, ainsi que lors de mon passage à Toulouse ont été particulièrement enrichissants.

Je remercie chaleureusement les professeurs Denis Bosq et Véronique Maume-Deschamps, ainsi que Rachid Senoussi de me faire l'honneur de faire partie de mon jury et d'avoir ainsi accepté de juger mon travail. Je suis particulièrement sensible à la présence de monsieur Bosq dans mon jury, qui m'a donné le goût de la statistique non paramétrique lors de mon passage à Paris 6. Merci à Véronique Maume-Deschamps et Rachid Senoussi pour leurs accueils chaleureux à l'ISFA et à l'INRA, ainsi que pour les multiples échanges constructifs que j'ai eus avec l'une comme avec l'autre. J'adresse à travers eux un salut aux équipes de l'ISFA (du plot 1 au plot 4) et du laboratoire BioSp pour leur chaleur et bonne humeur.

J'exprime mes remerciements sincères à Messieurs Philippe Bolle (Université d'Avi-

gnon), Denis Allard (INRA Avignon), Alois Kneip (Université de Bonn) et Paul Deheuvels (Université Paris 6), d'avoir mis à ma disposition les moyens matériels, qui m'ont permis de mener ce travail dans des bonnes conditions. Mes remerciements vont également à l'endroit des professeurs et maîtres de conférences du LANLG, d'avoir facilité mon installation au sein de leur équipe. Je n'oublie pas non plus les professeurs et maîtres de conférences du LSTA d'avoir assuré ma formation en Master, et d'avoir été pour moi des interlocuteurs concernés.

Je mesure la chance que j'ai eue d'être enseignant pendant 2 ans dans le département STID de l'IUT d'Avignon et d'avoir ainsi fait parti d'une équipe aussi sympathique et dynamique. Un énorme merci à tous mes anciens collègues notamment Edith Gabriel pour ses conseils et ses efforts permanents pour que je me sente bien au sein de l'équipe.

Cette thèse est financée par le ministère français des affaires étrangères, je tiens donc à remercier le service de coopération et d'action culturelle de l'ambassade de France à Moroni d'avoir soutenu mon projet, ainsi qu'à Égide d'avoir remarquablement facilité mon installation en France.

À mes amis et collègues qui m'ont soutenus et avec qui nous avons partagé des moments inoubliables. Merci donc aux jeunes docteurs, passés ou à venir, notamment : Amadou, Angélique, Aurélie, Boris, Céline, Claire, Clara, François-Xavier, Kaouthar, Mehdi, Mouna, Mohammed, Olivier, Véronique, Walid.

Je salue la communauté comorienne d'Avignon, avec une pensée particulière à Abdallah et sa femme, de m'avoir ouvert leur portes, sans oublier Samia, Asmahane, Ahmada et les autres.

Que mes chers parents enfin, situés à 10.000 kilomètres d'ici, pour toute la curiosité à laquelle ils n'ont eu de cesse pour m'éveiller, pour toute leur volonté de vouloir me pousser le plus loins possible dans mes études, pour tout ce qu'ils m'ont apporté et que je ne saurais leur rendre trouvent ici un bien modeste part de ma profonde reconnaissance !

Et enfin, un petit clin d'oeil à Faou, qui a la lourde tâche de supporter toujours mes caprices. Merci pour ton soutien et ta patience.

À Mohamed, mon petit frère disparu brusquement en 2002.

Résumé

Nous nous intéressons dans cette thèse aux méthodes d'estimation non paramétriques par noyaux récurrents ainsi qu'à leurs applications à la prévision. Nous introduisons dans un premier chapitre une famille d'estimateurs récurrents de la densité indexée par un paramètre $\ell \in [0, 1]$. Leur comportement asymptotique en fonction de ℓ va nous amener à introduire des critères de comparaison basés sur les biais, variance et erreur quadratique asymptotiques. Pour ces critères, nous comparons les estimateurs entre eux et aussi comparons notre famille à l'estimateur non récurrent de la densité de Parzen-Rosenblatt. Ensuite, nous définissons à partir de notre famille d'estimateurs de la densité, une famille d'estimateurs récurrents à noyau de la fonction de régression. Nous étudions ses propriétés asymptotiques en fonction du paramètre ℓ . Nous utilisons enfin les résultats obtenus sur l'estimation de la régression pour construire un prédicteur non paramétrique par noyau. Nous obtenons ainsi une famille de prédicteurs non paramétriques qui permettent de réduire considérablement le temps de calcul. Des exemples d'application sont donnés pour valider la performance de nos estimateurs.

Abstract

The aim of this thesis is to study methods of nonparametric estimation based on recursive kernel and their applications to forecasting. We introduce in the first chapter a family of recursive density estimators indexed by a parameter $\ell \in [0, 1]$. We study their asymptotic behavior according to ℓ , and then we introduce criteria of comparison based on bias, variance and asymptotic quadratic error. For these criteria, we compare our estimators in terms of ℓ , and also compare our family to the non-recursive density estimator of Parzen-Rosenblatt. As for density, we define a family of recursive kernel estimators of regression function. We study its asymptotic properties according to the parameter ℓ . Finally, results of regression estimation are applied to define a family of nonparametric predictors that reduce considerably the computing time and examples of application are given to validate the performance of our methods.

Lexique

Abréviations

p.s.	presque sûrement.
i.i.d	indépendantes et identiquement distribuées.
G.F.M.	Géométriquement Fortement Mélangeant.
E.Q.M.	Erreur Quadratique Moyenne.
M.B.	Mouvement Brownien.
B.C.	Biais au Carré.
L.L.I.	Loi du Logarithme Itéré.

Notations

$\mathbf{1}_A$	Indicatrice sur A .
A°	Intérieur de l'ensemble A .
A^c	Complémentaire de A .
λ_d	Mesure de Lebesgue de $\mathbb{R}^d$.
$\sigma(X)$	σ -algèbre des événements engendrés par X .
$c(f)$	Ensemble de points de continuité de f .
$f * g$	Produit de convolution de f et g .
$\overline{\lim}$	(resp. $\underline{\lim}$) Limite supérieure (resp. Limite inférieure).
$\ \cdot\ _\infty$	Norme infinie (ou norme sup).
$\ \cdot\ _2$	Norme L_2 .
$[x]$	partie entière d'un réel x .
$\ln \ln x$	est égale à $\ln^+ \ln^+ x$ avec $\ln^+ x := \ln \max(x, e)$.
x^+	est égale à $\max(x, 0)$.
Cste	constante positive dont la connaissance de la valeur exacte n'est pas importante.

Nous utilisons également les notations habituelles suivantes :

Soient $(u_n)_{n \geq 1}$ et $(v_n)_{n \geq 1}$ deux suites réelles positives.

* $u_n = o(v_n)$ (v_n est prépondérante sur u_n) : pour tout réel $\varepsilon > 0$ on a $u_n \leq \varepsilon v_n$ pour n assez grand.

* $u_n = O(v_n)$ (v_n domine u_n) : il existe un réel $\lambda > 0$ tel que $u_n \leq \lambda v_n$ pour n assez grand.

* $u_n \sim v_n$ (v_n et u_n sont asymptotiquement équivalentes) : $u_n/v_n \rightarrow 1$ lorsque $n \rightarrow \infty$.

Table des matières

1	Introduction générale	1
1.1	Présentation	1
1.2	Chapitre 2 : Estimateurs récursifs de la densité.	3
1.3	Chapitre 3 : Estimateurs récursifs de la régression.	6
1.4	Chapitre 4 : Application à la prévision et simulations	8
2	Estimateurs récursifs de la densité	11
2.1	Cadre d'étude	11
2.2	Estimation	12
2.2.1	Présentation de l'estimateur et des hypothèses	12
2.2.2	Estimation dans le cas iid	16
2.2.3	Estimation dans le cas dépendant	20
2.3	Comparaison d'estimateurs	24
2.4	Preuves	28
2.4.1	Preuve du Théorème 2.2.1	28
2.4.2	Preuve du Théorème 2.2.4	30
2.4.3	Preuve du Théorème 2.2.6	33
2.4.4	Preuve du Théorème 2.2.7	37
2.4.5	Preuve de la Proposition 2.2.9	42
2.4.6	Preuve du Théorème 2.2.10	45
2.4.7	Preuve du Théorème 2.3.1	49
2.4.8	Preuve du Théorème 2.3.2	51
3	Estimateurs récursifs de la régression	57
3.1	Cadre d'étude	57
3.1.1	Notations et hypothèses	58
3.2	Convergence presque sûre	59

TABLE DES MATIÈRES

3.2.1	Cas iid	60
3.2.2	Cas dépendant	61
3.3	Convergence en moyenne quadratique	62
3.4	Normalité asymptotique	63
3.5	Preuves	64
3.5.1	Preuve du Théorème 3.2.1	64
3.5.2	Preuve de Théorème 3.2.2	68
3.5.3	Preuve du Théorème 3.2.3	71
3.5.4	Preuve du Théorème 3.3.1	74
3.5.5	Preuve du Théorème 3.4.1	83
4	Application à la prévision et simulations	89
4.1	Prévision non paramétrique	89
4.1.1	Cadre d'étude	90
4.1.2	Présentation et étude de nos prédicteurs	90
4.2	Simulations	93
4.2.1	Estimation de la densité	94
4.2.2	Estimation de la régression	102
4.2.3	Prévision	114
A	Annexes	121
A.1	Lemmes utiles	121
A.1.1	Lemme de Bochner	121
A.1.2	Lemme de Toeplitz	122
A.1.3	Lemme de Volkonskii et Rozanov	122
A.1.4	Lemme de Borel-Cantelli	122
A.1.5	Lemme de couplage de Rio	123
A.2	Théorème d'approximation de Jain et al.	123
A.3	Loi du Logarithme Itéré pour un Mouvement Brownien	124
A.4	Quelques inégalités importantes	124
B	Table des équations, table des figures, liste des tableaux	125
	Bibliographie	133

Chapitre 1

Introduction générale

Sommaire

1.1	Présentation	1
1.2	Chapitre 2 : Estimateurs récursifs de la densité.	3
1.3	Chapitre 3 : Estimateurs récursifs de la régression.	6
1.4	Chapitre 4 : Application à la prévision et simulations	8

1.1 Présentation

Considérons un processus à temps discret $\{X_k : k \geq 1\}$. On suppose avoir observé les n réalisations $X_1, \dots, X_n$, à partir desquelles on souhaite prédire la valeur de X_{n+h} (prévision à l'horizon $h, h \in \mathbb{N}^*$). Ce problème est très simple dans sa formulation mais il en appelle d'autres, notamment les interrogations suivantes : quelle méthode de prévision peut-on utiliser ? Quel critère peut-on choisir pour mesurer la performance de notre prévision ? Plusieurs méthodes existent dans la littérature, cependant il n'existe pas de méthode universellement meilleure que les autres.

La méthode de prévision par k -moyennes mobiles, par exemple, consiste à prendre comme prévision la moyenne des observations des k périodes précédentes. La prévision est ainsi renouvelée de période en période. Cette méthode est simple d'utilisation avec pour avantage d'atténuer suffisamment les fluctuations de la série tout en préservant son allure générale, cependant elle s'avère moins performante lorsque l'on prend un nombre élevé de données. La méthode du lissage exponentiel, quant à elle, prend en compte la prévision de la période antérieure. À cette prévision, on augmente l'écart subi, pondéré d'un coefficient a compris

entre 0 et 1. Ceci se traduit, pour $h = 1$, par :

$$\widehat{X}_{n+1} = a\widehat{X}_n + (1-a)X_{n+1} \quad (0 < a < 1). \quad (1.1.1)$$

On constate à partir de (1.1.1) que la valeur prédite à l'instant $(n+1)$ est une moyenne pondérée entre la valeur estimée faite en n et la dernière observation de la série. L'avantage de cette récursivité est que l'on n'a pas à relisser de nouveau le processus lorsqu'une nouvelle observation s'ajoute à la série. Cela n'est pas négligeable car on réduit alors considérablement le temps de calcul, nous reviendrons d'ailleurs sur ce point qui constitue le point central de notre travail. Tout comme la méthode de prévision par moyennes mobiles, le lissage exponentiel est simple d'utilisation et facilement compréhensible mais son principal inconvénient réside sur le choix de la constante de lissage a . Elle est également moins efficace pour prédire des longues séries. Pour des prévisions à long terme on privilégiera la méthode de Box-Jenkins [19]. Cette dernière est ainsi utilisée pour la prévision des processus qui vérifient une forme prédéfinie comme les modèles ARMA ou ARIMA. Encore plus robustes que les méthodes de Box-Jenkins (dans la mesure où aucune forme particulière n'est imposée au processus) et ayant l'avantage d'une mise en pratique très facile, les prévisions dites non paramétriques sont plus récemment apparues pour tenter d'apporter un nouveau regard sur ce problème. Leur principe repose sur le fait que le problème de la prévision peut être vu comme un cas particulier de l'estimation de la régression dans le sens où si l'on suppose que le processus est Markovien d'ordre k et strictement stationnaire, alors le meilleur prédicteur probabiliste de X_{n+h} est donné par l'espérance conditionnelle :

$$E(X_{n+h} | X_n, \dots, X_{n-k+1})$$

que l'on peut estimer par :

$$\widehat{X}_{n+h} = \widehat{r}_{n-k-h+1}(X_{n-k+1}, \dots, X_n),$$

où $\widehat{r}_{n-k-h+1}(x)$ est l'estimateur à noyau de la régression basé sur les observations :

$$((X_i, \dots, X_{i+k-1}), (X_{i+k-h-1})) \text{ pour } i = 1, \dots, n - k - h + 1.$$

Carbon et Francq [21] montrent, par des exemples numériques, sur plusieurs types de séries chronologiques, l'efficacité des méthodes non paramétriques basées sur des estimateurs à noyau vis à vis de celles de Box-Jenkins.

L'objectif principal de cette thèse est d'améliorer encore les performances des prédicteurs non paramétriques par noyau, en réduisant leurs temps de calcul par l'utilisation de noyaux

récursifs.

Ayant en vue ce problème de la prévision, notre travail sera divisé en trois parties : estimation de la densité, de la régression et application à la prévision par des méthodes récursives.

1.2 Chapitre 2 : Estimateurs récursifs de la densité.

Le deuxième chapitre de cette thèse est consacré à l'estimation récursive de la densité de probabilité d'une suite de variables aléatoires de même loi (non nécessairement indépendantes). Soit donc $(X_t, t \in \mathbb{N})$ un processus à temps discret, tel que les X_t sont des vecteurs de $\mathbb{R}^d$ ayant la même loi, de densité de probabilité inconnue f .

L'estimation de la densité est un sujet qui a donné lieu à un grand nombre de travaux. Son champ d'application est très vaste et couvre divers domaines, comme l'analyse de la régression, des séries chronologiques et la théorie de la fiabilité. Par exemple Singh [85] utilise des résultats sur l'estimation de la densité, en particulier pour estimer l'information de Fisher.

Les principales méthodes non-paramétriques pour l'estimation de la densité sont la méthode du noyau introduite par Rosenblatt [81] et Parzen [69], la méthode des séries orthogonales étudiée entre autre par Schwartz [84] et Watson [92] et la méthode de l'histogramme introduite par Graunt puis développée par Scott, Tran [88], Carbon et Tran [22]. Parmi l'ensemble de ces estimateurs, l'un des plus utilisés reste l'estimateur à noyau défini par :

$$f_n^{\text{PR}}(x) := \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \forall x \in \mathbb{R}^d,$$

où K est un noyau défini dans $\mathbb{R}^d$, borné et intégrable par rapport à la mesure de Lebesgue et h_n une suite réelle appelée paramètre de lissage, tendant vers zéro à l'infini. D'autres conditions complémentaires sur K et h_n sont nécessaires pour l'étude de cet estimateur. L'étude de $f_n^{\text{PR}}(x)$ a donné lieu à une vaste littérature statistique, pour une représentation globale des résultats obtenus sur cet estimateur, notamment dans le cadre de données dépendantes, nous renvoyons aux livres de Prakasa-Rao [72], Bosq [14], Bosq, Lecoutre [17], Bosq et Blanke [15].

Nous nous intéressons aux versions récursives de $f_n^{\text{PR}}(x)$. Une première forme a été introduite par Wolverton et Wagner [97] :

$$f_n^{\text{WW}}(x) := \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right), \forall x \in \mathbb{R}^d.$$

Cet estimateur possède en effet, les mêmes propriétés asymptotiques que $f_n^{\text{PR}}(x)$ et peut également s'écrire sous la forme :

$$f_n^{\text{WW}}(x) = R_n \left[x, f_{n-1}^{\text{WW}}(x), X_n \right],$$

avec :

$$R_n(x, a, b) = \frac{n-1}{n} a + \frac{1}{h_n^d} K\left(\frac{x - b}{h_n}\right).$$

De nombreuses variantes récursives ont également été proposées et étudiées depuis. En particulier, Deheuvels [32, 33] s'est intéressé à la famille générale suivante :

$$f_n^H(x) := \left[\sum_{i=1}^n h_i H(h_i) \right]^{-1} \sum_{i=1}^n H(h_i) K\left(\frac{x - X_i}{h_i}\right).$$

L'estimateur obtenu dans l'expression de $f_n^H(x)$, pour $H(u) = 1$, est connu dans la littérature sous le nom d'estimateur de Deheuvels et s'écrit en dimension d sous la forme :

$$f_n^{\text{DHV}}(x) = \frac{1}{\sum_{i=1}^n h_i^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_i}\right).$$

Aussi, Wegman et Davies [95] étudient l'estimateur récursif suivant :

$$f_n^{\text{DW}}(x) := \frac{1}{n\sqrt{h_n^d}} \sum_{i=1}^n \frac{1}{h_i^{\frac{d}{2}}} K\left(\frac{x - X_i}{h_i}\right).$$

Leur idée consiste à partager le paramètre de lissage de l'estimateur à noyau habituel $f_n^{\text{PR}}(x)$ en deux puissances $\frac{1}{2}$. L'estimateur $f_n^{\text{DW}}(x)$ est asymptotiquement biaisé et son intégrale n'est pas égale à 1, mais il est très intéressant car une fois ces problèmes corrigés, sa variance asymptotique est plus petite que celle de $f_n^{\text{WW}}(x)$.

Dans cet même état d'esprit, nous proposons d'étudier dans cette thèse la famille paramétrique d'estimateurs récursifs à noyau définie par :

$$f_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}^d \quad (\ell \in [0, 1]) \quad (1.2.1)$$

qui correspond pour $d = 1$ au cas $H(u) = u^{-\ell}$ dans $f_n^H(x)$. Remarquons que la famille (1.2.1) contient les estimateurs les plus utilisés $f_n^{WW}(x)$, $f_n^{DHV}(x)$ et la version asymptotiquement non biaisée de $f_n^{DW}(x)$, qui correspondent respectivement aux cas $\ell = 0, \ell = 1$ et $\ell = 1/2$. Elle possède également l'avantage de s'exprimer sous une forme simple permettant la comparaison des différents estimateurs la composant.

La récursivité peut s'avérer cruciale lorsque l'on cherche à inférer sur des phénomènes qui évoluent dans le temps et qui nécessitent une mise à jour constante des estimations effectuées. En outre, les estimateurs récursifs peuvent s'avérer préférables aux versions non récursives du fait de leur plus faible variance asymptotique.

Dans le cas iid, Davies [30], Deheuvels [32, 33], Roussas [78], Wegman, Davies [95] et Wertz [96] étudient la famille $f_n^H(x)$ et les cas $\ell = 0$, $\ell = 1/2$ et $\ell = 1$. En particulier, en dimension $d = 1$, Deheuvels [32, 33] établit la convergence en moyenne quadratique de la famille $f_n^H(x)$ et donne des conditions nécessaires et suffisantes pour sa convergence presque sûre. Roussas [78], Wegman et Davies [95] établissent les vitesses de convergence presque sûre exactes dans les cas $\ell = 0$, $\ell = 1/2$ et $\ell = 1$. Aussi, Isogai [51] établit sous certaines conditions, la normalité asymptotique pour $\ell = 1$ dans le cas iid.

Dans le cas dépendant, seuls les cas $\ell = 1/2$ et $\ell = 1$ ont été étudiés dans la littérature. Les résultats sur la convergence en moyenne quadratique et la normalité asymptotique pour $\ell = 1/2$ et $\ell = 1$ sont établis par Masry [63], pour des processus stationnaires fortement mélangeants. La vitesse de convergence presque sûre ponctuelle pour $\ell = 1/2$ et $\ell = 1$ est étudiée par Takahata [86], Masry et Györfi [65], d'abord sous des conditions de ρ -mélangeance, ensuite pour $\ell = 1$, par Masry [64], pour des processus fortement mélangeants. Un résultat uniforme est également obtenu dans le cas $\ell = 1$ par Tran [87], sous des conditions de forte mélangeance. La normalité asymptotique pour $\ell = 1$ est également examinée par Lian et Baek [61] pour des suites de variables négativement associées. Les approches utilisées dans ces travaux, notamment pour la convergence en moyenne quadratique et la normalité asymptotique, ne se généralisent pas aisément, en dimension supérieure pour des valeurs plus petites de ℓ , alors qu'en particulier, le cas $\ell = 0$ est intéressant du fait de la faible variance de l'estimateur.

Nous présentons d'abord la famille d'estimateurs de la densité (1.2.1), étudions ses biais, variance et erreur quadratique moyenne (EQM) asymptotiques, ainsi que son comportement presque sûr pour des observations iid. Ensuite, nous généralisons ces résultats

à des processus fortement mélangeants et donnons la normalité asymptotique de la famille (1.2.1) dans ce cas. Enfin, nous comparons les estimateurs composant la famille (1.2.1) selon les valeurs de ℓ , ainsi qu'avec l'estimateur $f_n^{\text{PR}}(x)$, selon des critères de comparaison basés sur les biais, variance, EQM et la convergence presque sûr. Les preuves des résultats de ce chapitre sont regroupées en fin de chapitre.

1.3 Chapitre 3 : Estimateurs récursifs de la régression.

Dans le troisième chapitre, nous abordons le problème de l'estimation récursive de la régression. Nous considérons un processus stochastique bivarié $\{\zeta_t = (X_t, Y_t), t \in \mathbb{N}\}$ défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$, à valeurs dans $\mathbb{R}^d \times \mathbb{R}^{d'}$. À partir d'une suite de variables équidistribuées (X_n, Y_n) , nous cherchons à estimer une version de la fonction de régression définie par :

$$r(x) := \begin{cases} \mathbb{E}[m(Y_0) | X_0 = x] = \frac{\int_{\mathbb{R}^{d'}} m(y) f^*(x, y) dy}{f(x)} := \frac{\varphi(x)}{f(x)}, & \text{si } f(x) > 0 \\ \mathbb{E}m(Y_0), & \text{si } f(x) = 0, \end{cases}$$

où m est une fonction Borélienne de $\mathbb{R}^{d'}$ à valeurs dans $\mathbb{R}$ telle que $\omega \mapsto m^2(Y_t(\omega))$ soit P -intégrable et f^* est la densité de probabilité de ζ_n . Notons que la fonction m est choisie par le statisticien, les choix les plus courants étant les fonctions identité et les polynômes, pour l'estimation de la version usuelle de la régression et les moments conditionnels de Y_0 sachant X_0 .

Nadaraya [66] et Watson [92] proposent l'estimateur non récursif :

$$r_n^{\text{NW}}(x) := \frac{\sum_{i=1}^n m(Y_i) K_{h_n}(x, X_i)}{\sum_{i=1}^n K_{h_n}(x, X_i)}, \quad \forall x \in \mathbb{R}^d, \quad (1.3.1)$$

avec :

$$K_{h_n}(\cdot) := \frac{1}{h_n} K\left(\frac{\cdot}{h_n}\right).$$

Ainsi, on constate que K_{h_n} dépend du nombre n des observations. Dans certaines situations concrètes, la taille de l'échantillon est fluctuante, si la régression est estimée par la formule (1.3.1), une augmentation de cette taille, même de quelques observations, conduit à recalculer entièrement l'estimateur. Dans ce contexte multidimensionnel d'estimation de la régression, cela peut constituer une charge de calcul supplémentaire et une perte de temps non négligeable même pour des ordinateurs performants.

Les estimateurs récurrents de $r(x)$ considérés dans ce chapitre sont de la forme :

$$\frac{\sum_{i=1}^n m(Y_i) K_{h_i}^*(x, X_i)}{\sum_{i=1}^n K_{h_i}^*(x, X_i)},$$

et les plus populaires d'entre eux sont l'estimateur de Devroye et Wagner [38] et celui d'Ahmad et Lin [1], définis respectivement par :

$$r_n^{DW}(x) := \frac{\sum_{i=1}^n \frac{m(Y_i)}{h_i^d} K\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right)}$$

et

$$r_n^{AL}(x) := \frac{\sum_{i=1}^n m(Y_i) K\left(\frac{x - X_i}{h_i}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_i}\right)},$$

communément appelés estimateur récurrent et semi-récurrent respectivement. Ces deux estimateurs sont largement étudiés par Györfi et al. [49] dans le cas iid. La convergence au sens L^1 de $r_n^{DW}(x)$ est étudiée par Devroye et Wagner [38], la vitesse de convergence est établie par Krzyżak [57]. Ahmad et Lin [1] montrent la convergence uniforme presque sûre sans vitesse de l'estimateur $r_n^{AL}(x)$, pour des observations iid et dans Krzyżak et Pawlak [58, 59] sous des hypothèses légèrement différentes. Dans le cas dépendant, Wang et Liang [91] étudient la convergence uniforme presque sûre des versions tronquées des estimateurs $r_n^{DW}(x)$ et $r_n^{AL}(x)$, sous des conditions de φ -mélangeance. Roussas et Tran [79] établissent la normalité asymptotique de $r_n^{DW}(x)$ pour des processus fortement mélangeants.

Dans ce chapitre, nous introduisons et étudions une nouvelle famille générale d'estimateurs récurrents à noyau de la régression, définie par :

$$r_n^\ell(x) := \frac{\varphi_n^\ell(x)}{f_n^\ell(x)} \text{ où } \varphi_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{(1-\ell)d}} \sum_{i=1}^n \frac{m(Y_i)}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), \quad (1.3.2)$$

incluant les estimateurs populaires $r_n^{DW}(x)$ et $r_n^{AL}(x)$ qui correspondent respectivement aux cas $\ell = 1$ et $\ell = 0$.

Nous présentons d'abord le cadre de notre étude, et donnons notamment des hypothèses supplémentaires utilisées dans ce chapitre. Ensuite, nous étudions les vitesses de convergence presque sûre de (1.3.2) pour des observations iid, puis pour des processus

mélangeants. Enfin, nous étudions la convergence en moyenne quadratique et la normalité asymptotique de (1.3.2). Les preuves des résultats de ce chapitre sont présentées en fin de chapitre.

1.4 Chapitre 4 : Application à la prévision et simulations

Dans ce chapitre, nous appliquons les résultats des Chapitres 2 et 3 à la prévision d'un processus $(\xi_t, t \in \mathbb{N})$ observé jusqu'à un instant donné. Les domaines d'application sont nombreux en particulier en économie, ingénierie ou en contrôle aérien. On peut citer par exemple les séries correspondant à des températures moyennes journalières, à des cours de bourse, à des prix, à des taux, à des trajectoires d'avions etc... Les livres de Box, Jenkins [19], Pankratz [68], Gouriéroux et Monfort [47] présentent également plusieurs exemples. La prévision a fait l'objet d'une attention particulière chez de nombreux statisticiens, économistes, financiers etc., et différentes approches ont été proposées pour aborder cette question.

Nous nous intéressons à la prédiction de ξ_{n+1} sur la base des observations $\xi_1, \dots, \xi_n$, où le processus (ξ_t) est supposé Markovien d'ordre k et strictement stationnaire. Nous nous plaçons dans un cadre où la base des données contenant les ξ_t est mise à jour au fur et à mesure (la taille de l'échantillon passant ainsi progressivement de n à N , $N \geq n$) et l'on souhaite réactualiser notre prévision à chaque instant. La question est donc de faire les $N - n$ prévisions successives $\hat{\xi}_{n+p}$ à partir des $N - n$ suites d'observations $\xi_1, \dots, \xi_{n+p-1}$, respectivement, pour $p = 1, \dots, N - n$.

Pour illustrer nos propos, voici un exemple. Un des problèmes qui intéressent les contrôleurs aériens est de prédire les trajectoires pour différents types d'avions en vol (plus précisément leurs altitudes de vol), sur la base de leurs positions observées à partir du sol à des instants successifs, afin de résoudre d'éventuels conflits en vol. En fait, il faudra être capable de prédire en permanence l'altitude d'un avion, le plus précisément possible, jusqu'à la fin de son trajet en se basant sur l'historique (mise à jour au cours du vol) de ses altitudes antérieures. Le gain de temps de calcul s'avère crucial dans ce contexte.

Pour faire de la prévision, deux grandes familles de méthodes sont utilisées : les méthodes paramétriques et les méthodes non paramétriques. Dans le cadre paramétrique, on postule qu'une forme prédéfinie caractérise la série, et la prédiction est donc basée sur

l'existence d'un nombre fini de paramètres inconnus à estimer. Le cadre non paramétrique, s'affranchit de toute forme prédéfinie sur le modèle générant les ξ_t . Prédire ξ_{n+p} revient alors à chercher sa meilleure approximation sur la base des observations $\xi_1, \dots, \xi_{n+p-1}$. Le prédicteur naturel de ξ_{n+p} est ainsi donné par :

$$E(\xi_{n+p} | \xi_1, \dots, \xi_{n+p-1}).$$

Cette espérance conditionnelle est inconnue, mais puisque le processus (ξ_t) est Markovien, il suffit d'estimer :

$$E(\xi_{n+p} | \xi_{n+p-k}, \dots, \xi_{n+p-1})$$

par :

$$\hat{\xi}_{n+p} = r_{n+p-k-1}^\ell(\xi_{n+p-k}, \dots, \xi_{n+p-1}), \quad (1.4.1)$$

où $r_n^\ell(x)$ est l'estimateur récursif de la régression défini en (1.3.2) avec les choix :

$$X_i = (\xi_i, \dots, \xi_{i+k-1}) \text{ et } Y_i = \xi_{i+k}, \text{ pour } i = 1, \dots, n + p - k - 1.$$

Notons que d'autres prédicteurs non paramétriques à noyau basés sur le mode ou les quantiles conditionnels peuvent être considérés à la place de l'espérance conditionnelle. Ce type de prédicteurs sont étudiés entre-autres par Berline et al. [8, 7].

Plusieurs travaux, parmi lesquels on peut citer Carbon et Delecroix [20], Carbon et Francq [21], Gannoun [44] mettent en concurrence les méthodes paramétriques et non paramétriques. Tous ces auteurs concluent à la supériorité des secondes vis à vis des premières.

Les travaux sur la prévision non paramétrique basée sur l'estimateur à noyau sont nombreux, certains d'entre eux peuvent être trouvés dans les documents de synthèse suivants : Bosq [13, 14], Bosq et Lecoutre [18], Bosq et Blanke [15], Collomb [26, 27] et Robinson [77]. L'ensemble des travaux cités ci-dessus se situent dans un cadre purement non récursif.

Pour réduire les temps de calcul, nous considérons dans ce chapitre, le prédicteur (1.4.1) basé sur l'estimateur à noyau récursif. Son comportement asymptotique presque sûr est étudié dans ce chapitre. Pour conclure nous présentons, enfin de chapitre, quelques éléments de simulation mettant en avant l'efficacité des méthodes récursives du point de

vue du temps de calcul. Enfin nous concluons cette thèse par une appendice regroupant des résultats et outils classiques utilisés dans cette thèse et donnons quelques perspectives de recherche à venir.

Chapitre 2

Estimateurs récurrents de la densité

Sommaire

2.1	Cadre d'étude	11
2.2	Estimation	12
2.2.1	Présentation de l'estimateur et des hypothèses	12
2.2.2	Estimation dans le cas iid	16
2.2.3	Estimation dans le cas dépendant	20
2.3	Comparaison d'estimateurs	24
2.4	Preuves	28
2.4.1	Preuve du Théorème 2.2.1	28
2.4.2	Preuve du Théorème 2.2.4	30
2.4.3	Preuve du Théorème 2.2.6	33
2.4.4	Preuve du Théorème 2.2.7	37
2.4.5	Preuve de la Proposition 2.2.9	42
2.4.6	Preuve du Théorème 2.2.10	45
2.4.7	Preuve du Théorème 2.3.1	49
2.4.8	Preuve du Théorème 2.3.2	51

2.1 Cadre d'étude

Soit $(X_t, t \in \mathbb{N})$ un processus à temps discret défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$, tel que les X_t sont des vecteurs de $\mathbb{R}^d, d \geq 1$, ayant la même loi, de densité de probabilité inconnue f relativement à la mesure de Lebesgue λ_d .

Nous rappelons quelques définitions utiles dans ce chapitre en particulier, mais aussi dans l'ensemble de ce travail. Nous commençons par le coefficient de mélange fort introduit par Rosenblatt [80].

Définition 2.1. Soit $(\Omega, \mathcal{A}, \mathcal{P})$ un espace probabilisé. Le coefficient de mélange fort entre deux sous-tribus $\mathcal{B}_1$ et $\mathcal{B}_2$ de $\mathcal{A}$ est défini par :

$$\alpha(\mathcal{B}_1, \mathcal{B}_2) := \sup_{B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2} |P(B_1 \cap B_2) - P(B_1)P(B_2)|.$$

Nous donnons ensuite une définition de la notion de forte mélangeance pour un processus.

Définition 2.2. Un processus $(X_t, t \in \mathbb{N})$ est fortement mélangeant ou α -mélangeant si le coefficient de mélange fort de X défini pour tout $u \geq 0$, par :

$$\alpha(u) := \sup_{t \in \mathbb{N}} \alpha(\sigma(X_s, s \leq t), \sigma(X_s, s \geq t + u)), \quad (2.1.1)$$

est tel que $\alpha(u) \downarrow 0$ lorsque $u \rightarrow +\infty$.

Pour mesurer la dépendance de nos observations, nous utiliserons le coefficient de 2- α -mélange $\alpha^{(2)}$ légèrement moins restrictive que α .

Définition 2.3. Un processus $(X_t, t \in \mathbb{N})$ est 2-fortement mélangeant ou 2- α -mélangeant si le coefficient de 2-mélange fort de X défini, pour tout $u \geq 0$, par :

$$\alpha^{(2)}(u) := \sup_{t \in \mathbb{N}} \alpha(\sigma(X_t), \sigma(X_t + u)), \quad (2.1.2)$$

est tel que $\alpha^{(2)}(u) \downarrow 0$ lorsque $u \rightarrow +\infty$.

2.2 Estimation

2.2.1 Présentation de l'estimateur et des hypothèses

Pour estimer la densité f , nous proposons la famille paramétrique d'estimateurs récursifs à noyau (1.2.1) définie ci-dessous :

$$f_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}^d \quad (\ell \in [0, 1]).$$

Notre famille d'estimateurs peut se calculer de manière récursive par :

$$f_{n+1}^\ell(x) = \frac{\sum_{i=1}^n h_i^{d(1-\ell)}}{\sum_{i=1}^{n+1} h_i^{d(1-\ell)}} f_n^\ell(x) + K_{n+1}^\ell(x - X_{n+1}), \quad (2.2.1)$$

avec : $K_i^\ell(\cdot) := \frac{1}{h_i^{d\ell} \sum_{j=1}^i h_j^{d(1-\ell)}} K\left(\frac{\cdot}{h_i}\right).$

La construction de $f_n^\ell(x)$ est basée sur la généralisation de l'idée de Wegman et Davies [95] qui consiste à partager le paramètre de lissage en deux puissances. Nous avons donc pensé à partager la fenêtre en deux puissances ℓ et $1 - \ell$, $\ell \in [0, 1]$. Mais contrairement à eux, pour avoir des estimateurs asymptotiquement sans biais et d'intégrale égale à 1, nous les normalisons par la quantité :

$$B_{n,d(1-\ell)} := \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^{d(1-\ell)}.$$

On a donc :

$$f_n^\ell(x) = \frac{B_{n,d(1-\ell)}^{-1}}{nh_n^{d(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right).$$

Dans cette section, nous supposons qu'en plus d'être équidistribuées, les variables X_i sont indépendantes et nous donnons les biais, variance et EQM asymptotiques exacts de $f_n^\ell(x)$, en fonction de ℓ . Nous étudions ensuite la convergence presque sûre ponctuelle de notre famille d'estimateurs récursifs.

Pour établir nos résultats, nous avons besoin de faire quelques hypothèses. Nous supposons que la densité f est une fonction appartenant à $C_d^2(b)$ ($b > 0$), où $C_d^2(b)$ désigne l'ensemble des fonctions $\psi : \mathbb{R}^d \mapsto \mathbb{R}$ telles que $\psi^{(2)}$ existe pour toute dérivée partielle d'ordre 2, continue et $\|\psi^{(2)}\| \leq b$.

La condition imposée à la densité f est classique dans ce domaine, utilisée par exemple par Bosq [14] pour l'estimation de la densité avec l'estimateur de Parzen-Rosenblatt.

Nous considérons également des noyaux K vérifiant les hypothèses :

Hypothèses H.1.

- (i) : $K : \mathbb{R}^d \mapsto \mathbb{R}$ est une densité de probabilité, strictement positive, symétrique et bornée ;

(ii) :

$$\lim_{\|x\| \rightarrow +\infty} \|x\|^d K(x) = 0, \quad \forall x \in \mathbb{R}^d;$$

(iii) :

$$\int_{\mathbb{R}^d} |v_i v_j| K(v) dv < \infty, \quad i, j = 1, \dots, d.$$

Les hypothèses **H.1** sont classiques en estimation non paramétrique, vérifiées en particulier par les noyaux d'Epanechnikov, Gaussien etc.

La fenêtre h_n vérifie les conditions suivantes :

Hypothèses H.2.

(i) : $h_n \downarrow 0$ et $nh_n^{d+2} \rightarrow \infty$ lorsque $n \rightarrow \infty$;

(ii) : Pour tout $r \in]-\infty, d+2]$,

$$B_{n,r} := \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n} \right)^r \rightarrow \beta_r < \infty \text{ lorsque } n \rightarrow \infty.$$

La condition **H.2(i)** est classique en estimation récursive. La condition **H.2(ii)** est très utile dans nos calculs, et est également propre à la récursivité. Elle est souvent utilisée dans la littérature notamment par Yamato [98] pour $r = -1$, Wegman et Davies [95] pour $r = \frac{1}{2}$, Masry [63] pour $r = 1, \dots, k$ où k désigne le nombre de dérivées de la densité f . Le cas où $r \leq 0$ est utilisé par Samanta et Mugisha [82] puis repris par Isogai [52].

Enfin, il est important de noter que la condition **H.2(ii)** a pour but d'assouplir la complexité des calculs liée au noyaux récursifs. En effet, dans le cas non récursif, pour chaque n fixé, le processus $\left\{ \frac{1}{h_n^d} K\left(\frac{x-X_t}{h_n}\right), t \in \mathbb{N} \right\}$ est stationnaire si $(X_t, t \in \mathbb{N})$ est stationnaire, ceci offre beaucoup de simplifications dans les calculs, alors que pour les estimateurs récursifs ces simplifications ne sont pas possibles, car la stationnarité du processus $(X_t, t \in \mathbb{N})$ n'implique pas celle du processus $\left\{ \frac{1}{h_i^d} K\left(\frac{x-X_t}{h_i}\right), t \in \mathbb{N} \right\}$.

Donnons maintenant des choix de fenêtres typiques vérifiant les hypothèses **H.2**.

1. Si :

$$h_n = C_n n^{-\nu}, C_n \downarrow c > 0, \quad 0 < \nu < 1,$$

alors les conditions **H.2** sont satisfaites (pour tout réel $r \leq d+2$) pour tout ν tel que $\nu r < 1$ (ce qui est vrai dès lors que $\nu < \frac{1}{d+2}$) avec :

$$\beta_r = \frac{1}{1 - r\nu}.$$

En effet h_n vérifie clairement **H.2**(i) et pour ce qui est de **H.2**(ii), nous avons :

$$B_{n,r} = \frac{1}{n^{1-\nu r}} \sum_{i=1}^n i^{-\nu r}.$$

- si $\nu r > 1$, alors $\sum_{i=1}^n i^{-\nu r}$ converge et donc $\beta_r = \infty$;
- si $\nu r = 1$, alors :

$$\sum_{i=1}^n i^{-\nu r} \sim \ln n,$$

et donc $\beta_r = \infty$;

- si $\nu r < 1$, alors :

$$\sum_{i=1}^n i^{-\nu r} \sim \frac{n^{1-\nu r}}{1-\nu r},$$

par suite :

$$\beta_r = \frac{1}{1-\nu r}.$$

En particulier, si :

$$h_n = C_n n^{-\frac{1}{2p+d}}, C_n \downarrow c > 0,$$

avec $p \geq 2$ un entier naturel, alors **H.2**(ii) est vérifiée puisque :

$$\frac{1}{2p+d} < \frac{1}{d+2}.$$

2. Si :

$$h_n = C_n \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

alors **H.2** sont également satisfaites. En effet, h_n est décroissante par définition, et pour $\alpha = \frac{1}{d+4}$, on a pour tout $n \geq 1$, si $0 \leq r \leq d+2$:

$$\begin{aligned} B_{n+1,r} - B_{n,r} &= \left(\frac{n}{\ln \ln n} \right)^{\alpha r} \sum_{i=1}^n \left(\frac{\ln \ln i}{i} \right)^{\alpha r} \left\{ \frac{(n+1)^{\alpha r-1}}{n^{\alpha r}} \left[\frac{\ln \ln n}{\ln \ln(n+1)} \right]^{\alpha r} - \frac{1}{n} \right\} + \frac{1}{n+1} \\ &\geq n \left[\frac{1}{n+1} - \frac{1}{n} + \frac{1}{n(n+1)} \right] = 0. \end{aligned}$$

Ensuite :

$$B_{n,r} = \frac{n^{\alpha r-1}}{(\ln \ln n)^{\alpha r}} \sum_{i=1}^n \left(\frac{\ln \ln i}{i} \right)^{\alpha r} \leq n^{\alpha r-1} \sum_{i=1}^n i^{-\alpha r} \rightarrow \frac{1}{1-\alpha r},$$

lorsque $n \rightarrow \infty$. Maintenant si $r < 0$, de façon analogue que ci-dessus, nous avons pour tout $n \geq 1$:

$$B_{n,r} \geq n^{\alpha r-1} \sum_{i=1}^n i^{-\alpha r} \rightarrow \frac{1}{1-\alpha r},$$

lorsque $n \rightarrow \infty$ et :

$$B_{n+1,r} - B_{n,r} \leq 0.$$

Ainsi, si :

$$0 \leq r \leq d+2 \text{ (resp. si } r < 0),$$

alors $B_{n,r}$ est une suite croissante (rep. décroissante) et majorée (resp. minorée) par une suite convergente. Par suite $B_{n,r}$ converge.

3. De façon similaire on montre que le choix :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^\nu, \quad C_n \downarrow c > 0, \quad 0 < \nu < \frac{1}{d+2},$$

satisfait **H.2**.

2.2.2 Estimation dans le cas iid

Convergence en moyenne quadratique

Nous pouvons maintenant établir les biais, variance et EQM asymptotiques exacts de $f_n^\ell(x)$ en fonction du paramètre ℓ .

Théorème 2.2.1. *Sous les hypothèses **H.1** et **H.2**, on a :*

(a) pour tout $\ell \in [0, 1]$,

$$h_n^{-4} \left[E f_n^\ell(x) - f(x) \right]^2 \longrightarrow \left[\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \right]^2 b_f^2(x),$$

lorsque $n \rightarrow \infty$, avec :

$$b_f(x) := \frac{1}{2} \sum_{1 \leq i, j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v) dv. \quad (2.2.2)$$

(b) pour tout $\ell \in [0, 1]$,

$$n h_n^d \text{Var} f_n^\ell(x) \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$, avec :

$$\sigma_\ell^2(x) := \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(x) dx. \quad (2.2.3)$$

(c) Si :

$$h_n = C_n n^{-\frac{1}{d+4}}, \quad C_n \downarrow c > 0,$$

alors :

$$n^{\frac{4}{d+4}} E \left[f_n^\ell(x) - f(x) \right]^2 \longrightarrow c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_f^2(x) + \frac{(4 + d\ell)^2 f(x) \|K\|_2^2}{2c^d(4 + d)(2 + d\ell)},$$

lorsque $n \rightarrow \infty$, en tout point x où $f(x) > 0$.

Le théorème 2.2.1 généralise en dimension d quelconque, les propositions suivantes établies par Deheuvels [34], dans le cas univarié pour les estimateurs $f_n^H(x)$. Nous les adaptons à la famille $(f_n^\ell(x))$ avec nos propres notations en vue de les comparer avec le résultat précédent.

Proposition 2.2.2. (Deheuvels [34]) On suppose que :

- K est une densité positive bornée sur $\mathbb{R}$ et il existe deux réels $\alpha \geq 2$, $\beta \geq 0$ tels que :

$$\lim_{t \rightarrow \infty} t^\beta \int_{|y|>t} |y|^\alpha K(y) dy = 0.$$

- f est 2 fois uniformément différentiable sur $\mathbb{R}$.

