

République Algérienne Démocratique Et Populaire
Ministère De L'Enseignement Supérieur Et De La Recherche Scientifique
Département De Mathématiques
Faculté Des Sciences Exactes
Université Constantine 1

N° d'ordre
N° de série

THÈSE DE DOCTORAT PRÉSENTÉE POUR L'OBTENTION DU
GRADE DE DOCTEUR EN SCIENCES
OPTION : MATHÉMATIQUES APPLIQUÉES

Thème

**INFÉRENCE STATISTIQUE DANS LES MODÈLES DE
RÉGRESSION NON PARAMÉTRIQUES**

Présentée par Mme. Kenza Assia Mezhoud

Soutenue devant le jury

Président : Pr. MESSACI Fatiha	U.Constantine 1
Directeur : Pr. MOHDEB Zaher	U.Constantine 1
Examineur : Pr. ADJABI Smail	U. Béjaia
Examineur : Pr. AISSANI Djamil	U. Béjaia
Examineur : Pr. LOUHICHI Sana	U. Grenoble
Examineur : Pr. RAHMANI Fouad Lazhar	U. Constantine 1

Date de la soutenance : 28.09.2014

Je dédie ce modeste travail à la mémoire de ma mère,
à mon père qui m'a inspiré la volonté et le courage.....,
à Chahinez, à mon mari,
à ma famille, à mes amis.

Remerciement

Merci infiniment au Pr. Mohdeb Zaher qui m'a fait l'honneur d'encadrer ce travail, pour ses encouragements, ses motivations, je tiens à lui témoigner ma gratitude.

Je tiens à exprimer ma grande gratitude aussi au Pr. Sana Louhichi pour son accueil chaleureux à Paris et à Grenoble, pour son aide, sa disponibilité et sa grande patience, les mots ne sont guère suffisants pour la remercier. J'adresse mes sincères remerciements au Pr. Messaci Fatiha d'avoir accepté de présider le jury, de me faire découvrir un jour la passion des mathématiques à travers le module de mesure.

Un grand merci aux examinateurs, Pr. Adjabi Smail, Pr. Aissani Djamil de l'université de Béjaia, et au Pr. Rahmani Fouad Lazhar de l'université de Constantine, de me faire l'honneur de faire partie de mon jury, et d'avoir eu la gentillesse et la patience de lire en détails cette thèse.

0.1 Abréviations

- *p.co* : presque complètement.
- *a.s* : almost surely.
- i.i.d. indépendantes et identiquement distribuées.
- r.v.s. random variable sequence.
- *CV* : cross validation.
- *UCV* : unbiased cross validation.
- *BCV* : biased cross validation.
- *MSE* : mean squared error.
- *MISE* : mean integrated squared error.
- *T.C.L* : théorème central limite.

0.2 Notations

- $[f] = \int f(x)dx$.
- $\text{Lip}(g) = \sup_{x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|_1}$.
- $\|x\|_1 = \sum_{i=1}^u |x_i|$.
- $\|x\|_\infty$: norme supérieure.
- \mathbb{I}_A : indicatrice de l'ensemble A .
- (\mathbb{R}_m, B_m) : l'espace euclidien de dimension m muni de sa tribu borélienne B_m .
- W_k : ensemble des fonctions de densité $f(x)$ ayant des dérivées d'ordre $k \geq 2$ telles que $f^{(k)}(x)$ soient bornées, continues et de carré intégrable.
- $a \wedge b$: minimum des deux réels a, b .
- $a \vee b$: maximum des deux réels a, b .
- $a_n = O(b_n)$: $a_n \leq Cb_n$, avec $C > 0$.
- $a_n = o(b_n)$: $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 0$.

Table des matières

0.1	Abréviations	4
0.2	Notations	5
Introduction		
I	Cas des variables indépendantes	15
1	Estimation de la densité par la méthode du noyau	17
1.1	Introduction	17
1.2	Propriétés asymptotiques	19
1.2.1	Convergence presque complète	19
1.2.2	Normalité asymptotique	23
1.2.3	Optimalité de l'estimateur de la densité	24
1.3	Quelques méthodes de sélection de la fenêtre de lissage	26
1.3.1	Méthodes classiques	26
1.3.2	Approche Bayésienne	30
1.4	Simulation	32
2	Estimation de la régression par la méthode du noyau	35
2.1	Introduction	35
2.2	Propriétés asymptotiques	36
2.2.1	Convergence presque complète	36
2.2.2	Normalité asymptotique	39
2.2.3	Optimalité de l'estimateur de la régression	40
2.2.4	Validation croisée de la régression	41
2.3	Cas multivarié	42
2.3.1	Fléau de dimension et modèles additifs	43
2.4	Simulation	45

3	Lois uniformes des dérivées de la densité et de la régression	47
3.1	Introduction	47
3.2	Lois uniformes des dérivées de la densité et de la régression . .	49
3.3	Application : Modèle de régression gaussien	51
3.3.1	Bandes de confiance de la densité et de la régression ($k = 0$)	53
3.4	Simulation	56
II	Cas des variables dépendantes	61
4	Kernel estimation under weak dependence	63
4.1	Introduction	63
4.2	Definition and examples	64
4.2.1	Bernoulli shift with independent inputs	65
4.2.2	Bernoulli shift with dependent inputs	66
4.2.3	The bilinear process	67
4.2.4	Chaotic Volterra models	67
4.3	Properties of kernel density and regression estimators under η -dependence	67
4.3.1	Convergence in mean square error	68
4.3.2	Asymptotic normality	73
4.3.3	Bernstein's type inequality	79
4.3.4	Almost complete convergence	82
5	Recursive kernel estimation of the density under weak dependence	83
5.1	Introduction	83
5.2	Assumptions	85
5.3	Results	86
5.3.1	Convergence in mean squared error	86
5.3.2	Asymptotic normality	88
5.4	Application	89
5.4.1	Bernoulli shift	89
5.4.2	Bilinear process	90
5.5	Simulation	91
5.6	Proofs	97
	Perspectives de recherche	
	Bibliographie	

Introduction

L'objet de cette thèse est l'étude des propriétés asymptotiques des estimateurs à noyau dans le cadre des modèles de régression non paramétriques sous différentes hypothèses portant sur la nature des variables sous étude. Notre cadre de travail est donc multiple regroupant des variables indépendantes, mais aussi des variables faiblement dépendantes observées en temps discret.

Contrairement à l'estimation paramétrique, l'estimation non-paramétrique consiste à estimer à partir des observations, une fonction inconnue élément d'une certaine classe de fonctions qui n'est pas en bijection avec un espace fini-dimensionnel.

En fait, depuis l'avènement de la statistique moderne, cette dernière et notamment l'estimation par la méthode du noyau a été intensément traitée par plusieurs auteurs.

Rosenblatt (1956) donne l'expression de l'erreur quadratique moyenne de l'estimateur de la densité basé sur le noyau uniforme dans le cas des variables indépendantes et identiquement distribuées.

Parzen (1962) établit la normalité asymptotique de cet estimateur, le cas multivarié a été traité après entre autres par Cacoullos (1966).

Ces travaux ont été étendus à l'estimateur à noyau de la régression, on identifie deux papiers publiés la même année, Nadaraya (1964) et Watson (1964). D'autres articles ont été développés sur l'étude des propriétés de cet estimateur, la consistance faible et forte a été étudiée entre autres par Colomb (1981), Colomb et Hardle (1986), nous renvoyons aussi à Bosq et Lecoutre (1987), Gyofri et al.(1989), Roussas (1990), Prakasa (1983).

L'estimateur à noyau dépend de deux paramètres, à savoir le noyau, fonction déterminant la forme du voisinage contenant les observations en étude, et la fenêtre de lissage qui contrôle la taille de ce voisinage. La perte L^2 sera

utilisée afin de déterminer les paramètres optimaux, au vu de cela, il s'avère que le choix de cette fenêtre est crucial pour la performance de l'estimation. La fenêtre optimale relativement à cette perte dépend généralement d'une quantité inconnue. Pour contourner ce problème délicat qu'est le choix de la largeur de la fenêtre, des travaux de recherche se sont attachés à introduire des méthodes alternatives dites : méthodes de sélection.

Stone (1985) et Marron (1987) ont étudié l'optimalité de l'estimateur à noyau de la densité basé sur la fenêtre calculée par la méthode de validation croisée. Les méthodes de ré-injection ont été étudiées par Hall et al.(1991), leur principe repose sur l'estimation de la quantité inconnue dépendant de la dérivée seconde dans l'expression du paramètre de lissage optimal.

L'optimalité de l'estimateur de la régression a été étudiée par Hardle et Marron (1985), la validation croisée a été aussi proposée pour calculer la fenêtre de cet estimateur. Des méthodes de type combinatoire ont été également proposées par Zougab et Adjabi (2007). Toutes les méthodes précédentes révèlent des inconvénients théoriques ou pratiques.

Une méthode alternative récente semble résoudre ce problème de sélection de la fenêtre, il s'agit de l'approche Bayésienne proposée et étudiée par Zhang (2002). Cette méthode considère le paramètre de lissage comme étant une variable aléatoire, sa loi de probabilité dite la loi a priori nous dote de plus d'informations, permettant via la loi de Bayes de calculer la loi a posteriori, et définir ensuite la moyenne a posteriori comme étant le paramètre de lissage. Cette approche offre des résultats plus stables sans conditions difficiles à vérifier.

D'autres propriétés de l'estimateur à noyau de la densité et de la régression sont étudiées sous l'hypothèse de l'indépendance. Au début des années 80, Stute utilise la théorie des processus empiriques afin d'établir les lois uniformes du logarithme de ces estimateurs. En s'appuyant sur ces travaux, Einmahl et Mason (2000), Deheuvels et Mason (2004), et Einmahl et Mason (2005) étudient les lois uniformes de l'estimateur à noyau basée sur le paramètre de lissage local, Blondin (2004) étudie les lois uniformes des dérivées de la densité et de la régression. Dans le même contexte, Nemouchi et Mohdeb (2010) ont établi les lois uniformes de la densité et de la régression dans le cadre d'un modèle gaussien. Ces lois uniformes permettent d'obtenir des bandes de confiance des fonctions à estimer.

L'hypothèse d'indépendance est dans beaucoup de cas incapable de décrire

certains phénomènes, les observations dépendantes sont plus adéquates à refléter la réalité. Outre le fait que cela permet de traiter le problème de prévision dans les séries temporelles, ce qui débouche sur de nombreuses applications telle que la prévision des valeurs boursières, prévision d'infractus,...etc. Compte tenu de ces arguments, des notions de dépendance sont introduites dans la littérature, s'exprimant en termes de coefficients de mélange entre les tribus engendrées par le passé \mathcal{F}_1^n et le futur \mathcal{F}_{n+r}^∞ du processus. Entre autres, on définit principalement le coefficient du α -mélange

$$\alpha_r = \sup_{A \in \mathcal{F}_1^n, B \in \mathcal{F}_{n+r}^\infty} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \rightarrow 0, \text{ quand } r \rightarrow \infty.$$

Ce coefficient mesure l'écart entre les tribus engendrées par le futur et le passé des variables.

En revanche, beaucoup de processus ne vérifient pas la condition du mélange. Le processus $X_n = \sum_{k=0}^n 2^{-(k+1)} \epsilon_{n-k}$, où $(\epsilon_i)_{i \in \mathbb{Z}}$ sont des innovations i.i.d, par exemple, n'est pas α -mélangeant, car l'évènement $A = \{X_k \leq \frac{1}{2}\}$ appartient à la tribu du passé $\sigma(X_t, t \leq k)$ et celle du futur $\sigma(X_t, t \geq n)$ et le coefficient du α -mélange ne décroît pas vers 0, en effet

$$\alpha_n \geq |\mathbb{P}(A \cap A) - \mathbb{P}(A)\mathbb{P}(A)| = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

Dans cet esprit, Doukhan et Louhichi (1999) ont généralisé le mélange en introduisant une notion de dépendance faible plus simple couvrant plus de processus en s'appuyant sur le contrôle de la covariance entre le passé $(X_{s_1}, \dots, X_{s_u})$ et le futur du processus $(X_{t_1}, \dots, X_{t_v})$ via des fonctions lipschitziennes g_1, g_2 tel que

$$|\text{Cov}(g_1(X_{s_1}, \dots, X_{s_u}), g_2(X_{t_1}, \dots, X_{t_v}))| \leq \psi(g_1, g_2, u, v) \epsilon_{t_1 - s_u},$$

avec

- ψ est une fonction qui dépend des arguments (g_1, g_2, u, v) .
- $\epsilon_{t_1 - s_u}$ est une suite décroissante vers 0.

Si $\psi(g_1, g_2, u, v) = (u \|g_2\|_\infty \text{Lip}(g_1) + v \|g_1\|_\infty \text{Lip}(g_2))$, on parle de η -dépendance faible.

Contrairement aux variables indépendantes, la variance de la somme n'est pas la somme des variances individuelles des variables faiblement dépendantes, on obtient alors une majoration du terme des covariances qui apparaît dans ce cas en utilisant l'inégalité ci-dessus, ce qui nous mènera à dire que la perte

par rapport au cas indépendant est asymptotiquement non-significative.

La littérature s'est enrichie grâce à cette dépendance d'outils techniques offrant aux estimateurs non paramétriques des propriétés asymptotiques proches à celles des variables indépendantes, telles que l'inégalité de type Bernstein et l'inégalité de Rosenthal qui sert à contrôler les moments d'ordre q d'une suite de variables aléatoires.

Le théorème de la limite centrale a pris sa part de généralisation sous l'hypothèse de dépendance faible, on cite principalement la normalité asymptotique de l'estimateur de Parzen établie par la méthode des blocks de Bernstein (2001), la normalité de l'estimateur de Nadaraya par la méthode de Lindberg (2002), en montrant que

$$\Delta_n = \mathbb{E}|\varphi(S_n) - \varphi(Y_n)| \rightarrow 0, \text{ quand } n \rightarrow \infty,$$

telle que S_n est une suite de variables faiblement dépendantes, φ est une fonction trois fois dérivable, Y_n est une suite de variables indépendantes et gaussiennes.

Bardet et Doukhan (2008) généralisent le théorème de Lindberg en proposant une version qui s'adapte à la fois avec les suites indépendantes et faiblement dépendantes.

Pour une meilleure compréhension de la dépendance faible, et de tous les outils probabilistes qui accompagnent cette nouvelle notion établie par Doukhan et Louhichi, nous renvoyons à Doukhan P. et Louhichi S. (1999), et Dedecker et al. (2007).

Lors de l'étude des séries temporelles, on a souvent besoin de mettre à jour l'expression de l'estimateur de la densité ou de la régression avec l'arrivée de nouvelles observations, l'utilisation d'un estimateur récursif figure importante voire préférable dans ce cas, elle permet un gain de temps de calcul et de variance de l'estimation. Divers estimateurs récursifs ont été proposés et étudiés, on cite entre autres, l'estimateur de Wolverton et Wagner (1969)

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

Deheuvels (1973) étudie aussi la convergence en moyenne quadratique et la convergence presque sûre dans le cas indépendant de l'estimateur récursif modifié suivant

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

Une autre version récursive a été introduite par Wegman et Davies (1979), ils établissent la convergence presque sûre dans le cas indépendant de l'estimateur

$$f_n^{DW}(x) = \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \frac{1}{\sqrt{h_i}} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}.$$

Sa variance améliore la variance de l'estimateur f_n^{WW} .

Ces estimateurs récursifs ont été aussi étudiés sous des hypothèses de mélange par Masry (1986), et d'association par Lian et Baek (2004).

Amiri (2010) introduit l'estimateur à noyau récursif de la densité dépendant d'un paramètre $\ell \in [0, 1]$ suivant

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

permettant ainsi la généralisation d'autres estimateurs récursifs tels que l'estimateur de Deheuvels pour $\ell = 0$, celui de Wolverton et Wagner pour $\ell = 1$, et Wegman et Davies pour $\ell = \frac{1}{2}$. Il étudie la convergence en moyenne quadratique, et la normalité asymptotique de ces estimateurs sous l'hypothèse de l'indépendance et du α -mélange.

Dans notre travail, nous étudions la convergence en moyenne quadratique et la normalité asymptotique de l'estimateur récursif $f_n^\ell(x)$ sous l'hypothèse de la dépendance faible au sens de Doukhan et Louhichi (1999).

Nous avons décidé de structurer notre travail en deux parties, selon à ce que les variables sous étude soient indépendantes ou dépendantes.

Dans le premier et le deuxième chapitre, nous rappelons les propriétés asymptotiques des estimateurs à noyau de la densité et de la régression sous l'indépendance, dans ce rappel nous avons choisi d'exposer l'approche la plus récente, celle de Ferraty et Vieu (2002). Les méthodes de sélection de la fenêtre sont étudiées, des simulations effectuées sous le logiciel R sont présentées aussi.

Les lois uniformes des dérivées de la densité et de la régression sont étudiées dans le chapitre trois en combinant les travaux de Blondin (2004) et Deheuvels (2004). Dans le cas d'un modèle de régression Gaussien, des simulations sous Matlab montrent la validité de ces bandes de confiance de la densité et de la régression étudiées par Nemouchi et Mohdeb (2010).

La deuxième partie de cette thèse est consacrée à la dépendance faible au sens de Doukhan et Louhichi.

Dans le chapitre quatre, nous présentons des conditions pour que les théorèmes limites obtenus dans le cas des variables indépendantes s'étendent aux

variables dépendantes.

Dans le chapitre cinq, nous présentons notre contribution sur l'estimation récursive de la densité dans le cas η -dépendant, des applications et des simulations effectuées sous Matlab sur des modèles faiblement dépendants montrent l'avantage de la récursivité par rapport à l'estimation classique en terme de temps de calcul et de variance.

Première partie

Cas des variables indépendantes

Chapitre 1

Estimation de la densité par la méthode du noyau

1.1 Introduction

Etant données les observations X_1, \dots, X_n indépendantes et identiquement distribuées à valeurs dans \mathbb{R} , de densité inconnue f supposée continue. L'estimation par l'histogramme suggère de subdiviser un intervalle borné de \mathbb{R} en k sous intervalles de la forme

$$I_k =]x + kh_n, x + (k + 1)h_n[, \quad x \in \mathbb{R}, \quad h_n \in \mathbb{R}_+^*, \quad k \in \mathbb{Z},$$

et indiquer pour chaque sous intervalle la fréquence empirique correspondante.

On améliore cet estimateur en considérant un voisinage de tout point $x \in \mathbb{R}$ au lieu de la fenêtre fixe du type I_k , on définit ainsi l'estimateur de la densité de Parzen-Rosenblatt par

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right). \quad (1.1)$$

h_n étant le paramètre de lissage contrôlant la taille du voisinage de x et vérifiant

- (H.1) $\lim_{n \rightarrow \infty} h_n = 0, \lim_{n \rightarrow \infty} nh_n = \infty.$
- (H.2) $\lim_{n \rightarrow \infty} \frac{nh_n}{\ln n} = \infty.$

K étant le noyau déterminant la forme du voisinage et satisfaisant

- (K.1) K est borné, symétrique, et à support compact.
- (K.2) $\int_{\mathbb{R}} K(u) du = 1$.
- (K.3) $\int_{\mathbb{R}} u^i K(u) du = 0$, $i \leq k - 1$, $[u^k K] = \int_{\mathbb{R}} u^k K(u) du < \infty$.
- (K.4) $[K^2] = \int_{\mathbb{R}} K^2(u) du < \infty$.

Remarque 1. – Si le noyau K est positif, la fonction f_n définie dans (1.1) est une densité de probabilité.

- Les noyaux rectangulaire $K(x) = \frac{1}{2}\mathbb{I}_{\{|x| \leq 1\}}$, et d'Epanechnikov $K(x) = \frac{3}{4}(1 - x^2)\mathbb{I}_{\{|x| \leq 1\}}$ vérifient les hypothèses ci-dessus.