Si :

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n h_i^{1-2\ell}}{\left(\sum_{i=1}^n h_i^{1-\ell} \right)^2} = 0 \text{ ou plus généralement } \sum_{i=1}^{\infty} h_i^{1-\ell} = \infty,$$

alors :

$$E f_n^\ell(x) - f(x) = \frac{B_{n,2-\ell}}{B_{n,1-\ell}} h_n f'(x) \int_{\mathbb{R}} y K(y) dy + \frac{B_{n,3-\ell}}{B_{n,1-\ell}} h_n^2 f''(x) \int_{\mathbb{R}} y^2 K(y) dy + o(h_n^2).$$

Proposition 2.2.3. (Deheuvels [34]) On suppose que :

- K est une densité positive bornée sur $\mathbb{R}$ et il existe deux réels $\alpha \geq 2$, $\beta \geq 0$ tels que :

$$\lim_{t \rightarrow \infty} t^\beta \int_{|y|>t} |y|^\alpha K(y) dy = 0.$$

Si :

$$\int_{|y|>t} K^p(y) dF(x + uy) - \int_{|y|>t} K^p(y) f(x) dx = o(1)$$

et :

$$\int_{|y| \leq t} K^p(y) [f(x + uy) - f(x)] dy = o(1),$$

$\forall p > 1$ uniformément en $u > 0$ et x , alors :

$$\text{Var} f_n^\ell(x) = \frac{B_{n,1-2\ell}}{B_{n,1-\ell}^2 n h_n} f(x) \int_{\mathbb{R}} K^2(y) dy - \frac{B_{n,2(1-\ell)}}{n B_{n,1-\ell}^2} f(x)^2 + o(n),$$

pourvu que :

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n h_i^{1-2\ell}}{\left(\sum_{i=1}^n h_i^{1-\ell}\right)^2} = 0.$$

Pour donner les valeurs asymptotiques exactes du biais et de la variance, nous avons introduit l'hypothèse **H.2(ii)**. Celle-ci entraîne les conditions utilisées par Deheuvels [34]. On peut noter que dans le cas iid, le passage en dimension $d \geq 1$ n'est pas contraignant pour la convergence en moyenne quadratique et se fait de façon naturelle, mais nous verrons dans la section 2.2.6 que cela n'est pas le cas pour des observations dépendantes.

Convergence presque sûre ponctuelle

Nous avons le résultat suivant qui établit la vitesse de convergence presque sûre de notre famille d'estimateurs de la densité.

Théorème 2.2.4. *Sous les hypothèses **H.1** et **H.2**, si pour tout $\alpha \geq 0$:*

$$\lim_{n \rightarrow \infty} \frac{n h_n^d}{(\ln \ln n)^{2(\alpha+1)} \ln n} = \infty \text{ et } \lim_{n \rightarrow \infty} \frac{\ln h_n}{\ln n} < \infty, \quad (2.2.4)$$

(a) alors pour tout x tel que $f(x) > 0$:

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{n h_n^d}{\ln \ln n}} [f_n^\ell(x) - E f_n^\ell(x)] = \sigma_\ell \sqrt{2} \text{ p.s. },$$

où σ_ℓ^2 est défini dans (2.2.3).

(b) De plus le choix :

$$h_n = C_n \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

implique que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4}} [f_n^\ell(x) - f(x)] = \sigma_\ell \sqrt{2c^d} + \frac{c^2 \beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_f(x) \text{ p.s.},$$

pour tout x tel que $f(x) > 0$, où $b_f(x)$ est défini en (2.2.2).

Notons que dans le cas non récursif, les vitesses de convergence obtenues (c.f. Giné et Guillaoux [45], etc.) sont du type $\left(\frac{\ln n}{n}\right)^{\frac{2}{d+4}}$. Nous verrons aussi dans la section 2.2.3, qu'en situation de dépendance, il est difficile d'atteindre une borne exacte comme ici, et que la vitesse de convergence est identique à celle obtenue pour l'estimateur à noyau classique.

Dans le cas univarié ($d = 1$), pour $\ell = 1$, on retrouve à partir du (a) du Théorème 2.2.4, un résultat comparable à celui obtenu par Wertz [96], et pour $\ell = \frac{1}{2}$, on retrouve le résultat suivant de Wegman et Davies [95] :

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{nh_n}{\ln \ln n}} [f_n^{DW}(x) - \mathbb{E}f_n^{DW}(x)] = \sqrt{2\beta_{1/2}f(x)} \|K\|_2 \text{ p.s.}$$

En outre, notre résultat (a) du Théorème 2.2.4 généralise au cadre multivarié, celui établi par Deheuvels [33] pour la famille plus générale $(f_n^H(x))$, dans le cas univarié, toutefois ici la preuve repose sur le théorème de Jain et al. [53] au lieu du lemme de Prohorov [73]. En effet, en adaptant les conditions de Deheuvels [33] à la famille $(f_n^\ell(x))$, son résultat peut s'écrire sous des conditions de régularité sur K et pour h_n vérifiant les conditions :

$$\sum_{n=1}^{\infty} h_n^{1-\ell} = \infty, \sum_{n=1}^{\infty} h_n^{1-2\ell} = \infty \text{ et } h_n^{-\ell} = o\left(\frac{\sum_{i=1}^n h_i^{1-2\ell}}{\ln \ln \sum_{i=1}^n h_i^{1-2\ell}}\right)^{\frac{1}{2}}, \quad (2.2.5)$$

sous la forme :

$$\overline{\lim}_{n \rightarrow \infty} \left(\text{resp. } \underline{\lim}_{n \rightarrow \infty} \right) \frac{\sum_{i=1}^n h_i^{1-\ell} [f_n^\ell(x) - \mathbb{E}f(x)]}{\sqrt{2 \|K\|_2^2 \sum_{i=1}^n h_i^{1-2\ell} \ln \ln \sum_{i=1}^n h_i^{1-\ell}}} = \pm \sqrt{f(x)}.$$

On peut montrer que les condition (2.2.5) se déduisent facilement des hypothèses **H.2**.

Pour clore cette section nous donnons le corollaire suivant qui établit le comportement asymptotique presque sûre de $f_n^\ell(x)$ avec un choix de fenêtre spécifié.

Corollaire 2.2.5. *Sous les hypothèses du Théorème 2.2.4, le choix*

$$h_n = n^{-\nu}, 0 < \nu < \frac{1}{d+2}$$

entraîne que, pour tout x tel que $f(x) > 0$,

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{n^{1-\nu d}}{\ln \ln n}} [f_n^\ell(x) - f(x)] = \left\{ \frac{2[1 - \nu d(1 - \ell)]^2}{c^d [1 - \nu d(1 - 2\ell)]} f(x) \right\}^{\frac{1}{2}} \|K\|_2 \text{ p.s.}$$

2.2.3 Estimation dans le cas dépendant

Dans cette section, nous nous proposons de généraliser les résultats obtenus dans le cadre iid aux processus fortement mélangeants. Ainsi, nous établissons les variance et EQM asymptotiques de $f_n^\ell(x)$ en fonction du paramètre ℓ , puisque le biais ne dépend pas de la dépendance ou non du processus. Nous étudions également la vitesse de convergence presque sûre de la famille d'estimateurs $f_n^\ell(x)$.

Convergence en moyenne quadratique

Pour établir la convergence en moyenne quadratique de $f_n^\ell(x)$ dans le cadre dépendant, nous avons besoin d'hypothèses supplémentaires sur le processus (X_t) .

Hypothèses H.3.

(i) : Le processus (X_t) est $2 - \alpha$ -mélangeant avec :

$$\alpha^{(2)}(k) \leq \gamma k^{-\rho}, \quad k \geq 1,$$

pour deux constantes strictement positives γ et ρ .

(ii) : Pour chaque couple (s, t) , $s \neq t$, le vecteur aléatoire (X_s, X_t) admet une densité $f_{(X_s, X_t)}$ telle que :

$$\sup_{|s-t| \geq 1} \|g_{s,t}\|_\infty < \infty, \quad \text{où } g_{s,t} := f_{(X_s, X_t)} - f \otimes f.$$

Les hypothèses **H.3** sont classiques dans ce domaine, en particulier, il n'y a pas d'hypothèse de stationnarité de second ordre statué sur le processus.

Nous pouvons maintenant déterminer les variance et EQM asymptotiques de notre famille d'estimateurs.

Théorème 2.2.6. *Sous les hypothèses **H.1** – **H.3** :*

(a) Pour tout $\ell \in \left[\left(\frac{d-2}{2d}\right)^+, 1\right]$:

$$nh_n^d \text{Var} f_n^\ell(x) \rightarrow \sigma_\ell^2,$$

lorsque $n \rightarrow \infty$, si $\rho > 2$, avec σ_ℓ^2 défini dans (2.2.3).

(b) Si $d \geq 3$ et $\ell \in \left[0, \frac{d-2}{2d}\right]$, la conclusion du (a) reste encore vraie si $\rho > \frac{d+2}{2}$.

(c) Pour tout $\ell \in \left[\left(\frac{d-4}{2d} \right)^+, 1 \right]$, (resp. pour tout $\ell \in \left[0, \frac{d-4}{2d} \right[$, si $d \geq 5$), si $\rho > 2$
 (resp. $\rho > \frac{d+2}{2}$), le choix :

$$h_n = C_n n^{-\frac{1}{d+4}}, \quad C_n \downarrow c > 0,$$

entraîne que :

$$n^{\frac{4}{d+4}} E \left[f_n^\ell(x) - f(x) \right]^2 \longrightarrow c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_f^2(x) + \frac{(4 + d\ell)^2 f(x) \|K\|_2^2}{2c^d(4 + d)(2 + d\ell)},$$

lorsque $n \rightarrow \infty$, pour les valeurs respectives de ℓ en tout point où $f(x) > 0$, où $b_f(x)$ est défini en (2.2.2).

Nous constatons que contrairement au cas iid, dans le cadre dépendant, le passage en dimension supérieure ou égale à 3 a une influence sur le comportement asymptotique de la variance des estimateurs de la famille $(f_n^\ell(x))$, construits avec les petites valeurs de ℓ . En effet pour l'obtenir, nous imposons une décroissance rapide vers 0, du coefficient de mélange du processus. Notons par ailleurs, que le choix de fenêtre :

$$h_n = C_n n^{-\frac{1}{d+4}}, \quad C_n \downarrow c > 0,$$

nous permet d'assouplir cette contrainte pour les dimensions 3 et 4. Ce résultat complète ainsi le manque de résultat sur l'estimateur $f_n^0(x)$ (pourtant connu dans la littérature et ayant une variance plus faible que les autres estimateurs récursifs), pour des observations multivariées dépendantes.

Pour des raisons de simplification, nous supposons dans la suite de notre travail que le processus $(X_t, t \in \mathbb{N})$ est géométriquement fortement mélangeant (G.F.M.). On rappelle la définition d'un processus G.F.M.

Définition 2.4. Un processus $(X_t, t \in \mathbb{N})$ est dit G.F.M. si le coefficient de mélange (2.1.1), défini à la page 12, vérifie :

$$\alpha(u) \leq \gamma e^{-\rho u}, \quad u \geq 0,$$

avec $\gamma, \rho > 0$.

Normalité asymptotique

Étudions maintenant la normalité asymptotique de notre famille d'estimateurs.

Théorème 2.2.7. *Supposons les hypothèses **H.1** – **H.3** satisfaites et que pour toutes suites d'entiers u_n et v_n , la suite h_n vérifie la condition :*

$$u_n \sim v_n \Rightarrow h_{u_n} \sim h_{v_n}. \quad (2.2.6)$$

S'il existe un réel positif $\varsigma_0 > 4$ tel que :

$$\frac{nh_n^{3d}}{(\ln n)^{\varsigma_0}} \rightarrow +\infty, \quad (2.2.7)$$

lorsque $n \rightarrow \infty$, alors pour tout x tel que $f(x) > 0$, on a :

$$\sqrt{nh_n^d} [f_n^\ell(x) - Ef_n^\ell(x)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\ell^2),$$

lorsque $n \rightarrow \infty$ avec σ_ℓ^2 défini dans (2.2.3).

Notre résultat généralise ceux obtenus par Masry [63] (pour $\ell = \frac{1}{2}$ et $\ell = 1$) et par Liang et Baek [61] (pour $\ell = 1$) à notre famille d'estimateurs de la densité. Notons que, pour simplifier, nous avons supposé que le processus est GFM, mais le résultat du Théorème 2.2.7 pourrait être obtenu sous une condition de mélangeance arithmétique.

REMARQUE 1. Dans le Théorème 2.2.7, les conditions (2.2.6) et (2.2.7) sont vérifiées par le choix :

$$h_n = C_n n^{-\nu}, 0 < \nu \leq \frac{1}{3d}, C_n \downarrow c > 0,$$

et (2.2.6) n'est plus nécessaire dans le cas où $\ell \geq \frac{1}{2}$.

Le Théorème 2.2.7 combiné avec le Théorème 2.2.1 (a) permettent d'écrire :

Corollaire 2.2.8. *Sous les hypothèses du Théorème 2.2.7, si h_n vérifie la condition :*

$$nh_n^{d+4} \rightarrow 0,$$

lorsque $n \rightarrow \infty$, alors :

$$\sqrt{nh_n^d} [f_n^\ell(x) - f(x)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_\ell^2),$$

lorsque $n \rightarrow \infty$.

Remarquons que pour $d = 1$, le choix de fenêtre :

$$h_n = C_n n^{-\nu}, C_n \downarrow c > 0, \frac{1}{4} < \nu < \frac{1}{3}$$

est un exemple de paramètre de lissage qui satisfait les conditions du Corollaire 2.2.8.

Convergence presque sûre

Dans ce paragraphe nous étudions la convergence presque sûre de notre famille d'estimateurs. La proposition suivante établit le comportement asymptotique presque sûr ponctuel de $f_n^\ell(x)$.

Proposition 2.2.9. *Sous les hypothèses **H.1** – **H.3**, si :*

$$\frac{nh_n^d}{(\ln n)^3} \rightarrow +\infty,$$

lorsque $n \rightarrow \infty$

(a) *alors pour tout x tel que $f(x) > 0$:*

$$\overline{\lim}_{n \rightarrow \infty} \sqrt{\frac{nh_n^d}{\ln n}} |f_n^\ell(x) - Ef_n^\ell(x)| \leq 1 + \sigma_\ell^2 \text{ p.s.}$$

(b) *De plus, le choix :*

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

implique que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x) - f(x)| \leq 2c^{-\frac{d}{2}} (1 + \sigma_\ell^2) + \frac{c^2 \beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} |b_f(x)|,$$

avec $b_f(x)$ et σ_ℓ^2 définis respectivement dans (2.2.2) et (2.2.3).

Pour $\ell = 1$ et $\ell = 1/2$, (b) est comparable aux résultats obtenus par Masry [63] concernant la convergence presque sûre. Un résultat similaire à (b) est également établi par Bosq et Blanke [15] pour l'estimateur non récursif $f_n^{\text{PR}}(x)$, avec la borne :

$$2c^{-\frac{d}{2}} \sqrt{f(x)} \|K\|_2 + c^2 |b_f(x)|.$$

Pour finir ce paragraphe, nous donnons un résultat uniforme pour la convergence presque sûre de notre famille d'estimateurs.

Théorème 2.2.10. *Sous les hypothèses **H.1** – **H.3**, avec f une densité bornée, soit des constantes $c_d > 0, \mu \geq 0$. Si K satisfait une condition Lipschitzienne alors :*

(a) *le choix :*

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0$$

implique que :

$$\sup_{\|x\| \leq c_d n^\mu} |f_n^\ell(x) - f(x)| = O \left[\left(\frac{\ln n}{n} \right)^{\frac{2}{d+4}} \right].$$

- (b) De plus si K est à support compact, $f(x)$ décroissante à partir d'un certain x , $E\|X_0\| < \infty$ et $\mu > 2$, alors :

$$\sup_{x \in \mathbb{R}^d} |f_n^\ell(x) - f(x)| = O \left[\left(\frac{\ln n}{n} \right)^{\frac{2}{d+4}} \right].$$

Ce théorème généralise à la famille $(f_n^\ell(x))$, le résultat obtenu par Tran [87] concernant la convergence uniforme de $f_n^1(x)$ sur un compact de $\mathbb{R}$.

2.3 Comparaison d'estimateurs

Maintenant que les biais, variance et EQM asymptotiques exacts de nos estimateurs sont établis en fonction d'un paramètre $\ell \in [0, 1]$, la question est de savoir comment choisir ce paramètre pour obtenir le meilleur estimateur selon des critères de comparaison donnés et si selon les mêmes critères, il y a un compromis ou non, en utilisant un noyau récursif à la place du noyau classique. Les critères de comparaison utilisés ici sont ponctuels et basés sur les biais, variance, EQM, et la convergence presque sûre. Ils sont définis de la manière suivante :

Définition 2.5. Soient $f_n(x)$ et $g_n(x)$ deux estimateurs à noyau de $f(x)$.

- (i) On dira que $f_n(x)$ est préférable à $g_n(x)$ au point x , au sens de la variance et on notera :

$$f_n(x) \prec_v g_n(x)$$

si :

$$0 \leq \lim_{n \rightarrow \infty} \frac{\text{Var} f_n(x)}{\text{Var} g_n(x)} < 1.$$

- (ii) On dira que $f_n(x)$ est préférable à $g_n(x)$ au point x , au sens du biais et on notera :

$$f_n(x) \prec_b g_n(x)$$

si :

$$0 \leq \lim_{n \rightarrow \infty} \frac{[E f_n(x) - f(x)]^2}{[E g_n(x) - f(x)]^2} < 1.$$

(iii) On suppose que $f(x) > 0$ et on choisit :

$$h_n = C_n n^{-\frac{1}{d+4}}, \quad C_n \downarrow c > 0,$$

avec :

$$c = c_{\min}[f_n(x)] \text{ (resp. } c = c_{\min}[g_n(x)] \text{)},$$

pour l'estimateur $f_n(x)$ (resp. $g_n(x)$), où $c_{\min}(\diamond)$ désigne la constante qui minimise l'EQM asymptotique de l'estimateur $\diamond$.

Sous ces conditions, on dira que $f_n(x)$ est préférable à $g_n(x)$ au point x , au sens de l'EQM et on notera :

$$f_n(x) \prec_{L^2} g_n(x)$$

si :

$$0 \leq \lim_{n \rightarrow \infty} \frac{\mathbb{E}[f_n(x) - f(x)]^2}{\mathbb{E}[g_n(x) - f(x)]^2} < 1.$$

(iv) Si $f(x) > 0$ et $h_n = n^{-\nu}$, $0 < \nu < \frac{1}{d+2}$, on dira que $f_n(x)$ est préférable à $g_n(x)$ au point x , au sens de la convergence presque sûre et on notera :

$$f_n(x) \prec_{ps} g_n(x)$$

si :

$$0 \leq \frac{\overline{\lim}_{n \rightarrow \infty} |f_n(x) - f(x)|}{\overline{\lim}_{n \rightarrow \infty} |g_n(x) - f(x)|} < 1 \text{ p.s.}$$

Les critères (i) et (iv) d'où sont inspirés les critères (ii) et (iii) ont été respectivement introduits par Banon [4] et par Deheuvels [34] pour comparer des estimateurs récurrents à noyau.

Notre premier résultat de cette partie permet de classer nos estimateurs selon les valeurs de ℓ par les critères de comparaisons définis ci-dessus.

Théorème 2.3.1. *On suppose que les hypothèses **H.1** – **H.3** sont vérifiées. On choisit*

$$h_n = C_n n^{-\nu}, \quad C_n \downarrow c > 0, \quad 0 < \nu < \frac{1}{d+2}.$$

Pour tous $\ell_1, \ell_2 \in [0, 1]$:

(a) *si $\ell_1 < \ell_2$, alors :*

$$f_n^{\ell_1}(x) \prec_v f_n^{\ell_2}(x) \left(\text{resp. } f_n^{\ell_2}(x) \prec_b f_n^{\ell_1}(x) \right).$$

(b) Si $f(x) > 0$, et $\nu = \frac{1}{d+4}$, et si $\ell_1 < \ell_2$, alors :

$$f_n^{\ell_2}(x) \prec_{L^2} f_n^{\ell_1}(x).$$

(c) Dans le cas où les X_t sont iid, si $\ell_1 < \ell_2$, alors :

$$f_n^{\ell_1}(x) \prec_{ps} f_n^{\ell_2}(x).$$

On constate que $f_n^0(x)$ est optimal dans la famille $(f_n^\ell(x))$, pour les critères de la variance et de la convergence presque sûre. Ceci est sans surprise car Deheuvels [34] et Wertz [96] avaient déjà montré que cet estimateur était optimal respectivement pour la variance et pour la convergence presque sûre sur la famille $f_n^H(x)$, mais la nouveauté de notre résultat par rapport aux références citées précédemment est que nous ordonnons la famille $(f_n^\ell(x))$ en fonction du paramètre ℓ selon les critères de comparaison.

Le Théorème 2.3.1 montre en revanche, que l'estimateur $f_n^0(x)$ est le "moins préférable" de la famille $(f_n^\ell(x))$ au sens du biais et de l'EQM. Pour ces critères, c'est $f_n^1(x)$ qui est optimal.

Notons que Banon [4] a construit "au prix d'une complexité accrue", l'estimateur :

$$f_n^{BAN}(x) := \frac{1}{\sum_{i=1}^n h_i} \sum_{i=1}^n \frac{h_i}{\sum_{k=1}^i h_k} \sum_{k=1}^i K\left(\frac{x - X_i}{h_k}\right),$$

vérifiant la relation récursive :

$$f_{n+1}^{BAN}(x) = \frac{\sum_{i=1}^n h_i}{\sum_{i=1}^{n+1} h_i} f_n^{BAN}(x) + \frac{h_n}{\left(\sum_{i=1}^{n+1} h_i\right)^2} \sum_{k=1}^{n+1} K\left(\frac{x - X_{n+1}}{h_k}\right),$$

qui, bien qu'il n'appartienne pas à la famille d'estimateurs $f_n^H(x)$, demeure jusqu'ici, pour $d = 1$, le seul estimateur récursif de la densité qui soit meilleur que $f_n^0(x)$ au sens de la variance.

Notre deuxième résultat compare notre famille d'estimateurs à l'estimateur à noyau usuel.

Théorème 2.3.2. *On se place sous les hypothèses du Théorème 2.3.1. Alors :*

- (a) tous les estimateurs $f_n^\ell(x)$, $\ell \in [0, 1]$ sont préférables à $f_n^{PR}(x)$ au sens de la variance.
- (b) Aucun estimateur $f_n^\ell(x)$, $\ell \in [0, 1]$ n'est préférable à $f_n^{PR}(x)$ au sens du biais.
- (c) Si $f(x) > 0$, et $\nu = \frac{1}{d+4}$, $f_n^{PR}(x)$ est préférable à tous les estimateurs $f_n^\ell(x)$, $\ell \in [0, 1]$ au sens de l'EQM pour les choix "optimaux" respectifs de c .

En conclusion, la récursivité améliore l'estimateur à noyau habituel au sens de la variance. Ainsi, l'augmentation du biais apporté par la récursivité n'est pas compensée par la diminution de la variance et se traduit par une légère détérioration de l'EQM. Toutefois les estimateurs récursifs atteignent bien les vitesses optimales et conservent un avantage décisif en terme de rapidité de calcul.

2.4 Preuves

Nos preuves utilisent entre autres des résultats classiques présentés en annexe. Il s'agit des résultats dont les numéros de référence commencent par A. Pour prouver nos résultats, nous utiliserons souvent le Lemme suivant :

Lemme 2.4.1. *Soit $(w_n)_{n \geq 1}$ une suite de nombres réels tendant vers w . Si les hypothèses **H.2** sont vérifiées pour tout $r \in]-\infty, d+2]$ alors :*

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n} \right)^r w_i \rightarrow \beta_r w,$$

lorsque $n \rightarrow \infty$.

Le Lemme 2.4.1 est une conséquence immédiate du Lemme de Toeplitz (Lemme A.1.2). En effet, si l'on pose :

$$a_{n,i} = \begin{cases} \frac{1}{n} \left(\frac{h_i}{h_n} \right)^r & \text{si } i \leq n \\ 0 & \text{si } i > n \end{cases},$$

alors :

- (i) pour tout $i \geq 1$, $a_{n,i} \leq \frac{h_i^r}{nh_n^r} \rightarrow 0$, lorsque $n \rightarrow \infty$, grâce à l'hypothèse **H.2**(i).
- (ii) Grâce à l'hypothèse **H.2**(ii), nous avons aussi :

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n a_{n,i} = \lim_{n \rightarrow \infty} B_{n,r} = \beta_r < \infty.$$

- (iii) Il existe $C > 0$ tel que pour tout $n \geq 1$, $\sum_{i=1}^{\infty} |a_{n,i}| < C < \infty$, grâce à la convergence de $B_{n,r}$.

On peut donc appliquer le Lemme de Toeplitz, et on obtient le résultat.

2.4.1 Preuve du Théorème 2.2.1

(a) Nous avons :

$$\begin{aligned} \mathbb{E}f_n^\ell(x) - f(x) &= \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-1} \sum_{i=1}^n h_i^{-d\ell} \int_{\mathbb{R}^d} K\left(\frac{x-u}{h_i}\right) f(u) du - f(x) \\ &= \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-1} \sum_{i=1}^n h_i^{d(1-\ell)} \left[\int_{\mathbb{R}^d} \frac{1}{h_i^d} K\left(\frac{x-u}{h_i}\right) f(u) du - f(x) \right] \\ &= \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-1} \sum_{i=1}^n h_i^{d(1-\ell)} \int_{\mathbb{R}^d} [f(x - h_i v) - f(x)] K(v) dv. \end{aligned}$$

Donc, par un développement de Taylor, il existe un réel θ , avec $0 < \theta < 1$ tel que :

$$h_n^{-2} \left[\mathbb{E} f_n^\ell(x) - f(x) \right] = \frac{B_{n,d(1-\ell)}^{-1}}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n} \right)^{d(1-\ell)+2} \frac{1}{2} \int_{\mathbb{R}^d} K(v) \sum_{1 \leq i,j \leq d} v_i v_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta h_i v) dv.$$

Puisque $f \in C_d^2(b)$, alors grâce aux hypothèses **H.1(ii)**, **(iii)**, le théorème de convergence dominée entraîne que :

$$\frac{1}{2} \int_{\mathbb{R}^d} K(v) \sum_{1 \leq i,j \leq d} v_i v_j \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta h_i v) dv \rightarrow \frac{1}{2} \sum_{1 \leq i,j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v) dv,$$

lorsque $i \rightarrow \infty$. Ensuite puisque :

$$d(1 - \ell) + 2 \leq d + 2,$$

le Lemme 2.4.1 donne le résultat sous les hypothèses **H.2**.

(b) Puisque les X_i sont indépendantes on a :

$$\begin{aligned} \text{Var} f_n^\ell(x) &= \frac{1}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \sum_{i=1}^n \frac{1}{h_i^{2d\ell}} \int_{\mathbb{R}^d} K^2 \left(\frac{x-u}{h_i} \right) f(u) du \\ &\quad - \frac{1}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \sum_{i=1}^n \frac{1}{h_i^{2d\ell}} \left[\int_{\mathbb{R}^d} K \left(\frac{x-u}{h_i} \right) f(u) du \right]^2 \\ &=: S_1 - S_2. \end{aligned}$$

Le premier terme s'écrit :

$$S_1 = \frac{1}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \sum_{i=1}^n h_i^{d(1-2\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K^2 \left(\frac{x-u}{h_i} \right) f(u) du.$$

Donc :

$$n h_n^d S_1 = \frac{B_{n,d(1-\ell)}^{-2}}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n} \right)^{d(1-2\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K^2 \left(\frac{x-u}{h_i} \right) f(u) du.$$

On a par application du Lemme de Bochner, grâce aux hypothèses **H.1(i)** et **H.2(i)** :

$$\int_{\mathbb{R}^d} \frac{1}{h_i^d} K^2 \left(\frac{x-u}{h_i} \right) f(u) du \rightarrow f(x) \int_{\mathbb{R}^d} K^2(u) du,$$

lorsque $i \rightarrow \infty$. Donc les hypothèses **H.2** et le Lemme 2.4.1 entraînent que :

$$n h_n^d S_1 \rightarrow \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(u) du, \quad (2.4.1)$$

lorsque $n \rightarrow \infty$. Ensuite puisque h_n est décroissante, on a :

$$S_2 \leq \frac{h_1^{d(1-\ell)}}{\left[\sum_{i=1}^n h_i^{d(1-\ell)}\right]^2} \sum_{i=1}^n h_i^{d(1-\ell)} \left[\int_{\mathbb{R}^d} \frac{1}{h_i^d} K\left(\frac{x-u}{h_i}\right) f(u) du \right]^2.$$

Par suite :

$$nh_n^d S_2 \leq \frac{h_1^{d(1-\ell)}}{n B_{n,d(1-\ell)}^2} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^{d(1-\ell)} \left[\int_{\mathbb{R}^d} \frac{1}{h_i^d} K\left(\frac{x-u}{h_i}\right) f(u) du \right]^2 \times h_n^{d\ell} \rightarrow 0, \quad (2.4.2)$$

lorsque $n \rightarrow \infty$, par application du Lemme de Bochner et du Lemme 2.4.1, car :

$$d(1-\ell) < d+2.$$

Ainsi (2.4.1) et (2.4.2) impliquent que :

$$nh_n^d \text{Var} f_n^\ell(x) = nh_n^d (S_1 - S_2) \rightarrow \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(u) du,$$

lorsque $n \rightarrow \infty$.

(c) Cette partie découle de la décomposition : EQM=BC+Variance, en se servant aussi du fait que si :

$$h_n = C_n n^{-\frac{1}{4+d}}, C_n \downarrow c > 0,$$

alors pour tout $r \leq d+2$, on a :

$$\beta_r = \frac{d+4}{d+4-r}.$$

Il vient que :

$$\left[\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \right]^2 = \left(\frac{4+d\ell}{2+d\ell} \right)^2 \quad \text{et} \quad \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} = \frac{(4+d\ell)^2}{2(4+d)(2+d\ell)}.$$

■

2.4.2 Preuve du Théorème 2.2.4

(a) Nous appliquons le théorème de Jain et al. [53] (Théorème A.2.1) aux variables Z_i définies par :

$$Z_i = \frac{1}{h_i^{d\ell}} \left[K\left(\frac{x-X_i}{h_i}\right) - \mathbb{E}K\left(\frac{x-X_i}{h_i}\right) \right]. \quad (2.4.3)$$

Posons :

$$S_n = \sum_{i=1}^n Z_i,$$

on a :

$$f_n^\ell(x) - \mathbb{E}f_n^\ell(x) = \frac{S_n}{nB_{n,d(1-\ell)}h_n^{d(1-\ell)}}. \quad (2.4.4)$$

Ensuite si l'on pose :

$$V_n = \sum_{i=1}^n \mathbb{E}Z_i^2,$$

alors nous avons grâce à l'indépendance des X_i :

$$\begin{aligned} V_n &= \sum_{i=1}^n h_i^{-2d\ell} \mathbb{E} \left[K \left(\frac{x-X_i}{h_i} \right) - \mathbb{E}K \left(\frac{x-X_i}{h_i} \right) \right]^2 \\ &= \text{Var} \sum_{i=1}^n h_i^{-d\ell} \left[K \left(\frac{x-X_i}{h_i} \right) \right] \\ &= nB_{n,d(1-\ell)}^2 h_n^{d(1-2\ell)} n h_n^d \text{Var} f_n^\ell(x). \end{aligned}$$

Soit S une fonction aléatoire définie sur $[0, +\infty[$ en posant :

$$\text{pour } t \in [V_n, V_{n+1}[, S(t) = S_n.$$

Le Théorème 2.2.1 (b), nous permet donc d'écrire :

$$V_n \sim n h_n^{d(1-2\ell)} \beta_{d(1-2\ell)} f(x) \int_{\mathbb{R}^d} K^2(u) du \rightarrow \infty, \quad (2.4.5)$$

lorsque $n \rightarrow \infty$. La condition (2.2.4) implique que :

$$\frac{n h_n^d}{\ln [n h_n^{d(1-2\ell)}] \left[\ln \ln (n h_n^{d(1-2\ell)}) \right]^{2(\alpha+1)}} \rightarrow \infty,$$

lorsque $n \rightarrow \infty$. Donc il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, on ait :

$$\frac{n h_n^{d(1-2\ell)}}{\ln [n h_n^{d(1-2\ell)}] \left[\ln \ln (n h_n^{d(1-2\ell)}) \right]^{2(\alpha+1)}} > \|K\|_\infty^2 h_n^{-2d\ell} \geq Z_n^2.$$

Ainsi l'événement :

$$\left\{ Z_n^2 > \frac{n h_n^{d(1-2\ell)}}{\ln [n h_n^{d(1-2\ell)}] \left\{ \ln \ln [n h_n^{d(1-2\ell)}] \right\}^{2(\alpha+1)}} \right\}$$

est un événement impossible pour $n \geq n_0$. Par suite, nous avons grâce à (2.4.5), que :

$$\sum_{n=1}^{\infty} \frac{(\ln \ln V_n)^\alpha}{V_n} \mathbb{E} \left(Z_n^2 \mathbf{1}_{\left\{ Z_n^2 > \frac{V_n}{\ln V_n (\ln \ln V_n)^{2(\alpha+1)}} \right\}} \right) < \infty.$$

Le Théorème A.2.1 entraîne donc qu'il existe un Mouvement Brownien (M.B.) W tel que pour $t \in [V_n, V_{n+1}[$:

$$\left| \frac{S(t) - W(t)}{(2t \ln \ln t)^{\frac{1}{2}}} \right| = o \left[(\ln \ln t)^{-\frac{\alpha}{2}} \right].$$

Mais puisque le M.B. vérifie la Loi de Logarithme Itérée (L.L.I.), alors :

$$\overline{\lim}_{t \rightarrow \infty} \frac{S(t)}{(2t \ln \ln t)^{\frac{1}{2}}} = \overline{\lim}_{t \rightarrow \infty} \left[\frac{S(t) - W(t)}{(2t \ln \ln t)^{\frac{1}{2}}} + \frac{W(t)}{(2t \ln \ln t)^{\frac{1}{2}}} \right] = 1 \text{ p.s.}$$

Ainsi, par l'équivalence (2.4.5), la définition de $S(t)$, le fait que $t \in [V_n, V_{n+1}[$ et que $\frac{V_{n+1}}{V_n} \rightarrow 1$, et en utilisant (2.4.4), on arrive à :

$$\overline{\lim}_{n \rightarrow \infty} \frac{B_{n,d(1-\ell)} n h_n^{d(1-\ell)} [f_n^\ell(x) - E f_n^\ell(x)]}{(2V_n \ln \ln V_n)^{\frac{1}{2}}} \cdot \frac{\left\{ n h_n^{d(1-2\ell)} \ln \ln [n h_n^{d(1-2\ell)}] \right\}^{\frac{1}{2}}}{\left\{ n h_n^{d(1-2\ell)} \ln \ln [n h_n^{d(1-2\ell)}] \right\}^{\frac{1}{2}}} = 1 \text{ p.s.}$$

Mais nous avons :

$$\frac{\left\{ n h_n^{d(1-2\ell)} \ln \ln [n h_n^{d(1-2\ell)}] \right\}^{\frac{1}{2}} B_{n,d(1-\ell)}}{(2V_n \ln \ln V_n)^{\frac{1}{2}}} \rightarrow \frac{\beta_{d(1-\ell)}}{\left[2\beta_{d(1-2\ell)} f(x) \int_{\mathbb{R}^d} K^2(u) du \right]^{\frac{1}{2}}},$$

lorsque $n \rightarrow \infty$. Il vient alors que :

$$\overline{\lim}_{n \rightarrow \infty} \left\{ \frac{n h_n^d}{\ln \ln [n h_n^{d(1-2\ell)}]} \right\}^{\frac{1}{2}} [f_n^\ell(x) - E f(x)] = \sigma_\ell \sqrt{2} \text{ p.s.}$$

Nous déduisons le résultat cherché à partir de cette dernière convergence, en utilisant la seconde partie de (2.2.4) qui implique que :

$$\ln \ln [n h_n^{d(1-2\ell)}] = (\ln \ln n) [1 + o(1)].$$

(b) Nous savons que le choix :

$$h_n = C_n \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0$$

vérifie **H.2** et par un calcul de routine, on peut montrer que ce choix vérifie également les conditions (2.2.4). En utilisant le (a) et le Théorème 2.2.1(a), on obtient (b). ■

2.4.3 Preuve du Théorème 2.2.6

(a) On décompose la variance de l'estimateur en un terme principal et un terme de covariance. On a :

$$\begin{aligned} \text{Var} f_n^\ell(x) &= \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{i=1}^n \text{Var} \left[\frac{1}{h_i^{d\ell}} K \left(\frac{x - X_i}{h_i} \right) \right] \\ &\quad + \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{i=1}^n \sum_{i \neq j}^n A_{i,j} \\ &=: I_1 + I_2, \end{aligned}$$

où l'on a noté :

$$A_{i,j} := \text{Cov} \left[h_i^{-d\ell} K \left(\frac{x - X_i}{h_i} \right), h_j^{-d\ell} K \left(\frac{x - X_j}{h_j} \right) \right].$$

Le terme principal I_1 ayant été traité dans le cas iid, nous avons juste à montrer que le terme des covariances I_2 est négligeable et contrôlé par les hypothèses **H.3**. Pour cela, on considère une suite c_n qui tend vers l'infini, dont on précisera ultérieurement la valeur exacte selon les cas. On a :

$$I_2 \leq 2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \left(\sum_{i=1}^n \sum_{j=1}^n |A_{i,j}| 1_{\{1 \leq i-j \leq c_n\}} + \sum_{i=1}^n \sum_{j=1}^n |A_{i,j}| 1_{\{c_n+1 \leq i-j \leq n-1\}} \right).$$

Pour $i > j$, on pose :

$$k = i - j \text{ et } p = j \Leftrightarrow i = k + p \text{ et } j = p.$$

Alors :

$$p \in \{1, \dots, n\} \cap \{1 - k, \dots, n - k\} \text{ et } k \in \{1, \dots, n - 1\}.$$

Donc :

$$I_2 \leq 2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \left(\sum_{k=1}^{c_n} \sum_{p=1}^n |A_{k+p,p}| + \sum_{k=c_n+1}^{n-1} \sum_{p=1}^n |A_{k+p,p}| \right) := J_1 + J_2. \quad (2.4.6)$$

D'une part, nous avons par l'inégalité de Billingsley (A.4.1) :

$$A_{k+p,p} \leq 4\alpha^{(2)}(k) \|K\|_\infty^2 h_{k+p}^{-d\ell} h_p^{-d\ell}.$$

Donc :

$$\begin{aligned}
 J_2 &\leq 8 \|K\|_\infty^2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{k=c_n+1}^{n-1} \sum_{p=1}^n \alpha^{(2)}(k) h_{p+k}^{-d\ell} h_p^{-d\ell} \\
 &\leq 8\gamma \|K\|_\infty^2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{k=c_n}^{n-1} \sum_{p=1}^n k^{-\rho} h_{p+k}^{-d\ell} h_p^{-d\ell} \\
 &\leq 8\gamma \|K\|_\infty^2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \frac{h_n^{-2d\ell} c_n^{-\rho+1}}{\rho-1} \sum_{p=1}^n \left(\frac{h_p}{h_n} \right)^{-d\ell} \\
 &= \frac{8\gamma \|K\|_\infty^2 c_n^{1-\rho} B_{n,-d\ell}}{n h_n^{2d} B_{n,d(1-\ell)}^2 (\rho-1)}.
 \end{aligned}$$

Ainsi :

$$n h_n^d J_2 = O(c_n^{1-\rho} h_n^{-d}). \quad (2.4.7)$$

On a d'autre part :

$$\begin{aligned}
 A_{k+p,p} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (h_{k+p} h_p)^{-d\ell} K\left(\frac{x-u}{h_{k+p}}\right) K\left(\frac{x-v}{h_p}\right) g_{k+p,p}(u,v) du dv \\
 &\leq \sup_{k \geq 1} \|g_{k+p,p}\|_\infty (h_{k+p} h_p)^{-d\ell} \int_{\mathbb{R}^d} K\left(\frac{x-u}{h_{k+p}}\right) du \int_{\mathbb{R}^d} K\left(\frac{x-v}{h_p}\right) dv \\
 &= (h_{k+p} h_p)^{d(1-\ell)} \sup_{k \geq 1} \|g_{k+p,p}\|_\infty.
 \end{aligned}$$

Donc :

$$\begin{aligned}
 J_1 &\leq 2 \sup_{k \geq 1} \|g_{k+p,p}\|_\infty \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{k=1}^{c_n} \sum_{p=1}^{n-k} h_{p+k}^{d(1-\ell)} h_p^{d(1-\ell)} \\
 &\leq 2 \sup_{k \geq 1} \|g_{k+p,p}\|_\infty c_n \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{p=1}^n h_p^{2d(1-\ell)}.
 \end{aligned} \quad (2.4.8)$$

Si :

$$\ell \in \left[\left(\frac{d-2}{2d} \right)^+, 1 \right],$$

alors :

$$2d(1-\ell) \leq d+2 \Rightarrow B_{n,2d(1-\ell)} \rightarrow \beta_{2d(1-\ell)} < \infty,$$

lorsque $n \rightarrow \infty$, grâce à l'hypothèse **H.2(ii)**. Alors :

$$J_1 \leq 2 \sup_{k \geq 1} \|g_{k+p,p}\|_\infty \frac{c_n B_{n,2d(1-\ell)}}{n B_{n,d(1-\ell)}^2},$$

c'est-à-dire que :

$$nh_n^d J_1 = O\left(c_n h_n^d\right).$$

Ainsi avec le choix :

$$c_n := \left[h_n^{-\frac{2d}{\rho}} \right],$$

nous avons en vertu de (2.4.7) que :

$$nh_n^d I_2 = O\left[h_n^{-\frac{d(2-\rho)}{\rho}} \right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$, car $\rho > 2$, ce qui prouve (a).

(b) Maintenant, on s'intéresse au cas où :

$$d \geq 3, \quad \ell \in \left[0, \frac{d-2}{2d} \right] \text{ et } \rho > \frac{d+2}{2}.$$

On reprend la décomposition (2.4.6). Le terme J_2 est étudié comme au (a). Pour J_1 , on reprend sa majoration en (2.4.8) et on choisit un réel ξ tel que :

$$\frac{1}{\rho-1} < \xi \leq \frac{2}{d}.$$

Notons qu'un tel choix est possible si la condition :

$$\rho > \frac{d+2}{2}$$

est vraie. Ainsi choisi, ξ vérifie la relation :

$$d(\xi+1) \leq d+2,$$

par suite :

$$B_{n,d(\xi+1)} \rightarrow \beta_{d(\xi+1)} < \infty,$$

lorsque $n \rightarrow \infty$, en vertu de l'hypothèse **H.2(ii)**. Ensuite puisque h_n est décroissante nous avons :

$$\sum_{i=1}^n h_i^{d(1-\ell)} \geq h_1^{-d\ell} \sum_{i=1}^n h_i^d.$$

On obtient, donc grâce à la décroissance de h_n et les inégalités

$$0 \leq \ell < \frac{d-2}{2d},$$

que :

$$c_n \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{p=1}^n h_p^{2d(1-\ell)} \leq \frac{c_n h_1^{d(1-\xi-2\ell)} n h_n^{d(\xi+1)} B_{n,d(\xi+1)}}{n^2 h_1^{-2d\ell} h_n^{2d} B_{n,d}^2} = \frac{c_n h_1^{d(1-\xi)} h_n^{d\xi} B_{n,d(\xi+1)}}{n h_n^d B_{n,d}^2}.$$

En effet :

$$0 \leq \ell < \frac{d-2}{2d} \Rightarrow 1 - \xi - 2\ell > 0, \text{ car } \xi \leq \frac{2}{d}.$$

Ainsi il vient grâce à la majoration (2.4.8) que :

$$n h_n^d J_1 = O \left(c_n h_n^{d\xi} \right).$$

En choisissant :

$$c_n := \left[h_n^{-\frac{d(\xi+1)}{\rho}} \right],$$

on déduit grâce à (2.4.7) que :

$$n h_n^d I_2 = O \left[h_n^{-\frac{d(1+\xi-\rho\xi)}{\rho}} \right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$, car $\xi > \frac{1}{\rho-1}$; ce qui prouve (b). Enfin la preuve du (c) est identique à celle du Théorème 2.2.1(c). Pour l'étude de la variance, nous reprenons seulement le contrôle du terme J_1 de (2.4.6), dans le cas où :

$$\ell \in \left[\left(\frac{d-4}{2d} \right)^+, 1 \right].$$

Le choix de fenêtre :

$$C_n n^{-\frac{1}{d+4}}, \quad C_n \downarrow c > 0$$

implique que :

$$c_n \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{p=1}^n h_p^{2d(1-\ell)} = \frac{c_n \sum_{p=1}^n p^{-\frac{2d(1-\ell)}{d+4}}}{\left[\sum_{i=1}^n i^{-\frac{d(1-\ell)}{d+4}} \right]^2} \sim c_n n^{-1}.$$

Il vient que :

$$nh_n^d J_1 = O\left(c_n n^{-\frac{d}{d+4}}\right).$$

Ainsi le choix :

$$c_n := \left\lfloor n^{\frac{2d}{\rho(d+4)}} \right\rfloor,$$

utilisé conjointement avec la relation (2.4.7) entraînent que :

$$nh_n^d I_2 = O\left[n^{\frac{d(2-\rho)}{\rho(d+4)}}\right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$. ■

2.4.4 Preuve du Théorème 2.2.7

Considérons les suites ς_n, τ_n , et r_n définies comme suit :

$$\tau_n := \lfloor \tau_0 \log n \rfloor, \quad \varsigma_n := \left\lfloor \frac{\tau_0 \sqrt{nh_n^d}}{(\log n)^{\varsigma_0}} \right\rfloor \quad \text{et} \quad r_n := \left\lfloor \frac{n}{\varsigma_n + \tau_n} \right\rfloor,$$

avec τ_0 un réel positif. Pour établir la normalité asymptotique de notre famille d'estimateurs nous utilisons la méthode classique proposée par Doob [40] qui consiste à partager le terme :

$$\sqrt{nh_n^d} \left[f_n^\ell(x) - \mathbb{E} f_n^\ell(x) \right]$$

en sommes de gros blocs séparés par des petits définis par :

$$T_{nm} = \sum_{j=k_m}^{k_m+\varsigma_n-1} \Psi_{nj} \text{ (gros blocs) }, \quad T'_{nm} = \sum_{j=l_m}^{l_m+\tau_n-1} \Psi_{nj} \text{ (petits blocs) },$$

$$T'_{nr_n+1} = \sum_{j=\bar{N}+1}^n \Psi_{nj} \text{ (terme de reste)}, \quad \text{avec : } \bar{N} := r_n(\tau_n + \varsigma_n),$$

où l'on a posé pour $m = 1, \dots, r_n$:

$$k_m := (m-1)(\varsigma_n + \tau_n) + 1, \quad l_m := (m-1)(\varsigma_n + \tau_n) + \varsigma_n + 1$$

et

$$\Psi_{nj} := \left[\frac{h_n^{d(2\ell-1)}}{n} \right]^{\frac{1}{2}} \frac{h_j^{-d\ell}}{B_{n,d(1-\ell)}} \left[K\left(\frac{x - X_j}{h_j}\right) - \mathbb{E} K\left(\frac{x - X_j}{h_j}\right) \right]. \quad (2.4.9)$$

Ensuite nous définissons les sommes partielles :

$$S_{n1} = \sum_{m=1}^{r_n} T_{nm}, \quad S_{n2} = \sum_{m=1}^{r_n} T'_{nm} \text{ et } S_{n3} = T'_{nr_n+1}.$$

Ainsi nous avons :

$$\sqrt{nh_n^d} [f_n^\ell(x) - \mathbb{E}f_n^\ell(x)] = S_{n1} + S_{n2} + S_{n3}.$$

Alors pour aboutir à notre résultat, il suffit de prouver que $\mathbb{E}S_{n2}^2 \rightarrow 0$, $\mathbb{E}S_{n3}^2 \rightarrow 0$ (car cela implique que ces sommes convergent en probabilité vers zéro) et que S_{n1} est asymptotiquement normal. D'abord, nous avons :

$$\begin{aligned} \mathbb{E}S_{n2}^2 &= \sum_{m=1}^{r_n} \text{Var}(T'_{nm}) + 2 \sum_{1 \leq i < j \leq r_n} \text{Cov}(T'_{ni}, T'_{nj}) \\ &= \sum_{m=1}^{r_n} \sum_{i=l_m}^{l_m+\tau_n-1} \text{Var}(\Psi_{ni}) + 2 \sum_{m=1}^{r_n} \sum_{l_m \leq i < j \leq l_m+\tau_n-1} \text{Cov}(\Psi_{ni}, \Psi_{nj}) \\ &\quad + 2 \sum_{1 \leq i < j \leq r_n} \sum_{s=l_i}^{l_i+\tau_n-1} \sum_{t=l_j}^{l_j+\tau_n-1} \text{Cov}(\Psi_{ns}, \Psi_{nt}) \\ &:= \Delta_1 + \Delta_2 + \Delta_3. \end{aligned}$$

Pour le premier terme de $\mathbb{E}S_{n2}^2$, nous avons :

$$\Delta_1 = \frac{h_n^{d(2\ell-1)}}{nB_{n,d(1-\ell)}^2} \sum_{m=1}^{r_n} \sum_{j=l_m}^{l_m+\tau_n-1} h_j^{-2d\ell} \text{Var}K\left(\frac{x - X_j}{h_j}\right) \leq \frac{\|K\|_\infty^2 h_n^{d(2\ell-1)}}{nB_{n,d(1-\ell)}^2} \sum_{m=1}^{r_n} \sum_{j=l_m}^{l_m+\tau_n-1} h_j^{-2d\ell}.$$

Puisque h_n est décroissante nous pouvons écrire grâce à la condition (2.2.7) que :

$$\Delta_1 \leq \frac{r_n \tau_n \|K\|_\infty^2}{nh_n^d B_{n,d(1-\ell)}^2} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Pour le second terme de $\mathbb{E}S_{n2}^2$, nous avons par l'inégalité de Cauchy-Schwartz et avec un raisonnement analogue que pour Δ_1 :

$$\Delta_2 \leq \frac{r_n \tau_n^2 \|K\|_\infty^2}{nh_n^d B_{n,d(1-\ell)}^2} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Enfin le dernier terme de $\mathbb{E}S_{n2}^2$ est traité avec l'inégalité de Billingsley, la condition (2.2.7) et l'hypothèse **H.3**(i). Nous écrivons :

$$\Delta_3 \leq \frac{2 \|K\|_\infty^2 h_n^{d(2\ell-1)}}{nB_{n,d(1-\ell)}^2} \sum_{k=1}^{r_n-1} \sum_{j=1}^{r_n} \sum_{s=l_j}^{l_j+\tau_n-1} \sum_{t=l_j}^{l_j+\tau_n-1} (h_s h_t)^{-d\ell} \alpha[k(\varsigma_n + \tau_n)] \leq \frac{2 \|K\|_\infty^2 r_n \tau_n^2}{nh_n^d B_{n,d(1-\ell)}^2} \sum_{k=1}^{r_n-1} e^{-k\tau_n}.$$

Ainsi :

$$\Delta_3 = O \left\{ \frac{r_n \tau_n^2 e^{-\tau_n}}{n h_n^d B_{n,d(1-\ell)}^2} \left[1 - e^{-\tau_n(r_n-1)} \right] \right\} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. De même nous allons maintenant montrer que $ES_{n3}^2 \rightarrow 0$ lorsque $n \rightarrow \infty$.
Nous avons :

$$ES_{n3}^2 = \sum_{j=\bar{N}+1}^n \text{Var}(\Psi_{nj}) + 2 \sum_{\bar{N}+1 \leq i < j \leq n} \text{Cov}(\Psi_{ni}, \Psi_{nj}) := \Theta_{n1} + \Theta_{n2}.$$

D'après (2.4.9) nous avons :

$$n h_n^d \text{Var} f_n^\ell(x) \sim \sum_{j=1}^n \text{Var}(\Psi_{nj}) = \sum_{j=1}^{\bar{N}} \text{Var}(\Psi_{nj}) + \Theta_{n1}.$$

Or nous pouvons écrire que :

$$\sum_{j=1}^{\bar{N}} \text{Var}(\Psi_{nj}) = \left(\frac{n h_n^d}{\bar{N} h_{\bar{N}}^d} \right) \bar{N} h_{\bar{N}}^d \text{Var} f_{\bar{N}}^\ell(x).$$

Puisque $\bar{N} \sim n$, la condition (2.2.6) entraîne que :

$$\left(\frac{n h_n^d}{\bar{N} h_{\bar{N}}^d} \right) \rightarrow 1,$$

lorsque $n \rightarrow \infty$. Par suite, la convergence :

$$n h_n^d \text{Var} f_n^\ell(x) \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$, (c.f. Théorème 2.2.6(b), page 20), implique que :

$$\sum_{j=1}^{\bar{N}} \text{Var}(\Psi_{nj}) \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$, par suite :

$$\Theta_{n1} = o(1),$$

car :

$$\sum_{j=1}^n \text{Var}(\Psi_{nj}) \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$. Concernant le terme de covariance Θ_{n2} , nous avons vu dans la preuve du Théorème 2.2.6(b) (page 33) que :

$$\Theta_{n2} = \frac{2nh_n^d}{\left[\sum_{i=1}^n h_i^{d(1-\ell)}\right]^2} \sum_{1 \leq i < j \leq n} |A_{i,j}| \rightarrow 0, \quad (2.4.10)$$

lorsque $n \rightarrow \infty$. Notons que dans le contrôle de Θ_{n1} , la condition (2.2.6) n'est pas nécessaire si $\ell \geq 1/2$. En effet, d'après (2.4.9), le terme de variance Θ_{n1} s'écrit :

$$\Theta_{n1} = \frac{B_{n,d(1-\ell)}^{-2}}{n} \sum_{i=\bar{N}+1}^n \left(\frac{h_i}{h_n}\right)^{d(1-2\ell)} h_i^{-d} \text{Var} K\left(\frac{x - X_i}{h_i}\right).$$

Puisque h_n est décroissante et $\ell \geq \frac{1}{2}$ alors la convergence :

$$h_i^{-d} \text{Var} K\left(\frac{x - X_i}{h_i}\right) \rightarrow f(x) \int_{\mathbb{R}^d} K^2(x) dx,$$

lorsque $i \rightarrow \infty$ implique que :

$$\Theta_{n1} \leq \frac{\text{Cste}(n - \bar{N})}{n B_{n,d(1-\ell)}^2},$$

où Cste désigne une constante dont la connaissance de la valeur exacte n'est pas importante. Puisque nous avons :

$$n - \bar{N} \leq \varsigma_n + \tau_n,$$

alors $\Theta_{n1} \rightarrow 0$ lorsque $n \rightarrow \infty$. Maintenant nous établissons la normalité asymptotique de S_{n1} . D'après la décomposition :

$$\sqrt{nh_n^d} [f_n^\ell(x) - \mathbb{E}f_n^\ell(x)] = S_{n1} + S_{n2} + S_{n3},$$

nous pouvons écrire :

$$\begin{aligned} \text{Var}(S_{n1}) &= nh_n^d \text{Var} f_n^\ell(x) - [\text{Var} S_{n2} + \text{Var} S_{n3}] \\ &\quad - 2[\text{Cov}(S_{n1}, S_{n2}) + \text{Cov}(S_{n1}, S_{n3}) + \text{Cov}(S_{n2}, S_{n3})]. \end{aligned}$$

En utilisant les deux derniers résultats établis pour S_{n2} et S_{n3} , le Théorème 2.2.6 (a,b) ainsi que l'inégalité de Cauchy-Schwartz impliquent que :

$$\text{Var}(S_{n1}) \sim nh_n^d \text{Var} f_n^\ell(x) \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$. Nous avons également pour les gros blocs :

$$|T_{nm}| \leq \frac{\varsigma_n \|K\|_\infty}{\sqrt{nh_n^d B_{n,d(1-\ell)}}}. \quad (2.4.11)$$

Considérons des variables aléatoires $Z_{n1}, \dots, Z_{nr_n}$ iid ayant la même loi de probabilité que T_{nm} . Nous avons donc $EZ_{n1} = 0$. Notons $\Phi_{T_{nm}}$ la fonction caractéristique (f.c) de T_{nm} . Alors $\Phi_{T_{nm}}^{r_n}$ est la f.c de la variable aléatoire $\sum_{m=1}^{r_n} Z_{nm}$. Établir la normalité asymptotique de S_{n1} , revient à montrer que les variables $\sum_{m=1}^{r_n} Z_{nm}$ et $\sum_{m=1}^{r_n} T_{nm}$ ont la même loi de probabilité et que cette dernière est gaussienne. Par le Lemme de Volkonskii-Rosanov [89] (Lemme A.1.3), nous avons :

$$\left| E \prod_{m=1}^{r_n} e^{itT_{nm}} - \prod_{m=1}^{r_n} E e^{itT_{nm}} \right| \leq 8(r_n - 1)\alpha(\tau_n) \leq \rho_0 r_n e^{-\rho_1 \tau_n} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Il vient que :

$$\left| E \prod_{m=1}^{r_n} e^{itT_{nm}} - \Phi_{T_n}^{r_n} \right| \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Par suite nous avons juste à montrer que $\Phi_{T_{nm}}^{r_n}$ converge vers la fonction caractéristique d'une variable aléatoire gaussienne. Pour cela nous procédons comme suit. Posons :

$$Z'_{nm} := \frac{Z_{nm}}{s_n},$$

où :

$$s_n^2 := \sum_{m=1}^{r_n} \text{Var} Z_{nm}.$$

Nous avons :

$$s_n^2 \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$. En effet :

$$s_n^2 = \sum_{m=1}^{r_n} \text{Var} T_{nm} \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$, car nous avons d'une part :

$$\text{Var} S_{n1} \rightarrow \sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$, et l'on peut montrer d'autre part, comme pour Δ_2 , que :

$$\sum_{1 \leq i < j \leq r_n} \text{Cov}(T_{ni}, T_{nj}) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Les variables Z'_{nm} sont iid, de plus :

$$\mathbb{E}Z'_{n1} = 0 \text{ et } \sum_{m=1}^{r_n} \text{Var}Z'_{nm} = 1.$$

Par les conditions de Lindeberg (c.f. Loève [62]), nous montrons seulement que pour tout $\varepsilon > 0$:

$$\sum_{m=1}^{r_n} \mathbb{E} \left(Z_{nm}'^2 \mathbf{1}_{\{|Z'_{nm}| > \varepsilon\}} \right) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. En appliquant l'inégalité de Markov nous avons grâce à (2.4.11) que :

$$\begin{aligned} \sum_{m=1}^{r_n} \mathbb{E} \left(Z_{nm}'^2 \mathbf{1}_{\{|Z'_{nm}| > \varepsilon\}} \right) &= \sum_{m=1}^{r_n} \mathbb{E} \left(\frac{T_{nm}^2}{s_n^2} \mathbf{1}_{\{|T_{nm}| > \varepsilon s_n\}} \right) \\ &\leq \frac{\zeta_n^2 \|K\|_\infty^2}{nh_n^d B_{n,d(1-\ell)}^2 s_n^2} \sum_{m=1}^{r_n} P(|T_{nm}| > \varepsilon s_n) \\ &\leq \frac{\zeta_n^2 \|K\|_\infty^2}{\varepsilon^2 s_n^2 nh_n^d B_{n,d(1-\ell)}^2} \\ &\leq \left[\frac{\zeta_n}{\sqrt{nh_n^d}} \cdot \frac{\|K\|_\infty \varepsilon^{-1}}{s_n B_{n,d(1-\ell)}} \right]^2 \rightarrow 0, \end{aligned}$$

lorsque $n \rightarrow \infty$, en vertu de la convergence $s_n^2 \rightarrow \sigma_\ell^2(x)$, de la relation (2.2.7) et des hypothèses **H.2**. ■

2.4.5 Preuve de la Proposition 2.2.9

(a) Nous utilisons la technique usuelle dite des blocs en posant :

$$Z_i = \frac{1}{B_{n,d(1-\ell)} h_n^{d(1-\ell)} h_i^{d\ell}} \left[K \left(\frac{x - X_i}{h_i} \right) - \mathbb{E} K \left(\frac{x - X_i}{h_i} \right) \right],$$

et considérons les suites p_n et q_n définies par :

$$p_n := \lfloor p_0 \ln n \rfloor, p_0 > 0 \text{ et } q_n := \left\lfloor \frac{n}{2p_n} \right\rfloor,$$

puis définissons enfin les sommes de blocs :

$$S'_n = \sum_{j=1}^{q_n} \tilde{V}_n(2j-1), \quad S''_n = \sum_{j=1}^{q_n} \tilde{V}_n(2j) \quad \text{et} \quad S'''_n = \frac{1}{n} \sum_{k=2p_n q_n + 1}^n Z_k,$$

avec :

$$\tilde{V}_n(j) = \frac{1}{n} \sum_{k=(j-1)p_n+1}^{jp_n} Z_k, j = 1, \dots, 2q_n. \quad (2.4.12)$$

Nous remarquons alors que :

$$S'_n + S''_n + S'''_n = f_n^\ell(x) - \mathbb{E}f_n^\ell(x).$$

L'influence du terme de reste S'''_n est négligeable dans les calculs, donc il suffit de montrer la convergence presque sûre de $S'_n + S''_n$. Ainsi, nous avons, pour tout $\varepsilon > 0$:

$$P\left(|S'_n + S''_n| > \varepsilon\right) \leq P\left(|S'_n| > \frac{\varepsilon}{2}\right) + P\left(|S''_n| > \frac{\varepsilon}{2}\right).$$

Les deux termes de droite de cette inégalité sont traités de la même façon, donc on étudie seulement le premier. Nous utilisons de manière récursive le Lemme de couplage de Rio [76] (Lemme A.1.5) pour approximer les variables $\tilde{V}_n(j)$ par des variables $V_n^*(j)$ iid de même loi que les $\tilde{V}_n(2j-1)$ vérifiant :

$$\mathbb{E}\left|V_n^*(2j-1) - \tilde{V}_n(2j-1)\right| \leq 4\left\|\tilde{V}_n(2j-1)\right\|_\infty \alpha(p_n). \quad (2.4.13)$$

Donc en vertu de (2.4.12), nous pouvons écrire, pour tous $\varepsilon, \kappa > 0$:

$$\begin{aligned} P\left(|S'_n| > \frac{\varepsilon}{2}\right) &\leq P\left[\left|\sum_{j=1}^{q_n} \tilde{V}_n(2j-1) - V_n^*(2j-1)\right| > \frac{\varepsilon\kappa}{2(1+\kappa)}\right] \\ &+ P\left[\left|\sum_{j=1}^{q_n} V_n^*(2j-1)\right| > \frac{\varepsilon}{2(1+\kappa)}\right]. \end{aligned}$$

Puisque K est une fonction positive et h_n une suite décroissante la relation (2.4.12) implique que :

$$\max_{1 \leq j \leq 2q_n} \left|\tilde{V}_n(2j-1)\right| \leq \frac{p_n \|K\|_\infty}{B_{n,d(1-\ell)} n h_n^d}.$$

Grâce au choix logarithmique de p_n , cette dernière inégalité utilisée conjointement avec l'inégalité de Markov, la relation (2.4.13), et l'hypothèse **H.3(i)** permettent de déduire que :

$$\begin{aligned} P\left[\left|\sum_{j=1}^{q_n} \tilde{V}_n(2j-1) - V_n^*(2j-1)\right| > \frac{\varepsilon\kappa}{2(1+\kappa)}\right] &\leq \frac{4\|K\|_\infty(1+\kappa)}{\varepsilon\kappa B_{n,d(1-\ell)} h_n^d} \alpha(p_n) \\ &\leq \frac{4\|K\|_\infty(1+\kappa)}{\varepsilon\kappa} \rho_0 \frac{e^{-\rho_1 p_0 \ln n}}{B_{n,d(1-\ell)} h_n^d}. \end{aligned} \quad (2.4.14)$$

Maintenant pour ce qui est du terme iid, posons pour tout $\eta > 0$:

$$\varepsilon_n = \eta \sqrt{\frac{\ln n}{n h_n^d}} \text{ et } \lambda_n = \sqrt{n h_n^d \ln n}. \quad (2.4.15)$$

Puisque h_n est décroissante et vérifie la condition :

$$\frac{n h_n^d}{(\ln n)^3} \rightarrow +\infty,$$

lorsque $n \rightarrow \infty$, alors par le choix logarithmique de p_n nous avons pour tout $j \geq 1$:

$$|\lambda_n V_n^*(j)| \leq \frac{\|K\|_\infty p_n}{B_{n,d(1-\ell)}} \left(\frac{\ln n}{n h_n^d} \right)^{\frac{1}{2}} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Ainsi pour n assez grand, on a :

$$|\lambda_n V_n^*(j)| \leq \frac{1}{2},$$

par suite :

$$\exp [\pm \lambda_n V_n^*(j)] \leq 1 \pm \lambda_n V_n^*(j) + [\lambda_n V_n^*(j)]^2.$$

Appliquons de nouveau l'inégalité de Markov :

$$\begin{aligned} P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\varepsilon_n}{2(1+\kappa)} \right] &= P \left\{ \exp \left[\sum_{j=1}^{q_n} \lambda_n V_n^*(2j-1) \right] > \exp \left[\frac{\lambda_n \varepsilon_n}{2(1+\kappa)} \right] \right\} \\ &+ P \left\{ \exp \left[- \sum_{j=1}^{q_n} \lambda_n V_n^*(2j-1) \right] > \exp \left[\frac{\lambda_n \varepsilon_n}{2(1+\kappa)} \right] \right\} \\ &\leq 2 \exp \left[- \frac{\lambda_n \varepsilon_n}{2(1+\kappa)} + \lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \right]. \end{aligned}$$

Mais nous avons :

$$\sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \leq \frac{1}{n^2} \left[\sum_{k=1}^n \text{Var} Z_k + \sum_{k=1}^n \sum_{\substack{k' \neq k \\ k'=1}}^n |\text{Cov}(Z_k, Z_{k'})| \right].$$

Nous avons vu dans la preuve du Théorème 2.2.6(b) que sous les hypothèses **H.1** – **H.3** :

$$\frac{h_n^d}{n^2} \sum_{k=1}^n \text{Var} Z_k \rightarrow \sigma_\ell^2(x) \quad \text{et} \quad \frac{h_n^d}{n^2} \sum_{k=1}^n \sum_{\substack{k' \neq k \\ k'=1}}^n |\text{Cov}(Z_k, Z_{k'})| \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Donc :

$$\lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \leq \sigma_\ell^2(x) \ln n [1 + o(1)].$$

Par conséquent :

$$P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\eta}{2(1+\kappa)} \sqrt{\frac{\ln n}{n h_n^d}} \right] \leq 2 \exp \left\{ \left[- \frac{\eta}{2(1+\kappa)} + \sigma_\ell^2(x) (1 + o(1)) \right] \ln n \right\}.$$

En vertu de (2.4.14), il vient que :

$$\begin{aligned} P \left(\frac{1}{n} \left| \sum_{i=1}^n Z_i - \mathbb{E} Z_i \right| > \eta \sqrt{\frac{\ln n}{n h_n^d}} \right) &\leq \frac{8 \|K\|_\infty (1+\kappa)}{\eta \kappa} \rho_0 \frac{n^{-\rho_1 p_0}}{B_{n,d(1-\ell)} h_n^d} \\ &+ 4 \exp \left\{ - \left[\frac{\eta}{2(1+\kappa)} - \sigma_\ell^2(x) (1 + o(1)) \right] \ln n \right\} \\ &:= u_n + v_n. \end{aligned}$$

Pour v_n , nous avons :

$$\sum_{n=1}^{\infty} v_n < \infty \text{ pour tout } \eta > 2(1 + \kappa) \left[1 + \sigma_{\ell}^2(x) \right].$$

Ensuite, la condition :

$$\frac{nh_n^d}{(\ln n)^3} \rightarrow +\infty,$$

lorsque $n \rightarrow \infty$, implique qu'il existe $n_0 \geq 0$ tel que pour tout $n \geq n_0$ on ait :

$$u_n \leq \frac{8\|K\|_{\infty}(1 + \kappa)}{\eta\kappa} \rho_0 \frac{n^{1-\rho_1 p_0}}{(\ln n)^3}.$$

Alors :

$$\sum_{n=1}^{\infty} u_n < \infty, \text{ si } p_0 > \frac{2}{\rho_1}.$$

Finalement (a) découle du fait que u_n et v_n sont les termes généraux de séries convergentes, par application du lemme de Borel-Cantelli.

(b) Ce résultat est une conséquence de (a) et du Théorème 2.2.1(a). ■

2.4.6 Preuve du Théorème 2.2.10

(a) Nous procédons de la même manière que Bosq et Blanke [15], pour la démonstration de la convergence presque sûre de l'estimateur non récursif de Parzen-Rosenblatt. Pour le terme de biais, d'une part, puisque $f \in C_d^2(b)$, par les hypothèses **H.1**, une application de la formule de Taylor donne :

$$\begin{aligned} |Ef_n^{\ell}(x) - f(x)| &\leq \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{k=1}^n h_i^{d(1-\ell)} \left| \int_{\mathbb{R}^d} K(z) (f(x - h_i z) - f(x)) dz \right| \\ &\leq \frac{h_n^2}{B_{n,d(1-\ell)}} \frac{1}{n} \sum_{k=1}^n \left(\frac{h_k}{h_n} \right)^{d(1-\ell)+2} \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta h_i) \int_{\mathbb{R}^d} |z_i z_j| K(z) dz \\ &\leq \frac{h_n^2}{B_{n,d(1-\ell)}} \frac{1}{n} \sum_{k=1}^n \left(\frac{h_k}{h_n} \right)^{d(1-\ell)+2} \frac{b}{2} \sum_{i,j=1}^d \int_{\mathbb{R}^d} |z_i z_j| K(z) dz. \end{aligned}$$

Donc :

$$\overline{\lim}_{n \rightarrow \infty} h_n^{-2} \sup_{x \in \mathbb{R}^d} |Ef_n^{\ell}(x) - f(x)| \leq \frac{b}{2} \left[\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \right] \sum_{1 \leq i,j \leq d} \int_{\mathbb{R}^d} |v_i v_j| K(v) dv. \quad (2.4.16)$$

Ensuite, pour le terme stochastique, on pose :

$$D_n := \left\{ x : \|x\| \leq \frac{c_{(d)} n^{\mu}}{M_n} \right\},$$

où M_n est un réel que l'on donnera la valeur par la suite. On couvre ce compact par M_n^d hypercubes $D_{k,n}$ centrées en $x_{k,n}$ définies par :

$$D_{k,n} = \left\{ x : \|x - x_{k,n}\| \leq \frac{c_{(d)} n^\mu}{M_n} \right\}, \text{ pour } 1 \leq k \leq M_n^d,$$

avec :

$$D_{k,n}^\circ \cap D_{k',n}^\circ \neq \emptyset, \text{ pour } 1 \leq k \neq k' \leq M_n^d.$$

On a :

$$\begin{aligned} \sup_{\|x\| \leq c_d n^\mu} |f_n^\ell(x) - \mathbb{E} f_n^\ell(x)| &\leq \max_{1 \leq k \leq M_n^d} \sup_{x \in D_{k,n}} |f_n^\ell(x) - f_n^\ell(x_{k,n})| \\ &\quad + \max_{1 \leq k \leq M_n^d} |f_n^\ell(x_{k,n}) - \mathbb{E} f_n^\ell(x_{k,n})| \\ &\quad + \max_{1 \leq k \leq M_n^d} \sup_{x \in D_{k,n}} |\mathbb{E} f_n^\ell(x_{k,n}) - \mathbb{E} f_n^\ell(x)| \\ &= A_1 + A_2 + A_3. \end{aligned}$$

Puisque K est Lipschitzienne, alors il existe $L > 0$ tel que :

$$|f_n^\ell(x) - f_n^\ell(x_{k,n})| \leq \frac{L \|x - x_{k,n}\|}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^{1+d\ell}}, \quad x \in D_{k,n}, 1 \leq k \leq M_n.$$

Donc :

$$\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} (A_1 + A_3) \leq \frac{2Lc_d n^{\frac{2}{d+4}+\mu} \sum_{i=1}^n h_i^{-(1+d\ell)}}{M_n (\ln n)^{\frac{2}{d+4}} \sum_{i=1}^n h_i^{d(1-\ell)}} = \frac{2Lc_d n^{\frac{2}{d+4}+\mu} B_{n, -(1+d\ell)}}{M_n (\ln n)^{\frac{2}{d+4}} h_n^{d+1} B_{n, d(1-\ell)}}.$$

On choisit alors :

$$M_n = \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} n^\mu h_n^{-(1+d)} \log_m n, \quad m \geq 1.$$

Donc :

$$\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} (A_1 + A_3) = o(1). \quad (2.4.17)$$

Pour le contrôle de A_2 , écrivons pour tout $\varepsilon > 0$:

$$P \left(\max_{1 \leq k \leq M_n^d} |f_n^\ell(x_{k,n}) - \mathbb{E} f_n^\ell(x_{k,n})| > \varepsilon \right) \leq \sum_{k=1}^{M_n^d} P \left(|f_n^\ell(x_{k,n}) - \mathbb{E} f_n^\ell(x_{k,n})| > \varepsilon \right).$$

Ensuite, nous reprenons les mêmes étapes de la preuve de la Proposition 2.2.9(a), mais cette fois pour contrôler la somme des moments d'ordre 2 des variables $V_n^*(2j-1)$, nous majorons par la borne uniforme suivante :

$$\frac{1}{n^2} \sum_{k=1}^n \text{Var} Z_k \leq \frac{\|K\|_2^2 \|f\|_\infty B_{n, d(1-2\ell)}}{n h_n^d B_{n, d(1-\ell)}^2}.$$

Il vient alors que :

$$\lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \leq \frac{\|K\|_2^2 \|f\|_\infty \beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} \ln n [1 + o(1)],$$

où $o(1)$ uniforme en $x_{k,n}$. Par conséquent :

$$\begin{aligned} P \left[\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x_{k,n}) - \mathbb{E} f_n^\ell(x_{k,n})| > \eta \right] &\leq \frac{8M_n^d \|K\|_\infty (1+\kappa) c^{\frac{d}{2}}}{\eta \kappa} \rho_0 \frac{n^{\frac{d+2}{d+4} - \rho_1 p_0}}{B_{n,d(1-\ell)} (\ln n)^{\frac{d+2}{d+4}}} \\ &+ 4M_n^d \exp \left\{ - \left[\frac{\eta}{2(1+\kappa)} - \frac{\|K\|_2^2 \|f\|_\infty \beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} (1 + o(1)) \right] \ln n \right\}, \end{aligned}$$

avec $o(1)$ uniforme en $x_{k,n}$. Donc le choix :

$$M_n = \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} n^\mu h_n^{-(1+d)} \log_m n, \quad m \geq 1,$$

implique que la série :

$$\sum_{n=1}^{\infty} P \left[\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} A_2 \right] < \infty,$$

si l'on choisit :

$$p_0 > \frac{\mu d + d + 3}{\rho_1} \text{ et } \eta > 2(1+\kappa) \left[\mu d + d + 1 + \frac{\|K\|_2^2 \|f\|_\infty \beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} \right].$$

On obtient ainsi le résultat (a) du Théorème 2.2.10 par application du Lemme de Borel-Cantelli et en utilisant les relations (2.4.16), (2.4.17).

(b) Nous avons :

$$\begin{aligned} \sup_{x \in \mathbb{R}^d} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x) - f(x)| &\leq \sup_{\|x\| \leq c_d n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x) - f(x)| \\ &+ \sup_{\|x\| > c_d n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x)| + \sup_{\|x\| > c_d n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} f(x) \\ &:= B_1 + B_2 + B_3. \end{aligned}$$

B_1 est étudié en (a). Pour le terme B_3 , nous avons :

$$B_3 \leq \frac{\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}}}{c_{(d)} n^\mu} \sup_{\|x\| > c_{(d)} n^\mu} \|x\| f(x).$$

D'une part puisque $\mu > 2$ alors :

$$\frac{\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}}}{c_{(d)} n^\mu} \rightarrow 0,$$

lorsque $n \rightarrow \infty$, et d'autre part puisque f est décroissante à partir d'un certain rang et intégrable alors :

$$\sup_{\|x\| > c_{(d)} n^\mu} \|x\| f(x) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Ainsi :

$$B_3 \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Enfin pour étudier B_2 , on considère les événements :

$$E_1 := \left\{ \forall \eta > 0, \sup_{\|x\| > c_{(d)} n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x)| > \eta \right\} \text{ et } E_2 := \left\{ \sup_{1 \leq j \leq n} \|X_j\| > \frac{c_{(d)} n^\mu}{2} \right\}.$$

On a pour tout $\eta > 0$:

$$\begin{aligned} \sum_{n=1}^{\infty} P \left[\sup_{\|x\| > c_{(d)} n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x)| > \eta \right] &= \sum_{n=1}^{\infty} P(E_1 \cap E_2) + \sum_{n=1}^{\infty} P(E_1 \cap E_2^c) \\ &:= P_1 + P_2. \end{aligned}$$

On note $[-c_K, c_K]$ le support de K . Puisque $c_{(d)} n^\mu \rightarrow \infty$, lorsque $n \rightarrow \infty$, alors il existe $n_0 \geq 0$ tel que pour tout $n \geq n_0$ on ait :

$$\frac{c_{(d)} n^\mu}{2h_1} > c_K.$$

Ainsi pour tout x tel que $\|x\| > c_{(d)} n^\mu$ et $n \geq n_0$, on a par l'inégalité triangulaire et la décroissance de h_n que :

$$\left\| \frac{x - X_j}{h_j} \right\| > \frac{c_{(d)} n^\mu}{2h_1} > c_K, \text{ si } \sup_{1 \leq j \leq n} \|X_j\| \leq \frac{c_{(d)} n^\mu}{2},$$

ce qui entraîne du fait que K est à support compact que pour tout $n \geq n_0$ et pour tout $1 \leq j \leq n$, que :

$$K \left(\frac{x - X_j}{h_j} \right) = 0.$$

Ainsi pour tout $n \geq n_0$, on a :

$$E_2^c \subset \left\{ \sup_{\|x\| > c_{(d)} n^\mu} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} |f_n^\ell(x)| = 0 \right\}.$$

Donc pour tout $n \geq n_0$, on a :

$$E_1 \cap E_2^c = \emptyset \text{ et } E_1 \subset E_2.$$

C'est-à-dire :

$$P_2 = \sum_{n=1}^{n_0} P(E_1 \cap E_2^c) < \infty \text{ et } P_1 \leq \sum_{n=1}^{n_0} P(E_1 \cap E_2) + \sum_{n=n_0}^{\infty} P(E_2).$$

Le premier terme du second membre de P_1 est fini, ensuite par l'inégalité de Markov nous avons :

$$\sum_{n=n_0}^{\infty} P(E_2) \leq \sum_{n=n_0}^{\infty} P\left[\|X_0\| > \frac{c_{(d)}n^{\mu-1}}{2}\right] \leq \frac{c_{(d)}E\|X_0\|}{2} \sum_{n=n_0}^{\infty} n^{1-\mu}.$$

Donc, il en résulte grâce au Lemme de Borel-Cantelli, que :

$$B_2 \rightarrow 0 \text{ p.s,}$$

lorsque $n \rightarrow \infty$, si $\mu > 2$, donc (b) est prouvé. ■

2.4.7 Preuve du Théorème 2.3.1

(a) Pour le choix :

$$h_n = C_n n^{-\nu}, C_n \downarrow c > 0, 0 < \nu < \frac{1}{d+2},$$

la partie variable en ℓ de la variance de nos estimateurs s'écrit :

$$\frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} = \frac{[1 - \nu d(1 - \ell)]^2}{1 - \nu d(1 - 2\ell)} = \frac{\nu^2 d^2 \ell^2 + 2\nu d(1 - \nu d)\ell + (1 - \nu d)^2}{2\nu d\ell + 1 - \nu d} =: F(\ell).$$

F est définie et dérivable sur $[0, 1]$ puisque $0 < \nu < \frac{1}{d+2}$ et pour tout $\ell \in [0, 1]$,

$$F'(\ell) = \frac{2\nu^2 d^2 \ell(\nu d\ell + 1 - \nu d)}{(2\nu d\ell + 1 - \nu d)^2} =: \frac{N(\ell)}{D(\ell)}.$$

$N(\ell)$ s'annule en 0 et en $\ell = \frac{\nu d - 1}{d\nu} < 0$, donc $N(\ell)$ est positif sur $[0, 1]$. Par suite $\text{Var} f_n^\ell(x)$ est une fonction croissante en ℓ et on conclut grâce à la Définition 2.5(i).

Pour ce qui est du critère basé sur le biais, la partie variable en ℓ de celui-ci s'écrit :

$$\frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} = \frac{1 - \nu d(1 - \ell)}{1 - \nu d(1 - \ell) - 2\nu} =: G(\ell).$$

Le dénominateur de $G(\ell)$ s'annule pour :

$$1 - \ell = \frac{2\nu + 1}{\nu d},$$

ce qui est impossible car :

$$\nu(d+2) < 1 \Rightarrow 2\nu + 1 > \nu d.$$

Donc la fonction G est définie, positive et dérivable sur $[0, 1]$ et pour tout $\ell \in [0, 1]$,

$$G'(\ell) = \frac{-2\nu^2 d}{(1 - \nu d(1 - \ell) - 2\nu)^2} < 0.$$

(a) est donc prouvé.

(b) On note BCA, VA et EQMA respectivement les biais au carré, variance et erreur quadratique moyenne asymptotiques. On a pour $f_n^\ell(x)$ comme pour $f_n^{\text{PR}}(x)$:

$$\text{EQMA} \sim h_n^4 \text{BCA} + \frac{\text{VA}}{nh_n^d}.$$

EQMA est donc minimal pour :

$$h_n \sim \left(\frac{d\text{VA}}{4\text{BCA}} \right)^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}.$$

Ainsi, pour tout $\ell \in [0, 1]$, la constante c qui minimise $\text{EQMA} f_n^\ell(x)$ vaut :

$$c_{\min} [f_n^\ell(x)] = \left[\frac{d(2 + d\ell)f(x) \|K\|_2^2}{8(4 + d)b_f^2(x)} \right]^{\frac{1}{d+4}}.$$

Ce qui entraîne pour ces choix de constantes que pour tout $\ell \in [0, 1]$:

$$\text{EQM} f_n^\ell(x) \sim \frac{(4 + d\ell)^2}{(2 + d\ell)^{\frac{2(2+d)}{d+4}}} [b_f^2(x)]^{\frac{d}{d+4}} \left[\frac{f(x) \|K\|_2^2}{2(4 + d)} \right]^{\frac{4}{d+4}} \left[\left(\frac{d}{4} \right)^{\frac{4}{d+4}} + \left(\frac{d}{4} \right)^{-\frac{d}{d+4}} \right] n^{-\frac{4}{d+4}}.$$

Ainsi, pour avoir le meilleur choix de ℓ dans $[0, 1]$ au sens de l'EQM, il faudra minimiser la fonction :

$$R(\ell) = (4 + d\ell)^2 (2 + d\ell)^{-\frac{2(2+d)}{d+4}}.$$

R est définie et dérivable sur $[0, 1]$ et :

$$R'(\ell) = \frac{4d^2(d\ell + 4)(\ell - 1)}{(d + 4)(d\ell + 2)^{\frac{8+3d}{d+4}}} \leq 0, \quad \forall \ell \in [0, 1].$$

Donc, R est décroissante sur $[0, 1]$. ■

2.4.8 Preuve du Théorème 2.3.2

(a) Rappelons tout d'abord le résultat classique suivant :

sous les hypothèses **H.1**, alors :

$$nh_n^d \text{Var} f_n^{\text{PR}}(x) \rightarrow f(x) \int_{\mathbb{R}^d} K^2(u) du,$$

lorsque $n \rightarrow \infty$.

On a donc pour le "pire" des estimateurs (au sens de la variance)

$$\lim_{n \rightarrow \infty} \frac{\text{Var} f_n^1(x)}{\text{Var} f_n^{\text{PR}}(x)} = \frac{\beta_{-d}}{\beta_0^2}.$$

où :

$$h_n = C_n n^{-\nu}, 0 < \nu \leq \frac{1}{d+2} \text{ et } \beta_r = \frac{1}{1 - \nu r}.$$

Donc :

$$\frac{\beta_{-d}}{\beta_0^2} = \frac{1}{1 + \nu d} < 1.$$

Ainsi, $f_n^1(x)$ est meilleur au sens de la variance que $f_n^{\text{PR}}(x)$. En utilisant le Théorème 2.3.1(a), on obtient le résultat.

(b) Nous avons aussi besoin du résultat classique suivant :

si **H.1**, **H.2** sont vraies, alors :

$$h_n^{-4} \left[\mathbb{E} f_n^{\text{PR}}(x) - f(x) \right]^2 \rightarrow b_f^2(x),$$

lorsque $n \rightarrow \infty$. On a alors pour le meilleur des estimateurs (au sens du biais) :

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E} f_n^1(x) - f(x)}{\mathbb{E} f_n^{\text{PR}}(x) - f(x)} = \frac{\beta_2}{\beta_0}.$$

Avec le choix :

$$h_n = C_n n^{-\nu}, 0 < \nu \leq \frac{1}{d+2},$$

nous obtenons :

$$\frac{\beta_2}{\beta_0} = \frac{1}{1 - \nu(d+2)} > 1.$$

Ainsi, $f_n^{\text{PR}}(x)$ est meilleur au sens du biais que $f_n^1(x)$. On conclut aussi grâce au Théorème 2.3.1(a).

(c) On reprend les notations de la preuve du Théorème 2.3.1(b). On a :

$$c_{min} [f_n^1(x)] = \left[\frac{d(2+d)f(x) \|K\|_2^2}{8(4+d)b_f^2(x)} \right]^{\frac{1}{d+4}} \text{ et } c_{min} [f_n^{PR}(x)] = \left[\frac{f(x) \|K\|_2^2 d}{4b_f^2(x)} \right]^{\frac{1}{d+4}}.$$

Ces choix de constantes entraînent que :

$$n^{\frac{4}{d+4}} \text{EQM} f_n^1(x) \rightarrow \frac{(4+d)^2}{(2+d)^{\frac{2(2+d)}{d+4}}} [b_f^2(x)]^{\frac{d}{d+4}} \left[\frac{f(x) \|K\|_2^2}{2(4+d)} \right]^{\frac{4}{d+4}} \left[\left(\frac{d}{4} \right)^{\frac{4}{d+4}} + \left(\frac{d}{4} \right)^{-\frac{d}{d+4}} \right]$$

et

$$n^{\frac{4}{d+4}} \text{EQM} f_n^{PR}(x) \rightarrow [b_f^2(x)]^{\frac{d}{d+4}} [f(x) \|K\|_2^2]^{\frac{4}{d+4}} \left[\left(\frac{d}{4} \right)^{\frac{4}{d+4}} + \left(\frac{d}{4} \right)^{-\frac{d}{d+4}} \right],$$

lorsque $n \rightarrow \infty$. Donc :

$$\lim_{n \rightarrow \infty} \frac{\text{EQM} f_n^1(x)}{\text{EQM} f_n^{PR}(x)} > 1 \Leftrightarrow F(d) := (d+2) \ln \left(\frac{d+4}{d+2} \right) - 2 \ln 2 > 0.$$

Ceci est vrai pour tout $d \geq 1$ puisque nous avons :

$$F(0) = 0, \quad F'(+\infty) = 0 \text{ et } F''(d) = -\frac{4}{(d+2)(d+4)} < 0.$$

Le résultat suit grâce au Théorème 2.3.1(b).

Notons que cette comparaison théorique de nos estimateurs est possible, grâce à la connaissance de la forme explicite des β_r , pour $h_n = n^{-\nu}$, ce qui n'est pas le cas pour d'autres choix de fenêtres.

En résumé, nous avons les représentations graphiques suivantes, des constantes asymptotiques en fonction de ℓ , faites avec le logiciel **R**, à partir des valeurs :

$$\beta_r = \frac{d+4}{d+4-r},$$

obtenus avec le choix de fenêtre :

$$h_n = cn^{-\frac{1}{d+4}}, c > 0.$$

* Biais au carré et variance pour $d = 1$:

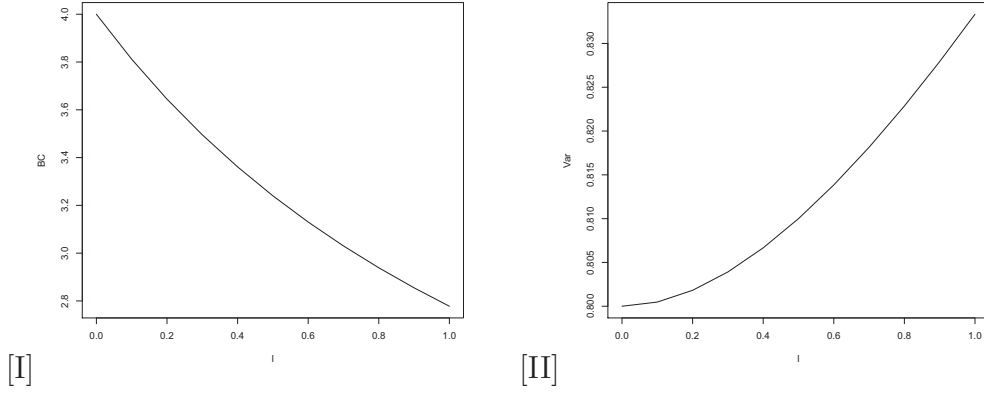


FIG. 2.1 – Constantes asymptotiques pour les Biais et variance.

[I] : Rep. graphique de $\left(\frac{\beta_{3-\ell}}{\beta_{1-\ell}}\right)^2$. [II] : Rep. graphique de $\frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2}$.

* EQM :

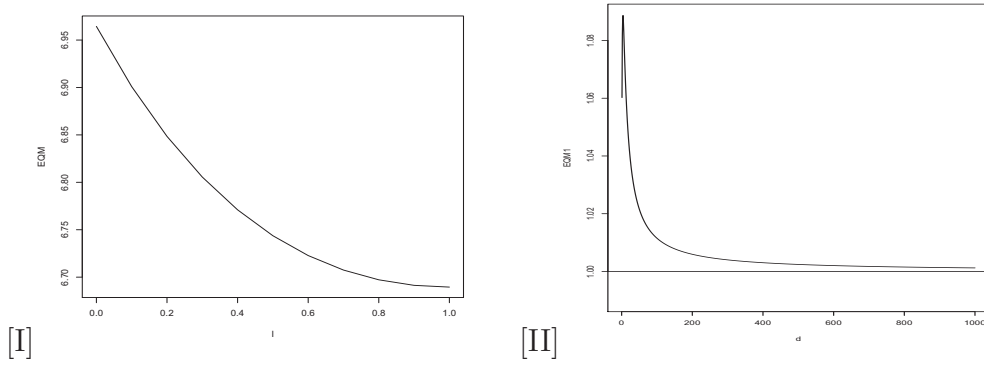


FIG. 2.2 – Constantes asymptotiques pour les EQM.