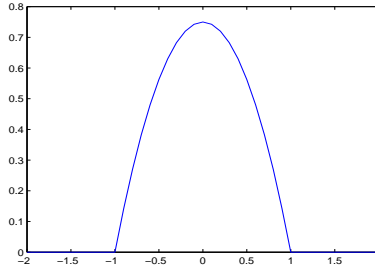


FIGURE 1.1 – Noyau d'Epanechnikov

- L'hypothèse (K.1) n'est pas indispensable, car l'utilisation d'un noyau symétrique peut nuire l'estimation, et notamment lorsqu'il s'agit d'estimer une densité à support borné, car il donne du poids à l'extérieur du support, cela mène à un estimateur biaisé au près du bord, ce qui est connu dans la littérature par l'effet du bord. Pour remédier à cela, on propose d'utiliser d'autres noyaux asymétriques. Lorsque la densité est à support $[0, 1]$ par exemple, Chen (1999) propose le noyau asymétrique beta donné par

$$K_{\frac{x}{h_n}+1, \frac{1-x}{h_n}+1}(u) = \frac{u^{\frac{x}{h_n}} (1-u)^{\frac{1-x}{h_n}}}{\beta(\frac{x}{h_n}+1, \frac{1-x}{h_n}+1)},$$

avec $K_{x, h_n}(u) = \frac{1}{h_n} K(\frac{x-u}{h_n})$, et

$$\beta(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt, \quad x > 0, \quad y > 0.$$

Ce noyau a été modifié pour obtenir un estimateur non biaisé

$$K^*(u) = \begin{cases} K_{\frac{x}{h_n}, \frac{1-x}{h_n}}(u) & \text{si } x \in [2h_n, 1 - 2h_n], \\ K_{\rho_{h_n}(x), \frac{1-x}{h_n}}(u) & \text{si } x \in [0, 2h_n], \\ K_{\frac{x}{h_n}, \rho_{h_n}(1-x)}(u) & \text{si } x \in [2h_n, 1], \end{cases} \quad (1.2)$$

tel que $\rho_{h_n}(x) = 2h_n^2 + 2.5 - \sqrt{4h_n^4 + 6h_n^2 + 2.25 - x^2 - \frac{x}{h_n}}$.

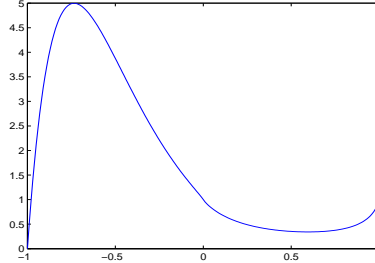


FIGURE 1.2 – Noyau beta

- L'utilisation d'un noyau d'ordre supérieur, c'est à dire vérifiant l'hypothèse (K.3) sert à réduire le biais, cependant, le choix d'un tel noyau d'ordre $k > 2$ positif entraîne une dégradation de la vitesse de convergence de l'estimateur.

Dans un premier temps, On rappelle quelques propriétés asymptotiques de l'estimateur (1.1).

1.2 Propriétés asymptotiques

1.2.1 Convergence presque complète

On présente dans ce qui suit la convergence presque complète de l'estimateur vers la densité f sous la condition de régularité suivante.

- (F.1) f est k -fois continûment dérivable autour de x .

Ce mode de convergence pas très répandu est plus fort que la convergence presque sûre et la convergence en probabilité.

Théorème 1. (Sarda et Vieu (2000))

Sous les hypothèses (H.1), (H.2), (K.1), (K.3) et (F.1), on a

$$|f_n(x) - f(x)| = O(h_n^k) + O_{p.co} \left(\sqrt{\frac{\ln n}{nh_n}} \right).$$

Si $h_n = c \left(\frac{n}{\ln n} \right)^{-\frac{1}{2k+1}}$, $c > 0$ alors

$$|f_n(x) - f(x)| = O_{p.co} \left(\left(\frac{n}{\ln n} \right)^{-\frac{k}{2k+1}} \right).$$

Preuve

Le développement de Taylor de f sous l'hypothèse (F.1) donne

$$\begin{aligned} |\mathbb{E}f_n(x) - f(x)| &= \left| \frac{(-1)^k h_n^k}{(k-1)!} \int \int_0^1 s^k K(s) t^{k-1} f^{(k)}(tsh_n + x - sh_n) ds dt \right| \\ &\sim \frac{f^{(k)}(x) h_n^k}{k!} \int s^k K(s) ds, n \rightarrow \infty. \end{aligned}$$

L'hypothèse (K.3) entraîne alors

$$|\mathbb{E}f_n(x) - f(x)| = O(h_n^k). \quad (2.3)$$

Pour le terme stochastique $|f_n(x) - \mathbb{E}(f_n(x))|$, le résultat découle immédiatement du lemme de Bernstein suivant.

Lemme 1. *Soient Z_1, \dots, Z_n des variables aléatoires réelles centrées indépendantes et identiquement distribuées telles qu'il existe deux réels positifs d et δ^2 vérifiant*

$$|Z_1| \leq d, \quad \mathbb{E}Z_1^2 \leq \delta^2,$$

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

$$\mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n Z_i \right| > \epsilon\right) \leq 2 \exp\left(-\frac{n\epsilon^2}{4\delta^2}\right).$$

On pose $S_n = \sum_{t=1}^n X_{n,t}$, avec

$$X_{n,t} = \frac{1}{h_n} \left(K\left(\frac{x - X_t}{h_n}\right) - \mathbb{E}\left(K\left(\frac{x - X_t}{h_n}\right)\right) \right).$$

Les conditions du lemme 1 sont vérifiées, en effet

$$|X_{n,1}| \leq \frac{C}{h_n}, \quad \mathbb{E}X_{n,1}^2 \leq \frac{C}{h_n}, \quad C > 0.$$

On obtient alors

$$\mathbb{P}(|f_n(x) - \mathbb{E}f_n(x)| \geq \epsilon) \leq 2 \exp\left(-\frac{nh_n\epsilon^2}{4C}\right).$$

Si $\epsilon = \epsilon_0 \sqrt{\frac{\ln n}{nh_n}}$, on peut choisir ϵ_0 tel qu'on ait la convergence suivante

$$\sum_{n=1}^{\infty} \mathbb{P}\left(|f_n(x) - \mathbb{E}f_n(x)| \geq \epsilon_0 \sqrt{\frac{\ln n}{nh_n}}\right) \leq \sum_{n=1}^{\infty} \frac{1}{n \frac{\epsilon_0^2}{4C}} < \infty,$$

c'est à dire

$$|f_n(x) - \mathbb{E}f_n(x)| = O_{p.co}\left(\sqrt{\frac{\ln n}{nh_n}}\right). \quad (2.4)$$

De (2.3) et (2.4), si $h_n = c\left(\frac{n}{\ln n}\right)^{\frac{-1}{2k+1}}$, $c > 0$, on aboutit à

$$|f_n(x) - f(x)| = O_{p.co}\left(\left(\frac{n}{\ln n}\right)^{\frac{-k}{2k+1}}\right).$$

□

Convergence presque complète uniforme

On peut montrer la convergence presque complète uniforme en supposant que le noyau K satisfait l'hypothèse de Lipschitz sur un compact I suivante.

$$- (K.5) \exists \beta > 0, \exists c < \infty, \forall x \in I, \forall y \in I : |K(x) - K(y)| \leq c|x - y|^\beta,$$

avec l'hypothèse de régularité de la densité

$$- (F.2) f \text{ est } k\text{-fois continûment dérivable autour du compact } I.$$

Théorème 2. (*Sarda et Vieu (2000)*)

Sous les hypothèses (H.1), (H.2), (K.1), (K.4), (K.5) et (F.2), il s'ensuit

$$\sup_{x \in I} |f_n(x) - f(x)| = O(h_n^k) + O_{a.co}\left(\left(\frac{\ln n}{nh_n}\right)^{\frac{1}{2}}\right).$$

Si $h_n = c\left(\frac{n}{\ln n}\right)^{\frac{-1}{2k+1}}$, $c > 0$ alors

$$\sup_{x \in I} |f_n(x) - f(x)| = O_{p.co}\left(\left(\frac{n}{\ln n}\right)^{\frac{-k}{2k+1}}\right).$$

Preuve

L'idée de la preuve repose sur le recouvrement du compact I par un nombre fini d'intervalles. En effet, on a

$$I \subset \cup_{i=1}^{\lambda_n}]t_i - \gamma_n, t_i + \gamma_n[,$$

où $\gamma_n = c\lambda_n^{-1}$, $c > 0$.

On pose $t_x = \arg \min_{\{t_1, \dots, t_{\lambda_n}\}} |x - t|$, on alors

$$\begin{aligned} & \mathbb{P}(\sup_{x \in I} |f_n(x) - \mathbb{E}(f_n(x))| \geq \epsilon) \\ & \leq \mathbb{P}(\sup_{x \in I} |f_n(t_x) - \mathbb{E}(f_n(t_x))| \geq \frac{\epsilon}{3}) + \mathbb{P}(\sup_{x \in I} |f_n(x) - f_n(t_x)| \geq \frac{\epsilon}{3}) \\ & \quad + \mathbb{P}(\sup_{x \in I} |\mathbb{E}f_n(t_x) - \mathbb{E}f_n(x)| \geq \frac{\epsilon}{3}). \end{aligned}$$

Pour le premier terme

$$\begin{aligned}
& \mathbb{P}(\sup_{x \in I} |f_n(t_x) - \mathbb{E}f_n(t_x)| \geq \frac{\epsilon}{3}) \\
&= \mathbb{P}(\max_{i=1:\lambda_n} |f_n(t_i) - \mathbb{E}(f_n(t_i))| \geq \frac{\epsilon}{3}) \\
&\leq \sum_{i=1}^{\lambda_n} P(|f_n(t_i) - \mathbb{E}f_n(t_i)| \geq \frac{\epsilon}{3}) \\
&\leq \sum_{i=1}^{\lambda_n} 2 \exp\left(\frac{-nh_n\epsilon^2}{12C}\right) \leq 2\gamma_n^{-1} \exp\left(\frac{-nh_n\epsilon^2}{12C}\right), \quad C > 0. \tag{2.5}
\end{aligned}$$

Pour le deuxième terme, puisque K est Lipchitzienne, on a

$$|f_n(x) - f_n(t_x)| \leq \frac{|\gamma_n|^\beta}{h_n^{1+\beta}},$$

on obtient alors sous l'hypothèse (H.1) pour $\gamma_n = n^{-p}$, $p > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|f_n(x) - f_n(t_x)| > \epsilon) = 0. \tag{2.6}$$

Idem pour le troisième terme

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\mathbb{E}f_n(x) - \mathbb{E}f_n(t_x)| > \epsilon) = 0. \tag{2.7}$$

En combinant (2.5), (2.6), (2.7), avec $\epsilon = \epsilon_0 \sqrt{\frac{\ln n}{nh_n}}$, on obtient

$$\mathbb{P}\left(\sup_{x \in I} |f_n(x) - \mathbb{E}(f_n(x))| \geq \epsilon_0 \sqrt{\frac{\ln n}{nh_n}}\right) \leq \frac{1}{n^{\frac{\epsilon_0^2}{12C}}}.$$

Un choix convenable de ϵ_0 permet de conclure ce qui suit

$$\sup_{x \in I} |f_n(x) - \mathbb{E}(f_n(x))| = O_{a.co}\left(\sqrt{\frac{\ln n}{nh_n}}\right).$$

de (2.3), on a aussi

$$\sup_{x \in I} |\mathbb{E}f_n(x) - f(x)| = O(h_n^k).$$

Si $h_n = c\left(\frac{n}{\ln n}\right)^{\frac{-1}{2k+1}}$, $c > 0$, on obtient donc le résultat. \square

Remarque 2. En supposant que $k = 0$ dans (F.1), on retrouve la convergence presque complète sous l'hypothèse de continuité à la lumière du lemme de Bochner.

Lemme 2. (Bochner)

Soit la fonction mesurable $K : (\mathbb{R}^m, B_m) \rightarrow (\mathbb{R}, B)$ bornée, intégrable, vérifiant

$$\lim_{z \rightarrow \infty} \|z\|^m |K(z)| = 0.$$

Etant donné aussi la fonction mesurable et intégrable $g : (\mathbb{R}^m, B_m) \rightarrow (\mathbb{R}, B)$. On définit la fonction

$$g_n(x) = \frac{1}{h_n^m} \int_{\mathbb{R}^m} K\left(\frac{z}{h_n}\right) g(x - z) dz,$$

où $h_n \rightarrow 0$ quand $n \rightarrow \infty$.

Si g est continue alors

$$\lim_{n \rightarrow \infty} g_n(x) = g(x) \int_{\mathbb{R}^m} K(z) dz.$$

Si g est uniformément continue alors la convergence ci-dessus est uniforme.

Théorème 3. – Sous les hypothèses (H.1), (H.2), (K.1), (F.1) avec $k = 0$, on a

$$\lim_{n \rightarrow \infty} f_n(x) = f(x) \text{ p.co.}$$

– Sous les hypothèses (H.1), (H.2), (K.1), (K.6), (F.2) avec $k = 0$,

$$\lim_{n \rightarrow \infty} \sup_{x \in I} |f_n(x) - f(x)| = 0 \text{ p.co.}$$

1.2.2 Normalité asymptotique

Afin d'établir un théorème central limite, plusieurs variantes du lemme de Lindberg existent dans la littérature.

Dans cette partie, on s'intéresse à la généralisation suivante établie par Bardet et al. (2008).

Lemme 3. (Bardet J.M. et al. (2008))

Soit $(X_i)_{i \in \mathbb{N}}$ une suite de variables aléatoires indépendantes à valeurs dans \mathbb{R}^d centrées satisfaisant l'hypothèse suivante

– $\exists 0 < \delta < 1, \forall i \in \mathbb{N}, \mathbb{E}(\|X_i\|^{2+\delta}) < \infty,$

alors

$$\Delta_k \leq 6 \|f^{(2)}\|_\infty^{1-\delta} \|f^{(3)}\|_\infty^\delta A_k,$$

où

- $A_k = \sum_{i=1}^k \mathbb{E}(\|X_i\|^{2+\delta}).$
- f est une fonction trois fois dérivable, de dérivée troisième continue.
- $\Delta_k = \mathbb{E}|f(X_1 + \dots + X_k) - f(Y_1 + \dots + Y_k)|.$
- $(Y_i)_{i \in \mathbb{N}}$ est une suite de variables aléatoires indépendantes gaussiennes centrées ayant la même variance que $(X_i)_{i \in \mathbb{N}}.$

Théorème 4. (Bardet J.M. et al.(2008))

On suppose que les hypothèses du lemme 3 soient satisfaites, si de plus

$$\begin{aligned} \exists \Sigma > 0, \text{ tel que } \sum_{i=1}^n \text{Cov}(X_i) \rightarrow \Sigma, \text{ quand } n \rightarrow \infty. \\ A_n \rightarrow 0, \text{ quand } n \rightarrow \infty. \end{aligned}$$

Alors

$$S_n = \sum_{i=1}^n X_i \sim \mathcal{N}(0, \Sigma), \text{ as } n \rightarrow \infty.$$

Le résultat qui suit, établi par Parzen (1962), peut être démontré aussi grâce au théorème 4.

Corollaire 1. (Parzen E. (1962))

Sous les hypothèses (H.1) (K.1), (K.4), on obtient

$$\sqrt{nh_n} \left(\frac{f_n(x) - \mathbb{E}f_n(x)}{\sqrt{\text{Var}f_n(x)}} \right) \sim \mathcal{N}(0, 1), \text{ quand } n \rightarrow \infty.$$

1.2.3 Optimalité de l'estimateur de la densité

Dans la pratique, pour utiliser l'estimateur (1.1), il faut choisir ses constituants à savoir la fenêtre de lissage h_n et le noyau K en optimisant un critère d'erreur d'estimation.

Soit alors l'erreur moyenne quadratique

$$\text{MSE}(f_n) = \mathbb{E}(f_n(x) - f(x))^2$$

et l'erreur moyenne quadratique intégrée

$$\text{MISE}(f_n) = \mathbb{E} \int_{\mathbb{R}} (f_n(x) - f(x))^2 dx,$$

qui peut être écrite grâce au théorème de Fubini

$$\text{MISE}(f_n) = \int_{\mathbb{R}} \mathbb{E}(f_n(x) - f(x))^2 dx.$$

La vitesse de décroissance vers 0 de cette quantité appelée vitesse de convergence, mesure la qualité de l'estimateur, plus la perte est petite, plus l'estimateur sera bon, cette vitesse dépend de la régularité de la densité. Ainsi, l'estimation d'une fonction régulière est plus facile que l'estimation d'une fonction qui fluctue beaucoup et rapidement. A cet effet, on se limitera aux fonctions appartenant à la classe de Hölder .

Soit alors W_k l'ensemble des fonctions de densité $f(x)$ ayant des dérivées d'ordre $k \geq 2$ telles que $f^{(k)}(x)$ soient bornées, continues et de carré intégrables. Sous les hypothèses (K.1), (K.3) et (K.4), on a

$$\text{MISE}(f_n) = \frac{1}{nh_n} [K^2] + \frac{h_n^{2k}}{(k!)^2} [u^k K]^2 \int (f^{(k)}(u))^2 du + o\left(\frac{1}{nh_n} + h_n^{2k}\right).$$

En ignorant les termes d'ordre supérieur, on obtient l'approximation *AMISE* suivante

$$\text{AMISE}(f_n) = \frac{1}{nh_n} [K^2] + \frac{h_n^{2k}}{(k!)^2} [u^k K]^2 \int (f^{(k)}(u))^2 du. \quad (2.8)$$

Le choix du noyau n'est pas déterminant pour la performance de l'estimation. Epanechnikov (1969) montre que le noyau optimal au sens de $\text{MISE}(f_n(x))$ est de la forme

$$K_{opt}(x) = \frac{3}{4}(1 - x^2)I_{\{|x| \leq 1\}}.$$

Pour comparer l'efficacité de cet estimateur avec d'autres estimateurs raisonnables, on introduit la mesure d'efficacité suivante

$$E(K) = \frac{\int K^2(x) dx}{\int K_{opt}^2(x) dx}.$$

En effet, le tableau suivant montre que les différents noyaux offrent presque la même efficacité.

$K(x)$	$E(K)$
$\frac{1}{2}\mathbb{I}_{\{ x \leq 1\}}$	1.07
$\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$	1.01
$(1- x)\mathbb{I}_{\{ x \leq 1\}}$	1.05
$\frac{\pi}{4}\cos(\frac{\pi}{2}x)\mathbb{I}_{\{ x \leq 1\}}$	1.01

TAB 3. Efficacité des noyaux par rapport à K_{opt} .

En revanche, le choix de la fenêtre est d'une importance capitale. La forme (2.8) nous indique que, la largeur de fenêtre ne doit pas être choisie trop grande puisque dans ce cas elle augmentera la composante proportionnelle à h_n^{2k} , ni trop petite car elle augmentera la composante proportionnelle à h_n^{-1} . La fenêtre optimale théorique au sens de $AMISE(f_n(x))$ est donc

$$h_{opt1} = \left(\frac{(k!)^2 [K^2]}{2k[u^2 K]^2 [(f^{(k)})^2]} \right)^{\frac{1}{1+2k}} n^{-\frac{1}{1+2k}}. \quad (2.9)$$

Idem pour l'erreur moyenne quadratique

$$MSE(f_n) = \frac{1}{nh_n} f(x)[K^2] + \frac{h_n^{2k}}{(k!)^2} (f^{(k)}(x))^2 [u^k K]^2 + o\left(\frac{1}{nh_n} + h_n^{2k}\right),$$

d'où la fenêtre optimale

$$h_{opt2} = \left(\frac{(k!)^2 f(x)[K^2]}{2k[u^2 K]^2 (f^{(k)}(x))^2} \right)^{\frac{1}{1+2k}} n^{-\frac{1}{1+2k}}. \quad (2.10)$$

Le meilleur choix théorique obtenu de cette fenêtre donc dépend de la régularité de la densité. Or cette régularité est représentée par un paramètre inconnu. Ce choix n'est donc pas utilisable en pratique.

Des méthodes alternatives dites : Méthodes de sélection ont été proposées pour remédier à cela.

1.3 Quelques méthodes de sélection de la fenêtre de lissage

1.3.1 Méthodes classiques

Parmi les travaux proposés pour remédier au problème de la fenêtre de lissage, on cite l'article de M.C. Jones, J.S. Marron et Sheather [3], on a choisi de rappeler brièvement les méthodes les plus importantes.