[I] : Rep. gr. ($d = 1$) de : $\frac{(4+\ell)^2}{(2+\ell)^{\frac{6}{5}}}$. [II] : Rep. gr. de : $\frac{\text{EQM}f_n^1(x)}{\text{EQM}f_n^{PR}(x)}$ en fct. de d .

Dans les deux premières figures, nous avons représenté les parties variables en ℓ des BC (FIGURE 2.4[I]) et variance (FIGURE 2.4[II]) asymptotiques de $f_n^\ell(x)$ en fonction de ℓ . Ainsi, on constate que pour ce qui est de la variance (resp. du BC), en plus de sa décroissance (resp. croissance), la courbe est au dessous (resp. au dessus) de 1. Donc notre famille est préférable (resp. non préférable) à l'estimateur à noyau usuelle au sens de la variance (resp. du biais), et son efficacité est décroissante (resp. croissante) en fonction de ℓ selon le critère de la variance (resp. du biais), ceci est en accord avec les résultats obtenus avec le calcul direct avec d quelconque. Nous retrouvons également à partir de la représentation graphique de la partie variable en ℓ de l'EQM (FIGURE 2.2[I]), la croissance de l'efficacité de $f_n^\ell(x)$ en fonction de ℓ . Enfin, la FIGURE 2.2[II] montre l'efficacité de l'estimateur de Parzen-Rosenblatt par rapport à $f_n^1(x)$ au sens de l'EQM. Ce dernier étant optimal sur la famille $(f_n^\ell(x))$ au sens de l'EQM, on déduit que l'estimateur de Parzen-Rosenblatt est meilleur au sens de l'EQM que la famille $(f_n^\ell(x))$. ■

Nous, nous proposons également d'étudier numériquement en fonction de n les valeurs :

$$\beta_r, \quad r \in \{-1, 1/2, 1, 2, 5/2, 3\},$$

obtenues avec le choix de fenêtre :

$$h_n = c \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}, \quad c > 0.$$

Ces valeurs sont celles prises par :

$$\beta_{3-\ell}, \beta_{1-\ell}, \beta_{1-2\ell}, \quad \text{pour } \ell = 0, 1/2, 1,$$

dans le cas i.i.d. et de la convergence p.s. de l'estimateur.

* Représentation graphiques des β_r

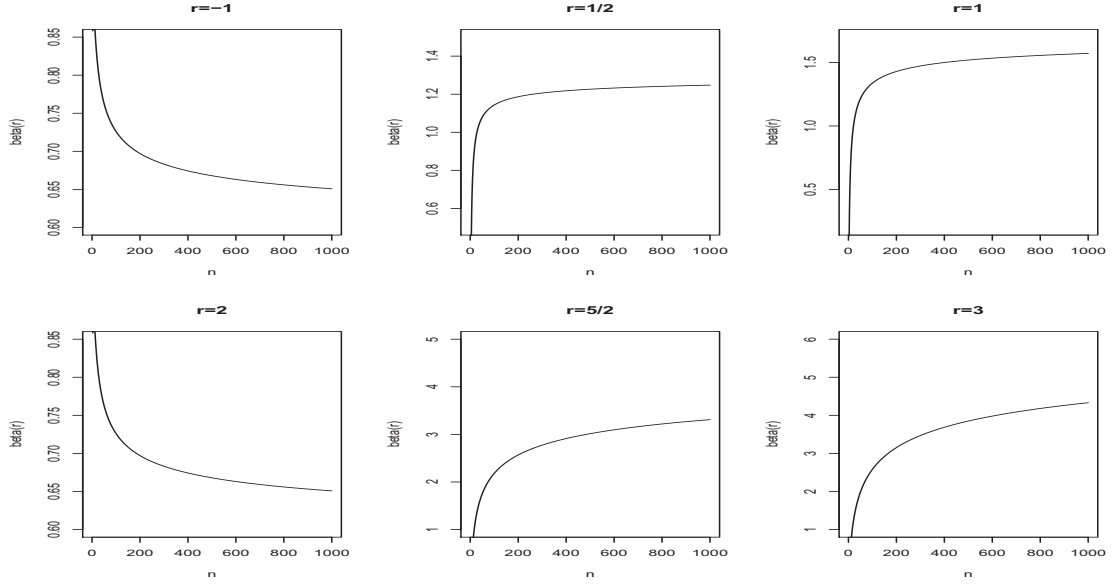


FIG. 2.3 – Représentation graphique des β_r .

Pour ce même choix de h_n , nous reprenons également les représentations graphiques faites à la FIGURE 2.4 en utilisant cette fois, des valeurs numériques des $\beta_{3-\ell}, \beta_{1-\ell}, \beta_{1-2\ell}$ obtenues avec $n = 10000$.

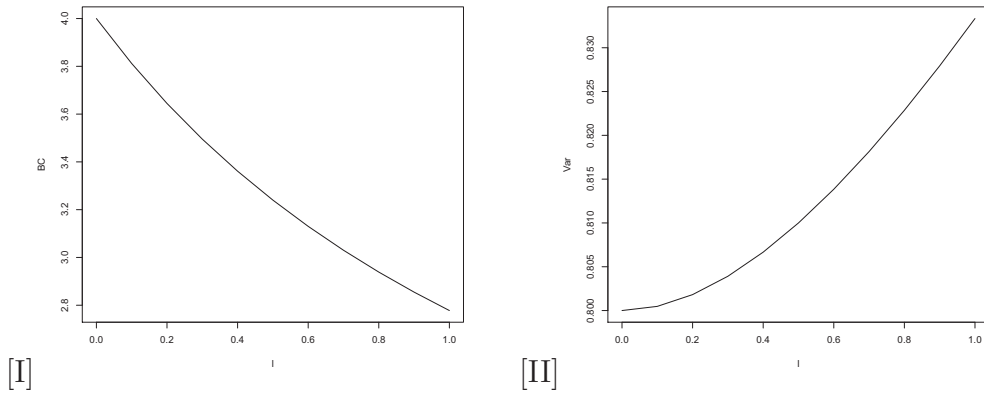


FIG. 2.4 – Constantes asymptotiques pour les Biais et variance.

[I] : Rep. graphique de $\left(\frac{\beta_{3-\ell}}{\beta_{1-\ell}}\right)^2$. [II] : Rep. graphique de $\frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2}$.

Chapitre 3

Estimateurs récursifs de la régression

Sommaire

3.1	Cadre d'étude	57
3.1.1	Notations et hypothèses	58
3.2	Convergence presque sûre	59
3.2.1	Cas iid	60
3.2.2	Cas dépendant	61
3.3	Convergence en moyenne quadratique	62
3.4	Normalité asymptotique	63
3.5	Preuves	64
3.5.1	Preuve du Théorème 3.2.1	64
3.5.2	Preuve de Théorème 3.2.2	68
3.5.3	Preuve du Théorème 3.2.3	71
3.5.4	Preuve du Théorème 3.3.1	74
3.5.5	Preuve du Théorème 3.4.1	83

3.1 Cadre d'étude

Soit $\{\zeta_t = (X_t, Y_t), t \in \mathbb{N}\}$ un processus stochastique bivarié, défini sur un espace de probabilité $(\Omega, \mathcal{A}, P)$, à valeurs dans $\mathbb{R}^d \times \mathbb{R}^{d'}$ ($d \geq 1, d' \geq 1$), tel que les (X_t, Y_t) ont la même densité de probabilité jointe notée f^* . Nous allons estimer de manière récursive une

version de la fonction de régression de $m(Y_0)$ en X_0 , définie par :

$$r(x) := \begin{cases} \mathbb{E}(m(Y_0) | X_0 = x) = \frac{\int_{\mathbb{R}^{d'}} m(y) f^*(x, y) dy}{f(x)} := \frac{\varphi(x)}{f(x)}, & \text{si } f(x) > 0 \\ \mathbb{E}m(Y_0), & \text{si } f(x) = 0, \end{cases}$$

où m est une fonction positive (choisie par le statisticien), Borélienne de $\mathbb{R}^{d'}$ à valeurs dans $\mathbb{R}$ telle que $\omega \mapsto m^2(Y_t(\omega))$ soit P -intégrable et f la densité marginale des X_i , en se basant sur les observations $\zeta_1, \dots, \zeta_n$. L'intérêt de faire cette étude de manière récursive a été expliqué précédemment. Nous verrons dans le prochain chapitre que la connaissance du paramètre de régression r est très utile pour prédire des valeurs futures de $m(Y)$ sachant $X = x$.

Dans la suite de ce chapitre, on considérera seulement les $x \in \mathbb{R}^d$ tels que $f(x) > 0$.

3.1.1 Notations et hypothèses

La fonction $r(x)$ est estimée par la famille d'estimateurs récursifs (1.3.2) rappelée ci-dessous :

$$r_n^\ell(x) := \frac{\varphi_n^\ell(x)}{f_n^\ell(x)} \text{ où } \varphi_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{(1-\ell)d}} \sum_{i=1}^n \frac{m(Y_i)}{h_i^{d\ell}} K\left(\frac{x - X_i}{h_i}\right), (\ell \in [0, 1]).$$

L'estimateur $r_n^\ell(x)$ peut se calculer de manière récursive par :

$$r_{n+1}^\ell(x) = \frac{\left[\sum_{i=1}^n h_i^{d(1-\ell)}\right] \varphi_n^\ell(x) + \left[\sum_{i=1}^{n+1} h_i^{d(1-\ell)}\right] m(Y_{n+1}) K_{n+1}^\ell(x - X_{n+1})}{\left[\sum_{i=1}^n h_i^{d(1-\ell)}\right] f_n^\ell(x) + \left[\sum_{i=1}^{n+1} h_i^{d(1-\ell)}\right] K_{n+1}^\ell(x - X_{n+1})},$$

où $K_i^\ell(\cdot)$ est défini dans (2.2.1) à la page 13.

En plus des hypothèses faites au Chapitre 2, nous introduisons des hypothèses supplémentaires pour obtenir le comportement asymptotique de notre famille d'estimateurs de la régression.

Hypothèses H.4.

(i) $f, \varphi \in C_d^2(b)$;

(ii) La fonction :

$$\mathbb{E} \left[m^2(Y_0) | X_0 = \cdot \right] f(\cdot)$$

est continue et bornée ;

(iii) Il existe $\lambda > 0, \chi > 0$ tels que :

$$\mathbb{E} \exp(\lambda |m(Y_0)|^\chi) < \infty;$$

(iv) Pour chaque $k \neq k'$, $(\zeta_k, \zeta_{k'})$ admet une densité de probabilité $f_{(\zeta_k, \zeta_{k'})}$, telle que :

$$G := \sup_{|k-k'| \geq 1} \sup_{(s,t) \in \mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}} \int_{\mathbb{R}^{d'}} |G_{k,k'}(s, u, t, v)| \, dudv < \infty,$$

où :

$$G_{k,k'}(s, u, t, v) = f_{(\zeta_k, \zeta_{k'})}(s, u, t, v) - f^*(s, u) f^*(t, v).$$

(v) (ξ_t) est un processus Géométriquement fortement mélangeant (GFM) avec :

$$\alpha_\xi(u) \leq \rho_0 e^{-\rho_1 u}, u \geq 1, \rho_0 > 0 \text{ et } \rho_1 > 0.$$

Les hypothèses **H.4** sont classiques en estimation non paramétrique de la régression. Elles sont utilisées par exemple par Bosq et Blanke [15] pour l'étude de l'estimateur de Nadaraya-Watson. La condition **H.4(iii)** est clairement vérifiée si m est bornée. Il est important de remarquer aussi que grâce à un argument de concavité de la fonction $(\ln x)^{\frac{p}{\chi}}$, la condition **H.4(iii)** entraîne que pour tous $p \geq 1, n \geq 2$:

$$\mathbb{E} \left(\max_{1 \leq i \leq n} |m(Y_i)|^p \right) = O \left[(\ln n)^{\frac{p}{\chi}} \right]. \quad (3.1.1)$$

Cette relation est également utilisée par Bosq et Cheze-Payaud [16] pour établir la convergence en moyenne quadratique de l'estimateur non récursif. La condition **H.4(iv)** est facilement satisfaite pour des densités f^* et f uniformément bornées. La condition **H.4(v)** est vérifiée par les processus linéaires et les chaînes de Markov. En outre, il est bien connu que les processus ARMA stationnaires sont GFM sous des conditions assez légères. Un exemple typique de processus GFM est le signal plus bruit défini par :

$$X_n = S_n + \varepsilon_n,$$

où S_n est choisi par exemple comme un processus GARCH et ε_n une suite de variables aléatoires indépendantes et indépendantes de S_n .

3.2 Convergence presque sûre

Dans ce paragraphe, nous établissons la convergence presque sûre de la famille d'estimateurs récursifs de la régression $r_n^\ell(x)$, en commençant par le cas iid.

3.2.1 Cas iid

Dans le théorème suivant, on montre que $r_n^\ell(x)$ converge presque sûrement avec la même vitesse que la famille d'estimateurs de la densité. Il s'agit d'un résultat similaire (avec une borne différente) au résultat (b) du Théorème 2.2.4. Nous reprenons également la notation :

$$\sigma_\ell^2(x) := \frac{\beta_{d(1-2\ell)}}{\beta_{d(1-\ell)}^2} f(x) \int_{\mathbb{R}^d} K^2(x) dx, \quad (3.2.1)$$

définie en (2.2.3).

Théorème 3.2.1. *Supposons que les hypothèses **H.1**, **H.2** et **H.4(i)** – (iii) sont satisfaites. Si pour tout $\alpha \geq 0$:*

$$\frac{nh_n^d}{(\ln n)^{1+\frac{2}{\alpha}} (\ln \ln n)^{2(\alpha+1)}} \rightarrow \infty, \text{ lorsque } n \rightarrow \infty, \text{ et } \lim_{n \rightarrow \infty} \frac{\ln h_n}{\ln n} < \infty, \quad (3.2.2)$$

et si K vérifie la condition :

$$\int_{\mathbb{R}^d} \|u\|^3 K(u) < \infty,$$

alors le choix :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

implique que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4}} \left[r_n^\ell(x) - r(x) \right] = C_0(x, K, \varphi, f, \ell) \text{ p.s.},$$

où :

$$C_0(x, K, \varphi, f, \ell) := \frac{1}{f(x)} \left[c^{\frac{-d}{2}} \sigma_\ell(x) \sqrt{2V(x)} + c^2 \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_{\varphi, f}(x) \right],$$

avec :

$$b_{\varphi, f}(x) := \frac{1}{2} \sum_{i,j=1}^d \left[\frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x) - r(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] \int_{\mathbb{R}^d} z_i z_j K(z) dz, \quad (3.2.3)$$

$$V(x) := E[m^2(Y_0) | X_0 = x] - r^2(x),$$

et σ_ℓ^2 défini dans (3.2.1).

Ce théorème généralise le résultat de Roussas [78] pour $\ell = 1/2$. Notons que dans le cas non récursif, la vitesse de convergence obtenue est du type $\left(\frac{\ln n}{n}\right)^{\frac{2}{d+4}}$. La section suivante établit que dans le cas dépendant, nous atteignons ce même type de vitesse.

3.2.2 Cas dépendant

Dans cette partie, le processus ζ est supposé GFM, on établit alors la convergence presque sûre de $r_n^\ell(x)$.

Théorème 3.2.2. *Sous les hypothèses **H.1** – **H.4**, si K vérifie la condition :*

$$\int_{\mathbb{R}^d} \|u\|^3 K(u) < \infty,$$

alors le choix :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

implique que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} \left| r_n^\ell(x) - r(x) \right| \leq C_1(x, K, \varphi, f, \ell) \text{ p.s.},$$

où :

$$C_1(x, K, \varphi, f, \ell) := \frac{1}{f(x)} \left\{ 2c^{-\frac{d}{2}} \left[1 + \sigma_\ell^2(x) V(x) \right] + \frac{c^2 \beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_{\varphi, f}(x) \right\},$$

avec σ_ℓ^2 , $b_{\varphi, f}(x)$ et $V(x)$ définies respectivement dans (3.2.1) et (3.2.3).

La vitesse de convergence obtenue dans ce théorème est la même que celle obtenue par Bosq et Blanke [15] pour l'estimateur de Nadaraya-Watson avec la borne :

$$\frac{1}{f(x)} \left\{ 2c^{-\frac{d}{2}} V(x) f(x) \|K\|_2^2 + c^2 b_{\varphi, f}(x) \right\}.$$

Nous donnons maintenant un résultat uniforme sur la convergence presque sûre de notre famille d'estimateurs de la régression.

Théorème 3.2.3. *Soit D un compact de $\mathbb{R}^d$. Sous les hypothèses **H.1**–**H.4**, si K est Lipschitzienne et satisfait la condition :*

$$\int_{\mathbb{R}^d} \|u\|^3 K(u) < \infty,$$

et si $\inf_{x \in D} f(x) > 0$, alors le choix :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0$$

entraîne que :

$$\sup_{x \in D} |r_n^\ell(x) - r(x)| = O \left[\left(\frac{\ln n}{n} \right)^{\frac{2}{d+4}} (\ln n)^{\frac{1}{\chi}} \right].$$

Notons que l'on peut obtenir un résultat analogue sous une condition plus faible que **H4(iv)**. Cette étude est effectuée au Chapitre 4, pour l'étude de la convergence des prédicteurs non paramétriques.

3.3 Convergence en moyenne quadratique

On s'intéresse maintenant à la convergence en moyenne quadratique de $r_n^\ell(x)$. Soient alors les hypothèses suivantes :

Hypothèses H'4.

(iv) : Pour chaque couple $k \neq k'$, on a :

$$\begin{aligned} (*) \quad G_1 &:= \sup_{|k-k'| \geq 1} \left\| f_{(X_k, X_{k'})|Y_{k'}}(s, t|v) - f(s) f_{X_{k'}|Y_{k'}}(t|v) \right\|_\infty < \infty; \\ (**) \quad G_2 &:= \sup_{|k-k'| \geq 1} \left\| f_{(X_k, X_{k'})|(Y_k, Y_{k'})}(s|u, t|v) - f_{X_k|Y_k}(s|u) f_{X_{k'}|Y_{k'}}(t|v) \right\|_\infty < \infty. \end{aligned}$$

(v) (ζ_t) est un processus $2 - \alpha$ -mélangeant avec :

$$\alpha_\zeta^{(2)}(k) \leq \rho_0 k^{-\rho_1}, \quad k \geq 1,$$

pour deux constantes :

$$\rho_0 > 0 \text{ et } \rho_1 > \max[8, 2(d+2)].$$

La condition **H'4(iv)** est également utilisée par Roussas et Tran [79] pour l'étude de la normalité asymptotique de l'estimateur de Devroye-Wagner (cas $\ell = 1$). Même si elle est plus restrictive par rapport à **H.4(iv)**, elle permet de contrôler les termes de covariance de $\varphi_n^\ell(x)$ sans avoir recours à sa version tronquée.

Le théorème suivant établit l'EQM asymptotique exacte de la famille $(r_n^\ell(x))$.

Théorème 3.3.1. *Sous les hypothèses **H.1-H.4**, et en remplaçant **H.4(iv)-(v)** par **H'.4(iv)-(v)**, le choix :*

$$h_n = C_n n^{-\frac{1}{d+4}}, C_n \downarrow c > 0$$

implique que pour tout $\ell \in [0, 1]$:

$$n^{\frac{4}{d+4}} E \left[r_n^\ell(x) - r(x) \right]^2 \rightarrow C_2(x, K, \varphi, f, \ell),$$

lorsque $n \rightarrow \infty$ avec :

$$C_2(x, K, \varphi, f, \ell) := \left\{ \frac{c^2 (4 + d\ell) [r(x)b_f(x) + b_\varphi(x)]}{f(x)(2 + d\ell)} \right\}^2 + \frac{(4 + d\ell)^2 \|K\|_2^2 V(x)}{2c^d(4 + d)(2 + d\ell)f(x)},$$

où :

$$b_\varphi(x) := \frac{1}{2} \sum_{1 \leq i, j \leq d} \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x) \int_{\mathbb{R}^d} v_i v_j K(v) dv,$$

$V(x)$ donné en (3.2.3) et où nous rappelons que $b_f(x)$ est obtenu en remplaçant φ par f dans l'expression de $b_\varphi(x)$.

3.4 Normalité asymptotique

Le théorème suivant établit la normalité asymptotique de $r_n^\ell(x)$.

Théorème 3.4.1. *Supposons les hypothèses **H.1 – H.4** satisfaites et que la suite h_n est telle que :*

$$nh_n^{d+4} \rightarrow 0 \text{ et } (\ln n)^{\frac{1}{x}} h_n^d \rightarrow 0,$$

lorsque $n \rightarrow \infty$, et pour toutes suites d'entiers u_n et v_n , on a :

$$u_n \sim v_n \Rightarrow h_{u_n} \sim h_{v_n}. \quad (3.4.1)$$

S'il existe un réel positif $\varsigma_0 > 4$ tel que :

$$\frac{nh_n^d}{(\ln n)^{\varsigma_0}} \rightarrow +\infty,$$

lorsque $n \rightarrow \infty$, alors :

$$\sqrt{nh_n^d} \left[r_n^\ell(x) - r(x) \right] \xrightarrow{\mathcal{L}} \mathcal{N} \left[0, \frac{\beta_{d(1-2\ell)} \|K\|_2^2 V(x)}{\beta_{d(1-\ell)}^2 f(x)} \right],$$

lorsque $n \rightarrow \infty$, où $V(x)$ est défini en (3.2.3).

3.5 Preuves

3.5.1 Preuve du Théorème 3.2.1

La preuve se fait en deux phases. D'abord, on montre que le terme résiduel entre l'estimateur et sa version tronquée est négligeable. Ensuite, on établit que la version tronquée de l'estimateur converge vers $r(x)$ avec la bonne vitesse. Pour cela, nous procédons comme Bosq et Blanke [15] en écrivant :

$$r_n^\ell(x) - r(x) = \frac{\tilde{\varphi}_n^\ell(x) - r(x)f_n^\ell(x)}{f_n^\ell(x)} + \frac{\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)}{f_n^\ell(x)},$$

où $\tilde{\varphi}_n^\ell(x)$ désigne la version tronquée de $\varphi_n^\ell(x)$ définie par :

$$\tilde{\varphi}_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{(1-\ell)d}} \sum_{i=1}^n \frac{m(Y_i)}{h_i^{d\ell}} \mathbf{1}_{\{|m(Y_i)| \leq b_n\}} K\left(\frac{x - X_i}{h_i}\right), \quad (3.5.1)$$

avec b_n une suite réelle tendant vers $+\infty$ lorsque $n \rightarrow \infty$. Pour étudier le terme résiduel, si l'on choisit :

$$b_n = (\delta \ln n)^{\frac{1}{\chi}},$$

alors nous avons pour tout $\varepsilon > 0$:

$$\begin{aligned} P \left[\left| \varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x) \right| > \varepsilon \left(\frac{\ln \ln n}{n} \right)^{\frac{2}{d+4}} \right] &\leq P \left[\frac{\sum_{i=1}^n |m(Y_i)| \mathbf{1}_{\{|m(Y_i)| > b_n\}} h_i^{-d\ell} K\left(\frac{x - X_i}{h_i}\right)}{\left(\frac{\ln \ln n}{n} \right)^{\frac{2}{d+4}} \sum_{i=1}^n h_i^{d(1-\ell)}} > \varepsilon \right] \\ &\leq P(\cup_{i=1}^n \{|m(Y_i)| > b_n\}) \\ &\leq \sum_{i=1}^n P(|m(Y_i)| > b_n) = nP(|m(Y_0)| > b_n) \\ &\leq \mathbb{E} e^{\lambda |m(Y_0)|^\chi} n^{1-\lambda\delta}. \end{aligned}$$

En utilisant l'hypothèse **H.4**(iii), il vient que pour tout $\varepsilon > 0$:

$$\sum_{n=1}^{\infty} P \left[\left| \varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x) \right| > \varepsilon \left(\frac{\ln \ln n}{n} \right)^{\frac{2}{d+4}} \right] < \infty,$$

si $\delta > \frac{2}{\chi}$ et le Lemme de Borel-Cantelli implique que :

$$\left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4}} \left| \varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x) \right| \rightarrow 0 \text{ p.s.},$$

lorsque $n \rightarrow \infty$. On en déduit que le terme résiduel est négligeable, en vertu de la convergence

$$f_n^\ell(x) \rightarrow f(x) \text{ p.s.},$$

Étudions maintenant la convergence du terme principal. Ce dernier se décompose de la manière suivante :

$$\begin{aligned}\tilde{\varphi}_n^\ell(x) - r(x)f_n^\ell(x) &= \left\{ \tilde{\varphi}_n^\ell(x) - r(x)f_n^\ell(x) - \mathbb{E} \left[\tilde{\varphi}_n^\ell(x) - r(x)f_n^\ell(x) \right] \right\} \\ &\quad + \left\{ \mathbb{E} \left[\tilde{\varphi}_n^\ell(x) - r(x)f_n^\ell(x) \right] \right\} \\ &=: I_1^* + I_2^*.\end{aligned}\tag{3.5.2}$$

On commence par étudier I_1^* . Pour cela, on pose :

$$Z_i^* = W_{n,i} - \mathbb{E}W_{n,i},$$

$$\text{avec : } W_{n,i} := \frac{K \left(\frac{x - X_i}{h_i} \right) [m(Y_i) - r(x)] \mathbf{1}_{\{|m(Y_i)| \leq b_n\}}}{h_i^{d\ell}},$$

et :

$$S_n^* = \sum_{i=1}^n Z_i^*.$$

Il est alors important de remarquer que :

$$I_1^* = \frac{S_n^*}{nB_{n,d(1-\ell)}h_n^{d(1-\ell)}}.$$

Soit enfin :

$$V_n^* = \sum_{i=1}^n \mathbb{E}Z_i^{*2}.\tag{3.5.3}$$

La première étape consiste à établir que :

$$\frac{S_n^*}{\sqrt{2V_n^* \ln \ln V_n^*}} \rightarrow 1 \text{ p.s.},$$

lorsque $n \rightarrow \infty$. Nous avons :

$$\begin{aligned}\frac{V_n^*}{nh_n^{d(1-2\ell)}B_{n,d(1-\ell)}} &= \frac{nh_n^d}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \sum_{k=1}^n \left\{ h_k^{-2d\ell} \mathbb{E}K^2 \left(\frac{x-X_0}{h_k} \right) [m(Y_0) - r(x)]^2 \right. \\ &\quad \left. + h_k^{-2d\ell} \mathbb{E}K^2 \left(\frac{x-X_0}{h_k} \right) m(Y_0) [2r(x) - m(Y_0)] \mathbf{1}_{\{|m(Y_0)| > b_n\}} \right\} \\ &\leq \frac{nh_n^d}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \sum_{k=1}^n \left[h_k^{d(1-2\ell)} \int_{\mathbb{R}^d} \frac{1}{h_k^d} K^2 \left(\frac{x-u}{h_k} \right) V(u) f(u) du \right] \\ &\quad + \frac{\|K\|_\infty^2 \left\{ \mathbb{E}m^2(Y_0) [2r(x) - m(Y_0)]^2 P(|m(Y_0)| > b_n) \right\}^{\frac{1}{2}} nh_n^d \sum_{k=1}^n h_k^{-2d\ell}}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \\ &\leq \frac{nh_n^d}{B_{n,d(1-\ell)}^2 n^2 h_n^d} \sum_{k=1}^n \left[\left(\frac{h_k}{h_n} \right)^{d(1-2\ell)} \int_{\mathbb{R}^d} \frac{1}{h_k^d} K^2 \left(\frac{x-u}{h_k} \right) V(u) f(u) du \right] \\ &\quad + \frac{\|K\|_\infty^2 \left\{ \mathbb{E}m^2(Y_0) [2r(x) - m(Y_0)]^2 P(|m(Y_0)| > b_n) \right\}^{\frac{1}{2}} B_{n,-2d\ell}}{h_n^d B_{n,d(1-\ell)}^2} \\ &:= D_1 + D_2.\end{aligned}$$

Les hypothèses **H.4(ii)**, **(iii)**, le théorème de convergence dominée et le lemme de Bochner impliquent d'une part que :

$$\int_{\mathbb{R}^d} \frac{1}{h_k^d} K^2 \left(\frac{x-u}{h_k} \right) V(u) f(u) du \rightarrow f(x) V(x) \|K\|_2^2,$$

lorsque $k \rightarrow \infty$, par suite l'hypothèse **H.2(ii)**, et le Lemme 2.4.1 permettent d'en déduire que :

$$D_1 \rightarrow \sigma_\ell^2(x) V(x),$$

lorsque $n \rightarrow \infty$. D'autre part, si l'on choisi :

$$b_n = (\delta \ln n)^{\frac{1}{\chi}} \text{ avec } \delta > \frac{2}{\lambda},$$

alors en utilisant la relation (3.1.1) et l'inégalité de Markov on obtient :

$$D_2 = O \left[\frac{\exp \left(-\frac{\lambda b_n^\chi}{2} \right) (\ln n)^{\frac{2}{\chi}} B_{n,-2d\ell}}{h_n^d B_{n,d(1-\ell)}^2} \right] = O \left[\frac{n^{\frac{d}{d+4} - \frac{\lambda \delta}{2}}}{(\ln n)^{\frac{d}{d+4} - \frac{2}{\chi}}} \right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Il vient alors que :

$$V_n^* \sim n h_n^{d(1-2\ell)} \beta_{d(1-2\ell)} f(x) V(x) \int_{\mathbb{R}^d} K^2(u) du, \quad (3.5.4)$$

lorsque $n \rightarrow \infty$. Maintenant par la condition (3.2.2), nous avons :

$$\frac{n h_n^d (\ln n)^{-\frac{2}{\chi}}}{\ln \left[n h_n^{d(1-2\ell)} \right] \left\{ \ln \ln \left[n h_n^{d(1-2\ell)} \right] \right\}^{2(\alpha+1)}} \rightarrow \infty,$$

lorsque $n \rightarrow \infty$. Par suite le choix :

$$b_n = (\delta \ln n)^{\frac{1}{\chi}}$$

implique qu'il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, on ait :

$$\frac{n h_n^d (\ln n)^{-\frac{2}{\chi}}}{\ln \left[n h_n^{d(1-2\ell)} \right] \left\{ \ln \ln \left[n h_n^{d(1-2\ell)} \right] \right\}^{2(\alpha+1)}} > \|K\|_\infty^2 h_n^{-2d\ell} (\delta \ln n)^{\frac{2}{\chi}} \geq Z_n^{*2}.$$

Avec les mêmes arguments que pour la preuve du Théorème 2.2.4(a), on applique le Théorème A.2.1 pour obtenir :

$$\frac{S_n^*}{(2V_n^* \ln \ln V_n^*)^{\frac{1}{2}}} \rightarrow 1 \text{ p.s.},$$

lorsque $n \rightarrow \infty$. La deuxième étape consiste à établir la convergence de I_1^* à partir de cette dernière convergence. En utilisant pour cela, la condition (3.2.2) et le fait que :

$$\frac{\left\{nh_n^{d(1-2\ell)} \ln \ln [nh_n^{d(1-2\ell)}]\right\}^{\frac{1}{2}} B_{n,d(1-\ell)}}{(2V_n^* \ln \ln V_n^*)^{\frac{1}{2}}} \rightarrow \frac{\beta_{d(1-\ell)}}{\left[2\beta_{d(1-2\ell)}f(x)V(x) \int_{\mathbb{R}^d} K^2(u)du\right]^{\frac{1}{2}}},$$

lorsque $n \rightarrow \infty$, on déduit que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{nh_n^d}{\ln \ln n} \right)^{\frac{1}{2}} I_1^* = \frac{\left[2\beta_{d(1-2\ell)}f(x)V(x) \int_{\mathbb{R}^d} K^2(u)du\right]^{\frac{1}{2}}}{\beta_{d(1-\ell)}} \text{ p.s.} \quad (3.5.5)$$

Il faut étudier maintenant le terme I_2^* de (3.5.2). Nous procédons de la manière suivante :

$$\begin{aligned} I_2^* &= \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n h_i^{-d\ell} \int_{\mathbb{R}^d} K\left(\frac{x-u}{h_i}\right) [r(u) - r(x)] f(u) du \\ &= \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{k=i}^n h_i^{d(1-\ell)} \int_{\mathbb{R}^d} [r(x - h_i z) - r(x)] f(x - h_i z) K(z) dz. \end{aligned}$$

Les hypothèses **H.4(i)**, (ii) ainsi que la formule Taylor donnent :

$$\begin{aligned} I_2^* &= \frac{h_n^2}{B_{n,d(1-\ell)} n} \sum_{k=1}^n \left(\frac{h_k}{h_n}\right)^{d(1-\ell)+2} \\ &\quad \times \frac{1}{2} \sum_{i,j=1}^d \int_{\mathbb{R}^d} \left[\frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x - \theta h_k z) - r(x - \theta h_k z) \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta h_k z) \right] z_i z_j K(z) dz, \end{aligned}$$

avec $0 < \theta < 1$. Ensuite soit :

$$w_k = \sum_{i,j=1}^d \int_{\mathbb{R}^d} \left\{ \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x - \theta h_k z) - r(x - \theta h_k z) \frac{\partial^2 f}{\partial x_i \partial x_j}(x - \theta h_k z) \right\} z_i z_j K(z) dz.$$

Grâce aux hypothèses **H.1(ii)**, (iii), le théorème de convergence dominée implique que :

$$w_k \rightarrow \sum_{i,j=1}^d \left\{ \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x) - r(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right\} \int_{\mathbb{R}^d} z_i z_j K(z) dz,$$

lorsque $k \rightarrow \infty$. Donc par les hypothèses **H.2**, le Lemme 2.4.1 entraîne que :

$$h_n^{-2} I_2^* \rightarrow \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_{\varphi,f}(x),$$

lorsque $n \rightarrow \infty$. Ainsi, cette convergence utilisée conjointement avec (3.5.5) et le choix :

$$h_n = C_n \left(\frac{\ln \ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

permettent de conclure que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln \ln n} \right)^{\frac{2}{d+4}} [\tilde{\varphi}_n^\ell(x) - r(x) f_n^\ell(x)] = c^{\frac{-d}{2}} \sigma_\ell(x) \sqrt{2V(x)} + c^2 \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_{\varphi,f}(x) \text{ p.s.},$$

ce qui donne le résultat recherché en vertu de la convergence $f_n^\ell(x) \rightarrow f(x)$ p.s. ■

3.5.2 Preuve de Théorème 3.2.2

La preuve se fait avec les mêmes étapes que celles de la preuve du Théorème 3.2.1. Le terme résiduel se traite exactement comme pour la preuve du Théorème 3.2.1, en remplaçant $\ln \ln n$ par $\ln n$. Ensuite, on reprend la décomposition (3.5.2) du terme principal. Il suffit alors d'étudier le terme I_1^* , car le terme I_2^* ne dépend pas de la dépendance ou non des observations. Soit alors :

$$Z_{i,n} = W_{n,i} - EW_{n,i},$$

$$\text{avec : } W_{i,n} := \frac{K\left(\frac{x - X_i}{h_i}\right) [m(Y_i) - r(x)] \mathbf{1}_{\{|m(Y_i)| \leq b_n\}}}{B_{n,d(1-\ell)} h_n^{d(1-\ell)} h_i^{d\ell}}.$$

On procède de la même façon que pour la preuve de la Proposition 2.2.9 (c.f. page 42), en utilisant les mêmes définitions des suites p_n et q_n et des sommes partielles, mais en remplaçant les blocs de variables $\tilde{V}_n(j)$ par :

$$V_n(j) = \frac{1}{n} \sum_{k=(j-1)p_n+1}^{jp_n} Z_{k,n}, \quad j = 1, \dots, 2q_n. \quad (3.5.6)$$

Avec cette nouvelle définition des blocs, on reprend la relation :

$$\begin{aligned} P\left(\left|S'_n\right| > \frac{\varepsilon}{2}\right) &\leq P\left[\left|\sum_{j=1}^{q_n} V_n(2j-1) - V_n^*(2j-1)\right| > \frac{\varepsilon\kappa}{2(1+\kappa)}\right] \\ &+ P\left[\left|\sum_{j=1}^{q_n} V_n^*(2j-1)\right| > \frac{\varepsilon}{2(1+\kappa)}\right], \end{aligned}$$

où les $V_n^*(j)$ sont des variables aléatoires iid de même loi que les $V_n(j)$. Pour l'étude du terme de mélange de cette dernière décomposition, nous avons d'après (3.5.6) et la décroissance de h_n , que :

$$\max_{1 \leq j \leq 2q_n} |V_n(2j-1)| \leq \frac{2p_n b_n \|K\|_\infty}{B_{n,d(1-\ell)} n h_n^d}.$$

On peut alors écrire que :

$$P\left[\left|\sum_{j=1}^{q_n} V_n(2j-1) - V_n^*(2j-1)\right| > \frac{\varepsilon\kappa}{2(1+\kappa)}\right] \leq \frac{4\|K\|_\infty(1+\kappa)}{\varepsilon\kappa} \rho_0 \frac{b_n e^{-\rho_1 p_0 \ln n}}{B_{n,d(1-\ell)} h_n^d}.$$

Le second membre de cette dernière inégalité est le terme général d'une série convergente, compte-tenu des choix :

$$b_n = (\delta \ln n)^{\frac{1}{\chi}} \text{ avec } \delta > \frac{2}{\lambda} \text{ et } h_n = C_n \left(\frac{\ln n}{n}\right)^{\frac{1}{d+4}}, \quad C_n \downarrow c > 0, \quad (3.5.7)$$

Maintenant pour le terme iid, si l'on considère λ_n et ε_n définis dans (2.4.15), les choix (3.5.7) impliquent que :

$$|\lambda_n V_n^*(j)| \leq \frac{\|K\|_\infty p_n b_n}{B_{n,d(1-\ell)}} \left(\frac{\ln n}{n h_n^d} \right)^{\frac{1}{2}} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Ce qui permet de retrouver la majoration suivante (détaillée à la page 44) :

$$P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\varepsilon_n}{2(1+\kappa)} \right] \leq 2 \exp \left[-\frac{\lambda_n \varepsilon_n}{2(1+\kappa)} + \lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \right].$$

Ensuite, le contrôle de la somme des moments d'ordre 2 des $V_n^*(2j-1)$, se fait en utilisant le Lemme suivant (démontré en fin de preuve) :

Lemme 3.5.1. *sous les hypothèses **H.1-H.4**, nous avons :*

$$\frac{h_n^d}{n} \sum_{k=1}^n \text{Var} Z_{k,n} \rightarrow \sigma_\ell^2(x) V(x) \quad \text{et} \quad \frac{h_n^d}{n} \sum_{k=1, k \neq k'}^n \sum_{k'=1}^n \text{Cov}(Z_{k,n}, Z_{k',n}) \rightarrow 0,$$

lorsque $n \rightarrow \infty$.

Ainsi, le Lemme 3.5.1 entraîne que :

$$\lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \leq \sigma_\ell^2(x) V(x) \ln n [1 + o(1)]$$

lorsque $n \rightarrow \infty$. Par conséquent :

$$P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\eta}{2(1+\kappa)} \sqrt{\frac{\ln n}{n h_n^d}} \right] \leq 2 \exp \left\{ \left[-\frac{\eta}{2(1+\kappa)} + \sigma_\ell^2(x) V(x) [1 + o(1)] \right] \ln n \right\}.$$

Donc on conclut, comme dans la preuve de la Proposition 2.2.9, que les choix (3.5.7) entraînent que :

$$\overline{\lim}_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} I_1^* \leq 2c^{-\frac{d}{2}} [1 + \sigma_\ell^2(x) V(x)] \quad \text{p.s.} \quad (3.5.8)$$

Pour finir la preuve, il faut montrer le Lemme 3.5.1.

Preuve du Lemme 3.5.1

Pour le terme de variance, il suffit de remarquer que :

$$\frac{h_n^d}{n} \sum_{k=1}^n \text{Var} Z_{k,n} = \frac{V_n^*}{nh^{d(1-\ell)} B_{n,d(1-\ell)}},$$

où V_n^* est défini en (3.5.3) (c.f. page 65), puis utiliser l'équivalence (3.5.4). En ce qui concerne le terme de covariance, la procédure est identique à celle détaillé dans la preuve du Théorème 2.2.6 (c.f. page 33). Soit c_n une suite tendant vers l'infini, écrivons :

$$\begin{aligned} \frac{1}{n^2} \sum_{k=1}^n \sum_{k' \neq k}^n \text{Cov} (Z_{k,n}, Z_{k',n}) &\leq \frac{2 \left[\sum_{i=1}^{c_n} \sum_{p=1}^n A_{i+p,p} + \sum_{i=c_n+1}^{n-1} \sum_{p=1}^n A_{i+p,p} \right]}{\left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} \\ &:= L_1 + L_2, \end{aligned}$$

avec :

$$A_{i+p,p} = \frac{\left| \text{Cov} \left[K \left(\frac{x - X_{i+p}}{h_{i+p}} \right) m(Y_{i+p}) \mathbf{1}_{\{|m(Y_{i+p})| \leq b_n\}}, K \left(\frac{x - X_p}{h_p} \right) m(Y_p) \mathbf{1}_{\{|m(Y_p)| \leq b_n\}} \right] \right|}{h_{i+p}^{d\ell} h_p^{d\ell}}.$$

Par l'inégalité de Billingsley, et avec un calcul similaire que pour la preuve du Théorème 2.2.6, on obtient d'une part que :

$$nh_n^d L_2 = O \left(b_n^2 c_n^{1-\rho_1} h_n^{-d} \right),$$

et d'autre part, pour le terme L_1 , on a :

$$\begin{aligned} A_{i+p,p} &= \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^{d'}} \int_{\mathbb{R}^{d'}} K \left(\frac{x-s}{h_{i+p}} \right) K \left(\frac{x-t}{h_p} \right) \frac{m(u) \mathbf{1}_{\{|m(u)| \leq b_n\}} m(v) \mathbf{1}_{\{|m(v)| \leq b_n\}} G_{i+p,p}(s,u,t,v)}{(h_{i+p} h_p)^{d\ell} \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^2} ds dt du dv \right| \\ &\leq \text{Cste } b_n^2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} (h_{i+p} h_p)^{d(1-\ell)}. \end{aligned}$$

Ainsi, comme pour la preuve du Théorème 2.2.6 on en déduit que :

$$L_1 \leq \text{Cste } b_n^2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{k=1}^{c_n} \sum_{p=1}^{n-k} h_{p+k}^{d(1-\ell)} h_p^{d(1-\ell)} \leq \text{Cste } b_n^2 c_n \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{p=1}^n h_p^{2d(1-\ell)}.$$

Puisque b_n est logarithmique et :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

les mêmes techniques et choix de c_n suivant la dimension indiqués dans la preuve du Théorème 2.2.6 permettent de conclure que :

$$\frac{h_n^d}{n} \sum_{k=1, k \neq k'}^n \sum_{k'=1}^n \text{Cov}(Z_{k,n}, Z_{k',n}) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. ■

3.5.3 Preuve du Théorème 3.2.3

(a) Le terme résiduel entre l'estimateur et sa version tronquée se contrôle exactement comme pour le cas ponctuel. Pour l'étude de la version tronquée, considérons la décomposition suivante :

$$\sup_{x \in D} |\tilde{r}_n^\ell(x) - r(x)| \leq \frac{\sup_{x \in D} |\tilde{\varphi}_n^\ell(x) - \varphi(x)|}{\inf_{x \in D} f_n^\ell(x)} + \frac{\sup_{x \in D} |r(x)| \sup_{x \in D} |f_n^\ell(x) - f(x)|}{\inf_{x \in D} f_n^\ell(x)}.$$

Le second terme du second membre de cette dernière décomposition est déjà étudié au Théorème 2.2.10, en remarquant que $\sup_{x \in D} |r(x)|$ existe et $\inf_{x \in D} f(x) > 0$. Le premier terme se décompose à son tour en un terme de biais plus un terme stochastique, on obtient alors l'inégalité suivante :

$$|\tilde{\varphi}_n^\ell(x) - \varphi(x)| \leq |E\tilde{\varphi}_n^\ell(x) - \varphi(x)| + |\tilde{\varphi}_n^\ell(x) - E\tilde{\varphi}_n^\ell(x)|.$$

D'une part, le terme de biais s'écrit :

$$\begin{aligned} |E\tilde{\varphi}_n^\ell(x) - \varphi(x)| &\leq \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n h_i^{-d\ell} \left| \int_{\mathbb{R}^d} K\left(\frac{x-u}{h_i}\right) [\varphi(u) - \varphi(x)] du \right| \\ &\quad + \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{i=1}^n h_i^{-d\ell} E \left[|m(Y_0)| \mathbf{1}_{\{|m(Y_0)| \geq b_n\}} K\left(\frac{x-X_0}{h_i}\right) \right] \\ &:= J_1 + J_2. \end{aligned}$$

Par application de la formule de Taylor on a :

$$\begin{aligned} J_1 &= \frac{1}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{k=i}^n h_i^{d(1-\ell)} \left| \int_{\mathbb{R}^d} [\varphi(x - \theta h_i z) - \varphi(x)] K(z) dz \right| \\ &= \frac{h_n^2}{B_{n,d(1-\ell)}} \frac{1}{n} \sum_{k=1}^n \left(\frac{h_k}{h_n} \right)^{d(1-\ell)+2} \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 \varphi}{\partial x_i \partial x_j}(x - \theta h_i) \left| \int_{\mathbb{R}^d} z_i z_j K(z) dz \right| \\ &\leq \frac{h_n^2}{B_{n,d(1-\ell)}} \frac{1}{n} \sum_{k=1}^n \left(\frac{h_k}{h_n} \right)^{d(1-\ell)+2} \frac{b}{2} \sum_{i,j=1}^d \left| \int_{\mathbb{R}^d} z_i z_j K(z) dz \right|. \end{aligned}$$

Donc l'hypothèse **H.2.(ii)** et le Lemme 2.4.1 impliquent que :

$$\lim_{n \rightarrow \infty} h_n^{-2} \sup_{x \in \mathbb{R}^d} J_1 \leq \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \frac{b}{2} \sum_{i,j=1}^d \left| \int_{\mathbb{R}^d} z_i z_j K(z) dz \right|.$$

Ensuite :

$$\begin{aligned} J_2 &\leq \frac{b_n^{-1}}{\sum_{i=1}^n h_i^{d(1-\ell)}} \sum_{k=1}^n h_i^{d(1-\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K\left(\frac{x-u}{h_i}\right) E[m^2(Y_0)|X_0 = u] f(u) du \\ &= \frac{b_n^{-1}}{B_{n,d(1-\ell)}} \frac{1}{n} \sum_{k=1}^n \left(\frac{h_i}{h_n}\right)^{d(1-\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K\left(\frac{x-u}{h_i}\right) E[m^2(Y_0)|X_0 = u] f(u) du \\ &\leq b_n^{-1} \|f(x) E[m^2(Y_0)|X_0 = x]\|_{\infty}. \end{aligned}$$

Ainsi, par les hypothèses **H.2.(ii)**, **H.4.(ii)**, le théorème de convergence dominée, le Lemme de Bochner et le Lemme 2.4.1 nous avons :

$$b_n \lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}^d} J_2 \leq \|f(x) E[m^2(Y_0)|X_0 = x]\|_{\infty}.$$

Si l'on choisit :

$$b_n = (\delta \ln n)^{\frac{1}{\chi}}, \text{ avec } \delta > 0,$$

alors :

$$\lim_{n \rightarrow \infty} h_n^{-2} \sup_{x \in \mathbb{R}^d} |E\tilde{\varphi}_n^{\ell}(x) - \varphi(x)| \leq \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} \frac{b}{2} \sum_{i,j=1}^d \int_{\mathbb{R}^d} |z_i z_j| K(z) dz,$$

grâce à l'hypothèse **H.4(ii)** et le fait que $h_n b_n^{-2} \rightarrow 0$ lorsque $n \rightarrow \infty$. Maintenant, il faut étudier le terme stochastique :

$$|\tilde{\varphi}_n^{\ell}(x) - E\tilde{\varphi}_n^{\ell}(x)|.$$

On couvre le compact D par M_n^d hypercubes $D_{k,n}$ centrés en $x_{k,n}$ définis par :

$$D_{k,n} = \left\{ x : \|x - x_{k,n}\| \leq \frac{1}{M_n} \right\}, \text{ pour } 1 \leq k \leq M_n^d,$$

avec :

$$D_{k,n}^{\circ} \cap D_{k',n}^{\circ} \neq \emptyset, 1 \leq k \neq k' \leq M_n^d.$$

Ainsi :

$$\begin{aligned} \sup_{x \in D} |\tilde{\varphi}_n^{\ell}(x) - E\tilde{\varphi}_n^{\ell}(x)| &\leq \max_{1 \leq k \leq M_n^d} \sup_{x \in D_{k,n}} |\tilde{\varphi}_n^{\ell}(x) - \tilde{\varphi}_n^{\ell}(x_{k,n})| \\ &\quad + \max_{1 \leq k \leq M_n^d} |\tilde{\varphi}_n^{\ell}(x_{k,n}) - E\tilde{\varphi}_n^{\ell}(x_{k,n})| \\ &\quad + \max_{1 \leq k \leq M_n^d} \sup_{x \in D_{k,n}} |E\tilde{\varphi}_n^{\ell}(x_{k,n}) - E\tilde{\varphi}_n^{\ell}(x)| \\ &= A_1 + A_2 + A_3. \end{aligned}$$

De façon analogue que pour la preuve du Théorème 2.2.10 (avec $\mu = 0$), nous avons :

$$\left(\frac{n}{\ln n}\right)^{\frac{2}{d+4}} b_n (A_1 + A_3) \leq \frac{2b_n^2 L c_d n^{\frac{2}{d+4}} B_{n, -(1+d\ell)}}{M_n (\ln n)^{\frac{2}{d+4}} h_n^{d+1} B_{n, d(1-\ell)}} = o(1), \quad (3.5.9)$$

si l'on choisit :

$$M_n = \left(\frac{n}{\ln n}\right)^{\frac{2}{d+4}} h_n^{-(1+d)} b_n^2 \log_m n, \quad m \geq 1.$$

L'étude de A_2 , est identique à la preuve du Théorème 3.2.2, en construisant les blocs $V_n(j)$ à partir des variables $\tilde{Z}_i$ définies par :

$$\tilde{Z}_{i,k} = \tilde{W}_{i,k} - \mathbb{E} \tilde{W}_{i,k},$$

$$\text{où : } \tilde{W}_{i,k} = \frac{m(Y_i) \mathbf{1}_{\{|m(Y_i)| \leq b_n\}}}{B_{n, d(1-\ell)} h_n^{d(1-\ell)} h_i^{d\ell}} K\left(\frac{x_{k,n} - X_i}{h_i}\right),$$

Avec ces nouvelles définition, nous avons grâce au choix logarithmique de b_n que :

$$\max_{1 \leq j \leq 2q_n} |V_n(2j-1)| \leq \frac{2p_n b_n \|K\|_\infty}{B_{n, d(1-\ell)} n h_n^d},$$

ce qui permet de contrôler le terme de mélange. Pour ce qui est du terme iid, en redéfinissant les suites ε_n et λ_n de la manière suivante :

$$\varepsilon_n = \eta b_n \sqrt{\frac{\ln n}{n h_n^d}} \text{ et } \lambda_n = b_n^{-1} \sqrt{n h_n^d \ln n},$$

il vient que pour tout j :

$$|\lambda_n V_n^*(j)| \leq \frac{\|K\|_\infty p_n}{B_{n, d(1-\ell)}} \left(\frac{\ln n}{n h_n^d}\right)^{\frac{1}{2}} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. On retrouve ainsi la majoration :

$$P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\varepsilon_n}{2(1+\kappa)} \right] \leq 2 \exp \left[-\frac{\lambda_n \varepsilon_n}{2(1+\kappa)} + \lambda_n^2 \sum_{j=1}^{q_n} \mathbb{E} V_n^{*2}(2j-1) \right].$$

(c.f. page 44 pour les détails). Il faut maintenant contrôler, de manière uniforme, la somme des moments d'ordre 2 des $V_n^*(2j-1)$, qui intervient dans cette dernière inégalité. Pour cela, nous avons la majoration uniforme :

$$\frac{h_n^d}{n} \sum_{k=1}^{q_n} \text{Var} \tilde{Z}_{k,n} \leq \frac{\|K\|_2^2 \|f\|_\infty b_n^2}{B_{n, d(1-\ell)}^2} \times \frac{1}{n h_n^{d(1-2\ell)}} \sum_{i=1}^{q_n} h_i^{d(1-2\ell)},$$

et le terme de covariance converge vers 0 uniformément en $x_{k,n}$. Il est important de remarquer que, par définition de q_n et du fait que h_n est une suite positive, le facteur :

$$\frac{1}{nh_n^{d(1-2\ell)}} \sum_{i=1}^{q_n} h_i^{d(1-2\ell)},$$

qui intervient dans l'inégalité précédente converge vers une limite finie. Il vient alors que :

$$\lambda_n^2 \sum_{j=1}^{q_n} \text{EV}_n^{*2}(2j-1) \leq \text{Cste} \ln n [1 + o(1)],$$

où $o(1)$ uniforme en $x_{k,n}$. Par conséquent :

$$P \left[\left| \sum_{j=1}^{q_n} V_n^*(2j-1) \right| > \frac{\eta}{2(1+\kappa)} b_n \sqrt{\frac{\ln n}{nh_n^d}} \right] \leq 2 \exp \left\{ \left[-\frac{\eta}{2(1+\kappa)} + \text{Cste} [1 + o(1)] \right] \ln n \right\},$$

où $o(1)$ uniforme en $x_{k,n}$. On conclut alors comme pour la Preuve de la Proposition 2.2.9 que le terme :

$$\left(\frac{n}{\ln n} \right)^{\frac{2}{d+4}} b_n A_2^*,$$

est borné pour n assez grand presque sûrement. Le résultat recherché découle de cette dernière convergence et (3.5.9).

3.5.4 Preuve du Théorème 3.3.1

Nous procédons comme Bosq [14] et Cheze-Payaud [24] en décomposant l'EQM de $r_n^\ell(x)$ de la façon suivante :

$$\begin{aligned} \mathbb{E} \left[r_n^\ell(x) - r(x) \right]^2 &= \frac{r^2(x)}{f^2(x)} \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^2 - \frac{2r(x)}{f^2(x)} \mathbb{E} \left\{ \left[f_n^\ell(x) - f(x) \right] \left[\varphi_n^\ell(x) - \varphi(x) \right] \right\} \\ &\quad + \frac{1}{f^2(x)} \mathbb{E} \left[\varphi_n^\ell(x) - \varphi(x) \right]^2 + \frac{1}{f^2(x)} \mathbb{E} \left\{ \left[r_n^{\ell^2}(x) - r^2(x) \right] \left[f_n^\ell(x) - f(x) \right]^2 \right\} \\ &\quad - \frac{2}{f^2(x)} \mathbb{E} \left\{ \left[r_n^\ell(x) - r(x) \right] \left[\varphi_n^\ell(x) - \varphi(x) \right] \left[f_n^\ell(x) - f(x) \right] \right\} \\ &:= E_1 - E_2 + E_3 + E_4 - E_5. \end{aligned}$$

Le Théorème 2.2.6 implique que :

$$n^{\frac{4}{d+4}} E_1 \rightarrow \frac{r^2(x)}{f^2(x)} \left[c^4 \left(\frac{4+d\ell}{2+d\ell} \right)^2 b_f^2(x) + \frac{(4+d\ell)^2 f(x) \|K\|_2^2}{2c^d(4+d)(2+d\ell)} \right], \quad (3.5.10)$$

lorsque $n \rightarrow \infty$.

* **Etude de E_2 :**

On a :

$$\mathbb{E} \left[f_n^\ell(x) - f(x) \right] \left[\varphi_n^\ell(x) - \varphi(x) \right] = \left[\mathbb{E} f_n^\ell(x) - f(x) \right] \left[\mathbb{E} \varphi_n^\ell(x) - \varphi(x) \right] + \text{Cov} \left[f_n^\ell(x), \varphi_n^\ell(x) \right].$$

De façon similaire que pour le Théorème 2.2.1 (a), en remplaçant f par φ , grâce à l'hypothèse **H.4**(i), il vient que :

$$h_n^{-2} \left[\mathbb{E} \varphi_n^\ell(x) - \varphi(x) \right] \rightarrow \frac{\beta_{d(1-\ell)+2}}{\beta_{d(1-\ell)}} b_\varphi(x),$$

lorsque $n \rightarrow \infty$. Ainsi, cette convergence utilisée conjointement avec le Théorème 2.2.1 (a) entraînent grâce au choix :

$$h_n = C_n n^{-\frac{1}{d+4}}, C_n \downarrow c > 0,$$

que :

$$n^{\frac{4}{d+4}} \left[\mathbb{E} f_n^\ell(x) - f(x) \right] \left[\mathbb{E} \varphi_n^\ell(x) - \varphi(x) \right] \rightarrow c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_f(x) b_\varphi(x), \quad (3.5.11)$$

lorsque $n \rightarrow \infty$. Le terme de covariance est quant à elle contrôlé par le lemme suivant (démontré en fin de preuve) :

Lemme 3.5.2. *Sous les hypothèses **H.1-H.4**, en remplaçant **H.4(iv)** par **H'.4(iv)**, si :*

$$(\ln n)^{\frac{1}{x}} h_n^d \rightarrow 0,$$

lorsque $n \rightarrow \infty$, alors :

$$n h_n^d \text{Cov} \left[f_n^\ell(x), \varphi_n^\ell(x) \right] \rightarrow \sigma_\ell^2(x) r(x),$$

lorsque $n \rightarrow \infty$, avec $\sigma_\ell^2(x)$ défini en (3.2.1).

Par conséquent, le Lemme 3.5.2 avec le choix :

$$h_n = C_n n^{-\frac{1}{d+4}}, C_n \downarrow c > 0,$$

permettent d'écrire que :

$$n^{\frac{4}{d+4}} \text{Cov} \left[f_n^\ell(x), \varphi_n^\ell(x) \right] \rightarrow \frac{(4 + d\ell)^2 f(x) r(x) \|K\|_2^2}{2c^d (4 + d)(2 + d\ell)},$$

lorsque $n \rightarrow \infty$. Par suite, en utilisant la relation (3.5.11), nous avons :

$$E_2 \rightarrow \frac{2r(x)}{f^2(x)} \left[c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_f(x) b_\varphi(x) + \frac{(4 + d\ell)^2 f(x) r(x) \|K\|_2^2}{2c^d (4 + d)(2 + d\ell)} \right], \quad (3.5.12)$$

lorsque $n \rightarrow \infty$.

* **Etude de E_3 :**

Le terme E_3 , se décompose en un terme de biais au carré et un terme de variance. Le terme de biais au carré se traite exactement comme le biais de $f_n^\ell(x)$ en remplaçant f par φ et le terme de variance se traite par le lemme suivant (démontré en fin de preuve) :

Lemme 3.5.3. *Sous les hypothèses **H.1-H.4**, en remplaçant **H.4(iv)** par **H'.4(iv)**, si :*

$$(\ln n)^{\frac{2}{\chi}} h_n^d \rightarrow 0,$$

lorsque $n \rightarrow \infty$, alors :

$$nh_n^d \text{Var}\varphi_n^\ell(x) \rightarrow \sigma_\ell^2(x) [V(x) + r^2(x)],$$

lorsque $n \rightarrow \infty$, avec $\sigma_\ell^2(x)$ et $V(x)$ définis respectivement en (3.2.1) et (3.2.3).

Il vient alors que :

$$n^{\frac{4}{d+4}} E_3 \rightarrow \frac{1}{f^2(x)} \left[c^4 \left(\frac{4 + d\ell}{2 + d\ell} \right)^2 b_\varphi^2(x) + \frac{(4 + d\ell)^2 [V(x) + r^2(x)] f(x) \|K\|_2^2}{2c^d(4 + d)(2 + d\ell)} \right], \quad (3.5.13)$$

lorsque $n \rightarrow \infty$.

Reste maintenant à montrer que les deux derniers termes E_4 et E_5 sont négligeables et donner les preuves des Lemmes 3.5.2 et 3.5.3. Les termes E_4 et E_5 étant traités de façon identique, nous étudions seulement E_4 .

* **Etude de E_4 :**

Soient deux nombres réels ε, γ , on a :

$$\begin{aligned} E_4 &\leq \mathbb{E} \left\{ \left| r_n^\ell(x) - r(x) \right| \left| r_n^\ell(x) + r(x) \right| \left[f_n^\ell(x) - f(x) \right]^2 \mathbf{1}_{|r_n^\ell(x)| \leq n^\gamma} \right\} \\ &+ \mathbb{E} \left\{ \left[r_n^{\ell^2}(x) - r^2(x) \right] \left[f_n^\ell(x) - f(x) \right]^2 \mathbf{1}_{|r_n^\ell(x)| > n^\gamma} \right\} \\ &:= G_1 + G_2. \end{aligned} \quad (3.5.14)$$

Par l'inégalité de Cauchy-Schwartz, nous avons :

$$\begin{aligned} G_2 &\leq \mathbb{E} \left\{ \max_{0 \leq i \leq n} |m(Y_i)|^2 \left[f_n^\ell(x) - f(x) \right]^2 \mathbf{1}_{|r_n^\ell(x)| > n^\gamma} \right\} \\ &\leq \left\{ \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^4 \mathbb{E} \left[\max_{1 \leq i \leq n} |m(Y_i)|^4 \right] P \left[|r_n^\ell(x)| > n^\gamma \right] \right\}^{\frac{1}{2}}. \end{aligned}$$

D'abord le Théorème 2.2.6(c) entraîne que :

$$\left\{ \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^4 \right\}^{\frac{1}{2}} \leq \frac{\|K\|_\infty h_n^{-d} B_{n,-d\ell}}{B_{n,d(1-\ell)}} \left\{ \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^2 \right\}^{\frac{1}{2}} = O \left[\frac{n^{\frac{d-2}{d+4}} B_{n,-d\ell}}{B_{n,d(1-\ell)}} \right].$$

Ensuite d'après (3.1.1) nous avons :

$$\mathbb{E} \left[\max_{0 \leq i \leq n} |m(Y_i)|^4 \right] = O \left[(\ln n)^{\frac{4}{x}} \right].$$

Enfin :

$$P \left[|r_n^\ell(x)| > n^\gamma \right] \leq P \left[\bigcup_{i=1}^n \{|m(Y_i)| > n^\gamma\} \right] \leq n \exp(-\lambda n^{\chi\gamma}) \mathbb{E} [\exp(\lambda |m(Y_0)|^\chi)].$$

Par suite :

$$n^{\frac{4}{d+4}} G_2 = O \left[\frac{n^{\frac{d-8}{2(d+4)}} (\ln n)^{\frac{2}{x}} B_{n,-d\ell}}{\exp\left(\frac{\lambda n^{\gamma\chi}}{2}\right) B_{n,d(1-\ell)}} \right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Maintenant le terme G_1 de (3.5.14) est contrôlé par l'inégalité de Hölder, comme suit :

$$\begin{aligned} G_1 &\leq [n^\gamma + r(x)] \mathbb{E} \left\{ \left[f_n^\ell(x) - f(x) \right]^2 \left| r_n^\ell(x) - r(x) \right| \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma, |r_n^\ell(x) - r(x)| \leq n^{-(1+\varepsilon)\gamma}\}} \right\} \\ &\quad + [n^\gamma + r(x)] \mathbb{E} \left\{ \left[f_n^\ell(x) - f(x) \right]^2 \left| r_n^\ell(x) - r(x) \right| \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma, |r_n^\ell(x) - r(x)| > n^{-(1+\varepsilon)\gamma}\}} \right\} \\ &\leq n^{-(1+\varepsilon)\gamma} [n^\gamma + |r(x)|] \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^2 + [n^\gamma + |r(x)|] \left\{ \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^4 \right\}^{\frac{1}{2}} \\ &\quad \times \left\{ \mathbb{E} \left| r_n^\ell(x) - r(x) \right|^{2v} \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma\}} \right\}^{\frac{1}{2v}} \left\{ P \left[|r_n^\ell(x) - r(x)| > n^{-(1+\varepsilon)\gamma}, |r_n^\ell(x)| \leq n^\gamma \right] \right\}^{\frac{1}{w}} \\ &:= H_1 + H_2, \end{aligned}$$

avec $\frac{1}{v} + \frac{1}{w} = 1$. Il est clair que :

$$n^{\frac{4}{d+4}} H_1 = O \left(n^{-\varepsilon\gamma} \right) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Concernant le terme H_2 , nous avons d'abord :

$$[n^\gamma + |r(x)|] \left\{ \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^4 \right\}^{\frac{1}{2}} = O \left[\frac{n^{\frac{d-2}{d+4} + \gamma} B_{n,-d\ell}}{B_{n,d(1-\ell)}} \right]. \quad (3.5.15)$$

Ensuite, puisque :

$$r_n^\ell(x) - r(x) = r_n^\ell(x) \frac{f_n^\ell(x) - f(x)}{f(x)} + \frac{\varphi_n^\ell(x) - \varphi(x)}{f(x)}$$

alors :

$$\begin{aligned} \mathbb{E} \left\{ \left[r_n^\ell(x) - r(x) \right]^2 \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma\}} \right\} &\leq \frac{2}{f^2(x)} \left\{ n^{2\gamma} \mathbb{E} \left[f_n^\ell(x) - f(x) \right]^2 + \mathbb{E} \left[\varphi_n^\ell(x) - \varphi(x) \right]^2 \right\} \\ &= O \left(n^{2\gamma - \frac{4}{d+4}} \right). \end{aligned}$$

Il en résulte que :

$$\begin{aligned} \left\{ \mathbb{E} \left[\left| r_n^\ell(x) - r(x) \right|^{2v} \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma\}} \right] \right\}^{\frac{1}{2v}} &= \left\{ \mathbb{E} \left[\left| r_n^\ell(x) - r(x) \right|^{2(v-1)+2} \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma\}} \right] \right\}^{\frac{1}{2v}} \\ &\leq n^{\frac{2\gamma(v-1)}{2v}} \mathbb{E} \left\{ \left[r_n^\ell(x) - r(x) \right]^2 \mathbf{1}_{\{|r_n^\ell(x)| \leq n^\gamma\}} \right\} \\ &= O \left[n^{\gamma - \frac{2}{v(d+4)}} \right]. \end{aligned} \tag{3.5.16}$$

Enfin, sur l'ensemble :

$$\left\{ |r_n^\ell(x)| \leq n^\gamma \right\},$$

on a :

$$\left| r_n^\ell(x) - r(x) \right| \leq \frac{1}{f(x)} \left\{ n^\gamma \left[|f_n^\ell(x) - f(x)| \right] + \left[|\varphi_n^\ell(x) - \varphi(x)| \right] \right\}.$$

Par conséquent :

$$\begin{aligned} P \left(\left| r_n^\ell(x) - r(x) \right| > n^{-(1+\varepsilon)\gamma}, |r_n^\ell(x)| \leq n^\gamma \right) &\leq P \left[n^\gamma \left(|f_n^\ell(x) - f(x)| \right) > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{2} \right] \\ &+ P \left[|\varphi_n^\ell(x) - \varphi(x)| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{2} \right]. \end{aligned}$$

Il suffit d'étudier seulement le second terme du second membre de cette inégalité, parce que le premier se déduit du deuxième en posant $m(y) = 1$. On a :

$$\begin{aligned} P \left[|\varphi_n^\ell(x) - \varphi(x)| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{2} \right] &\leq P \left[\left| \mathbb{E} \varphi_n^\ell(x) - \varphi(x) \right| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{4} \right] \\ &+ P \left[|\varphi_n^\ell(x) - \mathbb{E} \varphi_n^\ell(x)| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{4} \right]. \end{aligned} \tag{3.5.17}$$

Puisque :

$$\mathbb{E} \varphi_n^\ell(x) - \varphi(x) = o \left(h_n^2 \right),$$

si l'on choisit γ tel que pour $\varepsilon > 0$:

$$0 < (1 + \varepsilon)\gamma < \frac{2}{(d+2)},$$

alors le choix :

$$h_n = C_n n^{-\frac{(1+\varepsilon)\gamma}{2}}, C_n \downarrow c > 0$$

satisfait les hypothèses **H.2**. Il vient que le premier terme de (3.5.17) vaut zero. Le second s'écrit :

$$\begin{aligned} P\left(\left|\varphi_n^\ell(x) - \mathbb{E}\varphi_n^\ell(x)\right| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{4}\right) &\leq P\left[\left|\tilde{\varphi}_n^\ell(x) - \mathbb{E}\tilde{\varphi}_n^\ell(x)\right| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{8}\right] \\ &+ P\left[\left|\left[\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right] - \mathbb{E}\left[\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right]\right| > \frac{f(x)n^{-(1+\varepsilon)\gamma}}{8}\right] \\ &:= P_1 + P_2. \end{aligned} \quad (3.5.18)$$

Or on a :

$$\bigcap_{i=1}^n \{|m(Y_i)| \leq b_n\} \subset \left\{\left|\left[\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right] - \mathbb{E}\left[\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)\right]\right| = 0\right\}.$$

Il vient que :

$$P_2 \leq P\left(\bigcup_{i=1}^n \{|m(Y_i)| \leq b_n\}\right) \leq n \exp(-\lambda b_n^\chi) \mathbb{E} e^{\lambda |m(Y_0)|^\chi} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Concernant P_1 , on sait que :

$$\begin{aligned} P\left(\left|\tilde{\varphi}_n^\ell(x) - \mathbb{E}\tilde{\varphi}_n^\ell(x)\right| > \eta \sqrt{\frac{\ln n}{nh_n^d}}\right) &\leq \frac{8\|K\|_\infty(1+\kappa)}{\eta\kappa} \rho_0 \frac{n^{-\rho_1 p_0} b_n}{B_{n,d(1-\ell)} h_n^d} \\ &+ 4 \exp\left\{-\left[\frac{\eta}{2(1+\kappa)} - \sigma_\ell^2(x) [1 + o(1)]\right] \ln n\right\}. \end{aligned}$$

Le choix :

$$h_n = \left[\frac{64\eta^2 \ln n}{n^{1-2\gamma(1+\varepsilon)} f^2(x)}\right]^{\frac{1}{d}}$$

vérifie les hypothèses **H.2**, si l'on choisi γ et ε tels que :

$$\frac{1}{(d+2)} < (1+\varepsilon)\gamma < 1,$$

et implique que :

$$P_1 = O\left[\frac{n^{1-2\gamma(1+\varepsilon)-\rho_1 p_0} b_n}{B_{n,d(1-\ell)} \ln n}\right]. \quad (3.5.19)$$

Enfin, si l'on choisit $v = w = 2$, alors les relations (3.5.15), (3.5.16) et (3.5.19) entraînent que :

$$n^{\frac{4}{d+4}} H_2 = O \left[\frac{n^\varsigma b_n^{\frac{1}{2}} B_{n,-d\ell}^{\frac{1}{2}}}{B_{n,d(1-\ell)} (\ln n)^{\frac{1}{2}}} \right]$$

avec :

$$\varsigma := \frac{d+2}{d+4} + \gamma(1-\varepsilon) - \frac{\rho_1 p_0}{2}.$$

Finalement, grâce au choix logarithmique de b_n , nous avons :

$$n^{\frac{4}{d+4}} H_2 \rightarrow 0,$$

lorsque $n \rightarrow \infty$, si l'on choisit :

$$p_0 > \frac{2}{\rho_1}, \quad \varepsilon \geq 1 \quad \text{et} \quad \frac{1}{(d+2)(1+\varepsilon)} < \gamma < \frac{2}{(d+2)(1+\varepsilon)}.$$

Pour finir la preuve, il faut prouver les Lemmes 3.5.2 et 3.5.3.

Preuve du Lemme 3.5.2.

Nous avons la décomposition suivante :

$$\text{Cov} [f_n^\ell(x), \varphi_n^\ell(x)] = \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \left[\sum_{i=1}^n A_{ii} + \sum_{i=1}^n \sum_{i \neq j}^n A_{ij} \right] := F_1 + F_2.$$

où l'on a noté pour deux entiers s, t :

$$A_{s,t} := \text{Cov} \left[\frac{1}{h_s^{d\ell}} K \left(\frac{x - X_s}{h_s} \right), \frac{m(Y_t)}{h_t^{d\ell}} K \left(\frac{x - X_t}{h_t} \right) \right].$$

Nous commençons par étudier F_2 . Soit une suite c_n tendant vers $+\infty$, alors :

$$F_2 \leq 2 \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \left[\sum_{k=1}^{c_n} \sum_{p=1}^n |A_{p,k+p}| + \sum_{k=c_{n+1}}^{n-1} \sum_{p=1}^n |A_{p,k+p}| \right] := F_{21} + F_{22}.$$

Ensuite nous étudions le terme F_{22} . Ce terme se contrôle comme le terme L_2 dans la Preuve du Lemme 3.5.1 (c.f. page 70), par contre, ici nous avons à étudier $\varphi_n^\ell(x)$ et non pas sa version tronquée, comme ce fût le cas pour L_2 . Donc le Lemme de Billingsley remplacé par celui de Davydov, pour obtenir d'une part :

$$|A_{p,k+p}| \leq 8 \left[2\alpha^{(2)}(k) \right]^{\frac{1}{4}} \|K\|_{\infty}^2 h_{k+p}^{-d\ell} h_p^{-d\ell} \sqrt{\text{Em}^2(Y_0)}.$$

En utilisant la relation (3.1.1) et à l'aide de l'hypothèse **H.3**(i), on obtient :

$$\begin{aligned} F_{22} &\leq \text{Cst } (\ln n)^{\frac{1}{\chi}} \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{k=c_n+1}^{n-1} \sum_{p=1}^{n-k} k^{-\frac{\rho_1}{4}} h_{p+k}^{-d\ell} h_p^{-d\ell} \\ &\leq \text{Cst } (\ln n)^{\frac{1}{\chi}} \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} h_n^{-2d\ell} \frac{c_n^{-\frac{\rho_1}{4}+1}}{\frac{\rho_1}{4}-1} \sum_{p=1}^n \left(\frac{h_p}{h_n} \right)^{-d\ell} \\ &\leq \frac{\text{Cst } (\ln n)^{\frac{1}{\chi}} c_n^{1-\rho_1/4} B_{n,-d\ell}}{n h_n^{2d} B_{n,d(1-\ell)}^2}. \end{aligned}$$

Par suite :

$$n h_n^d F_{22} = O \left[(\ln n)^{\frac{1}{\chi}} c_n^{1-\rho_1/4} h_n^{-d} \right]. \quad (3.5.20)$$

D'autre part, pour F_{21} , on pose :

$$g_{k,k'}^* := f_{(X_k, X_{k'})|Y_{k'}} - f \otimes f_{X_{k'}|Y_{k'}},$$

alors en utilisant la condition **H.4**(iv), nous pouvons écrire :

$$\begin{aligned} A_{p,k+p} &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^{d'}} (h_{k+p} h_p)^{-d\ell} m(v) K \left(\frac{x-s}{h_{k+p}} \right) K \left(\frac{x-t}{h_p} \right) g_{k+p,p}^*(s, t|v) f_{Y_{k'}}(v) ds dt dv \\ &\leq \|K\|_{\infty}^2 h_{k+p}^{d(1-\ell)} h_p^{d(1-\ell)} \sup_{k \geq 1} \|g_{k+p,p}^*\|_{\infty} \mathbb{E} |m(Y_0)|. \end{aligned}$$

Ainsi, la décroissance de h_n , conduit à :

$$F_{21} \leq 2 \|K\|_{\infty}^2 \sup_{k \geq 1} \|g_{k+p,p}^*\|_{\infty} \mathbb{E} |m(Y_0)| c_n \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{p=1}^n h_p^{2d(1-\ell)}.$$

D'une part, si :

$$\ell \in \left[\left(\frac{d-2}{2d} \right)^+, 1 \right],$$

on montre que :

$$n h_n^d F_{21} = O \left[(\ln n)^{\frac{1}{\chi}} c_n h_n^d \right],$$

et en choisissant alors :

$$c_n := \left\lfloor h_n^{-\frac{8d}{\rho_1}} \right\rfloor,$$

et à l'aide de (3.5.20) on arrive à :

$$nh_n^d F_2 = O \left[(\ln n)^{\frac{1}{\chi}} h_n^{-\frac{d(8-\rho_1)}{\rho_1}} \right].$$

D'autre part, dans le cas où :

$$d \geq 3 \text{ et } \ell \in \left[0, \frac{d-2}{2d} \right],$$

on montre aussi qu'en choisissant un réel ξ tel que :

$$\frac{4}{\rho_1 - 4} < \xi \leq \frac{2}{d},$$

on a :

$$nh_n^d F_{21} = O \left[(\ln n)^{\frac{1}{\chi}} c_n h_n^{d\xi} \right].$$

Le choix :

$$c_n := \left[h_n^{-\frac{4d(\xi+1)}{\rho_1}} \right],$$

entraîne alors que :

$$nh_n^d F_2 = O \left[(\ln n)^{\frac{1}{\chi}} h_n^{-\frac{d(1+\xi-\rho_1\xi)}{\rho_1}} \right].$$

Dès lors que :

$$\rho_1 > 8 \text{ et } \xi > \frac{4}{\rho_1 - 4},$$

on en déduit alors que :

$$nh_n^d F_2 \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Maintenant pour F_1 on a :

$$\begin{aligned} nh_n^d F_1 &= nh_n^d \left[\sum_{j=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{i=1}^n h_i^{d(1-2\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K^2 \left(\frac{x-u}{h_i} \right) r(u) f(u) du \\ &- nh_n^d \left[\sum_{j=1}^n h_i^{d(1-\ell)} \right]^{-2} \sum_{i=1}^n h_i^{2d(1-\ell)} \int_{\mathbb{R}^d} \frac{1}{h_i^d} K \left(\frac{x-u}{h_i} \right) r(u) f(u) du \\ &\times \int_{\mathbb{R}^d} \frac{1}{h_i^d} K \left(\frac{x-v}{h_i} \right) f(v) dv. \end{aligned}$$

On montre alors en appliquant le Lemme 2.4.1, que le premier terme du second membre de $nh_n^d F_1$ converge vers $\sigma_\ell^2(x)r(x)$ et que le second est négligeable.

Preuve du Lemme 3.5.3.

Il suffit de décomposer $nh_n^d \text{Var} \varphi_n^\ell(x)$ en un terme principal et un terme de covariance. La suite de la preuve est identique à celle du Lemme 3.5.2, mais en remplaçant $g_{k,k'}^*$ par $g_{k,k'}^{**}$ défini par :

$$g_{k,k'}^{**} := f_{(X_k, X_{k'})|(Y_k, Y_{k'})} - f_{X_k|Y_k} \otimes f_{X_{k'}|Y_{k'}}$$

■

3.5.5 Preuve du Théorème 3.4.1

Nous montrons que la version tronquée de l'estimateur est asymptotiquement normale et que le terme résiduel est négligeable :

$$r_n^\ell(x) - r(x) = [\tilde{r}_n^\ell(x) - r(x)] + [r_n^\ell(x) - \tilde{r}_n^\ell(x)].$$

Pour le terme résiduel, on a pour tout $\varepsilon > 0$:

$$P\left(|\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)| > \varepsilon \sqrt{nh_n^d}\right) \leq E e^{\lambda|m(Y_0)|^x} n^{1-\lambda\delta}.$$

Donc pour tout $\varepsilon > 0$:

$$\sum_{n=1}^{\infty} P\left(|\varphi_n^\ell(x) - \tilde{\varphi}_n^\ell(x)| > \varepsilon \sqrt{nh_n^d}\right) < \infty.$$

Ainsi, par le lemme de Borel-Cantelli, on conclut que :

$$\sqrt{nh_n^d} [\tilde{r}_n^\ell(x) - r(x)] \rightarrow 0 \text{ p.s.},$$

lorsque $n \rightarrow \infty$. Maintenant il faut prouver que :

$$\sqrt{nh_n^d} [\tilde{r}_n^\ell(x) - r(x)] \xrightarrow{\mathcal{L}} \mathcal{N}\left[0, \frac{\beta_{d(1-2\ell)} \|K\|_2^2 V(x)}{\beta_{d(1-\ell)}^2 f(x)}\right],$$

lorsque $n \rightarrow \infty$. Le terme principal se décompose de la manière suivante :

$$\tilde{r}_n^\ell(x) - r(x) = \frac{f(x)\tilde{\varphi}_n^\ell(x) - \varphi(x)f_n^\ell(x)}{f(x)f_n^\ell(x)} = \frac{1}{f(x)f_n^\ell(x)} \begin{bmatrix} f(x) \\ -\varphi(x) \end{bmatrix}^T \begin{bmatrix} \tilde{\varphi}_n^\ell(x) - \varphi(x) \\ f_n^\ell(x) - f(x) \end{bmatrix}.$$

Mais puisque :

$$f_n^\ell(x) \rightarrow f(x),$$

presque sûrement, alors grâce au théorème de Cramer-Wold (c.f. Rao [75]), prouver le Théorème 2.2.7 revient à montrer que :

$$\sqrt{nh_n^d} \begin{bmatrix} \tilde{\varphi}_n^\ell(x) - \varphi(x) \\ f_n^\ell(x) - f(x) \end{bmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}_2 \left\{ 0, \sigma_\ell^2(x) \begin{bmatrix} V(x) + r^2(x) & r(x) \\ r(x) & 1 \end{bmatrix} \right\},$$

lorsque $n \rightarrow \infty$. Cela revient donc à prouver grâce à Cramer-Wold que pour tous $\lambda_1, \lambda_2 \in \mathbb{R}$ tels que $\lambda_1 + \lambda_2 \neq 0$ on a :

$$\sqrt{nh_n^d} \left\{ \lambda_1 [f_n^\ell(x) - f(x)] + \lambda_2 [\tilde{\varphi}_n^\ell(x) - \varphi(x)] \right\} \xrightarrow{\mathcal{L}} \mathcal{N} [0, \Sigma_\ell^2(x)], \quad (3.5.21)$$

lorsque $n \rightarrow \infty$, avec :

$$\Sigma_\ell^2(x) := \sigma_\ell^2(x) \left\{ \lambda_1^2 + 2\lambda_1\lambda_2 r(x) + \lambda_2^2 [V(x) + r^2(x)] \right\}.$$

Mais puisque $nh_n^{d+4} \rightarrow 0$, lorsque $n \rightarrow \infty$, alors le Théorème 2.2.1 (a) et le résultat concernant le biais de $\tilde{\varphi}_n^\ell(x)$ entraînent que la convergence (3.5.21) est équivalente à :

$$\sqrt{nh_n^d} \left\{ \lambda_1 [f_n^\ell(x) - \mathbb{E}f_n^\ell(x)] + \lambda_2 [\tilde{\varphi}_n^\ell(x) - \mathbb{E}\tilde{\varphi}_n^\ell(x)] \right\} \xrightarrow{\mathcal{L}} \mathcal{N} [0, \Sigma_\ell^2(x)],$$

lorsque $n \rightarrow \infty$. Pour démontrer cette dernière, on procède exactement comme pour la preuve du Théorème 2.2.7, en gardant les mêmes notations, et les mêmes définitions des sommes partielles S_{n1}, S_{n2} et S_{n3} , mais dans les définitions des blocs, le terme Ψ_{nj} défini dans (2.4.9) (c.f. page 37), est remplacé par $\tilde{\Psi}_{nj}$ défini par :

$$\tilde{\Psi}_{nj} := \lambda_1 \Psi_{nj} + \lambda_2 \Psi'_{nj}$$

avec :

$$\Psi'_{nj} := \left[\frac{h_n^{d(2\ell-1)}}{n} \right]^{\frac{1}{2}} \frac{h_j^{-d\ell}}{B_{n,d(1-\ell)}} (W_{nj} - \mathbb{E}W_{nj})$$

où :

$$W_{nj} := K \left(\frac{x - X_j}{h_j} \right) m(Y_j) \mathbf{1}_{\{|m(Y_j)| \leq b_n\}},$$

et Ψ_{nj} défini en (2.4.9). Nous avons alors l'égalité suivante :

$$\sqrt{nh_n^d} \left\{ \lambda_1 [f_n^\ell(x) - f(x)] + \lambda_2 [\tilde{\varphi}_n^\ell(x) - \varphi(x)] \right\} = S_{n1} + S_{n2} + S_{n3}.$$

La question est donc de montrer que dans ces nouvelles considérations, S_{n2} et S_{n3} tendent vers zero en moyenne quadratique et que S_{n1} est asymptotiquement normale. D'abord :

$$\begin{aligned}
 ES_{n2}^2 &= \sum_{m=1}^{r_n} \sum_{i=l_m}^{l_m+\tau_n-1} \text{Var} \tilde{\Psi}_{ni} + 2 \sum_{m=1}^{r_n} \sum_{l_m \leq i < j \leq l_m+\tau_n-1} \text{Cov}(\tilde{\Psi}_{ni}, \tilde{\Psi}_{nj}) \\
 &\quad + 2 \sum_{1 \leq i < j \leq r_n} \sum_{s=l_i}^{l_i+\tau_n-1} \sum_{t=l_j}^{l_j+\tau_n-1} \text{Cov}(\tilde{\Psi}_{ns}, \tilde{\Psi}_{nt}) \\
 &:= \Delta_1 + \Delta_2 + \Delta_3.
 \end{aligned} \tag{3.5.22}$$

Pour le premier terme de (3.5.22), nous avons :

$$\begin{aligned}
 \Delta_1 &= \sum_{m=1}^{r_n} \sum_{i=l_m}^{l_m+\tau_n-1} [\lambda_1^2 \text{Var} \Psi_{ni} + \lambda_2^2 \text{Var} \Psi'_{ni} + 2\lambda_1 \lambda_2 \text{Cov}(\Psi_{ni}, \Psi'_{ni})] \\
 &:= \Delta_{11} + \Delta_{12} + \Delta_{13}.
 \end{aligned}$$

Donc h_n étant décroissante, si b_n est logarithmique, alors comme pour la preuve du Théorème 2.2.7, on a :

$$\Delta_{11} + \Delta_{12} \leq \frac{r_n \tau_n (1 + b_n^2) \|K\|_\infty^2 \max(\lambda_1^2, \lambda_2^2)}{n h_n^d B_{n,d(1-\ell)}^2} \rightarrow 0,$$

lorsque $n \rightarrow \infty$ et

$$\Delta_{13} \leq \frac{2\lambda_1 \lambda_2 b_n r_n \tau_n \|K\|_\infty^2}{n h_n^d B_{n,d(1-\ell)}^2} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. De manière analogue et à l'aide de l'inégalité de Cauchy-Schwartz, on obtient que :

$$\Delta_2 \leq \frac{r_n \tau_n^2 (1 + b_n)^2 \|K\|_\infty^2 \max(\lambda_1^2, \lambda_2^2)}{n h_n^d B_{n,d(1-\ell)}^2} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Enfin, comme pour la preuve du Théorème 2.2.7, le dernier terme de (3.5.22) est contrôlé par l'inégalité de Billingsley à l'aide de l'hypothèse **H.3(i)**. Nous avons donc :

$$\begin{aligned}
 \Delta_3 &= 2 \sum_{1 \leq i < j \leq r_n} \sum_{s=l_i}^{l_i+\tau_n-1} \sum_{t=l_j}^{l_j+\tau_n-1} \{ \lambda_1^2 \text{Cov}(\Psi_{ns}, \Psi_{nt}) + \lambda_2^2 \text{Cov}(\Psi'_{ns}, \Psi'_{nt}) \\
 &\quad + \lambda_1 \lambda_2 [\text{Cov}(\Psi_{ns}, \Psi'_{nt}) + \text{Cov}(\Psi_{nt}, \Psi'_{ns})] \} \\
 &\leq \frac{2(1+b_n)^2 \|K\|_\infty^2 \max(\lambda_1^2, \lambda_2^2) h_n^{d(2\ell-1)}}{n B_{n,d(1-\ell)}^2} \sum_{k=1}^{r_n-1} \sum_{j=1}^{r_n} \sum_{s=l_j}^{l_j+\tau_n-1} \sum_{t=l_j}^{l_j+\tau_n-1} (h_s h_t)^{-d\ell} \alpha[k(\varsigma_n + \tau_n)] \\
 &\leq \frac{2(1+b_n)^2 \|K\|_\infty^2 \max(\lambda_1^2, \lambda_2^2) r_n \tau_n^2}{n h_n^d B_{n,d(1-\ell)}^2} \sum_{k=1}^{r_n-1} e^{-k\tau_n}.
 \end{aligned}$$

Ainsi :

$$\Delta_3 = O \left\{ \frac{b_n^2 r_n \tau_n^2 e^{-\tau_n}}{n h_n^d B_{n,d(1-\ell)}^2} [1 - e^{-\tau_n(r_n-1)}] \right\} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. De même, il faut maintenant montrer que $ES_{n3}^2 \rightarrow 0$ lorsque $n \rightarrow 0$. On a :

$$ES_{n3}^2 = \sum_{j=\bar{N}+1}^n \text{Var} \tilde{\Psi}_{nj} + 2 \sum_{\bar{N}+1 \leq i < j \leq n} \text{Cov}(\tilde{\Psi}_{ni}, \tilde{\Psi}_{nj}) := \Theta_{n1} + \Theta_{n2}.$$

Le terme de variance Θ_{n1} s'écrit :

$$\Theta_{n1} = \sum_{j=\bar{N}+1}^n \left[\lambda_1^2 \text{Var} \Psi_{nj} + \lambda_2^2 \text{Var} \Psi'_{nj} + 2\lambda_1 \lambda_2 \text{Cov} \left(\Psi_{nj}, \Psi'_{nj} \right) \right] := \lambda_1^2 \Theta_{n11} + \lambda_2^2 \Theta_{n12} + 2\lambda_1 \lambda_2 \Theta_{n13}.$$

Nous avons montré d'une part dans la preuve du Théorème 2.2.7, que :

$$\Theta_{n11} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Pour le terme Θ_{n12} , on reprend la démonstration détaillée à la page 66 concernant l'équivalence (3.5.4), en remplaçant les $W_{i,n}$ par :

$$W_{i,n} := \frac{1}{h_i^{d\ell}} K \left(\frac{x - X_i}{h_i} \right) \mathbf{1}_{\{|m(Y_i)| \leq b_n\}} m(Y_i).$$

On obtient ainsi la nouvelle équivalence :

$$V_n^* \sim n h_n^{d(1-2\ell)} \beta_{d(1-2\ell)} f(x) \left[V(x) + r^2(x) \right] \int_{\mathbb{R}^d} K^2(u) du,$$

lorsque $n \rightarrow \infty$. Il en résulte alors que :

$$n h_n^d \text{Var} \tilde{\varphi}_n^\ell(x) = \frac{V_n^*}{n h_n^{d(1-2\ell)} B_{n,d(1-2\ell)}} \rightarrow \sigma_\ell^2(x) \left[r^2(x) + V(x) \right],$$

lorsque $n \rightarrow \infty$. Or nous avons, grâce à l'équivalence $\bar{N} \sim n$:

$$n h_n^d \text{Var} \tilde{\varphi}_n^\ell(x) \sim \sum_{j=1}^n \text{Var} \Psi'_{nj} = \sum_{j=1}^{\bar{N}} \text{Var} \Psi'_{nj} + \Theta_{n12}.$$

Alors en vertu la condition (3.4.1), il vient que :

$$\Theta_{n12} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Enfin, le dernier terme Θ_{n13} , se contrôle par le lemme suivant :

Lemme 3.5.4. *Sous les hypothèses **H.1-H.4**, si :*

$$(\ln n)^{\frac{1}{\chi}} h_n^d \rightarrow 0,$$

lorsque $n \rightarrow \infty$, alors :

$$n h_n^d \text{Cov} \left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x) \right] \rightarrow \sigma_\ell^2(x) r(x).$$

Ainsi, puisque :

$$nh_n^d \text{Cov} \left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x) \right] \sim \sum_{i=1}^n \text{Cov}(\Psi_{ni}, \Psi'_{ni}) = \sum_{j=1}^{\bar{N}} \text{Cov}(\Psi_{ni}, \Psi'_{ni}) + \Theta_{n13},$$

le Lemme 3.5.4 entraîne que :

$$\Theta_{n13} \rightarrow 0,$$

lorsque $n \rightarrow \infty$, du fait que $\bar{N} \sim n$. Par suite :

$$\Theta_{n1} \rightarrow 0.$$

lorsque $n \rightarrow \infty$. Maintenant, on s'intéresse au terme Θ_{n2} de ES_{n3}^2 . Celui-ci s'écrit :

$$\Theta_{n2} = 2 \sum_{\bar{N}+1 \leq i < j \leq n} \left[\lambda_1^2 \text{Cov}(\Psi_{ni}, \Psi_{nj}) + \lambda_2^2 \text{Cov}(\Psi'_{ni}, \Psi'_{nj}) + 2\lambda_1 \lambda_2 \text{Cov}(\Psi_{ni}, \Psi'_{nj}) \right].$$

Comme pour (2.4.10), nous avons :

$$\sum_{\bar{N}+1 \leq i < j \leq n} \left[\lambda_1^2 \text{Cov}(\Psi_{ni}, \Psi_{nj}) + \lambda_2^2 \text{Cov}(\Psi'_{ni}, \Psi'_{nj}) \right] \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Ensuite :

$$\sum_{\bar{N}+1 \leq i < j \leq n} \text{Cov}(\Psi_{ni}, \Psi'_{nj}) \leq \sum_{1 \leq i < j \leq n} \text{Cov}(\Psi_{ni}, \Psi'_{nj}) \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Donc :

$$\Theta_{n13} \rightarrow 0,$$

lorsque $n \rightarrow \infty$. Reste maintenant à montrer que le terme S_{n1} est asymptotiquement normal. Toujours comme pour la preuve du Théorème 2.2.7, on a :

$$\text{Var} S_{n1} \sim nh_n^d \left\{ \lambda_1^2 \text{Var} f_n^\ell(x) + \lambda_2^2 \text{Var} \tilde{\varphi}_n^\ell(x) + 2\lambda_1 \lambda_2 \text{Cov} \left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x) \right] \right\} \rightarrow \Sigma_\ell^2(x),$$

lorsque $n \rightarrow \infty$. Comme pour (2.4.11), nous avons également :

$$|T_{nm}| \leq \frac{\varsigma_n \|K\|_\infty (1 + b_n)}{\sqrt{nh_n^d B_{n,d(1-\ell)}}}.$$

La suite de la preuve est identique à celle de la preuve du Théorème 2.2.7 en remarquant que :

$$\sum_{m=1}^{r_n} \mathbb{E} \left(Z_{nm}^2 \mathbf{1}_{\{|Z'_{nm}| > \varepsilon\}} \right) \leq \left[\frac{\varsigma_n (1 + b_n)}{nh_n^d} \cdot \frac{\|K\|_\infty \varepsilon^{-1}}{\varsigma_n B_{n,d(1-\ell)}} \right]^2 \rightarrow 0,$$

lorsque $n \rightarrow \infty$, à cause du choix logarithmique de b_n .