Règle du pouce

Le paramètre de lissage optimal par rapport à la moyenne quadratique intégrée contient le terme inconnu $f''(x)$. Cette méthode proposée par Deheuvels (1977) consiste à supposer que $f(x)$ est la densité de Gauss d'espérance 0 et de variance σ_n^2 , si on utilise le noyau de Gauss, on obtient la fenêtre

$$h_p = 1.06\hat{\sigma}_n n^{-\frac{1}{5}},$$

avec $\hat{\sigma}_n$ est l'estimateur empirique de σ_n .

Si la vraie densité n'est pas la densité de Gauss, cette estimation de la fenêtre ne donne pas de bons résultats.

Validation croisée des moindres carrés

La méthode est appelée aussi validation croisée non biaisée, elle a été introduite par Rudemo (1982) et Bowman (1985). L'idée consiste à minimiser l'estimateur sans biais de $\text{MISE}(h_n) - [f^2]$.

En effet, pour estimer l'erreur moyenne intégrée

$$\text{MISE}(f_n) = \mathbb{E} \int f_n^2 - 2\mathbb{E} \int f_n f + \int f^2,$$

le terme $[f^2]$ n'intervient pas car il ne dépend pas de h_n , le terme $\mathbb{E} \int f_n^2$ est estimé par son estimateur trivial $\int f_n^2$, donc il reste à estimer le terme $\mathbb{E} \int f_n f$.

Pour cela, on note

$$T_n = \frac{1}{n} \sum_{i=1}^n f_{h_n,i}(X_i),$$

où $f_{h_n,i}$ est l'estimateur 'Leave one out' défini par

$$f_{h_n,i}(x) = \frac{1}{(n-1)h_n} \sum_{1 \leq j \leq n, j \neq i}^n K\left(\frac{x - X_j}{h_n}\right).$$

D'une part, on a

$$\begin{aligned} \mathbb{E}T_n &= \mathbb{E}(f_{h_n,1}(X_1)) \\ &= \mathbb{E}\left(\frac{1}{(n-1)h_n} \sum_{j \neq 1} \int K\left(\frac{z - X_j}{h_n}\right) f(z) dz\right) \\ &= \frac{1}{h} \int f(x) \int K\left(\frac{z - x}{h_n}\right) f(z) dz dx. \end{aligned}$$

D'autre part,

$$\begin{aligned}\mathbb{E} \int f_n f &= \mathbb{E} \left(\frac{1}{nh_n} \sum_{j=1}^n \int K\left(\frac{z - X_j}{h_n}\right) f(z) dz \right) \\ &= \frac{1}{h} \int f(x) \int K\left(\frac{z - x}{h_n}\right) f(z) dz dx.\end{aligned}$$

Donc T_n est un estimateur sans biais de $\mathbb{E} \int f_n f$.

D'où l'estimateur du MISE(h_n) - $[f^2]$, qu'on note UCV(h_n)

$$\text{UCV}(h_n) = \int f_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{h_n,i}(X_i).$$

Validation croisée biaisée

Scott et Terrell [39] proposent de remplacer le terme $[(f'')^2]$ dans l'expression de l'erreur moyenne quadratique intégrée par son estimateur $[(f_n'')^2] - \frac{[(K'')^2]}{nh_n^5}$ avec

$$f_n''(x) = \frac{1}{nh_n^3} \sum_{i=1}^n K''\left(\frac{x - X_i}{h_n}\right).$$

En effet, on a

$$\begin{aligned}\mathbb{E}[(f_n'')^2] &= \mathbb{E} \frac{1}{n^2 h_n^6} \int \left(\sum_{i=1}^n K''\left(\frac{x - X_i}{h_n}\right) \right)^2 dx \\ &= \mathbb{E} \left(\frac{1}{n^2 h_n^6} \left(\int \sum_{i=1}^n K''^2\left(\frac{x - X_i}{h_n}\right) dx + \int \sum_{i=1}^n \sum_{j=1, j \neq i}^n K''\left(\frac{x - X_i}{h_n}\right) K''\left(\frac{x - X_j}{h_n}\right) dx \right) \right).\end{aligned}$$

En supposant que

$$\int K''(x) dx = 0, \quad \int x K''(x) dx = 0, \quad \int x^2 K''(x) dx = 2.$$

On obtient pour le premier terme

$$\begin{aligned}\mathbb{E} \left(\frac{1}{n^2 h_n^6} \int \sum_{i=1}^n K''^2\left(\frac{x - X_i}{h_n}\right) dx \right) \\ &= \frac{1}{nh_n^6} \int \int K''^2\left(\frac{x - u}{h_n}\right) f(u) du dx \\ &= \frac{1}{nh_n^5} \int K''^2(x) dx.\end{aligned}$$

Pour le deuxième terme

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{n^2 h_n^6} \int \sum_{i=1}^n \sum_{j=1, j \neq i}^n K'' \left(\frac{x - X_i}{h_n} \right) K'' \left(\frac{x - X_j}{h_n} \right) dx \right) \\ &= \frac{1}{n^2 h_n^6} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \int \left(\int K'' \left(\frac{x - u}{h_n} \right) f(u) du \right)^2 dx \\ &= \int f''^2(x) dx. \end{aligned}$$

Par conséquent, l'estimateur de MISE obtenu en injectant le terme $[(f''_n)^2] - \frac{[(K'')^2]}{nh_n^5}$ est donné par

$$BCV(h_n) = \frac{[K^2]}{nh_n} + \frac{h_n^4}{4} [x^2 K]^2 \left([f''^2] - \frac{[(K'')^2]}{nh_n^5} \right).$$

Méthode Plug in

En remarquant que

$$[(f'')^2] = \int f^{(4)}(x) f(x) dx = \mathbb{E}(f^{(4)}(X)),$$

Sheather et Jones (1991) proposent alors d'estimer le terme inconnu $[(f'')^2]$ par l'estimateur E_g suivant

$$\begin{aligned} E_g &= \frac{1}{n} \sum_{i=1}^n f_n^{(4)}(X_i) \\ &= \frac{1}{n(n-1)g^5} \sum_{i=1}^n \sum_{j=1}^n K^{(4)} \left(\frac{X_i - X_j}{g} \right). \end{aligned}$$

Ce qui est nouveau dans cette approche est que la fenêtre g de l'estimateur de $[(f'')^2]$ est différente du paramètre h_n en question. Soit l'erreur moyenne quadratique de l'estimateur E_g

$$\mathbb{E}(E_g - [(f'')^2])^2,$$

g est donc son minimum donné par l'expression

$$g = C_1 [f^{(3)}] C_2(K) n^{-\frac{1}{7}},$$

où encore $[f^{(3)}]$ est calculé en se référant à la densité de Gauss.

Fenêtre locale

L'estimation d'un paramètre de lissage optimal global, c'est à dire au sens du critère global MISE, est plus facile que celle d'un paramètre local (optimal au sens de MSE) car ce dernier est une fonction de x .

Cependant, la régularité de la fonction à estimer n'est pas homogène, elle varie en tout point x , l'utilisation d'une fenêtre globale ne permet pas l'adaptation ponctuelle à cette régularité. IL serait alors intéressant de pencher vers la fenêtre locale en chaque point x du type $h_n(x)$.

Pour pouvoir choisir entre les deux approches locale et globale, on définit l'efficacité relative suivante, (voir Cao R. (2001)).

$$RE = \frac{\text{MISE}(h_{opt1})}{\text{MISE}(h_{opt2})},$$

où h_{opt1} et h_{opt2} représentent les paramètres optimaux par rapport à MISE et MSE respectivement.

En utilisant les approximations AMISE et AMSE, pour $f \in W_2$, on obtient l'approximation de l'efficacité relative ARE suivante

$$\text{ARE} = \frac{\text{AMISE}(h_{opt1})}{\text{AMISE}(h_{opt2})} \quad (3.11)$$

$$= \frac{(\int (f''(x))^2 dx)^{\frac{1}{5}}}{\int |f''(x)|^{\frac{2}{5}} f(x)^{\frac{4}{5}} dx}. \quad (3.12)$$

D'après l'inégalité de Jensen, on a

$$\begin{aligned} \int |f''(x)|^{\frac{2}{5}} f(x)^{\frac{4}{5}} dx &= \int \left(\frac{f''(x)^2}{f(x)} \right)^{\frac{1}{5}} f(x) dx \\ &\leq \left(\int \frac{f''(x)^2}{f(x)} f(x) dx \right)^{\frac{1}{5}} \\ &= \left(\int (f''(x))^2 dx \right)^{\frac{1}{5}}. \end{aligned}$$

Donc $ARE > 1$, ce qui montre l'avantage de l'utilisation du paramètre local en terme de MISE par rapport au paramètre global.

1.3.2 Approche Bayésienne

Cette méthode récente consiste à doter le paramètre de lissage h , considéré ici comme une variable aléatoire, d'une loi de probabilité $\pi(h)$, ce qui donnera

plus d'information sur le paramètre inconnu h surtout pour de petites tailles d'échantillon.

(Voir les travaux de Brewer (1998), Zhang et al.(2006).)

Le paramètre global h peut être obtenu à partir de la formule de Bayes

$$\pi(h/x) = \frac{\pi(x/h)\pi(h)}{\pi(x)},$$

où

- $\pi(x/h)$ est la fonction de vraisemblance définie à partir de l'estimateur leave one out $f_{h,i}(x)$ par $\pi(x/h) = \prod_{i=1}^n f_{h,i}(X_i)$.
- $\pi(h)$ est la loi a priori de la fenêtre h .
- $\pi(h/x)$ est sa loi a posteriori.

On définit d'abord le risque bayésien quadratique

$$\mathbb{E}(\hat{h} - h)^2 = \int (\hat{h} - h)^2 \pi(h/x) dh.$$

L'estimateur de Bayes optimal \hat{h} du paramètre h au sens du risque bayésien est donc donné par la moyenne a posteriori

$$\hat{h} = \mathbb{E}(h/x) = \int h \pi(h/x) dh.$$

Le calcul de $\pi(x)$ n'est pas facile, à cet effet, une première alternative dite méthode Monté Carlo par Chaines de Markov MCMC a été proposée pour calculer la moyenne a posteriori $\mathbb{E}(h/x)$. Des résultats de simulation effectuée montrent que cette méthode n'est pas meilleure que les méthodes classiques. En revanche, l'approche locale étudiée par Gangopadhyay et Cheung (2002) donne de meilleurs résultats, la loi a posteriori de la fenêtre locale $h(x)$ dans ce cas est donnée par

$$\pi(h/x) = \frac{f_h(x)\pi(h)}{\int f_h(x)\pi(h)dh}.$$

$f_h(x)$ est la densité du paramètre local $h(x)$, elle peut être donnée par l'estimateur de Parzen $f_{n,h}(x)$. La loi a posteriori est estimée alors par

$$\hat{\pi}(h/x) = \frac{f_{n,h}(x)\pi(h)}{\int f_{n,h}(x)\pi(h)dh}.$$

Le choix de la loi a priori n'est pas important, Gangopadhyay et Cheng (2002) utilisent la loi suivante

$$\pi(h) = \frac{2}{\Gamma(\alpha)\beta^\alpha} \frac{1}{h^{2\alpha+1}} \exp\left(-\frac{1}{\beta h^2}\right),$$

où α et β sont des constantes positives.

Le paramètre de lissage local $h(x)$ est estimé donc par

$$\hat{h}(x) = \int h(x) \hat{\pi}(h/x) dh.$$

L'utilisation d'un noyau gaussien dans l'expression de l'estimateur de Parzen donne alors

$$\hat{h}(x) = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha + \frac{1}{2})} \frac{\sum_{i=1}^n \left(\frac{1}{(\beta(x-X_i)^2+2)}\right)^\alpha}{\sum_{i=1}^n \left(\frac{1}{(\beta(x-X_i)^2+2)}\right)^{\alpha+\frac{1}{2}}}.$$

1.4 Simulation

A partir de $N = 20$ réalisations d'un échantillon simulé de tailles différentes $n = 100, 500, 1000, 5000, 10000$ de loi

- Gaussienne $\mathcal{N}(0, 1)$.
- χ^2 de degré de liberté $df = 6$.

On étudie les performances de quelques méthodes de sélection étudiées ci-dessus à savoir la validation croisée biaisée, non-biaisée, la règle du pouce et la méthode plug-in.

Les résultats sont comparés avec le paramètre de lissage optimal au sens de l'erreur moyenne quadratique intégrée de l'estimateur à noyau de la densité $MISE(f_n)$ dans le cas $k = 2$.

On rappelle l'expression du paramètre optimal global pour $k = 2$,

$$h_{opt1} = \left(\frac{[K^2]}{[u^2 K]^2 [(f'')^2]}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Le noyau utilisé est le noyau d'Epanechnikov.

Les résultats de la simulation effectuée via le logiciel **R** sont tabulés ci-dessous

$n h$	h_{ucv}	h_{bcv}	h_p	h_{sj}	h_{opt}
100	0.4001	0.4196	0.4255	0.4485	0.4215
500	0.3150	0.3339	0.2994	0.2995	0.3055
1000	0.2971	0.2910	0.2605	0.2324	0.2660
5000	0.2012	0.2069	0.1925	0.1935	0.1927
10000	0.1611	0.1670	0.1682	0.1678	0.1678

TAB1 : Paramètres de lissage calculés par les méthodes de validation croisée biaisée et non-biaisée, règle du pouce et la méthode Plug-in pour un échantillon gaussien.

A partir de cette première simulation, on constate les points suivants

- h_{opt} décroît en augmentant la taille d'échantillon.
- h_p donne de bons résultats. Les valeurs de h_{sj} , et, h_{bcv} , h_{ucv} performant aussi bien.
- En augmentant la taille de l'échantillon, les résultats s'améliorent davantage.

$n h$	h_{ucv}	h_{bcv}	h_p	h_{sj}	h_{opt}
100	1.3906	1.672	1.5577	1.2961	0.7104
500	0.8868	0.8734	0.9977	0.8073	0.4514
1000	0.7965	0.7965	0.8548	0.755	0.4482
5000	0.458	0.4683	0.4741	0.3248	0.4112
10000	0.3597	0.2483	0.5546	0.3928	0.2828

TAB2 : Paramètres de lissage calculés par les méthodes de validation croisée biaisée et non-biaisée, règle du pouce et la méthode Plug-in pour un échantillon de χ^2 de degré de liberté $df = 6$.

Pour le deuxième échantillon issu de la loi χ^2 , on constate les points suivants

- h_{opt} décroît aussi en augmentant la taille d'échantillon.
- h_p ne donne plus de bons résultats même en augmentant la taille de l'échantillon.
- Pour les valeurs de h_{bcv} , h_{ucv} , les résultats ne sont pas stables entre les différents échantillons.
- h_{bcv} , h_{ucv} , et h_{sj} ne sont pas bons aussi mais ils sont meilleurs que les valeurs de h_p .

Remarque 3. A propos l'approche Bayésienne globale, des travaux de simulation effectués par Zoughab N. (2013) montrent que cette approche présente un inconvénient majeur dans le coût en terme de temps d'exécution à cause de la lenteur de la convergence de la méthode MCMC. En adoptant la version locale de cette approche, dans Gangopadhyay et Cheung (2002), on montre une bonne performance via des simulations sur des échantillons issus de différentes lois, avec le temps d'exécution plus réduit par rapport à l'approche Bayésienne globale.

Conclusion

La performance des méthodes de sélection classiques dépend des densités à estimer ce qui n'est pas valable car la densité est inconnue. Outre le fait que les méthodes de validation croisée montrent une grande variabilité des résultats et les conditions théoriques de consistance des méthodes Plug in sont difficiles à vérifier. L'approche Bayésienne locale semble avoir de bonnes performances théoriques et pratiques.

Chapitre 2

Estimation de la régression par la méthode du noyau

2.1 Introduction

On dispose d'un échantillon $(X_1, Y_1), \dots, (X_n, Y_n)$ de variables aléatoires indépendantes de même loi, de densité jointe $f(x, y)$ et densité marginale $f(x)$. On suppose aussi l'existence de la fonction de régression inconnue

$$r(x) = \mathbb{E}(Y/X = x),$$

qu'on propose d'estimer par l'estimateur à noyau suivant

$$r_n(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{si } \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) = 0, \end{cases} \quad (1.1)$$

tel que le paramètre de lissage $h_n \rightarrow 0$ vérifie

– (H.3) $n^\delta h_n^{\delta+2} \rightarrow \infty$, $\delta > 0$, $nh_n^5 \rightarrow 0$.

En plus des hypothèses suivantes

– (F.2) r et f sont k -fois continûment dérivables autour de x .

– (F.3) $\nu(x) = \mathbb{E}(Y^2/X = x)$ est continue.

– (R.1) $|Y| < M$, $M < \infty$.

Dans ce qui suit nous utilisons la notation

$$r_n(x) = \frac{g_n(x)}{f_n(x)}, \quad (1.2)$$

$g_n(x)$ est l'estimateur de la fonction $g(x) = f(x)r(x)$.

Remarque 4. Avant de présenter les propriétés de cet estimateur, il convient de faire un peu d'historique afin d'apprécier la présente méthode d'estimation.

Turkey (1961) a introduit d'abord le régressogramme

$$\hat{r}_{reg}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in B_j)}}{\sum_{i=1}^n \mathbb{I}_{(X_i \in B_j)}}, \quad x \in B_j$$

$B_j, j = 1, \dots, J$ est une partition du support de X .

Le choix du nombre J et des bornes des intervalles B_j est à l'origine de la complexité d'un tel estimateur, il a été remplacé ensuite par l'estimateur de la fenêtre mobile

$$\hat{r}_{mob}(x) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{(X_i \in [x-h, x+h])}}{\sum_{i=1}^n \mathbb{I}_{(X_i \in [x-h, x+h])}}.$$

Cependant, ce dernier n'est pas continu. Ainsi, il a été généralisé par l'estimateur (1.1).

2.2 Propriétés asymptotiques

2.2.1 Convergence presque complète

Le résultat suivant est obtenu sous la condition de dérivabilité sur les fonctions $r(x)$ et $f(x)$.

Théorème 5. (*Sarda et Vieu (2000)*)

Sous les hypothèses (F.1), (F.2), (H.1), (H.2), (K.1), (K.3), et (R.1), on a

$$|r_n(x) - r(x)| = O(h_n^k) + O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \quad p.co.$$

Preuve On utilise la décomposition suivante

$$r_n(x) - r(x) = \frac{g_n(x) - g(x)}{f_n(x)} + \frac{f(x) - f_n(x)}{f_n(x)} r(x). \quad (2.3)$$

Le résultat découle des points suivants

$$\begin{aligned} |\mathbb{E}(g_n(x)) - g(x)| &= O(h_n^k). \\ |\mathbb{E}(g_n(x)) - g_n(x)| &= O\left(\sqrt{\frac{\ln n}{nh_n}}\right), p.co. \\ \exists \delta > 0, \sum_{n=1}^{\infty} \mathbb{P}(f_n(x) \leq \delta) &< \infty. \end{aligned}$$

De la même manière que pour la densité, les deux premiers points sont démontrés en utilisant le développement de Taylor de la fonction g et l'inégalité de Bernstein .

Quant au dernier point, on a

$$|f_n(x)| \leq \frac{f(x)}{2} \Rightarrow |f_n(x) - f(x)| > \frac{f(x)}{2},$$

ceci entraîne

$$\mathbb{P}(|f_n(x)| \leq \frac{f(x)}{2}) \leq \mathbb{P}(|f_n(x) - f(x)| > \frac{f(x)}{2}).$$

Si on choisit $\delta = \epsilon = \frac{f(x)}{2}$, la convergence presque complète de l'estimateur de la densité donne

$$\exists \delta > 0, \sum_{n \geq 1} \mathbb{P}(|f_n(x)| \leq \delta) < \infty.$$

□

Convergence presque complète uniforme de r_n

En utilisant un noyau lipschitzien sur un compact I , avec l'hypothèse suivante

$$- (R.2) \exists \theta > 0, \inf_{x \in I} f(x) > \theta.$$

On obtient le résultat suivant.