Preuve du Lemme 3.5.4

Pour prouver le Lemme 3.5.4, la procédure est identique à celle du Lemme 3.5.2. On considère la décomposition :

$$\text{Cov} \left[f_n^\ell(x), \tilde{\varphi}_n^\ell(x) \right] = \left[\sum_{i=1}^n h_i^{d(1-\ell)} \right]^{-2} \left[\sum_{i=1}^n A_{ii} + \sum_{i=1}^n \sum_{i \neq j}^n A_{ij} \right] := F_1 + F_2.$$

où l'on a noté pour deux entiers s, t :

$$A_{s,t} := \text{Cov} \left[\frac{1}{h_s^{d\ell}} K \left(\frac{x - X_s}{h_s} \right), \frac{m(Y_t)}{h_t^{d\ell}} \mathbf{1}_{\{m(Y_i) \leq b_n\}} K \left(\frac{x - X_t}{h_t} \right) \right].$$

Ensuite, on procède exactement comme pour la preuve du Lemme 3.5.2 en remplaçant seulement l'inégalité de Davydov par celle Billingsley, pour obtenir que :

$$nh_n^d F_1 \rightarrow \sigma_\ell^2(x) r(x), \text{ et } nh_n^d F_2 \rightarrow 0,$$

lorsque $n \rightarrow \infty$. ■

Chapitre 4

Application à la prévision et simulations

Sommaire

4.1	Prévision non paramétrique	89
4.1.1	Cadre d'étude	90
4.1.2	Présentation et étude de nos prédictors	90
4.2	Simulations	93
4.2.1	Estimation de la densité	94
4.2.2	Estimation de la régression	102
4.2.3	Prévision	114

4.1 Prévision non paramétrique

Les méthodes non paramétriques pour la prédiction sont apparues ces dernières années, plusieurs auteurs ont comparé leur efficacité vis-à-vis des méthodes paramétriques et ont conclu à la supériorité des prédictors non paramétriques sur plusieurs types de processus. La littérature traitant le sujet est abondante, nous nous plaçons dans un cadre où la taille de l'échantillon à partir duquel la prédiction est faite est fluctuante, où l'utilisation de noyaux récurrents s'avère cruciale.

4.1.1 Cadre d'étude

Considérons un processus α -mélangeant, à temps discret $\{\xi_t : t \geq 1\}$. On suppose également que le processus est Markovien d'ordre k et strictement stationnaire. Il est bien connu que plusieurs processus usuels comme les processus ARMA sont Markoviens et fortement mélangeants (c.f. Ibragimov et Rosanov [50], Pham et Tran [70]).

On suppose avoir observé les n réalisations $\xi_1, \dots, \xi_n$, à partir desquelles on souhaite prédire la valeur de ξ_{n+h} (prévision à l'horizon $h, h \in \mathbb{N}^*$). Dans ce chapitre, nous nous limitons au cas $h = 1$, mais nous supposons que la base des données contenant les ξ_i est mise à jour au fur et à mesure (sa taille passant ainsi progressivement de n à $N, N \geq n$) et l'on souhaite réactualiser notre prévision à chaque instant. La question est donc de faire les $N - n$ prévisions successives $\hat{\xi}_{n+p}^\ell$ à partir des $N - n$ suites de données $\xi_1, \dots, \xi_{n+p-1}$, respectivement, pour $p = 1, \dots, N - n$.

L'efficacité des méthodes non paramétriques vis à vis des autres méthodes de prévision étant déjà étudiée par plusieurs auteurs, ce travail a pour objet d'améliorer ces méthodes du point de vue du temps de calcul.

4.1.2 Présentation et étude de nos prédicteurs

Nous construisons un processus associé :

$$\zeta_t = (X_t, Y_t) = ((\xi_t, \dots, \xi_{t+k-1}), \xi_{t+k}) \text{ pour } t = 1, \dots, n + p - k - 1,$$

et considérons la famille d'estimateurs récurrents de la régression $r_n^\ell(x)$ définie en (1.3.2) basée sur les observations :

$$\{\zeta_t, 1 \leq t \leq n + p - k - 1\}.$$

La prédiction de ξ_{n+p} est donnée alors par :

$$\hat{\xi}_{n+p}^\ell = r_{n+p-k-1}^\ell(\xi_{n+p-k}, \dots, \xi_{n+p-1}). \quad (4.1.1)$$

Nous supposons dans la suite, que le processus (ξ_t) est Markovien d'ordre 1. Le prédicteur s'écrit alors :

$$\hat{\xi}_{n+p}^\ell = \frac{\sum_{i=1}^{n+p-2} \frac{\xi_{i+1}}{h_i^{d_\ell}} K\left(\frac{\xi_{n+p-1}-\xi_i}{h_i}\right)}{\sum_{i=1}^{n+p-2} \frac{1}{h_i^{d_\ell}} K\left(\frac{\xi_{n+p-1}-\xi_i}{h_i}\right)},$$

et en utilisant la propriété de récursivité de l'estimateur $r_n^\ell(x)$, nous avons :

$$\widehat{\xi}_{n+p}^\ell = \frac{\left[\sum_{i=1}^{n+p-3} h_i^{d(1-\ell)}\right] \varphi_{n+p-3}^\ell(\xi_{n+p-1}) + \left[\sum_{i=1}^{n+p-2} h_i^{d(1-\ell)}\right] \xi_{n+p-1} K_{n+p-2}^\ell(\Delta_{n+p-2}\xi)}{\left[\sum_{i=1}^{n+p-3} h_i^{d(1-\ell)}\right] f_{n+p-3}^\ell(\xi_{n+p-1}) + \left[\sum_{i=1}^{n+p-2} h_i^{d(1-\ell)}\right] K_{n+p-2}^\ell(\Delta_{n+p-2}\xi)},$$

où $K_i^\ell(\cdot)$ défini dans (2.2.1) et :

$$\Delta_h \xi := \xi_{h+1} - \xi_h.$$

Pour étudier la convergence du prédicteur $\widehat{\xi}_{n+p+1}^\ell$, nous utilisons les résultats établis au Chapitre 3. Nous introduisons l'hypothèse suivante qui nous permet d'établir la convergence presque sûre de notre famille de prédicteurs.

H.4'''(iv) : Il existe un entier $r \geq 2$ tel que pour tous entiers m et m' tels que $|m - m'| \geq r$, le vecteur $(\zeta_m, \zeta_{m'})$ admet une densité de probabilité $f_{(\zeta_m, \zeta_{m'})}$, telle que :

$$G := \sup_{|m-m'| \geq r} \sup_{(s,t) \in \mathbb{R}^{2d}} \int_{\mathbb{R}^{d'}} \int_{\mathbb{R}^{d'}} |f_{(\zeta_m, \zeta_{m'})}(s, u, t, v) - f^*(s, u) f^*(t, v)| du dv < \infty,$$

Convergence presque sûre

La proposition suivante établit la convergence presque sûre de la famille de prédicteurs $\widehat{\xi}_{n+p+1}^\ell$.

Proposition 4.1.1. *Sous les hypothèses **H.1–H.4**, avec **H.4(iv)** remplacée par **H.4'''(iv)**, si le processus associé (ζ_t) est GFM à valeurs dans un compact D de $\mathbb{R}^d$ et telle que les fonctions f et φ sont bornées (avec $\inf_{x \in \mathbb{R}^d} f(x) > 0$) et appartiennent à $C_d^2(b)$, alors pour une fonction lipschitzienne K vérifiant la condition :*

$$\int_{\mathbb{R}^d} \|u\|^3 K(u) < \infty,$$

et un choix de fenêtre :

$$h_n = C_n \left(\frac{\ln n}{n} \right)^{\frac{1}{d+4}}, C_n \downarrow c > 0,$$

nous avons :

$$\left| \widehat{\xi}_{n+p}^\ell - r(\xi_{n+p-1}) \right| = O \left[\left(\frac{\ln n}{n} \right)^{\frac{2}{d+4}} (\ln n)^{\frac{1}{\chi}} \right] \text{ p.s.}$$

Preuve.

On a :

$$\left| \widehat{\xi}_{n+p}^\ell - r(\xi_{n+p-1}) \right| = \left| r_{n+p-1}^\ell(\xi_{n+p-1}) - r(\xi_{n+p-1}) \right| \leq \sup_{x \in D} \left| r_{n+p}^\ell(x) - r(x) \right|.$$

La suite de la preuve est identique à celle du Théorème 3.2.3, en reprenant seulement la majoration uniforme du terme :

$$\lambda_n^2 \sum_{j=1}^{q_n} \text{EV}_n^{*2}(2j-1),$$

où nous rappelons que :

$$\lambda_n = b_n^{-1} \sqrt{nh_n^d \ln n},$$

et :

$$\sum_{j=1}^{q_n} \text{EV}_n^{*2}(2j-1) = \frac{1}{n^2} \left[\sum_{m=1}^{q_n} \text{Var} \tilde{Z}_{m,k} + \sum_{m=1}^{q_n} \sum_{m' \neq m}^{q_n} \text{Cov}(\tilde{Z}_{m,k}, \tilde{Z}_{m',k}) \right],$$

avec :

$$\tilde{Z}_{i,k} = \tilde{W}_{i,k} - \mathbb{E} \tilde{W}_{i,k},$$

$$\text{où : } \tilde{W}_{i,k} = \frac{m(Y_i) \mathbf{1}_{\{|m(Y_i)| \leq b_n\}}}{B_{n,d(1-\ell)} h_n^{d(1-\ell)} h_i^{d\ell}} K\left(\frac{x_{k,n} - X_i}{h_i}\right).$$

Écrivons :

$$\begin{aligned} nh_n^d \sum_{j=1}^{q_n} \text{EV}_n^{*2}(2j-1) &= \frac{h_n^d}{n} \left[\sum_{m=1}^{q_n} \text{Var} \tilde{Z}_{m,k} + \sum_{m=1}^{q_n} \sum_{|m-m'| \geq r} \sum_{m'=1}^{q_n} \text{Cov}(\tilde{Z}_{m,k}, \tilde{Z}_{m',k}) \right] \\ &\quad + \frac{h_n^d}{n} \left[\sum_{m=1}^{q_n} \sum_{|m-m'| < r} \sum_{m'=1}^{q_n} \text{Cov}(\tilde{Z}_{m,k}, \tilde{Z}_{m',k}) \right] \\ &:= T_1 + T_2. \end{aligned} \tag{4.1.2}$$

Le terme T_1 de (4.1.2) est contrôlé comme dans la preuve du Théorème 3.2.3, grâce à l'hypothèse $H.4'''(iv)$. On obtient alors :

$$T_1 \leq \frac{\|K\|_2^2 \|f\|_\infty b_n^2}{B_{n,d(1-\ell)}^2} \times \frac{1}{nh_n^{d(1-2\ell)}} \sum_{i=1}^{q_n} h_i^{d(1-2\ell)} [1 + o(1)],$$

avec $o(1)$ uniforme en $x_{k,n}$. Le terme T_2 est quant à lui contrôlé par l'inégalité de Cauchy-Schwartz, en se servant de la décroissance de h_n . Ainsi nous avons (après un changement d'indice : $s = m - m'$, et $s' = m'$) :

$$\begin{aligned}
 T_2 &\leq \frac{2h_n^d}{n} \sum_{s=1}^r \sum_{s'=1}^{q_n} \left(\text{Var} \tilde{Z}_{s+s',k} \text{Var} \tilde{Z}_{s',k} \right)^{\frac{1}{2}} \leq \frac{2\|K\|_2^2 \|f\|_\infty b_n^2}{nB_{n,d(1-\ell)}^2 h_n^{d(1-2\ell)}} \sum_{s=1}^r \sum_{s'=1}^{q_n} (h_{s+s'} h_{s'})^{\frac{d(1-2\ell)}{2}} \\
 &\leq \frac{2r\|K\|_2^2 \|f\|_\infty b_n^2}{nB_{n,d(1-\ell)}^2 h_n^{d(1-\ell)}} \sum_{s'=1}^{q_n} h_{s'}^{d(1-\ell)}.
 \end{aligned}$$

En remarquant que le facteur :

$$\frac{\sum_{s'=1}^{q_n} h_{s'}^{d(1-\ell)}}{n h_n^{d(1-\ell)}}$$

converge vers une limite finie, nous obtenons alors la majoration uniforme suivante :

$$\lambda_n^2 \sum_{j=1}^{q_n} \text{EV}_n^{*2}(2j-1) \leq \text{Cste} \ln n [1 + o(1)],$$

où $o(1)$ uniforme en $x_{k,n}$. ■

4.2 Simulations

Dans cette partie, nous nous donnons des premiers éléments de simulation ayant pour but de présenter l'avantage prépondérant des estimateurs récursifs en terme de temps de calcul. Nous analysons les performances de nos estimateurs dans le cas où la dimension $d = 1$.

Choix du noyau et de la fenêtre.

Pour construire nos estimateurs, nous utilisons, parmi les noyaux proposés dans la littérature, le noyau gaussien :

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

qui est très compétitif, utilisé par plusieurs auteurs et facile à mettre en oeuvre. Pour le choix du paramètre de lissage, plusieurs auteurs parmi lesquels on peut citer Deheuvels [35], Carbon et Francq [21] recommandent le choix :

$$h_n = s_n n^{-\frac{1}{5}},$$

où s_n désigne l'écart-type estimé des observations. Pour une discussion approfondie sur ce choix, optimal dans le cas gaussien, nous nous référons à Deheuvels [35]. Notons que ce choix de fenêtre n'est pas optimal, notamment pour des lois non symétriques, mais il

est motivé par la fluctuation de la taille de notre échantillon, car la forme récursive de s_n permet une mise à jour immédiate de la fenêtre sans une augmentation importante du temps de calcul. Il existe cependant d'autres procédures plus performantes, de sélection de la largeur de la fenêtre, comme celle proposée par Sarda et Vieu [83], utilisant un critère de validation croisée ou celle développée par Berlinet et al.[6] basé sur la distance L^1 . Il serait intéressant dans l'avenir d'étendre ces méthodes au cas récursif.

4.2.1 Estimation de la densité

Dans ce paragraphe, nous analysons, par des considérations numériques, les performances de la famille d'estimateurs récursifs de la densité $(f_n^\ell(x))$ étudiée au Chapitre 2. Pour cela, on considère une suite (X_n) de variables aléatoires iid, de densité de probabilité f . On se place dans le cadre de l'estimation de la densité à partir d'une base de données mise à jour en permanence, dont la taille passe progressivement de n à N , avec n et N deux entiers naturels non nuls tels que $n < N$. Le problème consiste donc en l'estimation permanente de f sur un intervalle $I = [a, b]$, sur la base des k premières observations $X_1, \dots, X_k$, pour $k = n, \dots, N$. Plus précisément, soit ϖ une subdivision de l'intervalle I définie par :

$$\varpi = \{x_0 = a, x_1, \dots, x_J = b\}, J \geq 1,$$

et pour $k = n, \dots, N$. On calcule les valeurs $f_k^\ell(x_j)$ ($\ell \in \{0, 0.25, 0.5, 0.75, 1\}$) et $f_k^{PR}(x_j)$, en tout point x_j de ϖ .

Il est important de préciser que pour $k = n$, les estimateurs récursifs $f_n^\ell(x)$ sont calculés en un point x de ϖ , en utilisant la formule de récursivité (2.2.1), donnée en page 13, c'est à dire que nous avons initialisé l'algorithme à $f_1^\ell(x) = 0$, ensuite on a calculé $f_2^\ell(x)$, puis $f_3^\ell(x)$, en fonction de $f_2^\ell(x)$, et ainsi de suite jusqu'à $f_n^\ell(x)$. Quant à l'estimateur non récursif le calcul de sa valeur pour $k = n$ est fait en utilisant sa formule exacte définie au Chapitre 1 (c.f. page 3).

Les performances des estimateurs sont évaluées par une mesure adéquate de la distance entre l'estimateur et la vraie valeur de la densité et par son temps de calcul.

1. **Erreur quadratique moyenne empirique (EQME)** : elle est définie (pour un estimateur $\hat{f}_k$ de f) par :

$$EQME(\hat{f}_k) := \frac{1}{J+1} \sum_{j=0}^J [\hat{f}_k(x_j) - f(x_j)]^2.$$

2. **Erreur absolue maximum empirique (EAME)** : c'est une mesure qui évalue la distance absolue maximale entre la valeur de l'estimateur et celle de la vraie densité sur l'intervalle d'estimation. Cette erreur est définie, pour un estimateur $\hat{f}_k$ de f , par :

$$EAME(\hat{f}_k) := \max_{0 \leq j \leq J} |\hat{f}_k(t_j) - f(t_j)|.$$

3. **Temps de calcul** : On appelle temps de calcul d'un programme informatique le temps nécessaire à son exécution. On s'intéresse ici à l'évolution du temps de calcul d'un programme calculant un estimateur $\hat{f}_k$ (on l'appellera temps de calcul de l'estimateur) en fonction de la taille des données $X_1, \dots, X_k$, utilisées pour l'estimation. Cette définition est précise mais pas intrinsèque dans la mesure où celui-ci, exprimé par exemple en secondes, dépend de la machine utilisée (vitesse du processeur, temps d'accès à la mémoire, etc.), du logiciel utilisé et décroît rapidement avec les progrès de la technologie. Ici nous définissons notre temps de calcul avec la fonction **system.time** disponible sur logiciel **R**, à partir d'une machine dont les caractéristiques sont :

CPU : Duo T9300 2.5 GHz ; HD : 250GB and Memory : 3GB.

Notons néanmoins que des définitions alternatives (en terme du nombre d'instructions élémentaires) peuvent également être considérées.

Résultats :

Nous considérons deux cas : l'estimation d'une densité gaussienne et d'une densité exponentielle. Pour calculer les EQME (resp. EAME), nous procédons de la manière suivante : on simule T échantillons ($T \in \mathbb{N}^*$) de taille N suivant une loi de densité f . Ensuite, pour chacun de ces échantillons, nous calculons l'EQME (resp. EAME), correspondant à l'estimation de f à partir de l'échantillon considéré. Enfin, nous prenons la moyenne des T EQME (resp. EAME) obtenues.

Dans les deux cas étudiés, nous représentons d'abord la vraie densité avec ses estimateurs obtenus à partir du premier échantillon simulé (on aurait pu choisir un autre échantillon), pour différentes valeurs de k et ℓ . Ensuite, nous donnons les EQME et EAME de nos estimateurs. Enfin, nous donnons leur temps de calcul.

◇ Cas où f est la densité d'une loi normale centrée de variance σ^2 :

Les résultats suivants sont obtenus pour :

$n = 200$, $N = 500$, $\sigma^2 = 5$, $I = [-10, 10]$, $J = 400$ et $x_j = x_{j-1} + 0.05$, pour $j = 1, \dots, 400$.

Dans la FIGURE 4.1, les cinq premiers schémas représentent les estimateurs $f_k^\ell(x)$ pour $\ell \in \{0, 1/4, 1/2, 3/4, 1\}$ et $k = 200, 400$ avec la densité $f(x)$ sur l'intervalle $[-10, 10]$, tandis que la dernière représente l'estimateur non récursif de Parzen-Rosenblatt.

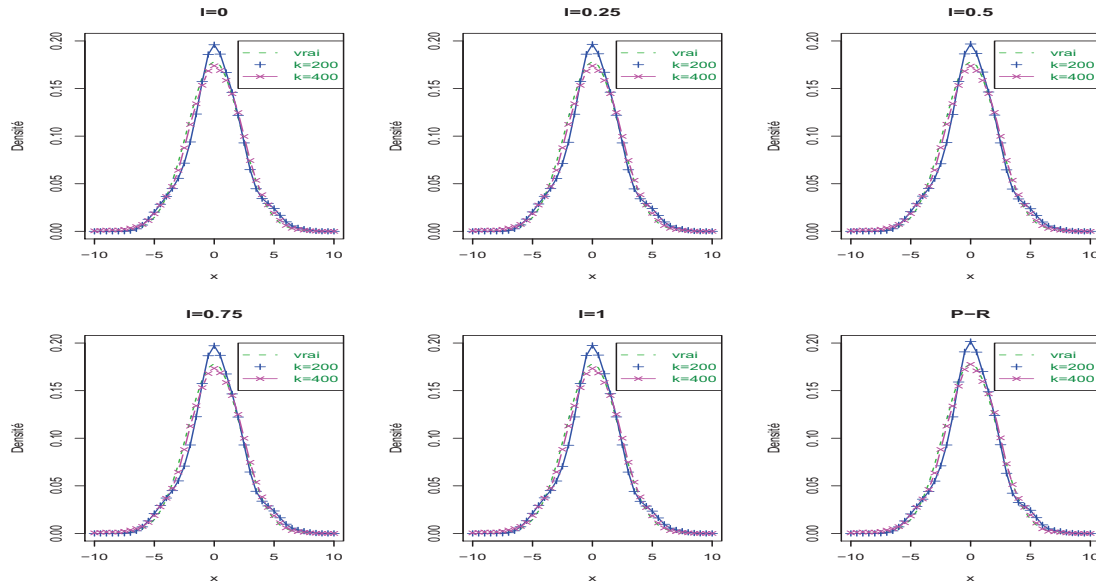
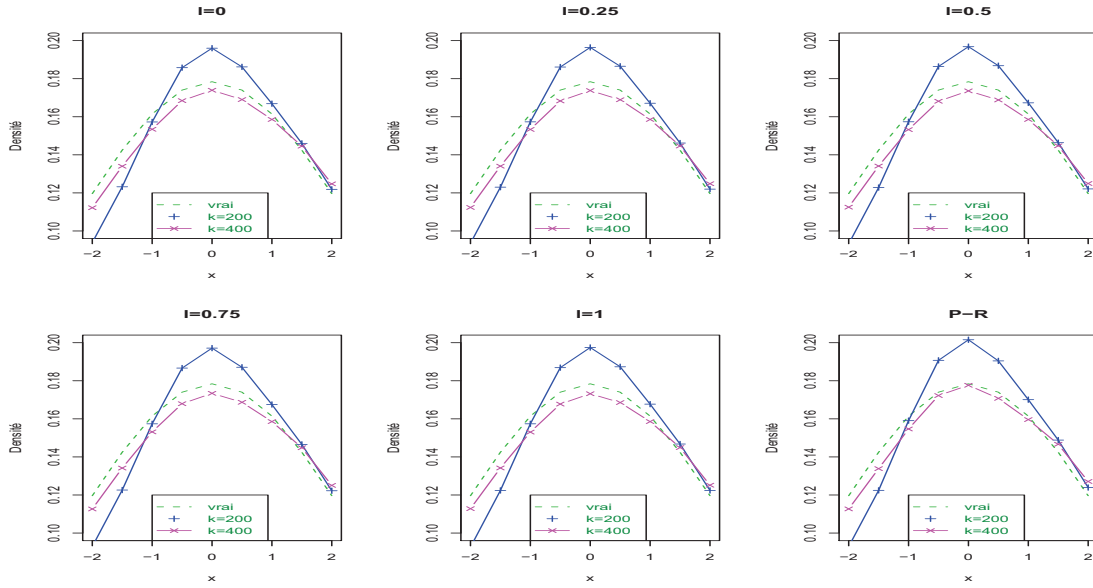


FIG. 4.1 – Estimateurs de la densité d'une loi normale sur $[-10, 10]$.

Nous remarquons que les six estimateurs sont tous compétitifs pour l'estimation de $f(x)$, mais nous allons faire un zoom au voisinage de 0, en réduisant l'intervalle de représentation à $[-2, 2]$, pour mieux visualiser l'évolution de nos estimateurs en fonction du nombre k d'observations considérées (FIGURE 4.2). Nous constatons une nette amélioration de nos estimateurs avec les 400 premières observations plutôt qu'avec les 200. En outre, nous remarquons qu'au voisinage de 0, l'estimateur non récursif estime mieux la densité pour $k = 400$. Nous cherchons donc à avoir une précision sur la proximité de ces estimateurs par rapport à la vraie densité, mais surtout évaluer la perte de précision due au passage du noyau classique au noyau récursif. Pour cela, nous étudions leurs EQME et EAME. Nous obtenons ainsi les résultats du TABLEAU 4.1, avec $T = 100$ échantillons. Les résultats du TABLEAU 4.1 sont représentés graphiquement à la FIGURE 4.3.

Le TABLEAU 4.1 et la FIGURE 4.3 donnent les erreurs quadratiques moyennes empiriques

FIG. 4.2 – Estimateurs de la densité d'une loi normale sur $[-2, 2]$.

k	200	250	300	350	400	450	500
$\ell = 0$	0.094	0.083	0.072	0.066	0.060	0.054	0.051
$\ell = 0.25$	0.090	0.079	0.069	0.063	0.057	0.051	0.049
$\ell = 0.5$	0.086	0.076	0.066	0.060	0.055	0.049	0.047
$\ell = 0.75$	0.083	0.074	0.064	0.059	0.053	0.048	0.046
$\ell = 1$	0.081	0.072	0.062	0.057	0.052	0.047	0.045
P-R	0.075	0.067	0.057	0.052	0.048	0.044	0.042

TAB. 4.1 – $\text{EQME} \times 10^3$ pour les estimateurs d'une densité gaussienne.

des estimateurs récurrents de la densité, pour $\ell \in \{0, 1/4, 1/2, 3/4, 1\}$ ainsi que celle de l'estimateur de Parzen-Rosenblatt. Dans cet exemple, on constate que l'efficacité de nos estimateurs est croissante en fonction de ℓ selon le critère de l'EQME, et que l'estimateur de Parzen-Rosenblatt est préférable aux estimateurs récurrents, ce qui est en accord avec les résultats théoriques obtenus aux Théorèmes 2.3.1 et 2.3.2. Bien entendu, ces théorèmes sont établis dans le cas d'une fenêtre non aléatoire. Ces premiers résultats numériques, obtenus pour une fenêtre aléatoire, sont encourageants et laissent augurer un bon comportement des estimateurs dans ce cadre. Nous projetons d'étudier ce cas dans des travaux ultérieurs. Par ailleurs, ces résultats montrent que ces EQME sont très proches pour $k \geq 200$ et ainsi, la détérioration de l'erreur quadratique due au passage du noyau clas-

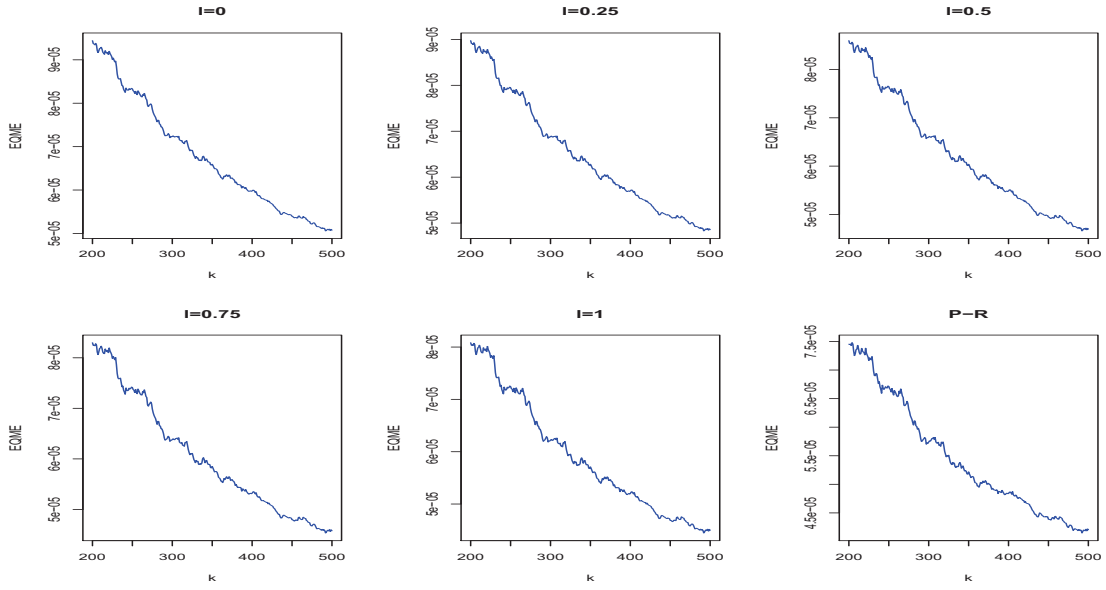


FIG. 4.3 – EQME pour les estimateurs d’une densité gaussienne.

sique au noyau récursif est très faible.

Le TABLEAU 4.2 et la FIGURE 4.4 donnent les erreurs absolues maximum empiriques de nos estimateurs récursifs de la densité pour les valeurs $\ell \in \{0, 1/4, 1/2, 3/4, 1\}$ ainsi que celle de l’estimateur de Parzen-Rosenblatt. Ici, on constate que les six estimateurs peuvent être recommandés pour l’estimation de $f(x)$, car la différence entre ces erreurs est très faible. Ainsi, cette similarité encourage donc notre intérêt pour les estimateurs récursifs. Notons que nous n’avons pas fait une comparaison théorique sur le critère basé sur l’erreur L^1 , comme ce fût le cas pour l’EQM, nous envisageons d’étendre notre étude à ce sujet.

Enfin, le TABLEAU 4.3 donne le temps CPU en secondes pour nos estimateurs récursifs,

k	200	250	300	350	400	450	500
$\ell = 0$	0.02402	0.02260	0.02161	0.02074	0.02014	0.01928	0.01875
$\ell = 0.25$	0.02357	0.02215	0.02115	0.02026	0.01971	0.01887	0.01843
$\ell = 0.5$	0.02322	0.02182	0.02078	0.01989	0.01936	0.01855	0.01819
$\ell = 0.75$	0.02296	0.02157	0.02052	0.01963	0.01908	0.01830	0.01803
$\ell = 1$	0.02332	0.02197	0.02069	0.01953	0.01890	0.01811	0.01791
P-R	0.02246	0.02153	0.02003	0.01923	0.01888	0.01823	0.01779

TAB. 4.2 – EAME pour les estimateurs d’une densité gaussienne.

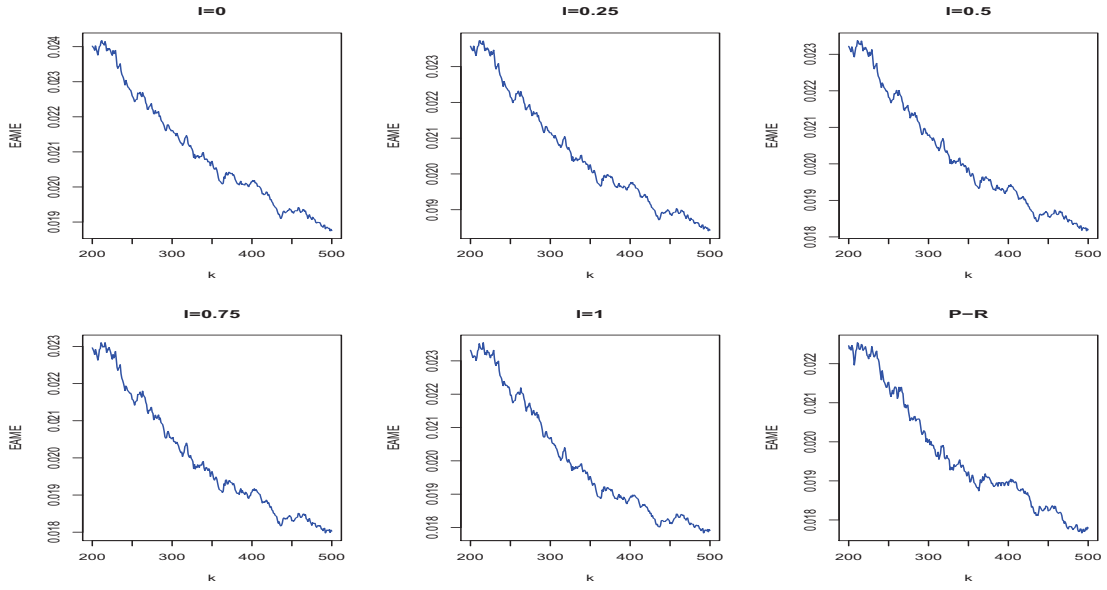


FIG. 4.4 – EAME pour les estimateurs d'une densité gaussienne.

pour $\ell \in \{0, 1/4, 1/2, 3/4, 1\}$ ainsi que pour l'estimateur de Parzen-Rosenblatt. Ici, on note par contre, qu'il faut beaucoup plus de temps pour calculer l'estimateur non récursif que les estimateurs récursifs. De plus, on constate qu'il n'y a pas d'influence du paramètre ℓ sur le temps de calcul, bien que le cas $\ell = 1$ présente un léger avantage de rapidité de calcul par rapport autres choix de paramètre.

k	200	250	300	350	400	450	500
$\ell = 0$	0.05	0.14	0.19	0.27	0.39	0.46	0.49
$\ell = 0.25$	0.06	0.16	0.27	0.28	0.39	0.53	0.53
$\ell = 0.5$	0.03	0.14	0.18	0.24	0.33	0.36	0.44
$\ell = 0.75$	0.03	0.18	0.27	0.27	0.40	0.42	0.60
$\ell = 1$	0.06	0.16	0.17	0.24	0.31	0.37	0.38
P-R	0.08	3.24	6.60	10.17	14.34	19.36	24.82

TAB. 4.3 – Temps CPU pour les estimateurs d'une densité gaussienne.

◇ Cas où f est la densité d'une loi exponentielle de paramètre λ :

Les résultats suivants sont obtenus pour :

$n = 200$, $N = 500$, $\lambda = 1$, $I = [0, 10]$, $J = 200$ et $x_j = x_{j-1} + 0.05$, pour $j = 1, \dots, 200$.

Notons que pour le cas de la densité exponentielle, nous avons délibérément choisi, de tronquer l'estimateur dans la partie positive du support de la loi. L'étude théorique de tels estimateurs tronqués doit encore être effectuée et pourra faire l'objet de travaux ultérieurs.

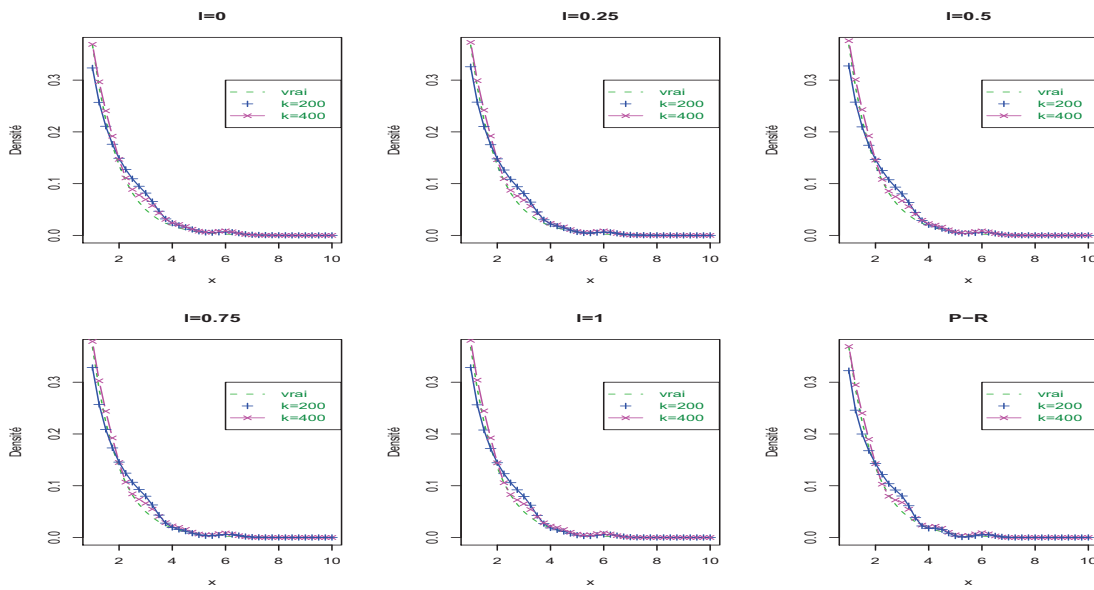
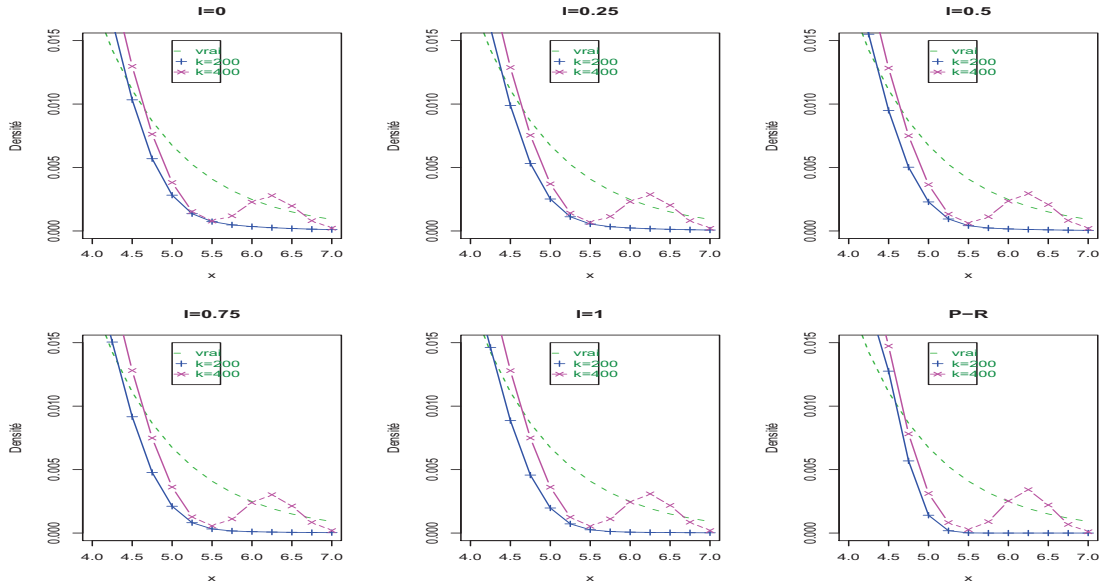


FIG. 4.5 – Estimateurs de la densité d'une loi exponentielle sur $[0, 10]$.

Dans la FIGURE 4.5, nous remarquons également, comme pour le cas gaussien, que les six estimateurs sont tous compétitifs pour l'estimation de $f(x)$, mais nous allons encore faire un zoom sur l'intervalle $[4, 7]$, pour mieux visualiser l'évolution de nos estimateurs en fonction du nombre k d'observations considérées (FIGURE 4.6). Notons tout de même que cette partie est délicate à estimer du fait de la faible valeur de la densité et donc du nombre peu élevé d'observations dans cette zone. Nous cherchons à avoir une précision sur la proximité de nos estimateurs par rapport à la vraie densité en étudiant leurs EQME et EAME (TABLEAUX 4.4 et 4.5).

Les résultats du TABLEAU 4.4 sont représentés graphiquement à la FIGURE 4.7. Nous obtenons des résultats similaires au cas gaussien, c'est-à-dire que l'on constate que l'efficacité de nos estimateurs est croissante en fonction de ℓ selon le critère de l'EQME, et

FIG. 4.6 – Estimateurs de la densité d'une loi exponentielle sur $[4, 7]$.

k	200	250	300	350	400	450	500
$\ell = 0$	0.78479	0.74634	0.72256	0.70738	0.68878	0.66958	0.65533
$\ell = 0.25$	0.76776	0.73051	0.70823	0.69414	0.67626	0.65748	0.64367
$\ell = 0.5$	0.75281	0.71670	0.69583	0.68275	0.66552	0.64711	0.63370
$\ell = 0.75$	0.73940	0.70440	0.68490	0.67278	0.65616	0.63808	0.62503
$\ell = 1$	0.72700	0.69314	0.67501	0.66383	0.64779	0.63004	0.61734
P-R	0.62625	0.59408	0.57637	0.56447	0.55076	0.53360	0.52300

TAB. 4.4 – $\text{EQME} \times 10^2$ pour les estimateurs d'une densité exponentielle.

que l'estimateur de Parzen-Rosenblatt est préférable aux estimateurs récursifs, ce qui est en accord avec les premiers résultats théoriques obtenus aux Théorèmes 2.3.1 et 2.3.2.

Les résultats du TABLEAU 4.5 sont représentés graphiquement à la FIGURE 4.8. Les conclusions faites sur l'EAME et le temps de calcul dans le cas gaussien se confirment également pour une densité exponentielle. En particulier, on retrouve l'efficacité (faible erreur d'estimation et rapidité de calcul) du cas $\ell = 1$ vis à vis des autres choix de ℓ . On note toujours une faible détérioration de l'erreur d'estimation et un important gain de temps de calcul en passant du noyau classique au noyau récursif.