Théorème 6. (*Sarda et Vieu (2000)*)

Sous les conditions (F.2), (F.2), (H.1), (H.2), (K.1), (K.3), (K.5), (R.1), et (R.2), on a

$$\sup_{x \in I} |r_n(x) - r(x)| = O(h_n^k) + O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \text{ p.co.}$$

Preuve On a

$$\sup_{x \in I} |r_n(x) - r(x)| \leq \frac{\sup_{x \in I} |g_n(x) - g(x)|}{\inf_{x \in I} |f_n(x)|} + \sup_{x \in I} |f_n(x) - f(x)| \frac{\sup_{x \in I} |r(x)|}{\inf_{x \in I} |f_n(x)|}.$$

En raisonnant comme pour la densité, on obtient

$$\sup_{x \in I} |g_n(x) - g(x)| = O(h_n^k) + O\left(\sqrt{\frac{\ln n}{nh_n}}\right), \text{ p.co.} \quad (2.4)$$

Maintenant, en utilisant l'hypothèse (R.2), on a

$$\inf_{x \in I} |f_n(x)| \leq \frac{\theta}{2},$$

cela implique

$$\sup_{x \in I} |f_n(x) - f(x)| > \frac{\theta}{2}.$$

On obtient alors

$$\mathbb{P}\left(\inf_{x \in I} |f_n(x)| \leq \frac{\theta}{2}\right) \leq \mathbb{P}\left(\sup_{x \in I} |f_n(x) - f(x)| > \frac{\theta}{2}\right),$$

la convergence presque complète uniforme de l'estimateur de la densité avec le choix $\delta = \epsilon = \frac{\theta}{2}$ entraîne alors

$$\exists \delta > 0, \sum_{n \geq 1} \mathbb{P}(\inf_{x \in I} |f_n(x)| \leq \delta) < \infty. \quad (2.5)$$

Maintenant de (2.4), (2.5), et le fait que r soit bornée on aboutit au résultat.

□

2.2.2 Normalité asymptotique

Schuster (1970) a établi la normalité asymptotique de l'estimateur à noyau de la régression sur des points distincts de \mathbb{R} . Nadaraya (1989) l'a généralisé sur \mathbb{R} de la manière suivante.

Théorème 7. (*Nadaraya (1989)*)

Soit $\mathbb{E}|Y|^{2+\delta} < \infty$ pour $\delta \in (\frac{1}{2}, 1]$.

Si (H.3), (F.2) pour $k = 2$, (F.3), (K.4) sont vérifiées, on obtient alors

$$\sqrt{nh_n}(r_n(x) - r(x)) \rightarrow_{n \rightarrow \infty} \mathcal{N}(0, \sigma(x)),$$

avec $\sigma^2(x) = \text{Var}(Y/X = x)f(x) \int K^2(u)du$.

Remarque 5. Dans ce qui suit, on considère le terme de centrage $\frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}$. D'après Nadaraya (1989), sous l'hypothèse (H.1), on a

$$\frac{\mathbb{E}(g_n(x))}{\mathbb{E}(f_n(x))} - \mathbb{E}r_n(x) = O\left(\frac{1}{nh_n}\right).$$

Preuve

On pose

$$r_n(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)} = \zeta_n(x)f_n(x)^{-1}(\mathbb{E}f_n(x))^{-1}, \quad (2.6)$$

avec

$$\begin{aligned} \zeta_n(x) &= \frac{1}{nh_n} \sum_{i=1}^n (\alpha_n Y_i - \beta_n) K\left(\frac{x - X_i}{h_n}\right), \\ \alpha_n &= \mathbb{E}(f_n(x)), \\ \beta_n &= \mathbb{E}(g_n(x)). \end{aligned}$$

On a

$$\mathbb{E}(\zeta_n(x)) = 0,$$

et

$$nh_n \text{Var}(\zeta_n(x)) \rightarrow_{n \rightarrow \infty} \text{Var}(Y/X = x) f^3(x) \int K^2(u)du.$$

Donc

$$L_n = \frac{\mathbb{E}|\zeta_n|^{2+\delta}}{n^{\frac{\delta}{2}} \text{Var}(\zeta_n)^{1+\frac{\delta}{2}}} = o\left(\frac{n^{-\frac{\delta}{2}}}{h_n^{1+\frac{\delta}{2}}}\right),$$

l'hypothèse (H.2) implique $L_n \rightarrow_{n \rightarrow \infty} 0$, donc

$$\frac{\zeta_n(x) - \mathbb{E}\zeta_n(x)}{\text{Var}(\zeta_n(x))} \rightarrow_{n \rightarrow \infty} \mathcal{N}(0, 1).$$

Maintenant, la convergence en probabilité de l'estimateur de la densité et (2.6) entraînent

$$\mathbb{P} \left(\sqrt{nh_n} \sigma^{-1}(x) (r_n(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}) < \lambda \right) \rightarrow \Phi(\lambda),$$

quand $n \rightarrow \infty$.

(Φ est la fonction de distribution de la loi de Gauss standard).

On conclut alors

$$\mathbb{P} \left(\sqrt{nh_n} \sigma^{-1}(x) (r_n(x) - \frac{g(x)}{f(x)}) < \lambda \right) \rightarrow \Phi(\lambda),$$

quand $n \rightarrow \infty$. \square

2.2.3 Optimalité de l'estimateur de la régression

Afin de déterminer le paramètre de lissage optimal, on s'intéresse à l'optimisation de l'erreur moyenne quadratique $\text{MSE}(r_n)$

$$\text{MSE}(r_n) = \mathbb{E}(r_n(x) - r(x))^2$$

et l'erreur moyenne quadratique intégrée $\text{MISE}(r_n)$

$$\text{MISE}(r_n) = \mathbb{E} \int (r_n(x) - r(x))^2 dx.$$

Sous la condition de régularité (F.2), on aboutit aux approximations suivantes.

$$\text{AMSE}(r_n) = \frac{h_n^{2k}}{k!^2} \left(\left(r^{(k)}(x) + kr^{(k-1)}(x) \frac{f'(x)}{f(x)} \right) [u^k K] \right)^2 + \frac{1}{nh_n} \frac{\sigma^2(x)}{f(x)} [K^2].$$

$$\text{AMISE}(r_n) = \frac{h_n^{2k}}{k!^2} \int \left(\left(r^{(k)}(x) + kr^{(k-1)}(x) \frac{f'(x)}{f(x)} \right) [u^k K] \right)^2 dx + \frac{[K^2]}{nh_n} \int \frac{\sigma^2(x)}{f(x)} dx,$$

avec $\sigma^2(x) = \frac{1}{f(x)} \int y^2 f(x, y) dy - r^2(x)$.

Le paramètre de lissage optimal au sens du MSE et du MISE respectivement pour $k = 2$ résulte immédiatement

$$h_{op1} = \left(\frac{\frac{\sigma^2(x)}{f(x)} [K^2]}{\left((r''(x) + 2r'(x) \frac{f'(x)}{f(x)}) [u^2 K] \right)^2} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \quad (2.7)$$

$$h_{op2} = \left(\frac{\int \frac{\sigma^2(x)}{f(x)} [K^2] dx}{\int \left((r''(x) + 2r'(x) \frac{f'(x)}{f(x)}) [u^2 K] \right)^2 dx} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (2.8)$$

2.2.4 Validation croisée de la régression

Les valeurs de la fenêtre de lissage (2.7) et (2.8) ne sont pas utilisables en pratique car elles dépendent aussi des paramètres inconnus. Pour surmonter ce problème, une méthode simple consiste à déterminer la valeur de la fenêtre h minimisant la variance estimée des résidus, c'est à dire la quantité donnée par

$$\hat{\sigma}_\epsilon^2(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - r_n(X_i))^2.$$

Cela permet d'obtenir la valeur de h de manière à ce que les données soient parfaitement ajustées.

En revanche, si K est à support compact, on a

$$\lim_{h \rightarrow 0} r_n(X_i) = \frac{K(0)Y_i}{K(0)} = Y_i.$$

Si l'on cherche alors h minimisant $\hat{\sigma}_\epsilon^2(h)$, on a aboutit à $h \rightarrow 0$, ce critère est donc sans intérêt.

Pour y parvenir, la méthode de validation croisée peut être introduite dans ce contexte aussi, son principe consiste à générer des échantillons par tirage et pour chacun de ces échantillons calculer l'estimateur 'Leave one out' de la régression défini par

$$r_{n,i}(x) = \frac{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h}\right) Y_j}{\sum_{j=1, j \neq i}^n K\left(\frac{x-X_j}{h}\right)}.$$

Le paramètre h est calculé de manière à ce qu'il minimise un estimateur sans biais de

$$\text{MISE}(r_n) = \int r(x)^2 f(x) dx,$$

cet estimateur est donné par la fonction de validation croisée

$$CV(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - r_{n,i}(X_i))^2. \quad (2.9)$$

Le paramètre de lissage calculé en minimisant la fonction de validation croisée CV

$$h_{cv} = \arg \min_{h>0} CV$$

est asymptotiquement optimal, (voir Hardle (1990)).

2.3 Cas multivarié

Soit (X, Y) un couple de vecteurs aléatoires de \mathbb{R}^p , $K : \mathbb{R}^p \rightarrow \mathbb{R}$ est une fonction noyau produit de noyaux univariés. L'estimateur de la régression dans ce cas est défini par

$$r_n(x) = \frac{g_n(x)}{f_n(x)}, \quad (3.10)$$

où

$$g_n(x) = \frac{1}{nh_n^p} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)$$

et

$$f_n(x) = \frac{1}{nh_n^p} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right).$$

Le calcul de l'erreur moyenne quadratique est étendu au cas multivarié, pour cela, on a besoin de l'hypothèse suivante

– (H.4) $h_n \rightarrow 0, nh_n^p \rightarrow \infty$, quand $n \rightarrow \infty$.

Théorème 8. *Sous les hypothèses (F.2) pour $k = 2$, (F.3), (H.4), si K est un noyau positif vérifiant (K.1), (K.6) on a*

$$\text{MSE}(r_n) = B^2 h_n^4 + \frac{[K^2] \sigma^2(x)}{nh_n^p} + o\left(h_n^4 + \frac{1}{nh_n^p}\right),$$

avec

$$B = \frac{1}{2f(x)} \sum_{j=1}^p T_k(j) \left(\left(\frac{\delta^k g}{\delta x_j^k} \right) - \left(r(x) \frac{\delta^k f}{\delta x_j^k} \right) \right).$$

$$T_k(j) = \int_{\mathbb{R}^p} u_j^k K(u_1, \dots, u_p) du_1 \dots du_p.$$

Donc le paramètre de lissage optimal au sens de L^2 est donné par

$$h_{opt} = \left(\frac{p[K^2]\sigma^2(x)}{4B^2} \right)^{\frac{1}{p+4}} n^{-\frac{1}{p+4}}.$$

2.3.1 Fléau de dimension et modèles additifs

Lorsqu'on estime la fonction de régression à l'aide d'un estimateur à noyau, la vitesse optimale de convergence au sens de la norme L^2 est de l'ordre $n^{-\frac{k}{2k+p}}$, pour $k \geq 2$.

Cette vitesse reste optimale pour de petites valeurs de p , mais elle devient moins bonne pour les grandes valeurs. Ce phénomène est connu dans la littérature par le fléau de dimension.

Pour pallier ce problème, des procédures ont été proposées pour réduire l'effet de la dimensionalité des données. Stone (1985) évoque par exemple les modèles additifs en écrivant la fonction de régression multivariée $r(x)$ comme la somme de fonctions univariées $r_\ell(x_\ell)$, ($\ell = 1, \dots, p$), c'est à dire

$$r(x) = \mu + \sum_{\ell=1}^p r_\ell(x), \quad (3.11)$$

avec $\mu = \mathbb{E}(Y)$.

$r_\ell(x)$, ($\ell = 1, \dots, p$) sont appelées composantes additives de la régression vérifiant $\mathbb{E}(r_\ell(x_\ell)) = 0$.

Afin d'estimer la fonction de régression inconnue $r(x)$, on propose d'estimer les composantes additives par la méthode d'intégration marginale introduite par Linton et Nielson (1995).

Méthode d'intégration marginale

Soient les fonctions de densité bornées et continues q_ℓ , ($\ell = 1, \dots, p$) et les produits

$$q(x) = \prod_{\ell=1}^p q_\ell(x_\ell), \quad q_{-\ell}(x_{-\ell}) = \prod_{j \neq \ell} q_j(x_j),$$

avec $x = (x_1, \dots, x_p) \in \mathbb{R}^p$, et $x_{-\ell} = (x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_p)$.

On introduit la quantité

$$\eta_\ell(x_\ell) = \int_{\mathbb{R}^{p-1}} r(x)q_{-\ell}(x_{-\ell})dx_{-\ell} - \int_{\mathbb{R}^p} r(x)q(x)dx. \quad (3.12)$$

En intégrant l'expression de la régression $r(x)$ définie dans (3.11) par rapport à $q(x)$, $q_{-\ell}(x_{-\ell})$ respectivement, on obtient

$$\begin{aligned} \int_{\mathbb{R}^p} r(x)q(x)dx &= \mu + \sum_{\ell=1}^p \int_{\mathbb{R}} r_\ell(z_\ell)q_\ell(z_\ell)dz_\ell, \\ \int_{\mathbb{R}^{p-1}} r(x)q_{-\ell}(x_{-\ell})dx_{-\ell} &= \mu + r_\ell(x_\ell) + \sum_{j \neq \ell} \int_{\mathbb{R}} r_j(z_j)q_j(z_j)dz_j. \end{aligned} \quad (3.13)$$

Donc

$$\begin{aligned} \int_{\mathbb{R}^{p-1}} r(x)q_{-\ell}(x_{-\ell})dx_{-\ell} - \int_{\mathbb{R}^p} r(x)q(x)dx \\ = r_\ell(x_\ell) - \int_{\mathbb{R}} r_\ell(z_\ell)q_\ell(z_\ell)dz_\ell. \end{aligned} \quad (3.14)$$

En combinant maintenant (3.12) et (3.14), on obtient

$$\eta_\ell(x_\ell) = r_\ell(x_\ell) - \int_{\mathbb{R}} r_\ell(z_\ell)q_\ell(z_\ell)dz_\ell. \quad (3.15)$$

Il est clair que les fonctions r_ℓ et η_ℓ sont égales à une constante près, c'est à dire que η_ℓ est aussi composante additive, l'expression (3.11) entraîne alors

$$r(x) = \sum_{\ell=1}^p \eta_\ell(x_\ell) + \sum_{\ell=1}^p \int_{\mathbb{R}} r_\ell(z)q_\ell(z)dz_\ell + \mu.$$

De (3.13) on a

$$r(x) = \sum_{\ell=1}^p \eta_\ell(x_\ell) + \int_{\mathbb{R}^p} r(z)q(z)dz.$$

De (3.12) les composantes $\eta_\ell(x_\ell)$ sont estimées via l'estimateur (3.10) par

$$\hat{\eta}_\ell(x_\ell) = \int_{\mathbb{R}^{p-1}} r_n(x)q_{-\ell}(x_{-\ell})dx_{-\ell} - \int_{\mathbb{R}^p} r_n(x)q(x)dx.$$

On s'attache à cette méthode car elle offre des avantages théoriques telle que la vitesse de convergence optimale obtenue dans la régression univariée.

2.4 Simulation

On propose de simuler la méthode de validation croisée qui s'impose grâce à son aspect direct pour approcher la valeur du paramètre de lissage optimal de l'estimateur à noyau de la régression.

Soit le modèle de régression $Y = r(X) + Z$, on génère $N = 10$ réalisations d'un échantillon de taille $n = 100$ de couple aléatoire $(X_i, Y_i)_{(i \leq n)}$, où $X_i \sim U[0, 1]$, $Z_i \sim \mathcal{N}(0, 1)$.

La régression $r(x) = (x+2)e^{-(x-0.5)^2}$ est estimée par $r_n(x)$ basé sur la fenêtre de lissage h_{cv} calculée en minimisant le critère de validation croisée (2.9).

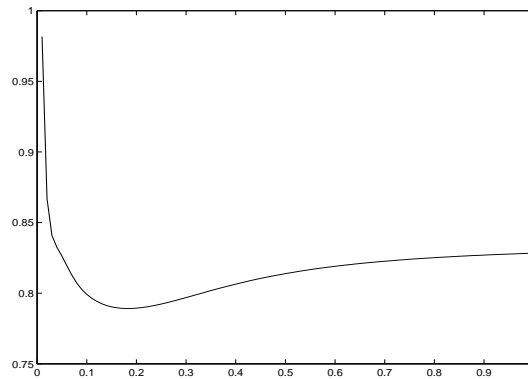


FIGURE 2.1 – Fonction de validation croisée pour $n = 100$.

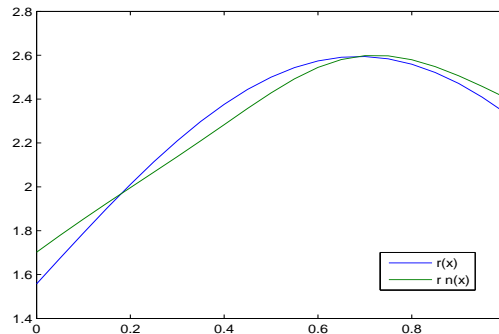


FIGURE 2.2 – Estimateur de la régression basé sur la fenêtre $h_{cv} = 0.2$ pour $n = 100$.

Même si cette méthode est pratique et facile à mettre en oeuvre, elle manque de stabilité. En effet, les résultats de la simulation montrent que la valeur du paramètre de lissage h_{cv} varie d'un échantillon à un autre, même en augmentant la taille de l'échantillon.

Chapitre 3

Lois uniformes des dérivées de la densité et de la régression

3.1 Introduction

Dans ce chapitre, on étudie les lois limites uniformes des estimateurs des dérivées de la densité et de la régression dépendant d'une fenêtre locale. Ces lois permettront ultérieurement de déduire les bandes de confiance des dérivées de la densité et de la régression.

Les lois uniformes de l'estimateur à noyau ont été étudiées dans divers travaux, Deheuvels et Mason (2004) établissent les lois uniformes de l'estimateur à noyau de la densité et de la régression dépendant d'une fenêtre locale, Nemouchi et Mohdeb (2010) étudient ces lois limites dans le cadre d'un modèle de régression Gaussien, ensuite Blondin (2004) étudie les lois uniformes des dérivées de la densité et de la régression dépendant d'une fenêtre globale. En utilisant les mêmes techniques, Debbarih (2007) investigate les lois uniformes de la régression additive dans le cas multivarié, et des composantes principales du modèle de régression additif. Notre travail est fortement inspiré et directement déduit de ces derniers, et notamment de Blondin (2004). Tous ces travaux sont basés sur la théorie des processus empiriques fondée par Stute (1980).

Soit $(X_1, Y_1) \cdots (X_n, Y_n)$ une suite de couples de variables aléatoires réelles indépendantes de même loi que le couple (X, Y) de densité jointe $f_{X,Y}(x, y)$ continue sur $J \times \mathbb{R}$, tel que $I = [a, b]$, $J = [c, d]$, $(-\infty < a < b < c < d < \infty)$ sont deux intervalles compacts de \mathbb{R} , et de densité marginale $f_X(x)$ stricte-

ment positive .

Soit la fonction de régression

$$r(x) =: E(\psi(Y)/X = x) = \frac{1}{f_X(x)} \int_{\mathbb{R}} \psi(Y) f_{X,Y}(x, y) dy =: \frac{g_\psi(x)}{f_X(x)}, \quad (1.1)$$

où $\psi(\cdot)$ est une fonction à valeurs réelles mesurable et bornée sur un intervalle compact de \mathbb{R} .

Pour établir les lois uniformes des dérivées de la densité et de la régression, on a besoin des hypothèses suivantes

- (F.1) $f_{X,Y}(x, y)$ et $f_X(x)$ sont k fois continûment différentiables sur $J \times \mathbb{R}$.
- (F.2) $YI_{(X \in J)}$ est bornée.
- (K'.1) $\forall x \in \mathbb{R}$, $K(\cdot)$ est une densité à support compact, à variations bornées et continue à droite sur \mathbb{R} , $K^{(k)}$ existe et à variations bornées.
- (H.4) $\frac{nh_n^{2k+1}}{\log \frac{1}{h_n}} \rightarrow \infty$ quand $n \rightarrow \infty$.

Soient les suites de fonctions mesurables $H_n(X_1, \dots, X_n; x)$, $\Theta_n(x)$.

On pose $H_n(x) = h_n C_n(x)$, $x \in J$. Pour les constantes $0 < c_1 \leq c_2 < \infty$, et les fonctions $C(x)$, $\Theta(x)$ positives continues et bornées sur J , on suppose ce qui suit

- (B.1) $P(c_1 h_n \leq \inf H_n(x) \leq \sup H_n(x) \leq c_2 h_n) \rightarrow 1$, quand $n \rightarrow \infty$.
- (B.2) $P(\sup_{x \in J} |\frac{H_n(x)}{h_n} - C(x)| \geq \epsilon) \rightarrow 0$, quand $n \rightarrow \infty$, $\forall \epsilon > 0$.
- (B.3) $P(\sup_{x \in J} |\frac{\Theta_n(x)}{\Theta(x)} - 1| \geq \epsilon) \rightarrow 0$, quand $n \rightarrow \infty$, $\forall \epsilon > 0$.

On définit les estimateurs des $k^{ièmes}$ dérivées de $f_X(x)$ et de $r(x)$ respectivement par

$$\hat{f}_{X,n}^{(k)}(x, h_n) = \frac{1}{nh_n^{k+1}} \sum_{i=1}^n K^{(k)}\left(\frac{x - X_i}{h_n}\right), \quad (1.2)$$

$$\hat{r}_{\psi,n}^{(k)}(x, h_n) = \sum_{j=0}^k C_j^k \hat{g}_{\psi,n}^{(j)}(x) (\hat{f}_{X,n}^{-1})^{(k-j)}(x), \quad (1.3)$$

où $\hat{g}_{\psi,n}^{(k)}(x, h_n) = \frac{1}{nh_n^{k+1}} \sum_{i=1}^n \psi(Y_i) K^{(k)}\left(\frac{x - X_i}{h_n}\right)$, avec $\hat{f}_{X,n}(x) \neq 0$.

3.2 Lois uniformes des dérivées de la densité et de la régression

Les théorèmes suivants établissent les lois limites de la déviation uniforme des estimateurs des dérivées de la densité et de la régression dépendant d'une fenêtre locale $H_n(x)$.

Théorème 9. (i) *Si les hypothèses (F.1), (K'.1), (H.4), (B.1), (B.3) sont vérifiées, alors on a*

$$\begin{aligned} & \sup_{x \in I} \pm \left(\frac{nh_n^{2k+1}(x)}{2 \log_{\theta, K} \frac{|I|}{H_n(x)}} \right)^{\frac{1}{2}} \Theta_n(x) \left(\hat{f}_{X,n}^{(k)}(x, H_n(x)) - \mathbb{E} \hat{f}_{X,n}^{(k)}(x, H_n(x)) \right) \\ & \xrightarrow{P} \left(\sup_{x \in I} \Theta^2(x) f_X(x) [(K^{(k)})^2] \right)^{\frac{1}{2}}. \end{aligned}$$

(ii) *Si de plus l'hypothèse (B.2) est vérifiée, alors*

$$\begin{aligned} & \left(\frac{nh_n^{2k+1}}{2 \log_{\theta, K} \frac{|I|}{h_n}} \right)^{\frac{1}{2}} \sup_{x \in I} \pm \Theta_n(x) \left(\hat{f}_{X,n}^{(k)}(x, H_n(x)) - \mathbb{E} \hat{f}_{X,n}^{(k)}(x, H_n(x)) \right) \\ & \xrightarrow{P} \left(\sup_{x \in I} \frac{\Theta^2(x) f_X(x)}{C(x)} [(K^{(k)})^2] \right)^{\frac{1}{2}}. \end{aligned}$$

Théorème 10. (i) *Sous les hypothèses (F.1), (F.2), (K'.1), (H.1), (H.4), (B.1) et (B.3), on a*

$$\begin{aligned} & \sup_{x \in I} \pm \left(\frac{nh_n^{2k+1}(x)}{2 \log_{\theta, K} \frac{|I|}{H_n(x)}} \right)^{\frac{1}{2}} \Theta_n(x) \left(\hat{r}_{\psi,n}^{(k)}(x, H_n(x)) - \mathbb{E} \hat{r}_{\psi,n}^{(k)}(x, H_n(x)) \right) \\ & \xrightarrow{P} \left(\sup_{x \in I} \frac{\Theta^2(x) \sigma_{\psi}^2(x)}{f_X(x)} [(K^{(k)})^2] \right)^{\frac{1}{2}}. \end{aligned}$$

(ii) *Sous l'hypothèse (B.2), on a*

$$\begin{aligned} & \left(\frac{nh_n^{2k+1}}{2 \log_{\theta, K} \frac{|I|}{h_n}} \right)^{\frac{1}{2}} \sup_{x \in I} \pm \Theta_n(x) \left(\hat{r}_{\psi,n}^{(k)}(x, H_n(x)) - \mathbb{E} \hat{r}_{\psi,n}^{(k)}(x, H_n(x)) \right) \\ & \xrightarrow{P} \left(\sup_{x \in I} \frac{\Theta^2(x) \sigma_{\psi}^2(x)}{C(x) f_X(x)} [(K^{(k)})^2] \right)^{\frac{1}{2}}, \end{aligned}$$

$$\text{où } \sigma_{\psi}^2(x) = \frac{1}{f(x)} \int (\psi(y) - r_{\psi}(x))^2 f_{X,Y}(x, y) dy.$$

Preuve

Deheuvels et Mason (2004) établissent les vitesses de convergence uniforme en probabilité pour la déviation stochastique des estimateurs à noyau de la densité et de la régression dépendant d'une fenêtre locale, en se basant sur le comportement asymptotique du processus empirique suivant

$$W_{n,\lambda}(x, \psi) = \sum_{i=1}^n \left[c(x)\psi(Y_i) + d(x) \right] K\left(\frac{x - X_i}{\lambda h_n}\right) - nE \left[(c(x)\psi(Y) + d(x)) K\left(\frac{x - X_i}{\lambda h_n}\right) \right],$$

où $c(\cdot)$ et $d(\cdot)$ sont deux fonctions bornées et continues sur J , $0 < \lambda \leq \infty$. Blondin (2004) procède d'une manière similaire, il utilise le processus

$$W_{n,k}(x, \psi) = \sum_{i=1}^n \left[c(x)\psi(Y_i) + d(x) \right] K^{(k)}\left(\frac{x - X_i}{h_n}\right) - nE \left[(c(x)\psi(Y) + d(x)) K^{(k)}\left(\frac{x - X_i}{h_n}\right) \right],$$

pour obtenir des lois uniformes des dérivées estimées à partir d'une fenêtre globale.

En combinant ces deux derniers travaux, on obtient des lois uniformes des dérivées dépendant d'une fenêtre locale en se basant sur le processus

$$W_{n,\lambda,k}(x, \psi) = \sum_{i=1}^n \left[c(x)\psi(Y_i) + d(x) \right] K^{(k)}\left(\frac{x - X_i}{\lambda h_n}\right) - nE \left[(c(x)\psi(Y) + d(x)) K^{(k)}\left(\frac{x - X_i}{\lambda h_n}\right) \right].$$

Le théorème 9 et le théorème 10 sont donc des résultats du théorème suivant

Théorème 11. *Sous les hypothèses (H1), (H.4), (K'.1), (K.4), (F.1), (F.2), (F.4) pour*

$0 < \lambda_1 \leq \lambda_2 \leq \infty$, quand $n \rightarrow \infty$, on a

$$\sup_{\lambda_1 \leq \lambda \leq \lambda_2} \left| \left(2n\lambda h_n \log \frac{1}{h_n} \right)^{-\frac{1}{2}} \sup_{x \in I} (\pm W_{n,\lambda,k}(x, \psi)) - \sigma_w(I) \right| = o_p(1),$$

avec

$$\sigma_w(I)^2 = \sup_{x \in I} E \left((c(x)\psi(y) + d(x))^2 / X = x \right) f_X(x) [(K^{(k)})^2].$$

Nous ne démontrons pas ce théorème car sa démonstration est assez technique, elle est essentiellement similaire à celle de Deheuvels et Mason (2004).

Remarque 6. Des lois uniformes ont été aussi obtenues dans le cadre des modèles additifs définis dans le chapitre deux, en utilisant les mêmes techniques de démonstration, Debbarih (2004) établit le résultat suivant

$$\sqrt{\frac{nh_n}{2|\log h_n|}} \sup_{x \in I} \pm(r_n(x) - r(x)) \xrightarrow{P} \sum_{\ell=1}^p \sigma_\ell,$$

où $\sigma_\ell = \sup_{x_\ell \in I_\ell} \sqrt{\frac{\phi(x_\ell)}{f_\ell(x_\ell)}} [K_\ell^2]$,

avec

– q_ℓ , ($\ell = 1, \dots, p$) sont des fonctions de densité bornées et continues et

$$q(x) = \prod_{\ell=1}^p q_\ell(x_\ell), \quad q_{-\ell}(x_{-\ell}) = \prod_{j \neq \ell} q_j(x_j).$$

– $\phi(x_\ell) = \int_{\mathbb{R}^{p-1}} \frac{\mathbb{E}(Y^2/X=x)}{f(x_{-\ell}/x_\ell)} q_{-\ell}(x_{-\ell}) dx_{-\ell}$, $x = (x_1, \dots, x_p)$, $x_{-\ell} = (x_1, \dots, x_{\ell-1}, x_{\ell+1}, \dots, x_p)$.

– f_ℓ est la densité marginale de X_ℓ .

– Le noyau produit $K(x) = \prod_{\ell=1}^p K_\ell(x_\ell)$.

– Le pavé $I = \prod_{\ell=1}^p I_\ell$.

Il a obtenu également les lois uniformes des composantes additives η_ℓ

$$\sqrt{\frac{nh_n}{2|\log h_n|}} \sup_{x_1 \in I_1} \pm(\hat{\eta}_1(x_1) - \eta_1(x_1)) \xrightarrow{P} \sum_{\ell=1}^p \sigma_\ell.$$

3.3 Application : Modèle de régression gaussien

Le théorème 9 permet de construire des bandes de confiance asymptotiques des dérivées d'ordre k de la densité. En effet, on a

$$\sup_{x \in I} \pm \frac{1}{G_n(x)} (\hat{f}_{X,n}^{(k)}(x, H_n(x)) - E(\hat{f}_{X,n}^{(k)}(x, H_n(x)))) \xrightarrow{P} 1, \quad (3.4)$$

où $G_n(x)$ est une fonction positive qui dépend des données.

Le biais de l'estimation vérifie

$$\sup_{x \in I} \frac{1}{G_n(x)} (E\hat{f}_{X,n}^{(k)}(x, H_n(x)) - f_X^{(k)}(x)) \xrightarrow{P} 0.$$

C'est à dire, pour $0 < \epsilon < 1$, lorsque $n \rightarrow \infty$, on a

$$\mathbb{P}\left(f^{(k)}(x) \in [\hat{f}_{X,n}^{(k)}(x, H_n(x)) - (1+\epsilon)G_n(x), \hat{f}_{X,n}^{(k)}(x, H_n(x)) + (1+\epsilon)G_n(x), \forall x \in I]\right) \rightarrow 1,$$

$$\mathbb{P}\left(f^{(k)}(x) \in [\hat{f}_{X,n}^{(k)}(x, H_n(x)) - (1-\epsilon)G_n(x), \hat{f}_{X,n}^{(k)}(x, H_n(x)) + (1-\epsilon)G_n(x), \forall x \in I]\right) \rightarrow 0.$$

Les bandes de confiance de la fonction $f^{(k)}(x)$ pour $\epsilon > 0$ sont donc données asymptotiquement par

$$[\hat{f}_{X,n}^{(k)}(x, H_n(x)) - G_n(x), \hat{f}_{X,n}^{(k)}(x, H_n(x)) + G_n(x)]. \quad (3.5)$$

D'une manière analogue, on obtient du théorème 10

$$\sup_{x \in I} \pm \frac{1}{T_n(x)} (\hat{r}_{X,n}^{(k)}(x, H_n(x)) - E(\hat{r}_{X,n}^{(k)}(x, H_n(x)))) \xrightarrow{P} 1,$$

avec

$$\sup_{x \in I} \frac{1}{T_n(x)} (E\hat{r}_{X,n}^{(k)}(x, H_n(x)) - r_X^{(k)}(x)) \xrightarrow{P} 0.$$

C'est à dire

$$\mathbb{P}\left(r^{(k)}(x) \in [\hat{r}_{X,n}^{(k)}(x, H_n(x)) - (1+\epsilon)T_n(x), \hat{r}_{X,n}^{(k)}(x, H_n(x)) + (1+\epsilon)T_n(x), \forall x \in I]\right) \rightarrow 1,$$

$$\mathbb{P}\left(r^{(k)}(x) \in [\hat{r}_{X,n}^{(k)}(x, H_n(x)) - (1-\epsilon)T_n(x), \hat{r}_{X,n}^{(k)}(x, H_n(x)) + (1-\epsilon)T_n(x), \forall x \in I]\right) \rightarrow 0.$$

Les bandes de confiance de la fonction $r^{(k)}(x)$ pour $\epsilon > 0$ sont données asymptotiquement par

$$[\hat{r}_{X,n}^{(k)}(x, H_n(x)) - T_n(x), \hat{r}_{X,n}^{(k)}(x, H_n(x)) + T_n(x)], \quad (3.6)$$

avec $T_n(x)$ est une fonction positive à déterminer.

Remarque 7. – Lorsque $k = 0$, $H_n(x) = h_n$, $C_n(x) = C(x) = 1$, $\Theta_n(x) = \Theta(x) = 1$, on obtient la version de la convergence presque sûre du théorème 9 et 10, établis par Einmahl et Mason(2000).

Le résultat du théorème 10 reste vrai presque sûrement si les hypothèses suivantes sont vérifiées

$$h_n \rightarrow 0, \quad nh_n \rightarrow \infty, \quad \frac{\ln(\frac{1}{h_n})}{\ln \ln n} \rightarrow \infty.$$

– Le facteur $\log_{\theta, K} \frac{|I|}{h_n}$ permet d'avoir des estimateurs invariants. En effet, on a pour $\tilde{K}(t) = \lambda K(\lambda t)$ et $\tilde{h} = \lambda h$,

$$\log_{\theta, K} \frac{|I|}{h} = \log_{\theta, \tilde{K}} \frac{|I|}{\tilde{h}}, \quad \log\left(\frac{|I|}{h}\right) \neq \log\left(\frac{|I|}{\tilde{h}}\right).$$

– Le choix de $\theta = 7$ est recommandé par Deheuvels et Mason (2004) car cela assure

$$(2 \log_{\theta, K}(u))^{\frac{1}{2}} \geq (2 \log 7)^{\frac{1}{2}} \simeq 1.97.$$

Les bandes de confiance simultanées obtenues à partir du théorème 9 données par

$$\mathbb{P}\left(f^{(k)}(x) \in \left[\hat{f}_{X,n}^{(k)}(x, h_n) - \sqrt{2 \log_{\theta, K} \frac{|I|}{h_n} \sup_{x \in I} \left(\frac{f(x)}{nh_n} \left([K^{(k)}]^2\right)^{\frac{1}{2}}\right)}, \right. \right. \\ \left. \left. \hat{f}_{X,n}^{(k)}(x, h_n) + \sqrt{2 \log_{\theta, K} \frac{|I|}{h_n} \sup_{x \in I} \left(\frac{f(x)}{nh_n} \left([K^{(k)}]^2\right)^{\frac{1}{2}}\right)}, \forall x \in I\right] \right) \approx 1,$$

incluent donc l'intervalle de confiance donné pour chaque $x_0 \in I$ par

$$\mathbb{P}\left(f^{(k)}(x_0) \in \left[\hat{f}_{X,n}^{(k)}(x_0, h_n) - 1.96 \times \left(\frac{f(x)}{nh_n} \left([K^{(k)}]^2\right)^{\frac{1}{2}}\right), \right. \right. \\ \left. \left. \hat{f}_{X,n}^{(k)}(x_0, h_n) + 1.96 \times \left(\frac{f(x)}{nh_n} \left([K^{(k)}]^2\right)^{\frac{1}{2}}\right)\right] \right) \approx 0.95.$$

D'autres choix de θ peuvent être traités ici, si on prend par exemple $\theta = 15$, on obtient

$$(2 \log_{\theta, K}(u))^{\frac{1}{2}} \geq (2 \log 7)^{\frac{1}{2}} \simeq 2.32725.$$

Dans ce cas, les bandes de confiance ci-dessus donnent des résultats comparables à l'intervalle de confiance à 99%, puisque le quantile correspondant de la loi normale standard $q = 2.32635$. Ce choix par contre n'est pas recommandé car il demande des tailles d'échantillon $n > 121839$.

3.3.1 Bandes de confiance de la densité et de la régression ($k = 0$)

On suppose que (X, Y) suit la loi normale d'espérance (μ_X, μ_Y) et de matrice de variance-covariance

$$\Gamma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix},$$

où

– ρ désigne le coefficient de corrélation entre X et Y .

– σ_X, σ_Y sont les écarts types de X et Y respectivement.

La fonction de régression est dans ce cas donnée par

$$r(x) = \mu_Y + \rho \frac{\sigma_X}{\sigma_Y} x, \quad (3.7)$$

avec la variance conditionnelle

$$\text{Var}(Y/X = x) = (1 - \rho^2)^2 \sigma_Y^2.$$

Lois uniformes de la densité :

D'après (2.10), l'erreur quadratique moyenne $\text{MSE}(f_n)$ atteint son minimum en

$$H_{n,1}(x) = n^{-\frac{1}{5}} \sigma_X \left(\frac{\sqrt{2\pi} [K^2] \exp\left(\frac{1}{2} \left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)}{[t^2 K]^2 \left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 1\right)^2} \right)^{\frac{1}{5}}.$$

Les valeurs de μ_X et σ_X sont inconnues, cette fenêtre est donc estimée par $\hat{H}_n(x)$ en injectant la moyenne empirique \bar{X} et la variance empirique $\hat{\sigma}_X$

$$\hat{H}_{n,1}(x) = n^{-\frac{1}{5}} \hat{\sigma}_X \left(\frac{\sqrt{2\pi} [K^2] \exp\left(\frac{1}{2} \left(\frac{x-\bar{X}}{\hat{\sigma}_X}\right)^2\right)}{[t^2 K]^2 \left(\left(\frac{x-\bar{X}}{\hat{\sigma}_X}\right)^2 - 1\right)^2} \right)^{\frac{1}{5}}. \quad (3.8)$$

où

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Les conditions (B.1), (B.2), et (B.3) sont vérifiées par $H_n(x) = \frac{H_{n,1}(x)}{\sigma_X}$, $\hat{H}_n(x) = \frac{\hat{H}_{n,1}(x)}{\hat{\sigma}_X}$, $\Theta_{n,1}(x)$, $\Theta_1(x)$ et $h_n = n^{-\frac{1}{5}}$ tels que

$$\Theta_{n,1}(x) = \sqrt{\frac{\hat{\sigma}_X}{[K^2]}} \left(\frac{\sqrt{(2\pi)^3} [K^2]^{\frac{1}{2}} \exp\left(\frac{3}{2} \left(\frac{x-\bar{X}}{\hat{\sigma}_X}\right)^2\right)}{[t^2 K] \left| \left(\frac{x-\bar{X}}{\hat{\sigma}_X}\right)^2 - 1 \right|} \right)^{\frac{1}{5}}, \quad (3.9)$$

$$\Theta_1(x) = \sqrt{\frac{\sigma_X}{[K^2]}} \left(\frac{\sqrt{(2\pi)^3} [K^2]^{\frac{1}{2}} \exp\left(\frac{3}{2} \left(\frac{x-\mu_X}{\sigma_X}\right)^2\right)}{[t^2 K] \left| \left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 1 \right|} \right)^{\frac{1}{5}}.$$

Le théorème 9 peut être appliqué donc pour $k = 0$, ce qui coïncide avec le résultat de Nemouchi et Mohdeb (2010).

Corollaire 2. (Nemouchi et Mohdeb (2010)).