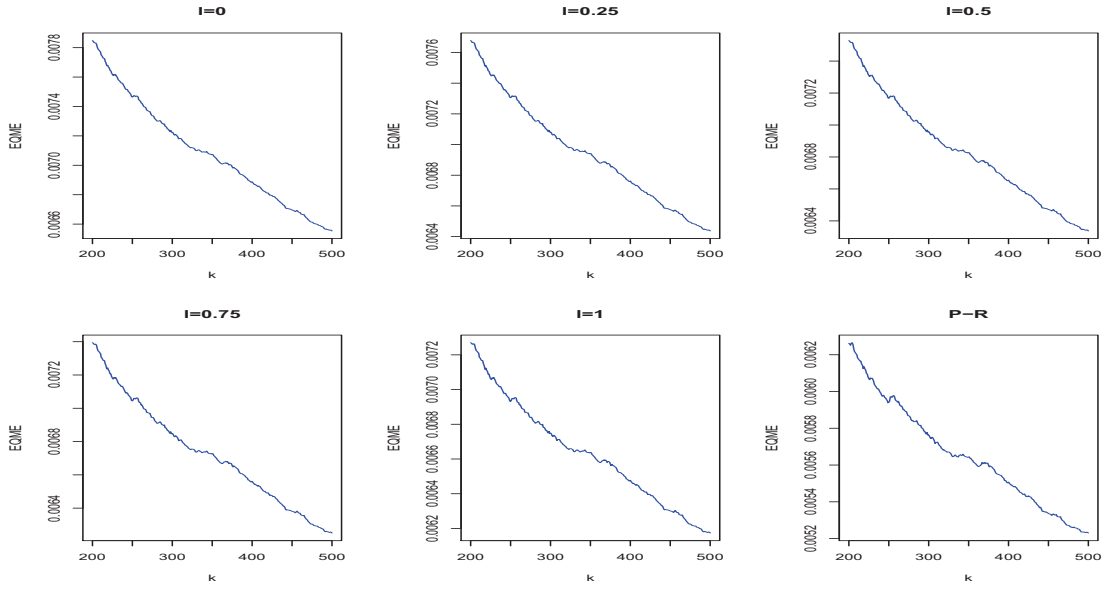


FIG. 4.7 – EQME pour les estimateur d'une densité exponentielle.

k	200	250	300	350	400	450	500
$\ell = 0$	0.64002	0.63437	0.63118	0.62951	0.62696	0.62380	0.62149
$\ell = 0.25$	0.63690	0.63142	0.62849	0.62703	0.62460	0.62147	0.61922
$\ell = 0.5$	0.63412	0.62881	0.62613	0.62488	0.62256	0.61946	0.61726
$\ell = 0.75$	0.63161	0.62646	0.62404	0.62299	0.62078	0.61770	0.61555
$\ell = 1$	0.62927	0.62430	0.62213	0.62128	0.61919	0.61613	0.61403
P-R	0.61119	0.60662	0.60469	0.60394	0.60213	0.59897	0.59711

TAB. 4.5 – EAME $\times 10^2$ pour les estimateurs d'une densité exponentielle.

k	200	250	300	350	400	450	500
$\ell = 0$	0.04	0.11	0.24	0.28	0.31	0.42	0.59
$\ell = 0.25$	0.05	0.12	0.21	0.29	0.33	0.50	0.56
$\ell = 0.5$	0.06	0.10	0.22	0.31	0.31	0.36	0.51
$\ell = 0.75$	0.06	0.11	0.20	0.28	0.42	0.42	0.61
$\ell = 1$	0.06	0.11	0.24	0.26	0.31	0.35	0.45
P-R	0.08	3.54	6.43	10.05	14.36	18.85	27.45

TAB. 4.6 – Temps CPU pour les estimateurs d'une densité exponentielle.

4.2.2 Estimation de la régression

Dans cette section, nous analysons par simulations les performances de l'estimateur récursif $r_n^\ell(x)$ étudié au Chapitre 3. Nous allons simuler N observations d'un processus

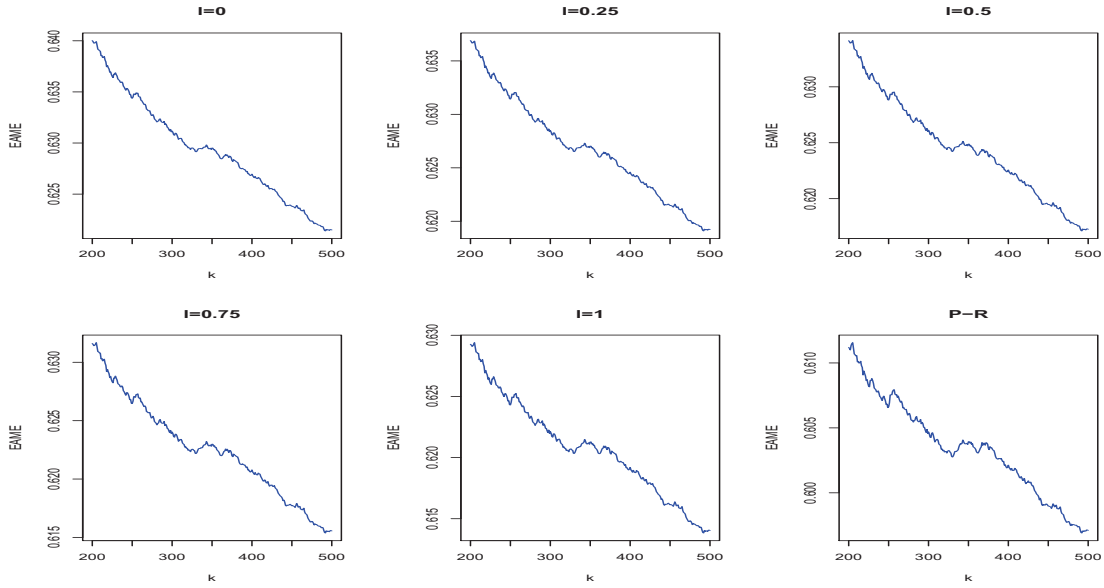


FIG. 4.8 – EAME pour les estimateur d'une densité exponentielle.

bivarié :

$$Z_t = (X_t, Y_t, 1 \leq t \leq N), N \in \mathbb{N}^*$$

avec :

$$r(x) = E(Y_t | X_t = x).$$

Pour cela, nous considérons une suite de N variables X_t iid suivant une loi normale centrée de variance σ^2 . À partir de l'échantillon transformé $r(X_t), t = 1, \dots, N$, nous construisons le processus bivarié :

$$Z_t = (X_t, r(X_t) + \varepsilon_t, t = 1 \dots, N), \text{ avec } \varepsilon_t \sim \mathcal{N}(0, \kappa^2).$$

Le but est d'estimer la fonction $r(x)$ sur un intervalle $I = [a, b]$, à partir des k premières observations $Z_1, \dots, Z_k$, pour $k = n, \dots, N$ avec n un entier non nul tel que $n < N$. Comme pour l'estimation de la densité, nous considérons une subdivision ϖ de I définie par :

$$\varpi = \{x_0 = a, x_1, \dots, x_J = b\}, J \geq 1,$$

et pour $k = n, \dots, N$, nous calculons les valeurs $r_k^\ell(x_j)$ ($\ell \in \{0, 0.25, 0.5, 0.75, 1\}$) et $r_k^{NW}(x_j)$, en tout point x_j de ϖ .

Nous utilisons les mêmes critères utilisés au paragraphe 4.2.1, pour évaluer les performances de nos estimateurs. Le calcul des erreurs d'estimation se fait également avec la même approche décrite au paragraphe 4.2.1.

Résultats :

Nous considérons deux cas : l'estimation d'une fonction de régression parabolique et d'une fonction de régression valeur absolue.

◇ **Cas où la fonction de régression est parabolique :** Les résultats suivant sont obtenus pour :

$n = 200$, $N = 500$, $\sigma^2 = 5$, $I = [-4, 4]$, $J = 160$, $r(x) = x^2 + x + 1$ et $x_j = x_{j-1} + 0.05$, $1 \leq j \leq 160$.

Nous représentons d'abord, à la FIGURE 4.9, la vraie fonction de régression avec ses estimateurs pour différentes valeurs de k et ℓ , dans le cas où $\kappa^2 = 0.5$. Ensuite, faisons

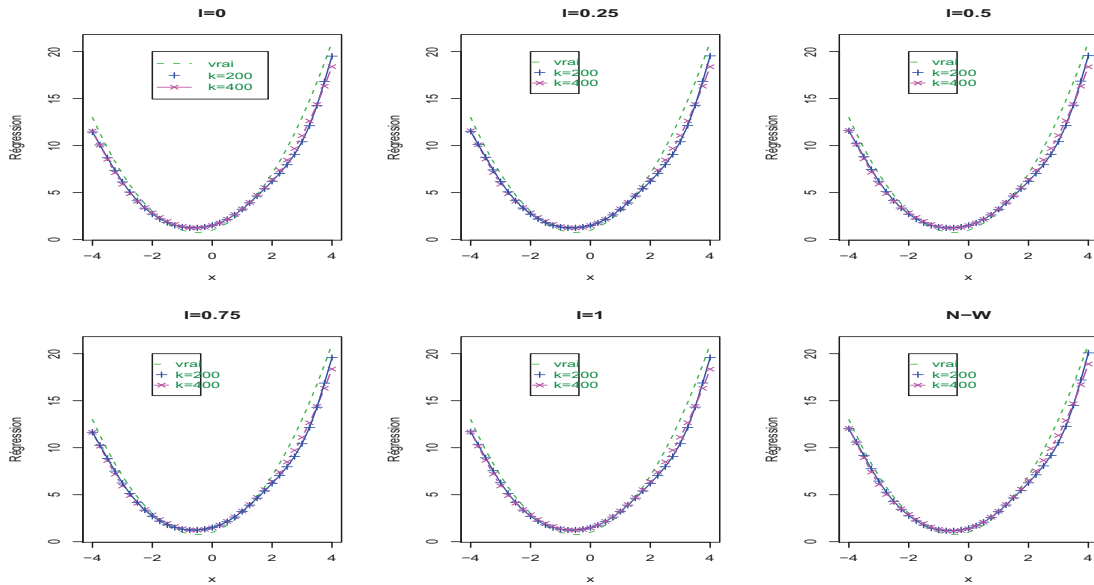
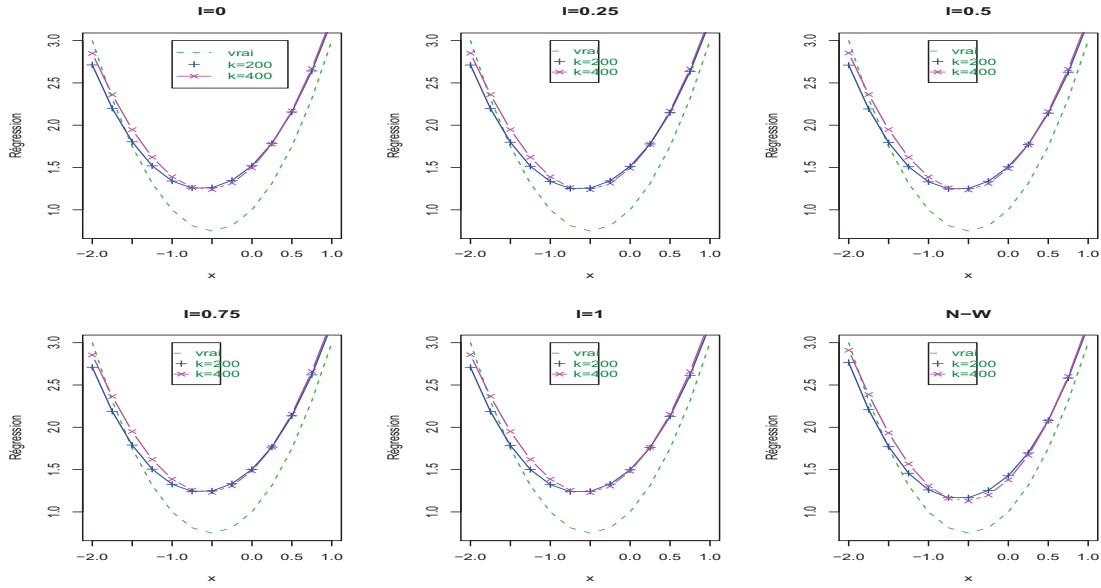


FIG. 4.9 – Estimateurs d'une régression parabolique sur $[-4, 4]$

un zoom au voisinage du sommet de la parabole (FIGURE 4.10). Bien qu'ici l'estimation est moins bonne par rapport à la densité, nous constatons dans la FIGURE 4.10, une amélioration de nos estimateurs avec les 400 premières observations plutôt qu'avec les 200. En outre, on retrouve le fait que l'estimateur non récursif estime mieux la régression pour $k = 400$. La question est donc d'avoir une précision sur la proximité de ces estimateurs par

FIG. 4.10 – Estimateurs d’une régression parabolique sur $[-2, 1]$

rapport à la vraie densité, mais surtout d’évaluer la perte de précision due au passage du noyau classique au noyau récursif. Pour cela, nous étudions leurs EQME et EAME. Nous obtenons ainsi les résultats présentés aux TABLEAUX 4.7 et 4.8, avec $T = 100$ échantillons et pour différentes valeurs de κ^2 . Dans cet exemple, on constate que l’efficacité de nos estimateurs est croissante en fonction de ℓ selon le critère de l’EQME, et que l’estimateur de Nadaraya-Watson est préférable aux estimateurs récursifs. Bien que nous n’ayons pas réalisé une étude théorique concernant la comparaison des estimateurs de la régression, ce constat est sans surprise grâce au Théorème 3.3.1, qui montre que les parties variables en ℓ des EQM asymptotiques de nos estimateurs de la régression sont identiques à celles des estimateurs de la densité (c.f. Théorème 2.2.6(c)).

Les résultats des TABLEAU 4.7 et 4.8 sont représentés graphiquement aux FIGURES 4.11 et 4.12 pour $\kappa^2 = 0.5$. Notons par ailleurs que, comme pour l’estimation de la densité, les EQME de nos estimateurs sont proches pour $k \geq 200$, bien qu’ici la détérioration de l’erreur quadratique due au passage du noyau classique au noyau récursif est moins faible que celle observée dans le cas de l’estimation de la densité.

Enfin, le TABLEAU 4.9 présente le temps de calcul de nos estimateurs de la régression. Dans cet exemple, nous constatons que le gain de temps dû au passage du noyau classique au noyau récursif est considérable.

CHAPITRE 4. Application à la prévision et simulations

	k	200	250	300	350	400	450	500
$\kappa^2 = 0$	$\ell = 0$	3.00745	2.66555	2.38530	2.17022	2.01224	1.87914	1.75416
	$\ell = 0.25$	2.93818	2.59119	2.31165	2.10231	1.94431	1.81494	1.69424
	$\ell = 0.5$	2.87695	2.52915	2.24852	2.04038	1.88330	1.75497	1.63527
	$\ell = 0.75$	2.81949	2.47122	2.18985	1.98311	1.82709	1.69999	1.58144
	$\ell = 1$	2.76545	2.41708	2.13530	1.93014	1.77531	1.64960	1.53230
	N-W	1.53149	1.28392	1.09015	0.96028	0.87034	0.78670	0.71280
$\kappa^2 = 0.5$	$\ell = 0$	3.14528	2.78423	2.50855	2.30359	2.15939	2.05767	1.93397
	$\ell = 0.25$	3.07657	2.71364	2.43788	2.23331	2.09187	1.99282	1.87021
	$\ell = 0.5$	3.01437	2.64998	2.37429	2.17010	2.03135	1.93490	1.81336
	$\ell = 0.75$	2.95783	2.59235	2.31686	2.11306	1.97691	1.88301	1.76248
	$\ell = 1$	2.90633	2.54008	2.26487	2.06146	1.92781	1.83638	1.71681
	N-W	1.66619	1.43676	1.26795	1.13733	1.05905	1.00531	0.93381
$\kappa^2 = 1$	$\ell = 0$	3.19881	2.86802	2.57585	2.37147	2.19237	2.06544	1.95558
	$\ell = 0.25$	3.13751	2.80621	2.51128	2.30806	2.12816	2.00218	1.89367
	$\ell = 0.5$	3.08054	2.74881	2.45137	2.24943	2.06890	1.94404	1.83702
	$\ell = 0.75$	3.02715	2.69523	2.39558	2.19509	2.01409	1.89054	1.78513
	$\ell = 1$	2.97692	2.64509	2.34357	2.14468	1.96338	1.84131	1.73760
	N-W	1.73268	1.49316	1.28177	1.15493	1.04319	0.96071	0.89734
$\kappa^2 = 5$	$\ell = 0$	3.22892	2.96587	2.60917	2.34679	2.20721	2.07679	1.96930
	$\ell = 0.25$	3.36280	2.92713	2.59530	2.37707	2.19007	2.01139	1.89688
	$\ell = 0.5$	3.30156	2.86147	2.52715	2.31107	2.12456	1.94607	1.83352
	$\ell = 0.75$	3.24453	2.80042	2.46398	2.25027	2.06447	1.88643	1.77592
	$\ell = 1$	3.19117	2.74349	2.40529	2.19415	2.00927	1.83191	1.72349
	N-W	1.97628	1.64167	1.40112	1.26373	1.14925	1.03104	0.96343
$\kappa^2 = 7$	$\ell = 0$	3.30056	3.06609	2.73727	2.56654	2.34196	2.20264	2.05205
	$\ell = 0.25$	3.25362	3.02033	2.68204	2.51228	2.28362	2.14224	1.99024
	$\ell = 0.5$	3.21301	2.98043	2.63261	2.46364	2.23074	2.08743	1.93416
	$\ell = 0.75$	3.17781	2.94562	2.58830	2.41998	2.18270	2.03763	1.88319
	$\ell = 1$	3.14741	2.91532	2.54855	2.38080	2.13904	1.99234	1.83682
	N-W	1.84001	1.68053	1.43217	1.33116	1.15931	1.06802	0.96989
$\kappa^2 = 10$	$\ell = 0$	3.34100	2.90384	2.65928	2.52851	2.26358	2.12566	2.05398
	$\ell = 0.25$	3.26020	2.82100	2.58084	2.45680	2.18858	2.05288	1.98537
	$\ell = 0.5$	3.18520	2.74437	2.50876	2.39147	2.12013	1.98683	1.92353
	$\ell = 0.75$	3.11542	2.67338	2.44250	2.33197	2.05767	1.92692	1.86784
	$\ell = 1$	3.05049	2.60763	2.38163	2.27786	2.00070	1.87263	1.81776
	N-W	1.80559	1.48215	1.33512	1.28240	1.09695	1.01065	0.98472

TAB. 4.7 – EQME pour les estimateurs d’une régression parabolique.

	k	200	250	300	350	400	450	500
$\kappa^2 = 0$	$\ell = 0$	4.09858	3.83213	3.61061	3.42434	3.29892	3.17187	3.01985
	$\ell = 0.25$	4.04305	3.76962	3.54374	3.35715	3.23011	3.10135	2.94602
	$\ell = 0.5$	3.98929	3.71029	3.48023	3.29378	3.16580	3.03564	2.87728
	$\ell = 0.75$	3.93741	3.65455	3.42115	3.23412	3.10569	2.97428	2.81394
	$\ell = 1$	3.88759	3.60383	3.36560	3.17814	3.04993	2.91750	2.75508
	N-W	2.74989	2.48017	2.26247	2.09108	1.99919	1.87329	1.72352
$\kappa^2 = 0.5$	$\ell = 0$	4.72216	4.43001	4.19203	3.99124	3.85568	3.74682	3.61955
	$\ell = 0.25$	4.65306	4.35747	4.11601	3.91216	3.77838	3.66920	3.54210
	$\ell = 0.5$	4.59221	4.29144	4.04838	3.84166	3.70901	3.60002	3.47266
	$\ell = 0.75$	4.53663	4.23125	3.98790	3.77818	3.64648	3.53756	3.41000
	$\ell = 1$	4.48656	4.17682	3.93319	3.72249	3.58970	3.48110	3.35380
	N-W	3.26703	3.01214	2.82928	2.64388	2.53107	2.43394	2.34552
$\kappa^2 = 1$	$\ell = 0$	4.96471	4.70513	4.47659	4.29361	4.15159	4.02833	3.88661
	$\ell = 0.25$	4.76298	4.53215	4.27563	4.12990	3.95074	3.85290	3.73843
	$\ell = 0.5$	4.70631	4.47179	4.21082	4.06583	3.88465	3.78494	3.67194
	$\ell = 0.75$	4.65292	4.41511	4.15186	4.00705	3.82354	3.72222	3.61069
	$\ell = 1$	4.60316	4.36401	4.09914	3.95342	3.76712	3.66473	3.55527
	N-W	3.34443	3.13341	2.91629	2.79458	2.62940	2.55343	2.46962
$\kappa^2 = 5$	$\ell = 0$	4.97337	4.66381	4.39176	4.18760	4.02391	3.85424	3.70874
	$\ell = 0.25$	4.91732	4.59799	4.31603	4.11142	3.94858	3.77660	3.62706
	$\ell = 0.5$	4.86498	4.53534	4.24575	4.03990	3.87940	3.70429	3.55122
	$\ell = 0.75$	4.81712	4.47631	4.18150	3.97399	3.81542	3.63711	3.48132
	$\ell = 1$	4.77187	4.42031	4.12276	3.91333	3.75573	3.57469	3.41663
	N-W	3.68481	3.32573	3.05368	2.86946	2.73227	2.54701	2.40723
$\kappa^2 = 7$	$\ell = 0$	4.97394	4.74539	4.50261	4.37304	4.19653	4.08109	3.93051
	$\ell = 0.25$	4.91896	4.69523	4.44663	4.31789	4.13483	4.01538	3.86248
	$\ell = 0.5$	4.88078	4.64928	4.39485	4.26645	4.07865	3.95486	3.80200
	$\ell = 0.75$	4.84819	4.60888	4.34742	4.21935	4.02835	3.90040	3.74650
	$\ell = 1$	4.82042	4.57445	4.30448	4.17700	3.98201	3.85086	3.69624
	N-W	3.62564	3.39146	3.12889	3.02900	2.85088	2.73432	2.62157
$\kappa^2 = 10$	$\ell = 0$	5.02434	4.70932	4.51557	4.38363	4.17785	4.05350	3.94531
	$\ell = 0.25$	4.95799	4.63293	4.44064	4.31099	4.10012	3.97475	3.86640
	$\ell = 0.5$	4.89591	4.56150	4.37124	4.24448	4.02766	3.90290	3.79535
	$\ell = 0.75$	4.83725	4.49501	4.30819	4.18272	3.96062	3.83716	3.73052
	$\ell = 1$	4.78165	4.43303	4.24968	4.12604	3.89903	3.77691	3.67104
	N-W	3.60823	3.29731	3.13773	3.04311	2.85004	2.73057	2.64798

TAB. 4.8 – EAME pour les estimateurs d'une régression parabolique.

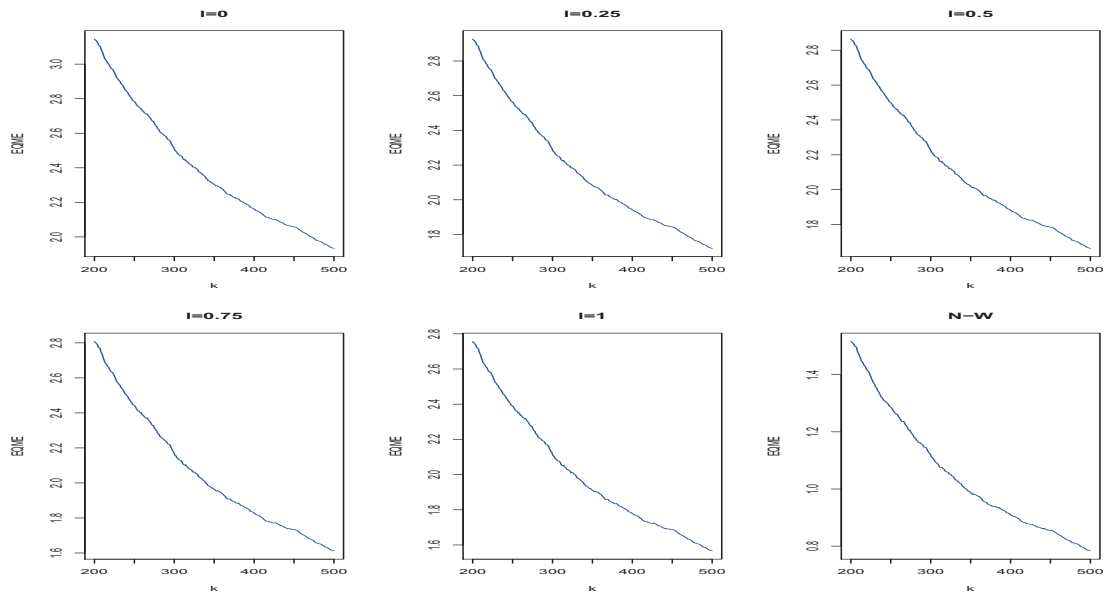


FIG. 4.11 – EQME pour les estimateurs d’une régression parabolique.

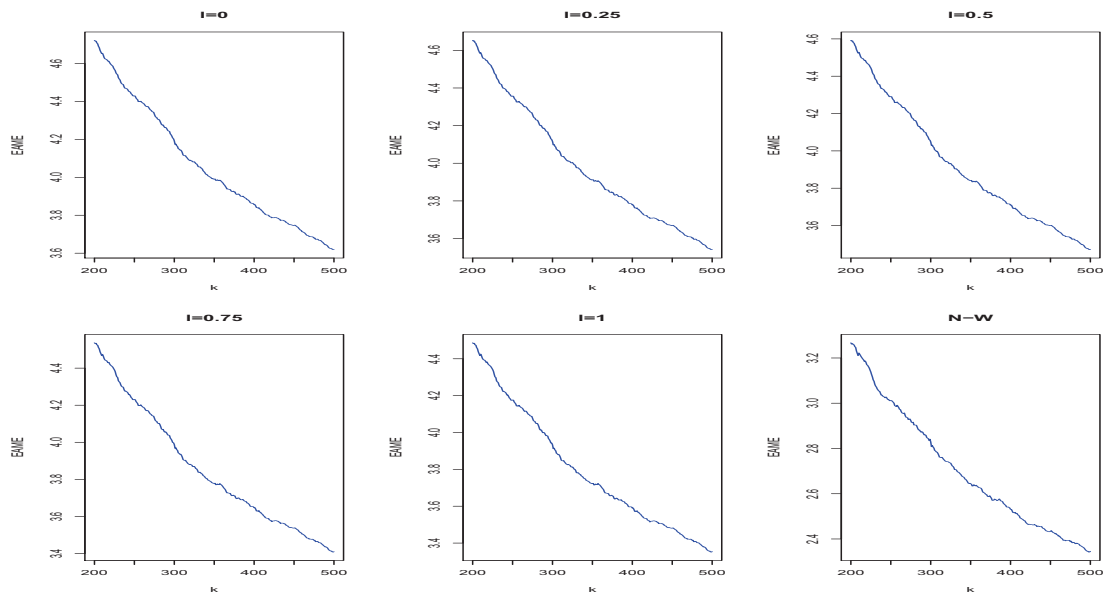


FIG. 4.12 – EAME pour les estimateurs d’une régression parabolique.

k	200	250	300	350	400	450	500
$\ell = 0$	0.17	0.20	0.30	0.39	0.49	0.61	0.71
$\ell = 0.25$	0.10	0.22	0.32	0.43	0.55	0.70	0.85
$\ell = 0.5$	0.11	0.20	0.30	0.38	0.48	0.60	0.70
$\ell = 0.75$	0.11	0.22	0.33	0.42	0.56	0.69	0.81
$\ell = 1$	0.10	0.21	0.28	0.36	0.46	0.55	0.65
N-W	0.24	9.20	18.87	29.65	41.65	54.46	68.50

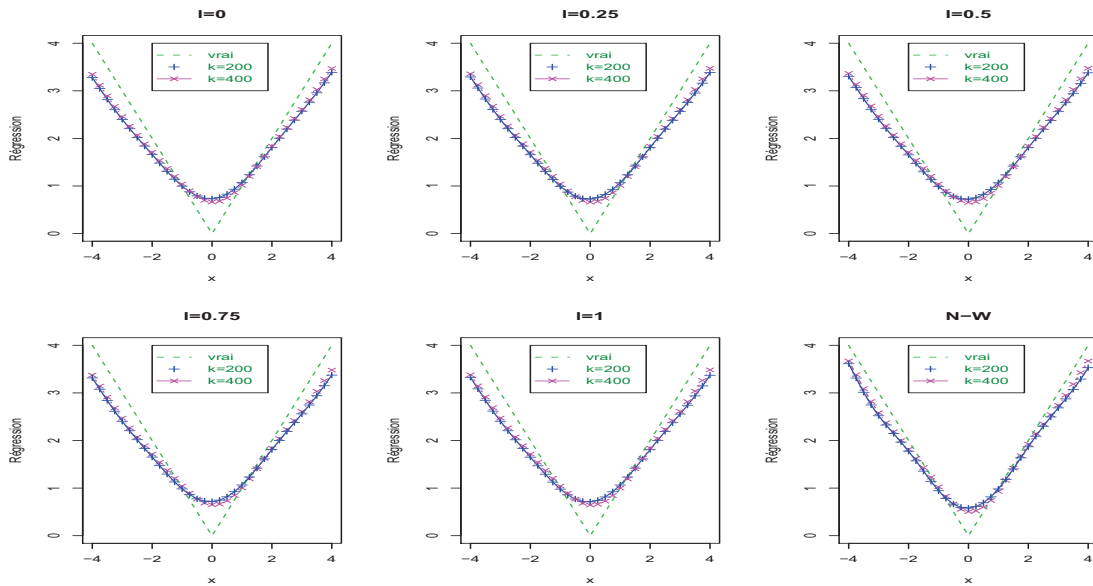
TAB. 4.9 – Temps CPU pour les estimateurs d'une régression parabolique.

◇ Cas où la fonction de régression est la valeur absolue :

Les résultats suivant sont obtenus pour :

$n = 200$, $N = 500$, $\sigma^2 = 5$, $I = [-4, 4]$, $J = 160$, $r(x) = |x|$ et $x_j = x_{j-1} + 0.05$, $1 \leq j \leq 160$.

Nous représentons d'abord, à la FIGURE 4.13, la vraie fonction de régression avec ses estimateurs dans le cas où $\kappa^2 = 0.5$. Ensuite, nous faisons un zoom sur l'intervalle $[-1, 1]$


 FIG. 4.13 – Estimateurs d'une régression absolue sur $[-4, 4]$.

(FIGURE 4.14). Les EQME et EAME de nos estimateurs sont présentées respectivement aux TABLEAUX 4.10 et 4.11, puis représentées graphiquement aux FIGURES 4.15 et 4.16 pour $\kappa^2 = 0.5$. Nous obtenons ainsi des résultats similaires au cas parabolique, c'est-à-dire

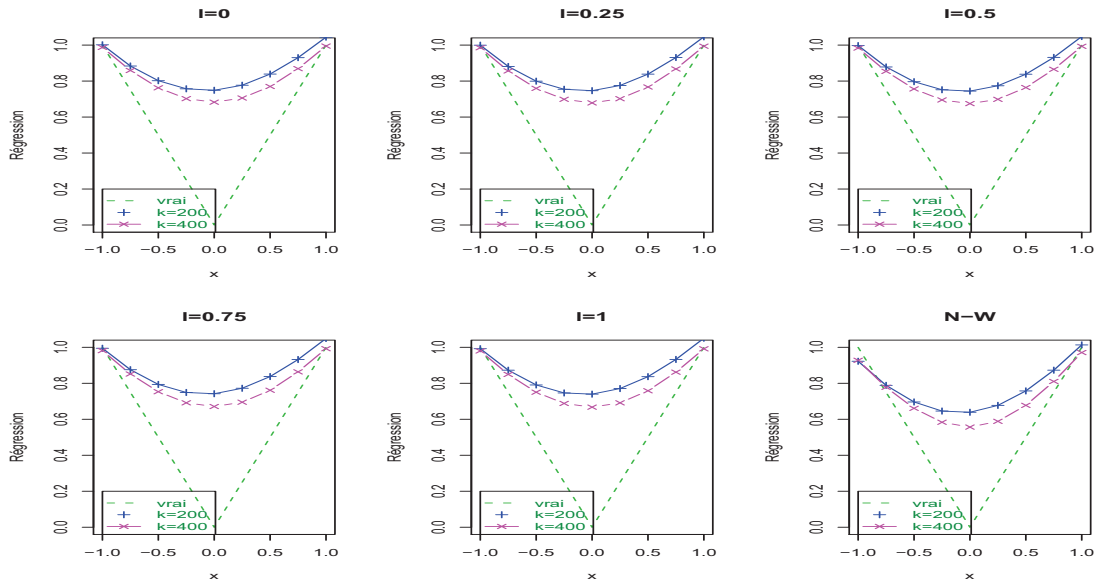


FIG. 4.14 – Estimateurs d'une régression absolue sur $[-1, 1]$.

que l'on constate que l'efficacité de nos estimateurs est croissante en fonction de ℓ selon le critère de l'EQME, et que l'estimateur de Nadaraya-Watson est préférable aux estimateurs récursifs.

Les temps de calcul de nos estimateurs sont donnés par le TABLEAU 4.12. Cet exemple montre que les conclusions faites sur l'EAME et le temps de calcul dans le cas parabolique se confirment également pour une fonction de régression valeur absolue. En particulier, on retrouve l'efficacité (faible erreur d'estimation et rapidité de calcul) du cas $\ell = 1$ vis à vis des autres choix de ℓ . On note toujours une faible détérioration de l'erreur d'estimation et un gain de temps de calcul très considérable en passant du noyau classique au noyau récursif.

Notons enfin que l'ensemble de ces résultats montrent que les erreurs d'estimation sont d'autant plus faibles que le nombre d'observations k augmente. Ceci exprime le fait qu'on enrichit la base des données de notre estimation en utilisant le maximum d'information disponible.

	k	200	250	300	350	400	450	500
$\kappa^2 = 0$	$\ell = 0$	0.16983	0.15197	0.13714	0.12602	0.11750	0.11014	0.10355
	$\ell = 0.25$	0.16594	0.14807	0.13322	0.12215	0.11371	0.10643	0.09990
	$\ell = 0.5$	0.16238	0.14451	0.12965	0.11864	0.11028	0.10308	0.09661
	$\ell = 0.75$	0.15911	0.14126	0.12640	0.11544	0.10716	0.10005	0.09364
	$\ell = 1$	0.15610	0.13827	0.12342	0.11252	0.10433	0.09730	0.09095
	N-W	0.08681	0.07457	0.06557	0.05881	0.05371	0.04931	0.04568
$\kappa^2 = 0.5$	$\ell = 0$	0.17819	0.15505	0.13846	0.12705	0.11804	0.11269	0.10566
	$\ell = 0.25$	0.17384	0.15058	0.13408	0.12277	0.11393	0.10879	0.10186
	$\ell = 0.5$	0.16993	0.14656	0.13016	0.11896	0.11027	0.10535	0.09851
	$\ell = 0.75$	0.16640	0.14295	0.12666	0.11557	0.10702	0.10230	0.09554
	$\ell = 1$	0.16322	0.13970	0.12353	0.11253	0.10412	0.09959	0.09291
	N-W	0.09471	0.07944	0.06876	0.06156	0.05623	0.05356	0.04942
$\kappa^2 = 1$	$\ell = 0$	0.24017	0.19179	0.18810	0.17283	0.16820	0.15845	0.14188
	$\ell = 0.25$	0.23407	0.18513	0.18513	0.16735	0.16316	0.15356	0.13684
	$\ell = 0.5$	0.22856	0.17913	0.17887	0.17250	0.15874	0.14928	0.13240
	$\ell = 0.75$	0.22361	0.17372	0.17306	0.16822	0.15486	0.14553	0.12849
	$\ell = 1$	0.21916	0.16885	0.16283	0.15444	0.15146	0.14225	0.12504
	N-W	0.14906	0.10886	0.10415	0.10078	0.09948	0.09312	0.08060
$\kappa^2 = 5$	$\ell = 0$	0.27262	0.24090	0.21826	0.22927	0.20197	0.19235	0.17637
	$\ell = 0.25$	0.26618	0.23507	0.21286	0.22483	0.19750	0.18778	0.17180
	$\ell = 0.5$	0.26086	0.23025	0.20842	0.22133	0.19391	0.18403	0.16801
	$\ell = 0.75$	0.21341	0.17583	0.15429	0.13520	0.13384	0.13912	0.11545
	$\ell = 1$	0.25292	0.22302	0.20175	0.21663	0.18880	0.17846	0.16225
	N-W	0.20567	0.18037	0.16334	0.17366	0.15100	0.14143	0.12956
$\kappa^2 = 7$	$\ell = 0$	0.44626	0.36819	0.34719	0.27715	0.25065	0.21869	0.19784
	$\ell = 0.25$	0.43202	0.35411	0.33459	0.26410	0.23850	0.20726	0.18705
	$\ell = 0.5$	0.42068	0.34257	0.32434	0.25318	0.22835	0.19777	0.17808
	$\ell = 0.75$	0.41178	0.33314	0.31602	0.24400	0.21984	0.18983	0.17059
	$\ell = 1$	0.40501	0.32552	0.30932	0.23625	0.21268	0.18319	0.16431
	N-W	0.32793	0.25706	0.24314	0.18127	0.16203	0.13810	0.12252
$\kappa^2 = 10$	$\ell = 0$	0.48508	0.38356	0.33996	0.34880	0.27343	0.25064	0.23404
	$\ell = 0.25$	0.47549	0.37316	0.32995	0.34030	0.26467	0.24282	0.22703
	$\ell = 0.5$	0.46840	0.36489	0.32186	0.33362	0.25758	0.23658	0.22150
	$\ell = 0.75$	0.46343	0.35837	0.31533	0.32840	0.25184	0.23159	0.21714
	$\ell = 1$	0.46034	0.35336	0.31012	0.32443	0.24721	0.22763	0.21374
	N-W	0.39337	0.29942	0.26181	0.27402	0.20952	0.19420	0.18348

TAB. 4.10 – EQME pour les estimateurs d'une régression absolue.

	k	200	250	300	350	400	450	500
$\kappa^2 = 0$	$\ell = 0$	0.79022	0.74892	0.71425	0.68185	0.66148	0.64553	0.63066
	$\ell = 0.25$	0.78016	0.73880	0.70383	0.67130	0.65183	0.63610	0.62176
	$\ell = 0.5$	0.77112	0.72971	0.69450	0.66217	0.64388	0.62797	0.61403
	$\ell = 0.75$	0.76322	0.72170	0.68628	0.65440	0.63679	0.62101	0.60731
	$\ell = 1$	0.75644	0.71453	0.67910	0.64756	0.63042	0.61498	0.60127
	N-W	0.60734	0.57578	0.55364	0.53456	0.51892	0.50641	0.49582
$\kappa^2 = 0.5$	$\ell = 0$	0.81597	0.76683	0.72815	0.69901	0.67481	0.65956	0.64787
	$\ell = 0.25$	0.80586	0.75517	0.71692	0.68728	0.66438	0.64920	0.63807
	$\ell = 0.5$	0.79704	0.74529	0.70747	0.67763	0.65513	0.64047	0.62969
	$\ell = 0.75$	0.78903	0.73660	0.69958	0.66961	0.64721	0.63298	0.62231
	$\ell = 1$	0.78186	0.72888	0.69271	0.66265	0.64042	0.62633	0.61590
	N-W	0.64736	0.60378	0.57420	0.54736	0.52873	0.51464	0.50757
$\kappa^2 = 1$	$\ell = 0$	0.93109	0.81254	0.77506	0.74637	0.72996	0.70265	0.66972
	$\ell = 0.25$	0.91520	0.79433	0.75957	0.73025	0.71383	0.68746	0.65647
	$\ell = 0.5$	0.90122	0.77875	0.74636	0.71596	0.69967	0.67405	0.64556
	$\ell = 0.75$	0.88856	0.76555	0.73476	0.70418	0.68735	0.66274	0.63635
	$\ell = 1$	0.87737	0.75501	0.72489	0.69454	0.67734	0.65360	0.62848
	N-W	0.72194	0.62327	0.59315	0.56323	0.54872	0.52879	0.51000
$\kappa^2 = 5$	$\ell = 0$	0.96211	0.89832	0.83874	0.83762	0.77494	0.75462	0.72147
	$\ell = 0.25$	0.94945	0.88709	0.82863	0.82651	0.76512	0.74427	0.71074
	$\ell = 0.5$	0.93877	0.87838	0.82056	0.81864	0.75816	0.73630	0.70256
	$\ell = 0.75$	0.92978	0.87108	0.81479	0.81322	0.75408	0.73005	0.69641
	$\ell = 1$	0.92278	0.86542	0.81029	0.81007	0.75100	0.72486	0.69135
	N-W	0.82454	0.77256	0.71891	0.71613	0.65758	0.63164	0.60154
$\kappa^2 = 7$	$\ell = 0$	1.32718	1.21484	1.17709	1.05372	0.99960	0.94676	0.90287
	$\ell = 0.25$	1.30139	1.18950	1.15531	1.02856	0.97423	0.92119	0.87858
	$\ell = 0.5$	1.28044	1.16835	1.13751	1.00683	0.95209	0.89974	0.85843
	$\ell = 0.75$	1.26324	1.15209	1.12416	0.98787	0.93386	0.88250	0.84223
	$\ell = 1$	1.24990	1.13968	1.11312	0.97161	0.91855	0.86905	0.82896
	N-W	1.13622	1.03433	1.00579	0.88142	0.82357	0.77885	0.74084
$\kappa^2 = 10$	$\ell = 0$	1.39914	1.23773	1.12082	1.11554	0.97766	0.94230	0.91264
	$\ell = 0.25$	1.38789	1.22212	1.10214	1.10028	0.95949	0.92650	0.90132
	$\ell = 0.5$	1.37938	1.20920	1.08727	1.08786	0.94396	0.91558	0.89264
	$\ell = 0.75$	1.37347	1.19823	1.07487	1.07838	0.93206	0.90726	0.88621
	$\ell = 1$	1.37015	1.18980	1.06475	1.07111	0.92398	0.90141	0.88189
	N-W	1.27434	1.11140	0.98727	0.99202	0.84446	0.82351	0.80441

TAB. 4.11 – EAME pour les estimateurs d’une régression absolue.

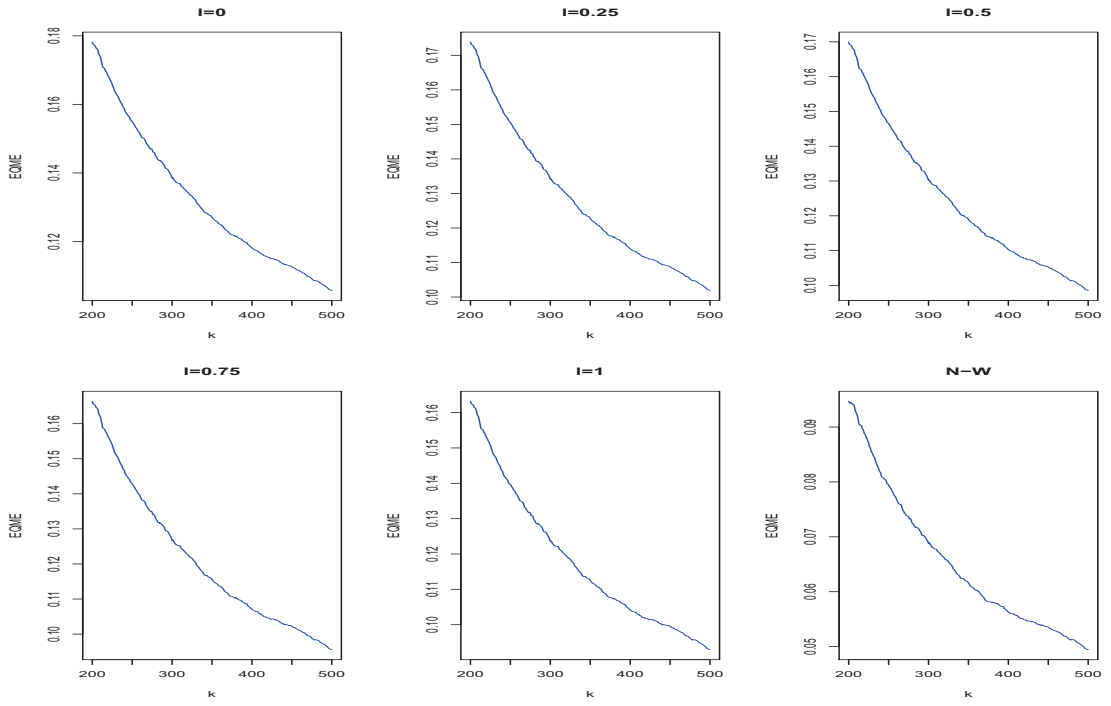


FIG. 4.15 – EQME pour les estimateurs d’une régression absolue sur $[-1, 1]$.

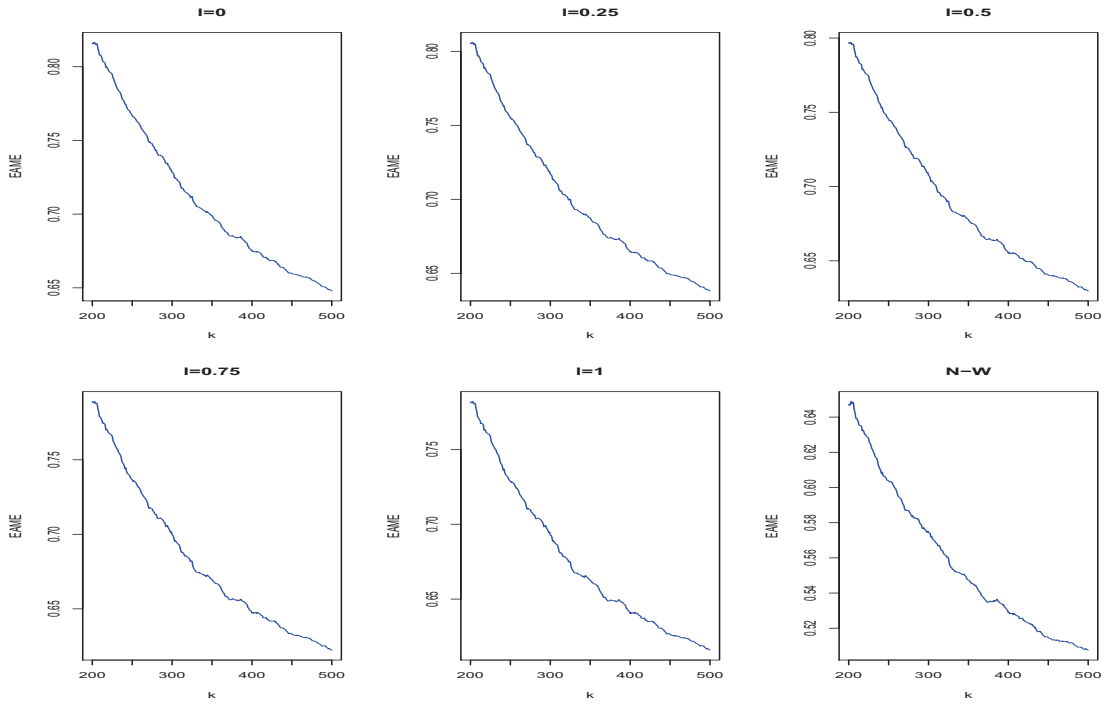


FIG. 4.16 – EQME pour les estimateurs d’une régression absolue sur $[-1, 1]$.

k	200	250	300	350	400	450	500
$\ell = 0$	0.14	0.20	0.29	0.39	0.49	0.60	0.74
$\ell = 0.25$	0.14	0.20	0.31	0.42	0.56	0.68	0.85
$\ell = 0.5$	0.14	0.19	0.28	0.38	0.50	0.58	0.72
$\ell = 0.75$	0.10	0.21	0.31	0.42	0.56	0.67	0.87
$\ell = 1$	0.09	0.19	0.27	0.37	0.47	0.57	0.69
N-W	0.23	8.99	18.77	29.75	41.59	54.60	71.49

TAB. 4.12 – Temps CPU pour les estimateurs d’une régression absolue.