Sous les mêmes hypothèses du théorème 9, on a

$$\left(\frac{n^{\frac{4}{5}} \hat{\sigma}_X}{2 \log_{\theta, K} \left(\frac{|I|}{n^{-\frac{1}{5}} \hat{\sigma}_X} \right)} \right)^{\frac{1}{2}} \sup_{x \in I} \pm \Theta_{n,1}(x) \left(\hat{f}_{X,n}(x, \hat{H}_{n,1}(x)) - f(x) \right) \xrightarrow{P} 1,$$

quand $n \rightarrow \infty$.

Lois uniformes de la régression

Dans ce cas, d'après (2.7), la fenêtre optimale au sens de l'erreur moyenne quadratique $MSE(r_n)$ est donnée par

$$H_{n,2}(x) = n^{-\frac{1}{5}} \sigma_X \left(\frac{\sqrt{2\pi} [K^2] (1 - \rho^2) \exp\left(\frac{1}{2} \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right)}{4\rho^2 x^2 [t^2 K]^2} \right)^{\frac{1}{5}}$$

son estimateur est donné par

$$\hat{H}_{n,2}(x) = n^{-\frac{1}{5}} \hat{\sigma}_X \left(\frac{\sqrt{2\pi} [K^2] (1 - \hat{\rho}^2) \exp\left(\frac{1}{2} \left(\frac{x - \bar{x}}{\hat{\sigma}_X}\right)^2\right)}{4\hat{\rho}^2 x^2 [t^2 K]^2} \right)^{\frac{1}{5}}, \quad (3.10)$$

tel que

$$\hat{\rho} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\hat{\sigma}_X \hat{\sigma}_Y},$$

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \hat{\sigma}_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2.$$

Soient les fonctions mesurables

$$\Theta_{n,2}(x) = \left(\frac{(1 - \hat{\rho}^2)^{-2} |x \hat{\rho}|^{-1} \exp\left(-\left(\frac{x - \bar{X}}{\hat{\sigma}_X}\right)^2\right)}{4\pi \hat{\sigma}_X^{3/2} \hat{\sigma}_Y^5 [t^2 K] [K]^2} \right)^{\frac{1}{5}}, \quad (3.11)$$

$$\Theta_2(x) = \left(\frac{(1 - \rho^2)^{-2} |x \rho|^{-1} \exp\left(-\left(\frac{x - \mu_X}{\sigma_X}\right)^2\right)}{4\pi \sigma_X^{3/2} \sigma_Y^5 [t^2 K] [K]^2} \right)^{\frac{1}{5}}.$$

Ces paramètres vérifient aussi les conditions (B.1), (B.2) et (B.3) donc le théorème 10 peut être appliqué pour $k = 0$.

Corollaire 3. (Nemouchi et Mohdeb (2010)).

Sous les mêmes hypothèses du théorème 10, on a

$$\left(\frac{n^{\frac{4}{5}} \hat{\sigma}_X}{2 \log_{\theta, K} \left(\frac{|I|}{n^{-\frac{1}{5}} \hat{\sigma}_X} \right)} \right)^{\frac{1}{2}} \sup_{x \in I} \pm \Theta_{n,2}(x) \left(\hat{r}_{X,n}(x, \hat{H}_{n,2}(x)) - r(x) \right) \xrightarrow{P} 1, \text{ quand } n \rightarrow \infty.$$

3.4 Simulation

Dans cette section, on a choisi de simuler les bandes de confiance de la densité et de la régression définies dans (3.5) et (3.6) afin de valider les résultats théoriques ci-dessus.

Les fonctions $G_n(x)$, $T_n(x)$ sont calculées à partir des données simulées du modèle de régression gaussien introduit dans (3.7).

Soit (X_i, Y_i) , $(1 \leq i \leq n)$ un n -échantillon de loi normale d'espérance inconnue (μ_X, μ_Y) estimée par la moyenne empirique (\bar{X}, \bar{Y}) et de matrice de variance-covariance estimée par $\hat{\Gamma}$ donné par

$$\hat{\Gamma} = \begin{pmatrix} \hat{\sigma}_X^2 & \hat{\rho}\hat{\sigma}_X\hat{\sigma}_Y \\ \hat{\rho}\hat{\sigma}_X\hat{\sigma}_Y & \hat{\sigma}_Y^2 \end{pmatrix},$$

où

- $\hat{\rho}$ est le coefficient de corrélation estimé.
- $\hat{\sigma}_X$, $\hat{\sigma}_Y$ sont les écarts types empiriques.

On représente d'abord l'estimateur de la densité basé sur la fenêtre locale $H_{n,1}(x)$ estimée par $\hat{H}_{n,1}(x)$, et le noyau d'Epanechnikov, pour un échantillon de taille $n = 10$, $n = 50$, $n = 200$, et $n = 500$.

La fonction $G_n(x)$ est dans ce cas de la forme

$$G_n(x) = \frac{1}{\theta_{n,1}(x)} \sqrt{\frac{2 \log_{\theta,K} \left(\frac{1}{n^{-\frac{1}{5}} \hat{\sigma}_X} \right)}{n^{\frac{4}{5}} \hat{\sigma}_X}}.$$

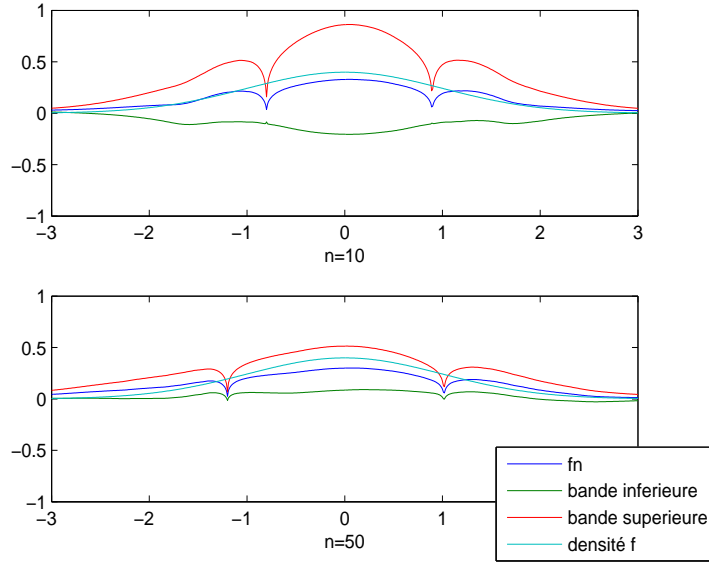
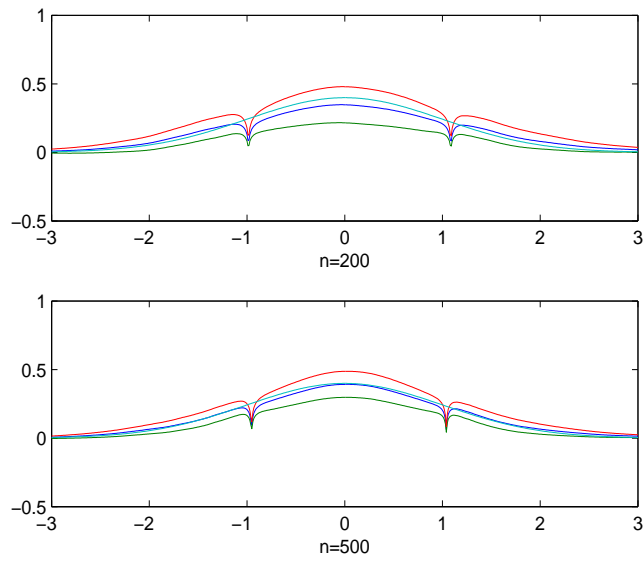
$\theta_{n,1}(x)$ est défini dans (3.9).

Les bandes de confiance de la densité définies dans (3.5) sont représentées dans Figure (3.1) et Figure (3.2), elles contiennent $f(x)$ même pour des tailles d'échantillon modérées.

Ces bandes s'approchent de plus en plus de la vraie densité en augmentant la taille d'échantillon.

Dans le cas de la régression, les bandes de confiance sont définies dans (3.6) à l'aide de la fonction $T_n(x)$ donnée par

$$T_n(x) = \frac{1}{\theta_{n,2}(x)} \sqrt{\frac{2 \log_{\theta,K} \left(\frac{1}{n^{-\frac{1}{5}} \hat{\sigma}_X} \right)}{n^{\frac{4}{5}} \hat{\sigma}_X}},$$

FIGURE 3.1 – Bandes de confiance de la densité $n=10$, $n=50$.FIGURE 3.2 – Bandes de confiance de la densité $n=200$, $n=500$.

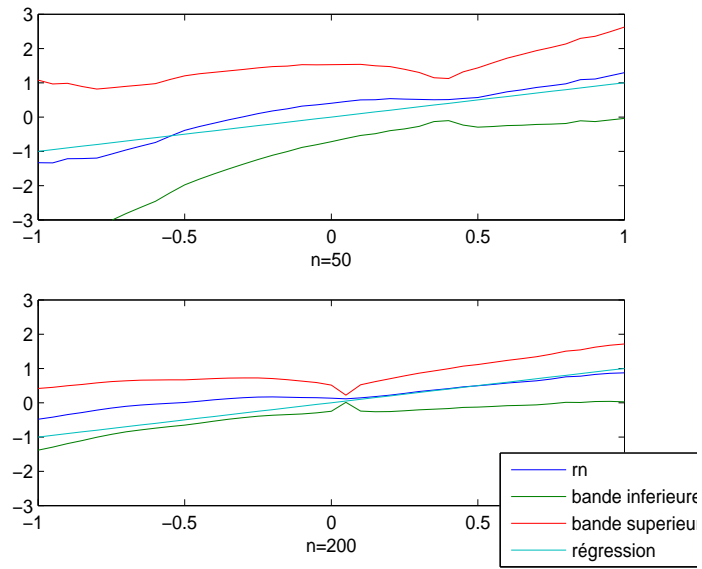


FIGURE 3.3 – Bandes de confiance de la régression.

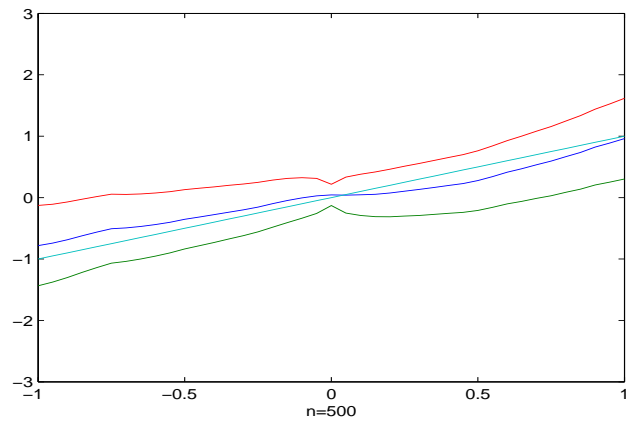


FIGURE 3.4 – Bandes de confiance de la régression.

avec $\theta_{n,2}(x)$ défini dans (3.11).

Les résultats de la simulation sont performants à partir de $n = 50$ dans le cas de la régression, les bandes de confiance (Figure (3.3), Figure (3.4)) contiennent la vraie régression, il est clair qu'elles s'approchent aussi de la vraie régression en augmentant la taille de l'échantillon.

Deuxième partie

Cas des variables dépendantes

Chapitre 4

Kernel estimation under weak dependence

4.1 Introduction

The case of the i.i.d. observations is the most detailed in nonparametric estimation, but in some applications where data are collected over time, the assumption of independence is often strong to be satisfied. Kernel estimator has been studied under different assumptions of dependence, this notion has been formalized in many ways, from the first papers of Rosenblatt under mixing until Doukhan and Louhichi works under a new and a more generalized approach of weak dependence.

Given a stationary weak dependent sequence $(Z_t)_{t \in \mathbb{Z}}$ with $Z_t = (X_t, Y_t)$, where $X_t, Y_t \in \mathbb{R}$, we will investigate the regression function $r(x) = \mathbb{E}(Y/X = x)$ and the marginal density $f(x)$ of X_t .

The properties of their kernel estimators are studied under η -weak dependence hypothesis.

Various mixing conditions are used to define the dependence, such as α -mixing, β -mixing and ϕ -mixing, these conditions are often difficult to check. We note also that, with some dependent processes, mixing conditions are not satisfied.

In fact the process $X_n = \sum_{k=0}^n 2^{-(k+1)} \epsilon_{n-k}$ does not satisfy mixing condition, the i.i.d. innovations ϵ_j , ($j \leq n$) are the digits of the binary expansion of X_n , note that X_n is a deterministic function of X_k for $k \leq n$, so the event $A = \{X_k \leq \frac{1}{2}\}$ belongs to the σ -algebra $\sigma(X_t, t \leq k)$ and the σ -algebra

$\sigma(X_t, t \geq n)$, in this case, the coefficient of α -mixing does not decrease to 0, in fact

$$\alpha(n) \geq |\mathbb{P}(A \cap A) - \mathbb{P}(A)\mathbb{P}(A)| = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}.$$

In this spirit, Doukhan and Louhichi (1999) were the first to generalize mixing, they introduced a concept of dependence based on covariance between Lipschitz functions of past and future. The literature is then enriched by probabilistic tools similar to those of the independence such as Rosenthal and Bernstein inequalities.

Later, they studied the almost sure convergence and the normality of kernel density and regression estimators, (see Doukhan and Louhichi (2001) and Ango Nze and al. (2002)).

In this chapter, we study the kernel estimator of the marginal density of the one η -weak dependent sequence $(X_t)_{t \in \mathbb{Z}}$ and also the regression in a two dimensional η -weak dependent model $Z_t = (X_t, Y_t)$. Similarly to independent case, the convergence and the normality are asserted.

4.2 Definition and examples

The sequence $(X_t)_{t \in \mathbb{Z}}$ is called η -weakly dependent, if it exists a decreasing sequence $(\eta_k)_{k \in \mathbb{N}} \rightarrow 0$ such that

$$|\text{Cov}(g_1(X_{s_1}, \dots, X_{s_u}), g_2(X_{t_1}, \dots, X_{t_v}))| \leq (u\|g_2\|_\infty \text{Lip}(g_1) + v\|g_1\|_\infty \text{Lip}(g_2))\eta_k, \quad (2.1)$$

with

* $s_1 \leq \dots \leq s_u \leq s_u + k \leq t_1 \leq \dots \leq t_v$ where $u, v \in \mathbb{N}^*$, $(s_1, \dots, s_u) \in \mathbb{Z}^u$ and $(t_1, \dots, t_v) \in \mathbb{Z}^v$.

* $g_1 : \mathbb{R}^u \rightarrow \mathbb{R}$, $g_2 : \mathbb{R}^v \rightarrow \mathbb{R}$ are two bounded functions with $\text{Lip}(g_l) < \infty$,

where $\text{Lip}(g_l) = \sup_{x \neq y} \frac{|g_l(x) - g_l(y)|}{\|x - y\|_1}$, $l = 1, 2$ and $\|x\|_1 = \sum_{i=1}^u |x_i|$.

Remark 1. – If stationary real valued process $(X_t)_{t \in \mathbb{Z}}$ is η -weakly dependent, then for Lipschitz and continuous real function F satisfying $\|F\|_\infty = M < \infty$ and $\text{Lip}F \leq 1$, the process $Y_t = F(X_t)$ is also stationary and η -weakly dependent.

– If $\text{Lip}F > 1$, the process is still weakly dependent with coefficients $\eta_r^Y = \text{Lip}F \eta_r$.

(See Doukhan and Neumann (2007)).

Examples

Let us discuss some examples of such weak dependent sequences known in the literature. (We refer the reader to Dedecker and al. (2007) for more details).

4.2.1 Bernoulli shift with independent inputs

Let $(\epsilon_i)_{i \in \mathbb{Z}}$ be a sequence of independent real valued r.v.s, and H be a measurable function defined on $\mathbb{R}^{\mathbb{Z}}$.

The sequence $(X_i)_{i \in \mathbb{Z}}$ defined by

$$X_i = H(\epsilon_{i-j}, j \in \mathbb{Z}),$$

is called a Bernoulli shift with independent inputs.

Remark 2. – For any positive integer p , define the sequence

$$\delta_p = \sup_{i \in \mathbb{Z}} \mathbb{E} |H(\epsilon_{i-j}, j \in \mathbb{Z}) - H(\epsilon_{i-j} \mathbb{1}_{|i| < p}, j \in \mathbb{Z})|.$$

If the function H satisfies the condition

$$|H(u_i, i \in \mathbb{Z}) - H(v_i, i \in \mathbb{Z})| \leq \sum_{i \in \mathbb{Z}} a_i |u_i - v_i|^b,$$

for some positive $(a_i)_{i \in \mathbb{Z}}$ satisfying $\sum_{i \in \mathbb{Z}} a_i < \infty$, $0 < b \leq 1$, and if $\mathbb{E}(|\epsilon_0|^p)$ exists then

$$\lim_{p \rightarrow \infty} \delta_p \leq \lim_{p \rightarrow \infty} \sum_{i \geq p} a_i \mathbb{E} |\epsilon_i|^b = 0.$$

– The function H depends on an infinite number of arguments, for that, it is often given in form of series, we consider for example $X_n = \sum_{i \in \mathbb{Z}} a_i \epsilon_{n-i}$, with real coefficients $(a_i)_{i \in \mathbb{Z}}$.

– The behaviour of the coefficient η_k depends on the coefficients a_i . If $a_i = O(i^{-A})$, with $A > 5$ and under some conditions of the moment of the innovations, one obtains $\eta_k = O(k^{1-A})$.

The geometric decay of the coefficient of η -dependence is also treated, (see Bardet and al.(2008)).

Proposition 1. (Doukhan and Louhichi (1999)).

The Bernoulli shift $X_i = H(\epsilon_{i-j}, j \in \mathbb{Z})$, with independent innovation process $(\epsilon_i)_{i \in \mathbb{Z}}$ is η -weakly dependent.

Proof

We define the truncated process $X_n^{(p)} = H(\epsilon_{n-i} \mathbb{1}_{|i| < p})$.

For two functions $g_1 : \mathbb{R}^u \rightarrow \mathbb{R}$, $g_2 : \mathbb{R}^v \rightarrow \mathbb{R}$, ($u, v \in \mathbb{N}^*$) satisfying

$$\|g_l\|_\infty < \infty, \text{Lip}(g_l) < \infty, \quad (l = 1, 2)$$

with

$$\begin{aligned} g_1 &= g_1(X_{s_1}, \dots, X_{s_u}), \quad g_2 = g_2(X_{t_1}, \dots, X_{t_v}), \quad t_1 - s_u \geq k, \\ g_1^{(p)} &= g_1(X_{s_1}^{(p)}, \dots, X_{s_u}^{(p)}), \quad g_2^{(p)} = g_2(X_{t_1}^{(p)}, \dots, X_{t_v}^{(p)}). \end{aligned}$$

We note that $(X_{s_1}^{(p)}, \dots, X_{s_u}^{(p)})$ and $(X_{t_1}^{(p)}, \dots, X_{t_v}^{(p)})$ are independent if $k > 2p$. Using Hölder inequality, and the fact that g_1 and g_2 are Lipschitzian, we get

$$\begin{aligned} |\text{Cov}(g_1, g_2)| &\leq |\text{Cov}(g_1 - g_1^{(p)}, g_2)| + |\text{Cov}(g_1^{(p)}, g_2 - g_2^{(p)})| \\ &\leq 2\|g_2\|_\infty \mathbb{E}|g_1 - g_1^{(p)}| + 2\|g_1\|_\infty \mathbb{E}|g_2 - g_2^{(p)}| \\ &\leq 2(\|g_2\|_\infty \text{Lip}g_1 \sum_{t=1}^u \mathbb{E}|X_{i_t} - X_{i_t}^{(p)}| + \|g_1\|_\infty \text{Lip}g_2 \sum_{t=1}^v \mathbb{E}|X_{j_t} - X_{j_t}^{(p)}|). \end{aligned}$$

This leads to

$$\text{Cov}(g_1, g_2) \leq 2(u\|g_2\|_\infty \text{Lip}g_1 + v\|g_1\|_\infty \text{Lip}g_2)\delta_p.$$

If we choose $\eta_k = 2\delta_{\lfloor \frac{k}{2} \rfloor}$, we get the result \square .