4.2.3 Prévision

Pour étudier les performances de nos méthodes non paramétriques basées sur des noyaux récurrents, nous reprenons un modèle ARMA :

$$\xi_t = 0.9\xi_{t-1} + 1000 + \varepsilon_t,$$

où ε_t est un bruit blanc fort $\mathcal{N}(0, 5)$. Ce modèle fait partie de ceux étudiés par Carbon et Francq [21], à partir duquel ces auteurs ont montré que les résultats obtenus par la méthode non paramétrique sont très proches de ceux obtenus par la méthode de Box Jenkins, méthode pourtant bien adaptée à ce type de séries. Pour cette série, nous comparons les prédictors non paramétriques récurrents et non récurrents entre eux ainsi qu’avec le prédictor obtenu par lissage exponentiel.

On se place dans le cadre de la section 4.1, où l’on cherche à faire les $N - n$ prévisions successives $\hat{\xi}_{n+p}^\ell$ de la série à partir des $N - n$ suites de données $\xi_1, \dots, \xi_{n+p-1}$, respectivement, pour $p = 1, \dots, N - n$ avec n et N deux entiers naturels tels que $n < N$.

Nous retenons trois critères pour évaluer les performances de nos prévisions : l’erreur quadratique moyenne empirique (EQME), l’erreur absolue relative moyenne (EARM), définies respectivement par :

$$EQME = \frac{1}{N - n} \sum_{p=1}^{N-n} \left(\hat{\xi}_{n+p} - \xi_{n+p} \right)^2,$$

et

$$EARM = \frac{1}{N - n} \sum_{p=1}^{N-n} \frac{|\hat{\xi}_{n+p} - \xi_{n+p}|}{|\xi_{n+p}|},$$

et le temps de calcul.

Pour la série étudiée, nous représentons graphiquement la série jusqu'à la valeur n , puis nous représentons, dans une même figure, entre n et N , les prévisions (pour les noyaux récursif non récursif et pour le lissage exponentiel) avec les vraies valeurs, et nous donnons les valeurs des EQME, EARM et le temps CPU.

Nous commençons par représenter à la FIGURE 4.17 les $n = 100$ premières valeurs de cette série.

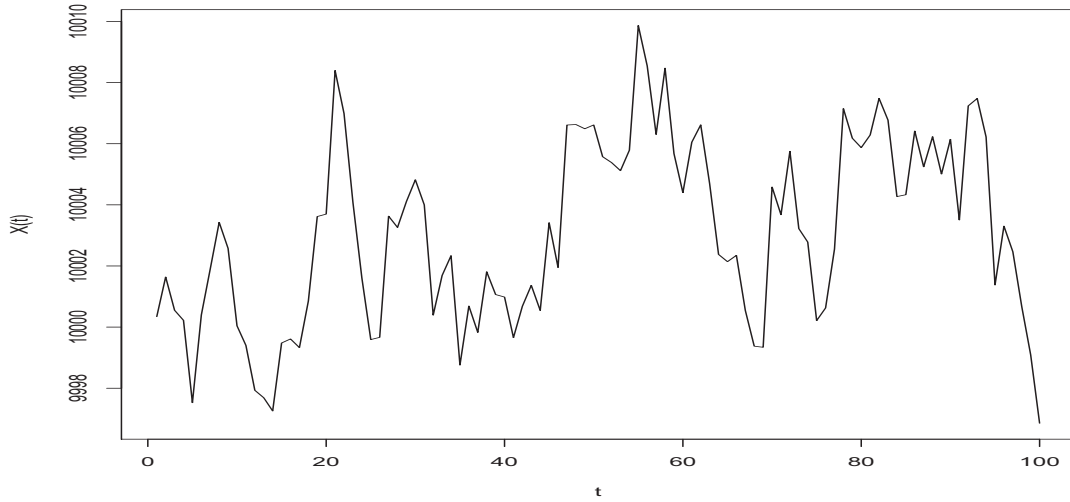


FIG. 4.17 – Représentation graphique des premières valeurs de la série.

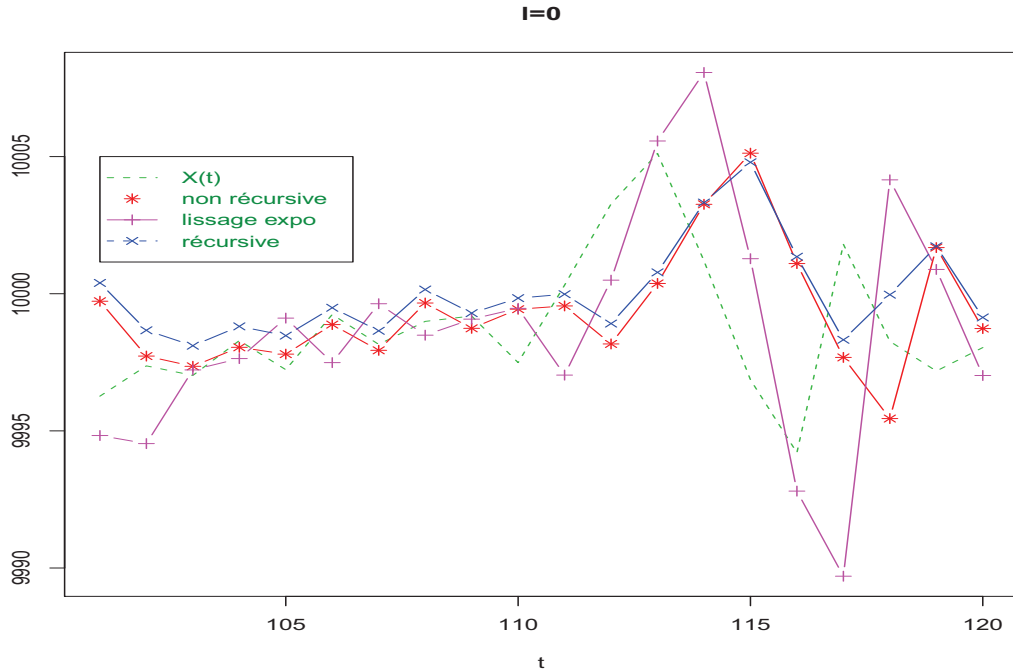
Ensuite le TABLEAU 4.13 donne les prévisions $\hat{\xi}_{100+p}$ basées sur les observations $\xi_1, \dots, \xi_{99+p}$. Les résultats du TABLEAU 4.13 sont représentés graphiquement aux FIGURES 4.18, 4.19 et 4.20.

Enfin le TABLEAU 4.14 fournit les temps de calcul des prévisions successives $\hat{\xi}_{101}, \dots, \hat{\xi}_{100+p}$ en fonction de p .

Notons que pour le lissage exponentiel, ces résultats sont obtenus avec un choix de la constante de lissage $a = 0.6965539$ (c.f. relation (1.1.1) page 2 pour la définition de a) fait automatiquement sous **R**. Comme nous l'avons évoqué en introduction, cet exemple

p	1	5	10	15	20	EQME	EARM
ξ_{100+p}	9996.266	9997.230	9997.488	9996.865	9998.037	0	0
$\hat{\xi}_{100+p}^{n-w}$	9999.726	9997.798	9999.436	10005.124	9998.730	11.6	0.2429‰
$\hat{\xi}_{100+p}^0$	10000.394	9998.472	9999.839	10004.791	9999.135	11.106	0.2484‰
$\hat{\xi}_{100+p}^{0.25}$	10000.374	9998.466	9999.829	10004.801	9999.127	11.116	0.2481‰
$\hat{\xi}_{100+p}^{0.5}$	10000.35	9998.46	9999.82	10004.81	9999.12	11.127	0.2479‰
$\hat{\xi}_{100+p}^{0.75}$	10000.333	9998.454	9999.809	10004.822	9999.113	11.139	0.2477‰
$\hat{\xi}_{100+p}^1$	10000.312	9998.449	9999.800	10004.832	9999.106	11.152	0.2474‰
$\hat{\xi}_{100+p}^{1.expo}$	9994.830	9999.111	9999.467	10001.277	9997.019	15.335	0.2738‰

TAB. 4.13 – Valeurs de la série et de ses prévisions.


 FIG. 4.18 – Représentation graphique des prévisions ($N = 120$) $\ell = 0$.

montre que le lissage exponentiel présente un grand avantage de rapidité de calcul par rapport aux méthodes non paramétriques, mais s'avère moins compétitif dès que p devient grand. Inversement, la méthode non paramétrique, non récursive commet des erreurs de prévision très petites, mais le calcul des prévisions demande beaucoup de temps. Enfin, on constate que les noyaux récurrents commettent des erreurs de prévision légèrement supérieures à celles du noyau non récursif mais ils réduisent considérablement l'écart de

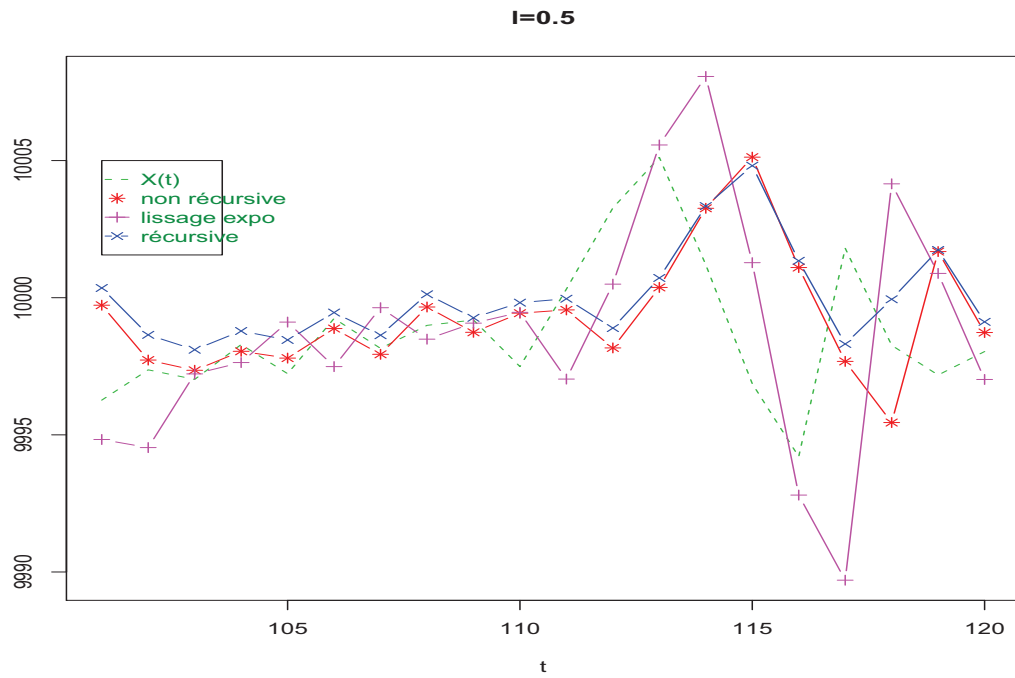


FIG. 4.19 – Représentation graphique des prévisions ($N = 120$) $\ell = 1/2$.

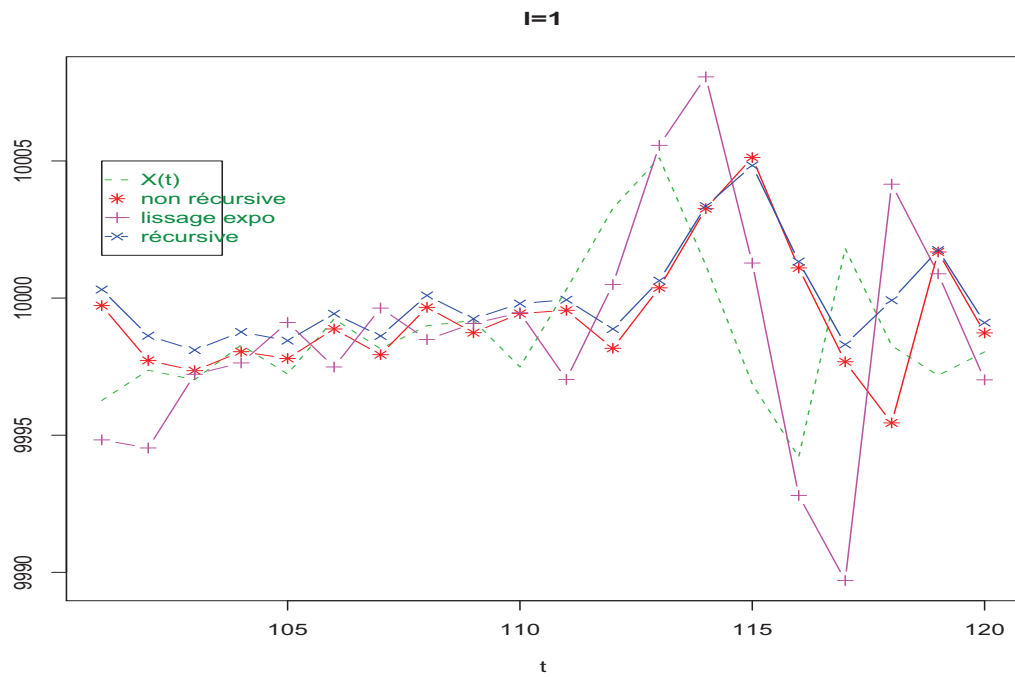


FIG. 4.20 – Représentation graphique des prévisions ($N = 120$) $\ell = 1$.

p	1	5	10	15	20
$\ell = 0$	0.19	0.64	1.42	2.13	2.98
$\ell = 0.25$	0.17	0.67	1.40	2.19	3.04
$\ell = 0.5$	0.19	0.64	1.33	2.09	2.94
$\ell = 0.75$	0.21	0.66	1.42	2.20	3.19
$\ell = 1$	0.17	0.60	1.36	2.03	2.87
Lis. expo	0.01	0.05	0.06	0.14	0.17
N-W	0.55	14.24	31.81	51.90	76.14

TAB. 4.14 – Temps de calcul des prédicteurs.

temps de calcul enregistré entre la méthode du lissage exponentiel et la méthode non paramétrique usuelle.

En conclusion, les résultats obtenus ici sont encourageants, nous comptons à terme poursuivre et approfondir ces simulations dans plusieurs voies : en variant les lois considérées (lois multi-modales, mélanges) et en regardant les problèmes de bords. Nous projetons également de chercher des méthodes plus sophistiquées pour le choix de la constante de la fenêtre (en essayant en particulier d'adapter les méthodes performantes existantes afin de les rendre récursives).

Conclusion et perspectives

Nous avons introduit et étudié dans cette thèse de nouvelles familles d'estimateurs récursifs de la densité et de la régression incluant les estimateurs récursifs usuels ainsi qu'une nouvelle famille de prédicteurs non paramétriques récursifs qui permettent de réduire considérablement le temps de calcul.

Notre étude a permis de constater que pour l'estimation de la densité et de la régression, le cas $\ell = 1$ est meilleur dans nos familles d'estimateurs. En effet, ce choix de paramètre minimise à la fois les erreurs d'estimation et le temps de calcul par rapport aux autres choix étudiés, bien que l'on peut considérer, au vu de ces résultats, que le choix de ℓ n'a pas d'influence majeure sur le temps de calcul. Ce constat est très important puisque les travaux réalisés sur le sujet mettent en avant l'efficacité du cas $\ell = 0$ à cause de la minimalité de sa variance.

Notre travail montre également qu'il y a une très légère détérioration des erreurs d'estimation, en passant du noyau classique au noyau récursif, mais au regard du gain de temps calcul réalisé, les noyaux récursifs présentent un avantage décisif.

Face à un problème de prévision, le souhait naturel du praticien est d'avoir des prévisions précises et faciles à mettre à jour. Ce travail montre qu'avec le problème de prévision posé dans cette thèse, le lissage exponentiel est certes très rapide, mais perd de sa précision avec l'augmentation de la durée de mise à jour. La méthode non paramétrique classique, quant à elle, reste toujours très compétitive mais présente une durée de mise à jour importante. L'augmentation de cette durée retarde considérablement les calculs avec cette méthode. Les méthodes non paramétriques récursives sont sans doute les plus appropriées à notre problème du fait qu'elles proposent une sorte d'équilibre entre la précision et la rapidité de calcul.

Le Chapitre 4 de cette thèse ouvre également sur des problèmes théoriques intéressants à étudier : comportement des estimateurs avec une fenêtre aléatoire, comportement d'estimateurs tronqués....

Ces simulations sont donc convaincantes et illustrent bien les résultats théoriques obtenus précédemment. Nous envisageons une étude beaucoup plus complète en appliquant les méthodes développées dans cette thèse à des données réelles. Nous envisageons aussi de réaliser une étude comparative sur les méthodes non paramétriques récursives et d'autres méthodes de prévision.

Enfin, il serait très intéressant d'étendre les méthodes récursives aux données fonctionnelles, c'est-à-dire le cas où les n observations considérées ne sont plus des vecteurs de $\mathbb{R}^d$, mais plutôt des courbes ou plus généralement des fonctions. L'usage des méthodes récursives s'avère crucial dans la mesure où en pratique, ces fonctions sont observées en des instants de discrétisation où le nombre de points de discrétisation est souvent très important et constitue donc une charge de calcul très lourde. Ce type de données sont apparues très récemment et se rencontrent de plus en plus fréquemment avec l'automatisation et l'informatisation des procédures de mesures. La liste des références sur les données fonctionnelles devient très longue mais, nous nous référons aux livres de Ferraty et Vieu [43], Ramsay et Silverman [74] et leurs bibliographies respectives, ainsi qu'entre autres aux travaux de Cardot, Crambes, Kneip et Sarda [23], ceux de Crambes, Delsol et Laksaci [28], ceux de Crambes, Sarda et Kneip [29], ceux de Ferraty, Goiab et Vieu [42], et ceux de Kneip, Li, MacGibbon et Ramsay [55].

Annexe A

Annexes

Sommaire

A.1	Lemmes utiles	121
A.1.1	Lemme de Bochner	121
A.1.2	Lemme de Toeplitz	122
A.1.3	Lemme de Volkonskii et Rozanov	122
A.1.4	Lemme de Borel-Cantelli	122
A.1.5	Lemme de couplage de Rio	123
A.2	Théorème d'approximation de Jain et al.	123
A.3	Loi du Logarithme Itéré pour un Mouvement Brownien . . .	124
A.4	Quelques inégalités importantes	124

A.1 Lemmes utiles

A.1.1 Lemme de Bochner

Lemme A.1.1. *Soit K une fonction bornée, $K \in L^1$ telle que :*

$$\int_{\mathbb{R}^d} K(x) dx = 1 \text{ et } \|xK(x)\| \longrightarrow 0,$$

lorsque $\|x\| \rightarrow \infty$. On pose :

$$K_\varepsilon := \frac{1}{\varepsilon^d} K\left(\frac{x}{\varepsilon}\right).$$

Alors, pour toute fonction $f \in L^1$, on a :

$$\forall x \in c(f), \quad f * K_\varepsilon(x) \longrightarrow f(x),$$

lorsque $\varepsilon \rightarrow 0$, où $f * K_\varepsilon$ désigne le produit de convolution de f et K_ε .

A.1.2 Lemme de Toeplitz

Le Lemme suivant dit de Toeplitz est rappelé entre autre par Masry [63].

Lemme A.1.2. *Soit $(a_{n,k})_{n \geq 1, k \geq 1}$ une suite réelle et $(w_n)_{n \geq 1}$ une suite qui converge vers w . On suppose que :*

- (i) *pour tout $k \geq 1$, $\lim_{n \rightarrow \infty} a_{n,k} = 0$;*
- (ii) *$\lim_{n \rightarrow \infty} \sum_{k=1}^{\infty} a_{n,k} = A < \infty$;*
- (iii) *Il existe une constante $C > 0$ telle que pour tout $n \geq 1$, $\sum_{k=1}^{\infty} |a_{n,k}| < C < \infty$.*

Alors nous avons :

$$\sum_{k=1}^{\infty} a_{n,k} w_k \rightarrow Aw,$$

lorsque $n \rightarrow \infty$.

A.1.3 Lemme de Volkonskii et Rozanov

Le lemme suivant est énoncé par Volkonskii et Rozanov [89].

Lemme A.1.3. *Soit $V_1, \dots, V_L$ des variables aléatoires mesurables par rapport aux tribus respectives $\mathcal{F}_{i_1}^{j_1}, \dots, \mathcal{F}_{i_L}^{j_L}$ avec $1 \leq i_1 < j_1 < \dots < j_L \leq T$, $i_{l+1} - j_l \geq \omega \geq 1$ et $|V_j| \leq 1$ pour $j = 1, \dots, L$ alors :*

$$\left| E \prod_{j=1}^{k-1} V_j - \prod_{j=1}^{k-1} EV_j \right| \leq 8(L-1)\alpha(w).$$

A.1.4 Lemme de Borel-Cantelli

Le lemme de Borel-Cantelli peut être consulté par exemple dans Neveu [67].

Lemme A.1.4. *Soit $(A_n)_{n \geq 1}$ une suite d'événements. Si $\sum_{n=1}^{\infty} P(A_n) < \infty$ (resp. $= \infty$ et si les A_n sont indépendantes), alors :*

$$P\left(\overline{\lim}_{n \rightarrow \infty} A_n\right) = 0 \text{ (resp. } = 1 \text{)}.$$

On rappelle que pour une suite d'événements A_n :

$$\overline{\lim}_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k = \lim_{n \rightarrow \infty} \downarrow \bigcup_{k \geq n} A_k,$$

ou encore :

$$\overline{\lim}_{n \rightarrow \infty} A_n = \{ \omega \in \Omega, \omega \in A_n, \text{ infiniment souvent} \} = \left\{ \omega \in \Omega, \sum_{n=1}^{\infty} \mathbf{1}_{A_n}(\omega) = \infty \right\}.$$

A.1.5 Lemme de couplage de Rio

Le lemme suivant est disponible dans le livre de Rio [76].

Lemme A.1.5. *Soit $\mathcal{A}$ une sous tribu de $(\Omega, \mathcal{F}, \mathcal{P})$ et X une variable aléatoire réelle, prenant ses valeurs dans un intervalle compact $[a, b]$. Soit U une variable Uniforme sur $[0, 1]$ indépendante de la tribu engendrée par X et $\mathcal{A}$. Alors il existe une variable aléatoire X^* , mesurable pour la tribu $\mathcal{A} \vee \sigma(X) \vee \sigma(U)$ indépendante de $\mathcal{A}$ et de même loi que X telle que*

$$E|X - X^*| \leq 2(b - a)\alpha(\mathcal{A}, \sigma(X)).$$

A.2 Théorème d'approximation de Jain et al.

Le théorème suivant est une conséquence du Théorème de Jain et al. [53] pour le cas indépendant. Il permet d'approximer une somme de variables aléatoires par un M. B. pour exploiter la Loi de du Logarithme Itéré vérifiée par le mouvement Brownien.

Théorème A.2.1. *Soit X_n une suite de variables aléatoires indépendantes définies sur un espace de probabilité $(\Omega, \mathcal{A}, P)$ telles que pour tout $n \geq 0$, EX_n^2 existe et $EX_n = 0$. Soit :*

$$S_n = \sum_{i=1}^n X_i, \quad S_0 = 0 \text{ et } V_n = \sum_{i=1}^n EX_i^2 \text{ si } n \geq 1, \quad V_0 = 0.$$

Pour tout $\alpha \geq 0$, on suppose que $V_n \rightarrow \infty$ et que :

$$\sum_{k=1}^{\infty} \frac{(\ln_2 V_k)^\alpha}{V_k} E \left(X_k^2 \mathbf{1}_{\left\{ X_k^2 > \frac{V_k}{\ln V_k (\ln_2 V_k)^{2(\alpha+1)}} \right\}} \right) < \infty.$$

Soit S une fonction aléatoire définie sur $[0, +\infty[$ telle que :

$$\text{pour } t \in [V_n, V_{n+1}[, S(t) = S_n.$$

Alors, en redéfinissant $\{S(t), t \geq 0\}$ si nécessaire sur un nouvel espace de probabilité, il existe un M. B. ξ tel que

$$|S(t) - \xi(t)| = o\left(t^{\frac{1}{2}} (\ln \ln t)^{\frac{1-\alpha}{2}}\right).$$

A.3 Loi du Logarithme Itéré pour un Mouvement Brownien

Le théorème suivant est une conséquence de la Loi du Logarithme Itéré énoncé par Levy [60].

Théorème A.3.1. *Si ξ est un Mouvement Brownien, alors on a :*

$$\overline{\lim}_{t \rightarrow \infty} \frac{\xi(t)}{\sqrt{2t \ln \ln t}} = 1 \text{ p.s.}$$

A.4 Quelques inégalités importantes

Les deux inégalités suivantes sont rappelées par Bosq et Blanke [15].

Inégalité de Billingsley

Si X et Y sont des variables aléatoires bornées, alors :

$$|\text{Cov}(X, Y)| \leq 4 \|X\|_{\infty} \|Y\|_{\infty} \alpha(\sigma(X), \sigma(Y)). \quad (\text{A.4.1})$$

Inégalité de Davydov

Si $X \in L^p$ et $Y \in L^r$ avec :

$$q > 1, r > 1 \text{ et } \frac{1}{q} + \frac{1}{r} < 1,$$

alors :

$$|\text{Cov}(X, Y)| \leq 2p [2\alpha(\sigma(X), \sigma(Y))]^{\frac{1}{p}} \|X\|_p \|Y\|_r, \quad (\text{A.4.2})$$

où :

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1.$$

Annexe B

Table des équations, table des figures, liste des tableaux

Table des équations

Numéro	Page
1.1.1	2
1.2.1	4
1.3.1	6
1.3.2	7
1.4.1	9
2.1.1	12
2.1.2	12
2.2.1	13
2.2.2	16
2.2.3	16
2.2.4	18
2.2.5	19
2.2.6	22
2.2.7	22
2.4.1	29
2.4.2	30
2.4.3	30

2.4.4	31
2.4.5	31
2.4.6	33
2.4.7	34
2.4.8	34
2.4.9	37
2.4.10	40
2.4.11	40
2.4.12	42
2.4.13	43
2.4.14	43
2.4.15	43
2.4.16	45
2.4.17	46
3.1.1	59
3.2.1	60
3.2.2	60
3.2.3	60
3.4.1	63
3.5.1	64
3.5.2	65
3.5.3	65
3.5.4	66
3.5.5	67
3.5.6	68
3.5.7	68
3.5.8	69
3.5.9	73
3.5.10	74
3.5.11	75
3.5.12	75
3.5.13	76
3.5.14	76
3.5.15	77
3.5.16	78
3.5.17	78
3.5.18	79

3.5.19		79
3.5.20		81
3.5.21		84
3.5.22		85
4.1.1		90
4.1.2		92

Table des figures

2.1	Constantes asymptotiques pour les Biais et variance.	53
2.2	Constantes asymptotiques pour les EQM.	53
2.3	Représentation graphique des β_r	55
2.4	Constantes asymptotiques pour les Biais et variance.	55
4.1	Estimateurs de la densité d'une loi normale sur $[-10, 10]$	96
4.2	Estimateurs de la densité d'une loi normale sur $[-2, 2]$	97
4.3	EQME pour les estimateurs d'une densité gaussienne.	98
4.4	EAME pour les estimateurs d'une densité gaussienne.	99
4.5	Estimateurs de la densité d'une loi exponentielle sur $[0, 10]$	100
4.6	Estimateurs de la densité d'une loi exponentielle sur $[4, 7]$	101
4.7	EQME pour les estimateur d'une densité exponentielle.	102
4.8	EAME pour les estimateur d'une densité exponentielle.	103
4.9	Estimateurs d'une régression parabolique sur $[-4, 4]$	104
4.10	Estimateurs d'une régression parabolique sur $[-2, 1]$	105
4.11	EQME pour les estimateurs d'une régression parabolique.	108
4.12	EAME pour les estimateurs d'une régression parabolique.	108
4.13	Estimateurs d'une régression absolue sur $[-4, 4]$	109
4.14	Estimateurs d'une régression absolue sur $[-1, 1]$	110
4.15	EQME pour les estimateurs d'une régression absolue sur $[-1, 1]$	113
4.16	EQME pour les estimateurs d'une régression absolue sur $[-1, 1]$	113
4.17	Représentation graphique des premières valeurs de la série.	115
4.18	Représentation graphique des prévisions ($N = 120$) $\ell = 0$	116
4.19	Représentation graphique des prévisions ($N = 120$) $\ell = 1/2$	117
4.20	Représentation graphique des prévisions ($N = 120$) $\ell = 1$	117

Liste des tableaux

4.1	EQME $\times 10^3$ pour les estimateurs d'une densité gaussienne.	97
4.2	EAME pour les estimateurs d'une densité gaussienne.	98
4.3	Temps CPU pour les estimateurs d'une densité gaussienne.	99
4.4	EQME $\times 10^2$ pour les estimateurs d'une densité exponentielle.	101
4.5	EAME $\times 10^2$ pour les estimateurs d'une densité exponentielle.	102
4.6	Temps CPU pour les estimateurs d'une densité exponentielle.	102
4.7	EQME pour les estimateurs d'une régression parabolique.	106
4.8	EAME pour les estimateurs d'une régression parabolique.	107
4.9	Temps CPU pour les estimateurs d'une régression parabolique.	109
4.10	EQME pour les estimateurs d'une régression absolue.	111
4.11	EAME pour les estimateurs d'une régression absolue.	112
4.12	Temps CPU pour les estimateurs d'une régression absolue.	114
4.13	Valeurs de la série et de ses prévisions.	116
4.14	Temps de calcul des prédicteurs.	118

Bibliographie

- [1] **Ahmad**, I. and **Lin**, P.E. (1976). Nonparametric sequential estimation of a multiple regression function, *Bull. Math. Statist.* 17, 63-75.
- [2] **Amiri**, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortement mélangeant, *C. R. Acad. Sci. Paris, Ser. I*, 347, 309-314.
- [3] **Amiri**, A. Rates of almost sure convergence of a family of recursive kernel estimators. *soumis*.
- [4] **Banon**, G. (1976). Sur un estimateur non paramétrique de la densité de probabilité, *Revue de statistique appliquée*. 24 (4), 61- 73.
- [5] **Berlinet**, A. (1999). How to get central limit theorems for global errors of estimates, *Appl. Math.* 44 (2), 81-96.
- [6] **Berlinet**, A., **Biau**, G. and **Rouvière**, L. (2005). Optimal L^1 bandwidth selection for variable kernel density estimates, *Statist. Probab. Lett.* 74 (2), 116-128.
- [7] **Berlinet**, A., **Gannoun**, A. et **Matzner-Løber**, E. (1998). Normalité asymptotique d'estimateurs convergents du mode conditionnel, *Canad. J. Statist.* 26 , 365-380.
- [8] **Berlinet**, A., **Gannoun**, A. and **Matzner-Løber**, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics* 35 (2), 139-169.
- [9] **Billingsley**, P. (1986). Probability and Measure, 2nd ed., New York : Wiley.
- [10] **Blanke**, D. (1997). Estimation non paramétrique pour des processus à temps continu bruités ou partiellement observés, *Thèse de doctorat université Paris 6*.
- [11] **Blanke**, D. (2004). Sample paths adaptative density estimation, *Math. Methods Statist.* 13 (2), 123-152.
- [12] **Bochner**, S. (1955). Harmonic analysis and the theory of probability, *University of California Press*.
- [13] **Bosq**, D. (1983). Sur la prédiction non paramétrique de variables aléatoires et mesures aléatoires, *Z. Wahrschein. Verw. Get.* 64, 541-553.

- [14] **Bosq**, D. (1998). Nonparametric statistics for stochastic processes lecture. Estimation and prediction, *Lecture notes in statistics, 2nd ed. Springer-Verlag, New York*.
- [15] **Bosq**, D. and **Blanke**, D. (2007). Inference and prediction in large dimensions, *Wiley series in probability and statistics*.
- [16] **Bosq**, D. and **Cheze-Payaud**, N. (1999). Optimal asymptotic quadratic error of non-parametric regression function estimates for a continuous-time process from sampled-data, *Statistics 32 (3)*, 229-247.
- [17] **Bosq**, D. et **Lecoutre**, J.P. (1987). Théorie de l'estimation fonctionnelle, *ed. Economica*.
- [18] **Bosq**, D. et **Lecoutre**, J.P. (1992). Analyse et prévision des séries chronologiques, *Masson*.
- [19] **Box**, G.E.P, **Jenkins**, G.M. and **Reinsel**, G.C. (1994). Time series analysis : forecasting and control, *3rd ed. Prentice-Hall, New Jersey*.
- [20] **Carbon**, M. and **Delecroix**, M. (1993). Nonparametric vs parametric forecasting in time series : a computational point view, *Applied Stochastic Models and Data Analysis 9*, 215-229.
- [21] **Carbon**, M. et **Francq**, C. (1995). Estimation non paramétrique de la densité et de la régression - Prévision non paramétrique, *La Revue de Modulad, ISSN 1145-895X*, 15, 1-25.
- [22] **Carbon**, M. and **Tran**, L. T. (1996). On histograms for linear processes, *J. Statist. Plann. Inference* , 53 (3), 403-419.
- [23] **Cardot**, H., **Crambes**, C., **Kneip**, A. and **Sarda**, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables, *Comput. Statist. Data Anal. 51* 4832-4848.
- [24] **Cheze-Payaud**, N. (1994). Régression, prédiction et discrétisation des processus en temps continu, *Thèse de doctorat université Paris 6*.
- [25] **Collomb**, G. (1981). Estimation non-paramétrique de la régression : Révue bibliographique. *Int. Statist. Rev. 49 (1)*, 75-93.
- [26] **Collomb**, G. (1981). From nonparametric regression to nonparametric prediction : Survey of the mean square error and original results on the predictogram, *Lecture notes in statistics 16*, 182-204.
- [27] **Collomb**, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau, *Z. Wahrschein. Verw. Get. 66*, 441-460.

-
- [28] **Crambes**, C., **Delsol**, L. and **Laksaci**, A. (2008). Robust nonparametric estimation for functional data, *J. Nonparametr. Stat.* 20 (7), 573-598.
- [29] **Crambes**, C., **Kneip**, A. and **Sarda**, P. (2009). Smoothing splines estimators for functional linear regression. *Ann. Statist.* 37 (1), 35-72.
- [30] **Davies**, H. I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist.* 15, 49-54.
- [31] **Davydov**, Y.A. (1970) The invariance principle for stationary processes, *Theory Probab. Appl.* 14, 487-498.
- [32] **Deheuvels**, P. (1973). Sur l'Estimation séquentielle de la densité, *C. R. Acad. Sci., Paris, Ser. A*, 276, 1119-1121.
- [33] **Deheuvels**, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C. R. Acad. Sci., Paris*, 278, 1217-1220.
- [34] **Deheuvels**, P. (1974). Estimation séquentielle de la densité, *Thèse de doctorat, université Paris 6*.
- [35] **Deheuvels**, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés, *Revue de Statistique Appliquée*, XXV, 5-42.
- [36] **Devroye**, L. (1979). On the pointwise and the integral convergence of recursive kernel estimates densities, *Util. Math.* 15, 113-128.
- [37] **Devroye**, L. (1981). On the almost everywhere convergence of nonparametric regression function estimates, *Ann. Statist.* 9, 1310-1319.
- [38] **Devroye**, L. and **Wagner**, T. J. (1980). On the L^1 convergence of kernel estimators of regression functions with application in discrimination, *Z. Wahrschein. Verw. Geb.* 51, 15-25.
- [39] **Devroye**, L. and **Wagner**, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation, *Ann. Statist.* 8, 231-239.
- [40] **Doob**, J. (1953). Stochastic process, *Wiley New York*.
- [41] **Feller**, W. (1946). The law of iterated logarithm for identically distributed random variables, *Ann. of Math.* 47 (4), 631-638.
- [42] **Ferraty**, F., **Goiab**, A. et **Vieu**, P. (2002). Régression non-paramétrique pour des variables aléatoires fonctionnelles mélangeantes, *C. R. Acad. Sci., Paris, Ser. I*, 334, 217-220.

- [43] **Ferraty**, F. and **Vieu**, P. (2006). Nonparametric functional data analysis : Methods, theory, applications and implementations. *Springer, London*.
- [44] **Gannoun**, A. (1990). Prédiction non paramétrique de la médiane conditionnelle : médianogramme et méthode du noyau. Application à la prévision des processus. *Ann. I.S.U.P. XXXV, 1, 11-22*.
- [45] **Giné**, E. and **Guillou**, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators, *Ann. Henri Poincaré 36 (6) 907-921*.
- [46] **Gorodetskii**, V.V. (1977). On the strong mixing property for linear processes, *Theory Probab. Appl. 22, 411-413*
- [47] **Gouriéroux**, C. et **Monfort**, A. (1990). Séries temporelles et modèles dynamiques, *Economica*.
- [48] **Greblecki**, W. and **Pawlak**, M. (1987). Necessary and sufficient consistency conditions for a recursive kernel regression estimate, *J. Multivariate Anal. 23, 67-76*.
- [49] **Györfi**, L., **Kholer**, M., **Krzyżak**, A. and **Walk**, H. (2002). A distribution-free theory of nonparametric regression, *Springer-Verlag New york*.
- [50] **Ibragimov**, I.A. et **Rozanov**, Y. (1974). Independent stationary sequences of random variables, *Wolters-Noordhof*.
- [51] **Isogai**, E. (1984). Joint asymptotic normality of nonparametric recursive density estimators at a finite number of distinct points, *J. Japan Statist. Soc. 14 (2), 125-135*.
- [52] **Isogai**, E. (1994). A Berry-Esseen-type bound for recursive estimators of a density and its derivatives, *J. Statist. Plann. Inference 40, 1-4*.
- [53] **Jain**, C., **Jogdeo**, K. and **Stout**, W. (1975). Upper and lower functions for martin-gales and mixing processes, *Ann. Probab. 3 (1), 119-145*.
- [54] **Joag-Dev**, K. and **Prochan**, F. (1983). Negative association of random variables, with applications, *Ann. Statist. 11 (1), 286-295*.
- [55] **Kneip**, A., **Li**, X. **MacGibbon**, K. B. and **Ramsay**, J. O. (2000). Curve registration by local regression, *Canad. J. Statist. 28 (1), 19-29*.
- [56] **Krzyżak**, A. (1986). The rates of convergence of kernel regression estimates and classification rules, *IEEE Trans. Inform. Theory 32, 668-679*.
- [57] **Krzyżak**, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation, *IEEE Trans. Inform. Theory 38, 1323-1338*.

-
- [58] **Krzyżak**, A. and **Pawlak**, M. (1983). Universal consistency results for Wolverton-Wagner regression function estimates with application in discrimination, *Probl. Contr. Inform. Theory* 12, 33-42.
 - [59] **Krzyżak**, A. and **Pawlak**, M. (1984). Almost everywhere convergence of a recursive regression function estimate and classification, *IEEE Trans. Inform. Theory* 30, 91-93.
 - [60] **Lévy**, P. (1965). Processus stochastiques et mouvement brownien, *Gauthier-Villars* (2nd éd).
 - [61] **Liang**, H.Y. and **Baek**, J. (2004). Asymptotic normality of recursive density estimates under some dependence assumptions, *Metrika* 60, 155-166.
 - [62] **Loève**, M. (1963). Probability theory, *Princeton, New Jersey Van Nostrand*.
 - [63] **Masry**, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, *IEEE Trans. Inform. Theory* 32, no 2, 254-267.
 - [64] **Masry**, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes, *Statist. Probab. Lett.* 5, 249-254.
 - [65] **Masry**, E. and **Györfi**, L. (1987). Strong consistency and rates for recursive probability density estimators of stationary processes, *J. Multivariate Anal.* 22, 79- 93.
 - [66] **Nadaraya**, E. (1964). On estimating regression, *Theory Probab. Appl.* 9, 141-142.
 - [67] **Neveu**, J. (1970). Bases mathématiques du calcul des probabilités, 2nd. ed. *Masson et Cie, Editeurs Paris*.
 - [68] **Pankratz**, A. (1983). Forecasting with univariate Box-Jenkins models : concept and cases, *Wiley*.
 - [69] **Parzen**, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* 33, 1065-1076.
 - [70] **Pham**, T.D. and **Tran**, L.T. (1985). Some strong mixing properties of time series models, *Stochastic Process. Appl.* 23, 291-300.
 - [71] **Poggi**, J.M. (1994). Prévision non paramétrique de la consommation électrique, *Revue de Statistique Appliquée XLII (4)* 83-98.
 - [72] **Prakasa Rao**, B.L.S. (1983), Nonparametric functional estimation, *New-York : Academic Press*.
 - [73] **Prohorov**, Y. (1949). Sur la loi forte des grands nombres, *Dokl. Akad. Nauk URSS* 69.

- [74] **Ramsay**, J. O. and **Silverman**, B. W. (2005). Applied functional data analysis, *2nd ed. Springer, New York*.
- [75] **Rao**, C.R. (1965). Linear statistical inference and its applications, *2nd ed, Wiley New York*.
- [76] **Rio**, E. (2000). Théorie asymptotique des processus aléatoires faiblement dépendants, *vol 31 Mathematics and applications Springer-Verlag, Berlin*.
- [77] **Robinson**, P.M. (1983). Nonparametric estimators for time series, *J. Time Ser. Anal.* 4, 185-207.
- [78] **Roussas**, G.G. (1992). Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation, *J. Nonparametr. Stat.* 1, 171-195.
- [79] **Roussas**, G.G. and **Tran**, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions, *Annals of Statist.* 20 (1), 98-120.
- [80] **Rosenblatt**, F. (1956). A central limit theorem and a strong mixing condition, *Proc. Natl. Acad. Sci. USA* 42, 43-47.
- [81] **Rosenblatt**, F. (1956). Remarks on some nonparametric estimates of a density function, *Ann. Math. Statist.* 38, 482-493.
- [82] **Samanta**, M. and **Mugisha**, R.X. (1981). On the class of estimates of the probability function and mode based on random number of observations, *Calcutta Statist. Assoc. Bull.* 117, 23-40.
- [83] **Sarda**, P. and **Vieu**, P.(1991). Smoothing parameter selection in hazard estimation, *Statist. Probab. Lett.* 11, 429-434.
- [84] **Schwartz**, S.C. (1967). Estimation of a probability density by an orthogonal series, *Ann. Math. Statist.* 38, 1261-1265.
- [85] **Singh**, R.S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 39, 357-363.
- [86] **Takahata**, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, *Bull. Tokyo Gakugei Univ. Nat. Sci. Ser. IV*, 11-32.
- [87] **Tran**, L.T (1989). Recursive density estimation under dependence. *IEEE Trans. Inform. Theory* 35 (5), 1103-1008.
- [88] **Tran**, L. T. (1994). Density estimation for time series by histograms, *J. Statist. Plann. Inference* , 40 (1), 61-79.

- [89] **Volkonskii**, V.A. and **Rozanov**, Yu.A. (1959). Some limit theorems for random functions, *Theory Probab. Appl.* 4, 178-197.
- [90] **Walk**, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Ann. Inst. Statist. Math.* 53 (4), 691-707.
- [91] **Wang**, L. and **Liang**, H.Y. (2004). Strong uniform convergence of the recursive regression estimators under ϕ -mixing conditions, *Metrika* 59, 245- 261.
- [92] **Watson**, G.S. (1964). Smooth regression analysis. *Shakhya Ser. A* 26, 359-372.
- [93] **Wegman**, E.J. (1972). Nonparametric probability density estimation, I. A summary of available methods. *Technometrics* 41, 533-546.
- [94] **Wegman**, E.J. (1972). Nonparametric probability density estimation, II. A comparison of density estimation methods. *J. Stat. Comput. Simul.* 1, 225-245.
- [95] **Wegman**, E.J. and **Davies**, H. I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Statist.* 7 (2), 316-327.
- [96] **Wertz**, W. (1985). Sequential and recursive estimators of the probability density, *Statistics* 16, 277-295.
- [97] **Wolverton**, C. and **Wagner**, T.J. (1969). Recursive estimates of probability densities, *IEEE Trans. Syst. Cybern.* 5, 307-308.
- [98] **Yamato**, H. (1972). Sequential estimation of a continuous propability density function and mode, *Bull. Math. Statist.* 14, 1-12.