4.2.2 Bernoulli shift with dependent inputs

Let $(\epsilon_i)_{i \in \mathbb{Z}}$ be a sequence of η -dependent real valued r.v.s. with coefficient $\eta_{\epsilon, r}$ and H be a measurable function defined on $\mathbb{R}^{\mathbb{Z}}$.

The sequence $(X_i)_{i \in \mathbb{Z}}$ defined by

$$X_i = H(\epsilon_{i-j}, j \in \mathbb{Z}),$$

is called a Bernoulli shift with dependent innovations.

Proposition 2. (*Doukhan and Winterberger (2006)*).

Bernoulli shift with dependent innovations is η -weakly dependent with coefficient η_k satisfying

$$\eta_k \leq \inf_{2r \leq k} \left(2 \sum_{|i| \geq r} b_i \|\epsilon_0\|_1 + (2r + 1)L\eta_{\epsilon, k-2r} \right).$$

4.2.3 The bilinear process

Let X_i be the bilinear process

$$X_i = \epsilon_i \left(a + \sum_{j \geq 1} a_j X_{i-j} \right) + b + \sum_{j \geq 1} b_j X_{i-j}, \quad (2.2)$$

with $(a_j)_{j \in \mathbb{N}^*}$ and $(b_j)_{j \in \mathbb{N}^*}$ are two sequences of real numbers, $a, b \in \mathbb{R}$, and $(\epsilon_i)_{i \in \mathbb{Z}}$ is a sequence of i.i.d random variables satisfying $(\mathbb{E}(|\epsilon_0|^p))^{\frac{1}{p}} < \infty$.

Proposition 3. (*Ragache N. and Wintenberger (2006)*).

The bilinear process (2.2) has a strictly stationary solution in L^p which is η -dependent if

$$\lambda = \|\epsilon_0\|_p \left(\sum_{j \geq 1} a_j + \sum_{j \geq 1} b_j \right) < 1.$$

Moreover if there exists an integer N such that $a_i = b_i = 0$, $i \geq N$, then the coefficient of dependence $\eta_k = O(\exp(-\lambda k))$, for some $\lambda > 0$.

4.2.4 Chaotic Volterra models

The Volterra process is defined by

$$X_t = v_0 + \sum_{k=1}^{\infty} V_{k;t}, \quad \text{where, } V_{k;t} = \sum_{i_1 < \dots < i_k} a_{k;i_1 \dots i_k} \epsilon_{t-i_1} \dots \epsilon_{t-i_k}, \quad (2.3)$$

v_0 is a constant, $(a_{k;i_1 \dots i_k})_{(i_1 \dots i_k) \in \mathbb{Z}^k}$ are real numbers for each $k \geq 1$, the sequence $(\epsilon_t)_{t \in \mathbb{Z}}$ is independent and identically distributed.

Proposition 4. (*Doukhan and Neumann (2007)*).

If $\mathbb{E}|\epsilon_0|^p < \infty$, $p \geq 1$ and $\sum_{k \geq 1} \sum_{i_1 < \dots < i_k} |a_{k;i_1 \dots i_k}|^p < \infty$, then the expression (2.3) converges in L^p and the process is η -weakly dependent.

4.3 Properties of kernel density and regression estimators under η -dependence

We introduce first the following assumptions.

- (H.1) $\exists \delta \in]0, 1[$, $\lim_{n \rightarrow \infty} n^\delta h_n = \infty$ as $n \rightarrow \infty$.
- (H.2) $\lim_{n \rightarrow \infty} \frac{nh_n}{\ln n} = \infty$.

- (K.1) $\int x^i K(x) dx = 0, i = 1, \dots, k-1, \int x^k K(x) dx \neq 0.$
 - (K.2) K is bounded, integrable, with compact support $I \subset \mathbb{R}.$
 - (K.3) $[K^2] = \int K^2(x) dx < \infty.$
 - (F.1) For $k = a + b, a \in \mathbb{N}, 0 \leq b < 1, \exists A > 0$ such as f is a times continuously derivable with $|f^{(a)}(x) - f^{(a)}(y)| \leq A|x-y|^b, \forall x, y \in I.$
 - (F.2) $f_{0,k}$ Joint densities of (X_0, X_k) are uniformly continuous.
 - (F.3) $\|f_{0,k}\|_\infty \leq C, \|f\|_\infty \leq C.$
 - (R.1) the function $r(\cdot) = \mathbb{E}(Y_0/X_0 = \cdot)$ exists and is continuous.
- For some $p \geq 1$ and all $m \leq p, m \in \mathbb{N},$
 $g_m(\cdot) = f(\cdot)\mathbb{E}(|Y_0|^m/X_0 = \cdot)$ exist and are continuous.
- (R.2) $\mathbb{E}(\exp(|Y_0|)) < \infty,$ or $\mathbb{E}(|Y_0|^s) < \infty,$ for some $s.$
 - (R.3) $\sup \|f_{0,k}\|_\infty \leq C,$ and $r(x, x') = \mathbb{E}(|Y_0 Y_k|/X_0 = x, X_k = x')$ are continuous, uniformly over all $k \in \mathbb{N}.$

4.3.1 Convergence in mean square error

To study the bias and the variance of the kernel estimator of the regression, we will need the next lemma which gives higher order moments of $g_n.$

Lemma 1. (*Ango NZE and al.(2002)*).

Let $(Z_t)_{t \in \mathbb{Z}}$ be an η -weak dependent stationary sequence satisfying assumptions (R.1) with $p = 2,$ (R.2), (R.3). If (H.1) and $\eta_r = O(r^{-a}), a > \max(q - 1, \frac{4+2\delta(q-1)}{(q-2)+\delta(4-q)}),$ for some integer $q \geq 3,$ hold then

$$\overline{\lim}_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} (nh)^{\frac{q}{2}} |\mathbb{E}(g_n(x) - \mathbb{E}g_n(x))^q| < \infty.$$

Variance

Theorem 1. (*Ango NZE and al.(2002)*).

Suppose that the stationary and the η - weakly dependent sequence $(Z_t)_{t \in \mathbb{Z}}$ satisfies condition (R.1) with $p = 2$ (R.2) and (R.3). Consider a positive kernel K satisfying (K.3). Let f, g satisfy (F.1) for some $k \in]0, 2]$ and (F.3). Then uniformly in x belonging to any compact subset of $\mathbb{R}.$

(i) If $\eta_r = O(r^{-a})$, with $a > 3$ then

$$\begin{aligned}\text{Var}(g_n(x)) &= \frac{1}{nh_n} g_2(x) \int K^2(x) dx + o\left(\frac{1}{nh_n}\right). \\ \text{Var}(f_n(x)) &= \frac{1}{nh_n} f(x) \int K^2(x) dx + o\left(\frac{1}{nh_n}\right).\end{aligned}$$

(ii) If $\eta_r = O(r^{-a})$, $a > \max(5, \frac{30\delta}{7-5\delta})$ then,

$$\text{Var}(r_n(x)) = \frac{G_2(x)}{nh_n f^2(x)} \int K^2(x) dx + o\left(\frac{1}{nh_n}\right), \text{ as } n \rightarrow \infty.$$

with $G_2(x) = f(x)\text{Var}(Y_0/X_0 = x)$.

Proof of Theorem 1. Consider $\tilde{g}(x)$ the truncated version of $g_n(x)$

$$\tilde{g}(x) = \frac{1}{nh_n} \sum_{i=1}^n T_i(x),$$

with

$$T_t(x) = Y_t \mathbb{I}_{\{|Y_t| \leq M(n)\}} K\left(\frac{x - X_t}{h_n}\right), \quad t = 1, \dots, n. \quad (3.4)$$

Under assumption (R.2), we have

$$\begin{aligned}nh_n \mathbb{E}(g_n - \tilde{g} - \mathbb{E}(g_n - \tilde{g}))^2(x) \\ \leq \frac{n^2}{nh_n} \mathbb{E}(Y_0^2 \mathbb{I}_{\{|Y_0| > M(n)\}} K^2\left(\frac{x - X_0}{h_n}\right)) \\ \leq CM^{2-s}(n)nh_n^{-1},\end{aligned}$$

if $M(n) = M_0 n^\gamma$, $M_0 > 0$, $\gamma \geq \frac{1+\delta}{s-2}$. we get

$$nh_n \mathbb{E}(g_n - \tilde{g} - \mathbb{E}(g_n - \tilde{g}))^2(x) \rightarrow_{n \rightarrow \infty} 0. \quad (3.5)$$

Let's calculate $\text{Var}(\tilde{g}(x))$

$$\text{Var}(\tilde{g}(x)) = \frac{1}{nh_n^2} \text{Var}(T_0(x)) + \frac{2}{(nh_n)^2} \sum_{r=1}^{n-1} (n-r) \text{Cov}(T_0, T_r).$$

According to [35]

$$\text{Var}(T_0(x)) = h_n g_2(x) \int K^2(x) dx + o(h_n),$$

On one hand, using assumption (F.3)

$$|\text{Cov}(T_0(x), T_r(x))| \leq Ch_n^2. \quad (3.6)$$

On other hand, the weak dependence leads to

$$|\text{Cov}(T_0(x), T_r(x))| \leq CM^2(n) \frac{1}{h_n} \eta_r. \quad (3.7)$$

Now, we have

$$\sum_{r=1}^{n-1} (n-r) \text{Cov}(T_0, T_r) = \sum_{r=1}^{\lfloor n^\xi \rfloor} (n-r) \text{Cov}(T_0, T_r) + \sum_{r=1+\lfloor n^\xi \rfloor}^{n-1} (n-r) \text{Cov}(T_0, T_r).$$

From [44], if $a > \frac{3s-4}{s-2} + \frac{2}{\delta(s-2)}$ then $\exists \xi \in [0, 1]$, $\frac{2\delta+2\gamma}{a-1} < \xi < \delta$, this gives from (3.6)

$$\frac{2}{nh_n} \sum_{r=1}^{\lfloor n^\xi \rfloor} (n-r) \text{Cov}(T_0, T_r) \leq Cn^{\xi-\delta}.$$

For the second term, using (3.7), we get

$$\begin{aligned} & \frac{2}{nh_n} \sum_{r=1+\lfloor n^\xi \rfloor}^{n-1} (n-r) \text{Cov}(T_0, T_r) \\ & \leq \frac{C}{h_n^2} M^2(n) \sum_{r>\lfloor n^\xi \rfloor} \eta_r \\ & \leq Cn^{\xi(1-a)+2\delta+2\gamma} \rightarrow 0, \text{ as } n \rightarrow \infty. \end{aligned}$$

The part (i) is then proved, one can choose $Y_i = 1$ to deduce $\text{Var}(f_n)$.

For the variance of regression estimator in part (iii), using the expansion (3.11) with $p = 1$, $u = f_n(x)$, $u_0 = \mathbb{E}(f_n(x))$, one obtain

$$\begin{aligned} & r_n(x) - \mathbb{E}(r_n(x)) \\ & = \frac{g_n(x)\mathbb{E}(f_n(x)) - f_n(x)\mathbb{E}(g_n(x))}{(\mathbb{E}(f_n(x)))^2} \\ & + \frac{(\mathbb{E}(g_n(x)) - g_n(x))(f_n(x) - \mathbb{E}(f_n(x))) + \mathbb{E}((g_n(x) - \mathbb{E}(g_n(x)))(f_n(x) - \mathbb{E}(f_n(x))))}{(\mathbb{E}(f_n(x)))^2} \\ & + \frac{r_n(x)(f_n(x) - \mathbb{E}(f_n(x)))^2 - \mathbb{E}(r_n(x)(f_n(x) - \mathbb{E}(f_n(x))))^2}{(\mathbb{E}(f_n(x)))^2}, \end{aligned}$$

then

$$\text{Var}(r_n(x)) = \frac{\mathbb{E}(g_n(x)\mathbb{E}(f_n(x)) - f_n(x)\mathbb{E}(g_n(x)))^2}{(\mathbb{E}(f_n(x)))^4} + o\left(\frac{1}{nh_n}\right). \quad (3.8)$$

In fact, because of weak dependence with $\eta_r = O(r^{-a})$ and $a > \max(3, \frac{27\delta}{5-2\delta}, \frac{9\delta}{1+2\delta})$, Lemma1 leads to

$$\begin{aligned} nh_n \mathbb{E}|(g_n(x) - \mathbb{E}(g_n(x)))|^3 &\leq ((nh)^{\frac{4}{3}} \mathbb{E}|(g_n(x) - \mathbb{E}(g_n(x)))|^4)^{\frac{3}{4}} \rightarrow 0, \\ nh_n \mathbb{E}|r_n(x)(g_n(x) - \mathbb{E}(g_n(x)))|^3 &\leq nh(\mathbb{E}r_n(x)^4)^{\frac{1}{4}} \mathbb{E}|(g_n(x) - \mathbb{E}(g_n(x)))|^4)^{\frac{3}{4}} \rightarrow 0, \end{aligned}$$

if $a > \max(1, \frac{6\delta}{1+\delta})$ then

$$\lim_{n \rightarrow \infty} nh_n (\text{Var}(g_n(x)))^2 = 0,$$

now if $a > \max(5, \frac{30\delta}{7-5\delta})$, we get

$$\lim_{n \rightarrow \infty} nh_n (\mathbb{E}(f_n(x) - \mathbb{E}(f_n(x)))^6)^{\frac{1}{2}} (\mathbb{E}(g_n(x) - \mathbb{E}(g_n(x)))^6)^{\frac{1}{6}} (\mathbb{E}|r_n(x)|^3)^{\frac{1}{3}} = 0.$$

Using (3.8), and following the same steps as in part (i), we get the result in (iii).

Bias

The study of the bias of the estimators f_n and g_n defined in (1.2) is independent of the structure of the weak dependence, then the classic results hold. In fact, assuming that f, g satisfy (F.1), using a kernel function K satisfying (K.1), we obtain

$$\mathbb{E}(f_n(x)) - f(x) \sim \frac{\int s^k K(s) ds}{k!} f^{(k)}(x) h_n^k = O(h_n^k). \quad (3.9)$$

$$\mathbb{E}(\hat{g}_n(x)) - g(x) \sim \frac{\int s^k K(s) ds}{k!} g^{(k)}(x) h_n^k = O(h_n^k). \quad (3.10)$$

Now, we give the bias of the kernel estimator of the regression.

Proposition 5. (*Ango NZE and al.(2002)*).

Suppose that the stationary and the η - weakly dependent sequence $(Z_t)_{t \in \mathbb{Z}}$ with $\eta_r = O(r^{-a})$, ($a > \max(3, 9\delta)$) satisfies conditions (R.1) with $p = 2$, (R.2) and (R.3). Consider a positive kernel K . Let f, g satisfy (F.1) for

some $k \in]0, 2]$, assume (H.1) satisfied.

Then uniformly in x belonging to any compact subset of \mathbb{R} ,

$$\mathbb{E}r_n(x) = r(x) + O\left(h_n^k + \frac{1}{nh_n}\right).$$

Proof of Proposition5

The expansion of Colomb with $p = 2$, $u = b_n$, $u_0 = \mathbb{E}b_n = 1$, gives

$$u^{-1} = \sum_{i=0}^p (-1)^i \frac{(u - u_0)^i}{u_0^{i+1}} + (-1)^{p+1} \frac{(u - u_0)^{p+1}}{u_0^{p+1}}, \quad (3.11)$$

this leads to

$$\begin{aligned} \mathbb{E}r_n(x) &= \mathbb{E}a_n - \mathbb{E}((a_n - \mathbb{E}a_n)(b_n - \mathbb{E}b_n)) + (\mathbb{E}a_n)\mathbb{E}(b_n - \mathbb{E}b_n)^2 + \\ &\mathbb{E}((a_n - \mathbb{E}a_n)(b_n - \mathbb{E}b_n)^2) - \mathbb{E}(r_n(x)(b_n - \mathbb{E}b_n)^3), \end{aligned}$$

$$\text{with } a_n = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}K\left(\frac{x-X_0}{h_n}\right)}, \text{ and } b_n = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}K\left(\frac{x-X_0}{h_n}\right)}.$$

From (3.9), we get

$$\mathbb{E}(a_n) = r(x) + O(h_n^k).$$

Now the expressions of the variance in Theorem1, and Cauchy Schwartz inequality lead to

$$\mathbb{E}(a_n - \mathbb{E}(a_n))(b_n - \mathbb{E}(b_n)) = o\left(\frac{1}{nh_n}\right).$$

The same argument gives

$$\mathbb{E}(a_n)\mathbb{E}(b_n - \mathbb{E}(b_n))^2 = o\left(\frac{1}{nh_n}\right).$$

Now using Cauchy Schwartz inequality again with lemma1, we get

$$\begin{aligned} &\mathbb{E}((a_n - \mathbb{E}a_n)(b_n - \mathbb{E}b_n)^2) \\ &\leq \left(\frac{h_n}{\mathbb{E}K\left(\frac{x-X_i}{h_n}\right)} \right)^3 (\mathbb{E}(f_n(x) - \mathbb{E}f_n(x))^4)^{\frac{1}{2}} (\mathbb{E}(g_n(x) - \mathbb{E}g_n(x))^2)^{\frac{1}{2}} = o((nh_n)^{-\frac{3}{2}}). \end{aligned}$$

For the last term, since

$$(\mathbb{E}(r_n(x))^4)^{\frac{1}{4}} \leq M_n + \mathbb{E} \left(\frac{\sum_{i=1}^n Y_i^4 I_{\{Y_i > M_n\}} K \left(\frac{x-X_i}{h_n} \right)}{\sum_{i=1}^n K \left(\frac{x-X_i}{h_n} \right)} \right) = O(\ln n + n^{2-\frac{M_0}{8}}).$$

We get in consequence by Cauchy Schwartz inequality

$$\mathbb{E}(r_n(x)(b_n - \mathbb{E}b_n)^3) = O(\ln n(nh_n)^{-\frac{3}{2}}).$$

□.

From the results above, one can deduce the approximations of $\text{MISE}(f_n)$ and $\text{MISE}(r_n)$. In fact, we have the following

$$\text{AMISE}(f_n) = \frac{1}{nh_n} [K^2] + \frac{h_n^{2k}}{(k!)^2} [u^k K]^2 \int (f^{(k)}(u))^2 du.$$

$$\text{AMISE}(r_n) = \frac{G_2(x)[K^2]}{nh_n f^2(x)} + \frac{h_n^{2k}}{k!^2} \int \left(\left(r^{(k)}(x) + kr^{(k-1)}(x) \frac{f'(x)}{f(x)} \right) [u^k K] \right)^2 dx.$$

with $G_2(x) = f(x)\text{Var}(Y_0/X_0 = x)$.

4.3.2 Asymptotic normality

Central limit theorems can be established using either Bernstein blocking technique or Lindberg's method. In this part, we focus on the latter.

Lindberg method for weakly dependent variables

Let X_1, \dots, X_n be a zero mean real random variables with $2 + \delta$ order finite moments for some $\delta > 0$, we consider independent random variables Y_1, \dots, Y_n independent of the variables $(X_i)_{1 \leq i \leq n}$ and such that $Y_i \sim \mathcal{N}(0, \text{Cov}(X_i))$.

Lindberg method consists to prove the following

$$\Delta_n = |\mathbb{E}(f(X_1 + \dots + X_n) - f(Y_1 + \dots + Y_n))| \rightarrow 0, \text{ as } n \rightarrow \infty,$$

with f is a bounded real function with bounded and continuous derivatives up to order 3.

Next lemma provides sufficient condition for establishing the CLT, by considering the particular and simpler choice $f(x) = \exp(itx)$, for some $t \in \mathbb{R}$.

Lemma 2. (*Bardet and al.(2008)*).

Let $(X_k)_{k \in \mathbb{N}^*}$ be a zero mean sequence with $2 + \delta$ order finite moments for some $0 < \delta < 1$, for the special case of complex exponential function $f(x) = \exp(itx)$, for some $t \in \mathbb{R}$,

$$\Delta_n \leq T(n) + 6|t|^{2+\delta} A_n,$$

with

$$T(n) = \sum_{j=1}^n |\text{Cov}(\exp\{it(X_1 + \dots + X_{j-1})\}, \exp\{itX_j\})|,$$

$$A_n = \sum_{k=1}^n \mathbb{E}(X_k^{2+\delta}).$$

Now we study the asymptotic behaviour of the sequence $S_n = \sum_{i=1}^n X_i$.

Theorem 2. (*See Bardet and al.(2008)*)

Assume that the sequence $(X_i)_{i \in \mathbb{N}}$ satisfies the assumptions of Lemma 2, and the following conditions

$$\exists \Sigma > 0, \text{ such that } \sum_{i=1}^n \text{Cov}(X_i) \rightarrow \Sigma, \text{ as } n \rightarrow \infty.$$

$$A_n \rightarrow 0, \text{ as } n \rightarrow \infty.$$

$$T(n) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Then

$$S_n = \sum_{i=1}^n X_i \sim \mathcal{N}(0, \Sigma), \text{ as } n \rightarrow \infty.$$

Proof of Theorem2

The result is obtained directly by applying Lemma 2, we obtain then

$$\lim_{n \rightarrow \infty} |\mathbb{E}(f(S_n) - f(Y_n))| = 0,$$

where $Y_n \sim \mathcal{N}(0, \text{Cov} X_n)$.

Note that $\lim_{n \rightarrow \infty} |\mathbb{E}(f(Y_n) - f(Y))| = 0$, with $Y \sim \mathcal{N}(0, \Sigma)$.

This leads to

$$\lim_{n \rightarrow \infty} |\mathbb{E}(f(S_n) - f(Y))| = 0.$$

□.

Remark 3. The condition $T(n) \rightarrow 0$, as $n \rightarrow \infty$ is related to the weak dependence structure, the other conditions are classical in the independent case.

Theorem 3. (Bardet and al.(2008)).

Let be $(X_n)_{n \in \mathbb{Z}}$ a stationary zero mean η - weakly dependent sequence with $\eta_r = O(r^{-a})$, $a > 3$. Suppose moreover that assumptions (H.1), (F.1), (F.3) and (K.2) hold, then

$$\sqrt{nh_n} \left(f_n(x) - E f_n(x) \right) \sim \mathcal{N}(0, \sigma^2),$$

as $n \rightarrow \infty$, with $\sigma^2 = f(x) \int K^2(x) dx$.

Proof of Theorem 3. Put

$$S_n = \sqrt{nh_n} (f_n(x) - E(f_n(x))) = \sum_{i=1}^n Z_i,$$

$$\text{with, } Z_i = \frac{1}{\sqrt{nh_n}} \left(K\left(\frac{x - X_i}{h_n}\right) - E\left(K\left(\frac{x - X_i}{h_n}\right)\right) \right) = u(X_i).$$

The first condition of Theorem2 is satisfied from Theorem 1. For the second condition, we have for $\delta > 0$,

$$\begin{aligned} A_n &= \sum_{i=1}^n E(\|Z_i\|^{2+\delta}), \\ &= (nh_n)^{-\frac{\delta}{2}} \frac{1}{h_n} \mathbb{E}(|K(\frac{x - X_i}{h_n}) - E(K(\frac{x - X_i}{h_n}))|^{2+\delta}), \\ &\leq 2^{2+\delta} (nh_n)^{-\frac{\delta}{2}} \mathbb{E}(|K(\frac{x - X_i}{h_n})|^{2+\delta}) \\ &\leq 2^{2+\delta} (nh_n)^{-\frac{\delta}{2}} \|f\|_\infty \int |K(s)|^{2+\delta} ds, \end{aligned}$$

then

$$\lim_{n \rightarrow \infty} A_n = 0, \text{ as } n \rightarrow \infty.$$

Now, using assumption (F.3), and (K.1) , we get

$$\begin{aligned} |\mathbb{E}(Z_i)| &\leq \frac{2h_n}{\sqrt{h_n n}} \|f\|_\infty \int |K(v)| dv \\ &\leq c \sqrt{\frac{h_n}{n}}. \end{aligned} \tag{3.12}$$

Now we have

$$\begin{aligned} T(n) &= \sum_{j=1}^{n-1} \text{Cov}(\exp(it(Z_1 + \dots + Z_{j-1})), \exp(it(Z_j))) \\ &= \sum_{k=1}^{j-1} \text{Cov}(\exp(itS_{j-1}), \exp(itZ_j)). \end{aligned}$$

Using the following decomposition (see [37]),

$$\exp(itS_{j-1}) = \sum_{k=1}^{j-1} (\exp(itS_k) - \exp(itS_{k-1})),$$

we obtain then

$$\text{Cov}(\exp(itS_{j-1}), \exp(itZ_j)) = \sum_{k=1}^{j-1} \text{Cov}((\exp(itS_k) - \exp(itS_{k-1})), \exp(itZ_j)).$$

with $S_k = \sum_{i=1}^k Z_i$, and $S_0 = 0$.

Consider Z_j^* a real random variable having the same distribution of Z_j and independent of (Z_1, \dots, Z_{k-1}) , using (3.12) we have

$$\begin{aligned} &|\text{Cov}(\exp(itS_k) - \exp(itS_{k-1}), \exp(itZ_j))| & (3.13) \\ &= |\mathbb{E}(\exp(itS_k) - \exp(itS_{k-1}))(\exp(itZ_j) - \exp(itZ_j^*))|, \\ &\leq |t|^2 \mathbb{E}(|Z_k|(|Z_j| + |Z_j^*|)). \\ &\leq \frac{Ch_n}{n}. \end{aligned}$$

Denote now

$$g_1(X_1, \dots, X_k) = \exp(itS_k) - \exp(itS_{k-1}) \text{ and } g_2(X_j) = \exp(itZ_j).$$

We have then

$$\begin{aligned} \|g_2\|_\infty &= 1, \\ \|g_1\|_\infty &\leq |t| \|S_k - S_{k-1}\|_\infty \\ &\leq |t| \|u\|_\infty \\ &\leq 2|t| \|K\|_\infty \frac{1}{\sqrt{nh_n}}, \end{aligned}$$

and

$$\begin{aligned} \text{Lip}g_1 &\leq 8|t|\text{Lip}K \frac{k}{h_n\sqrt{nh_n}}, \\ \text{Lip}g_2 &\leq 4|t|\text{Lip}K \frac{1}{h_n\sqrt{nh_n}}. \end{aligned}$$

Applying now the weak dependence property, one obtain

$$\text{Cov}(\exp(itS_k) - \exp(itS_{k-1}), \exp(itZ_j)) \leq \left(\frac{1}{nh_n^2} + \frac{1}{n^{\frac{1}{2}}h_n^{\frac{3}{2}}}\right)\eta_\ell, \quad (3.14)$$

with $\ell = j - k$.

From (3.13) and (3.14), one get

$$\text{Cov}(\exp(itS_k) - \exp(itS_{k-1}), \exp(itZ_j)) \leq \min\left(\frac{Ch_n}{n}, \max\left(\frac{k}{nh_n^2}, \frac{1}{n^{\frac{1}{2}}h_n^{\frac{3}{2}}}\right)\eta_\ell\right).$$

By consequence

$$\begin{aligned} T(n) &= \sum_{j=1}^n \sum_{k=1}^{j-1} \text{Cov}(\exp(itS_k) - \exp(itS_{k-1}), \exp(itZ_j)) \\ &\leq C \sum_{j=1}^n \sum_{\ell=1}^j \min\left(\frac{h_n}{n}, \max\left(\frac{k}{nh_n^2}, \frac{1}{n^{\frac{1}{2}}h_n^{\frac{3}{2}}}\right)\eta_\ell\right) \\ &\leq C \sum_{\ell=1}^n \min\left(h_n, n \max\left(\frac{k}{nh_n^2}, \frac{1}{n^{\frac{1}{2}}h_n^{\frac{3}{2}}}\right)\eta_\ell\right) \\ &\leq Ch_n^{1-\beta} n^\beta \sum_{\ell=1}^n \left(\max\left(\frac{k}{nh_n^2}, \frac{1}{n^{\frac{1}{2}}h_n^{\frac{3}{2}}}\right)\eta_\ell\right)^\beta, \quad \text{with, } \beta \geq 0. \end{aligned}$$

Assumption (H.1), (H.2) give then $\lim_{n \rightarrow \infty} T(n) = 0. \square$

Proposition 6. (Ango NZE and al. (2002)).

Suppose that the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ is η -weakly dependent with $\eta_r = O(r^{-a})$, $a > 3$. and satisfies condition (R.1) with $p = 2$, (R.2) and (R.3) .

Let f, g satisfy (F.1) for some $k \in]0, 2]$, then for all x belonging to any compact subset of \mathbb{R} ,

$$\sqrt{nh_n} \left(g_n(x) - \mathbb{E}g_n(x) \right) \sim \mathcal{N} \left(0, g_2(x) \int K^2(x) dx \right).$$

$$\sqrt{nh_n} \left(g_n(x) - r(x)f_n(x) \right) - \mathbb{E} \left(g_n(x) - r(x)f_n(x) \right) \sim \mathcal{N} \left(0, G_2(x) \int K^2(x) dx \right),$$

as $n \rightarrow \infty$.

Proof of Proposition 6

Set

$$S_k = \sum_{t=1}^k \xi_t,$$

with $\xi_t = \frac{1}{\sigma_n \sqrt{nh_n}} (T_t(x) - \mathbb{E}T_t(x))$, and $T_t(x)$ are defined in (3.4).

To prove this CLT, the following statements will be proved.

- i) $\lim_{n \rightarrow \infty} \sigma_n^2 = g_2(x) \int K^2(x) dx$
- ii) $\left(\sqrt{nh_n} \left(g_n(x) - \mathbb{E}g_n(x) \right) - \sigma_n S_n \right) \rightarrow 0$, as $n \rightarrow \infty$, in L^2 .
- iii) $S_n \sim \mathcal{N}(0, 1)$.

Statement (i) is the result of Theorem 1, and (ii) is already proved in (3.5).

Now for the statement (iii), Lindberg method is used, we will proceed exactly as in the proof of Theorem 3 with the centred variables ξ_t .

Replacing now Y_t by $Y_t - r(x)$ gives the second convergence. \square

Theorem 4. (Ango NZE and al. (2002)).

Suppose that the stationary sequence $(Z_t)_{t \in \mathbb{Z}}$ is η -weakly dependent with $\eta_r = O(r^{-a})$, and

$$a > \min \left(\max(3, 9\delta), \max \left(3 + \frac{1}{\delta}, \frac{2 + 6\delta}{1 + \delta} \right) \right).$$

satisfies condition (R.1) with $p = 2$, (R.2) and (R.3).

Let f, g satisfy (F.1) for some $k \in]0, 2]$, then for all x belonging to any compact subset of \mathbb{R}

$$\sqrt{nh_n} \left(r_n(x) - r(x) \right) \sim \mathcal{N} \left(0, \frac{G_2(x)}{nh_n f^2(x)} \int K^2(x) dx \right).$$

Proof of Theorem 4

Using the following decomposition

$$\begin{aligned} r_n(x) - r(x) &= \\ & \frac{g_n(x) - r(x)f_n(x)}{\mathbb{E}f_n(x)} - \left(r(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}\right) + \frac{\left(r(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}\right)(f_n(x) - \mathbb{E}f_n(x))}{\mathbb{E}f_n(x)} \\ & - \frac{(g_n(x) - \mathbb{E}g_n(x))(f_n(x) - \mathbb{E}f_n(x))}{(\mathbb{E}f_n(x))^2} + \frac{r_n(x)(f_n(x) - \mathbb{E}f_n(x))^2}{(\mathbb{E}f_n(x))^2} \\ & := I_1 + I_2 + I_3 + I_4 + I_5. \end{aligned}$$

For I_2 , from Proposition 5

$$\sqrt{nh_n} \left(r(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}\right) = O\left((nh_n^{1+2k})^{-\frac{1}{2}}\right).$$

Using both order of the bias in Proposition 5 and the moments of order 2 of $f_n(x)$, I_3 gives

$$\sqrt{nh_n} \mathbb{E} \left(\frac{\left(r(x) - \frac{\mathbb{E}g_n(x)}{\mathbb{E}f_n(x)}\right)(f_n(x) - \mathbb{E}f_n(x))}{\mathbb{E}f_n(x)} \right) = O((nh_n)^{-\frac{1}{2}}).$$

Using now the moments of order 2 of $f_n(x)$ and $g_n(x)$ in I_4 we get

$$\sqrt{nh_n} \mathbb{E} \left(\frac{(g_n(x) - \mathbb{E}g_n(x))(f_n(x) - \mathbb{E}f_n(x))}{(\mathbb{E}f_n(x))^2} \right) = O((nh_n)^{-\frac{1}{2}}).$$

Now, from assumption (R.2), since the function $\ln n$ is concave we get then for the term I_5

$$\sqrt{nh_n} \mathbb{E} \left(\frac{r_n(x)(f_n(x) - \mathbb{E}f_n(x))^2}{(\mathbb{E}f_n(x))^2} \right) = O\left(\left(\frac{nh_n}{\ln n}\right)^{-\frac{1}{2}}\right).$$

Application of Proposition 6 leads to the result. \square

4.3.3 Bernstein's type inequality

In our context of the weak dependence, and compared to Bernstein inequality stated in independent case, Doukhan and Louhichi used a combinatorial technique to prove an exponential inequality (see Doukhan and Louhichi

(1999)), it is obtained with \sqrt{t} rate. Bennett-type inequality was proved after by Dedecker and Prieur (see Dedecker and al.(2007)), which leads to Bernstein inequality with t^2 rate.

The following inequality established by Doukhan and Neumann (2007) extend previous works to more general conditions of weak dependence.

Theorem 5. (See Doukhan and Neumann (2007)).

Let be $S_n = X_1 + \dots + X_n$, where X_1, \dots, X_n are real valued random variables with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

Let be the function $\Psi(u, v) = u + v$ with $u, v \in \mathbb{N}$,

$s_1 \leq \dots \leq s_u \leq t_1 \leq \dots \leq t_v$, $(s_1, \dots, s_u) \in \mathbb{Z}^u$ and $(t_1, \dots, t_v) \in \mathbb{Z}^v$.

If the following condition holds

$$|\text{Cov}(X_{s_1} \dots X_{s_u}, X_{t_1} \dots X_{t_v})| \leq K^2 M^{u+v-2} ((u+v)!)^\nu \Psi(u, v) \rho(t_1 - s_u), \quad (3.15)$$

where

$$\sum_{s=0}^{\infty} (s+1)^k \rho(s) \leq L_1 L_2^k (k!)^\mu, \quad \forall k \geq 0, \quad (3.16)$$

and

$$E|X_t|^k \leq (k!)^\nu M^k, \quad \forall k \geq 0,$$

with $K, M, L_1, L_2 < \infty, \mu, \nu \geq 0$, and, $\rho(n)$ is a decreasing sequence to 0, we have then $\forall t \geq 0$,

$$\mathbb{P}(S_n \geq t) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{\frac{1}{\mu+\nu+2}} t^{\frac{2\mu+2\nu+3}{\mu+\nu+2}}}\right), \quad (3.17)$$

where

- $A_n \geq \sigma_n^2$, with σ_n^2 is the variance of S_n .
- $B_n = 2(K \vee M) L_2 \left(\left(\frac{2^{4+\mu+\nu} n K^2 L_1}{A_n} \right) \vee 1 \right)$.

Remark 4. - For the choice of the constant A_n , it can be taken equal to σ_n^2 , since σ_n^2 is usually of order $O(n)$, and B_n is then of order $O(1)$.

- If σ_n^2 is very small, it would be better to take $A_n > \sigma_n^2$. One can choose

$$A_n = 2^{1+\mu} n K^2 \Psi(1, 1) L_1, \quad B_n = 2(K \vee M) L_2 \left(\left(\frac{2^{3+\nu}}{\Psi(1, 1)} \right) \vee 1 \right),$$

because from (3.15), we have $\sigma_n^2 \leq 2^{1+\mu} n K^2 \Psi(1, 1) L_1$.

To relate now Bernstein type inequality in Theorem 5 with weak dependence, we state the following proposition.

Proposition 7. (*Doukhan and Neumann (2007)*).

If $(X_t)_{t \in \mathbb{Z}}$ is weakly dependent with $\|X_t\|_\infty \leq M$ then

$$|\text{Cov}(X_{s_1} \dots X_{s_u}, X_{t_1} \dots X_{t_v})| \leq M^{u+v} \Psi(M^{-1}, M^{-1}, u, v) \epsilon_{t_1 - s_u}.$$

Moreover, if $\epsilon_r = \exp(-ar)$, $a > 0$, then we may choose in inequality (3.16) $\mu = 1$, and $L_1 = L_2 = \frac{1}{1 - \exp(-a)}$.

Proof

Note that the definition of weak dependence (2.1) of the stationary sequence $(X_t)_{t \in \mathbb{Z}}$ with $\|X_t\|_\infty \leq M$ can be expressed as the condition (3.15) by considering the functions

$$\begin{cases} g_1(x_{s_1}, \dots, x_{s_u}) = \prod_{i=1}^u f\left(\frac{x_i}{M}\right) \\ g_2(x_{t_1}, \dots, x_{t_v}) = \prod_{i=1}^v f\left(\frac{x_i}{M}\right) \end{cases}$$

where $f(u) = u \vee (-1) \wedge 1$.

These functions satisfy

$$\text{Lip} g_l \leq \frac{1}{M}, \|g_l\|_\infty \leq 1, \quad (l = 1, 2).$$

For the second part of Proposition 7, we have

$$\begin{aligned} \sum_{s=0}^{\infty} (s+1)^k \exp(-as) &\leq \sum_{s=0}^{\infty} (s+1) \dots (s+k) \exp(-as) \\ &= \frac{d^k}{dp^k} \left[\left(\frac{1}{1-p} \right) \right]_{p=\exp(-a)} \\ &= k! \frac{1}{(1 - \exp(-a))^{k+1}}. \end{aligned}$$

The condition (3.16) is then checked for $\mu = 1$, and $L_1 = L_2 = \frac{1}{1 - \exp(-a)}$. \square

Remark 5. – From Proposition 7, one can deduce that, if $(X_t)_{t \in \mathbb{Z}}$ is η -weakly dependent, then condition (3.15) is checked with $\Psi(u, v) = u + v$, and $K^2 = M$.

– If $F : \mathbb{R}^d \rightarrow \mathbb{R}$ is Lipschitz and continuous with $\|F\|_\infty = M < \infty$ and $\text{Lip} F \leq 1$, the η -weakly dependent $Y_t = F(X_t)$ process satisfies also condition (3.15) with $\Psi(u, v) = u + v$, and $K^2 = M \text{Lip} F$.

4.3.4 Almost complete convergence

As an application of Theorem 5, Doukhan and Neumann (2007) derived the uniform almost sure convergence of density estimator f_n , they obtained

$$\sup_{x \in \mathbb{R}} |f_n(x) - \mathbb{E}f_n(x)| = O_{a.s.}\left(\left(\frac{\ln n}{nh_n}\right)^{\frac{1}{2}}\right).$$

This is exactly similar to the result of Giné and Guillou (2002) with independent variables.

To extend the result of almost complete convergence stated in the independent case both pointwisely and uniformly, we argue as Doukhan and Neumann (2007) to use Bernstein type inequality. We set

$S_n = \sum_{t=1}^n X_{n,t} = F(X_{n,t})$, with

$$X_{n,t} = \frac{1}{nh_n} \left(K\left(\frac{x - X_t}{h_n}\right) - \mathbb{E}\left(K\left(\frac{x - X_t}{h_n}\right)\right) \right).$$

We have $\|F\|_\infty \sim \frac{1}{nh_n}$, $LipF \leq \frac{LipK}{nh_n^2}$.

The constants $M = \|F\|_\infty$, $K^2 \sim \frac{1}{n^2 h_n^3}$, $A_n \sim \frac{f(x)}{nh_n} \int K^2(u) du$, $B_n \sim \frac{1}{nh_n^4}$, lead to the following result.

$$\sup_{x \in \mathbb{R}} |f_n(x) - \mathbb{E}f_n(x)| = O_{a.co.}\left(\left(\frac{\ln n}{nh_n}\right)^{\frac{1}{2}}\right).$$

Chapitre 5

Recursive kernel estimation of the density under weak dependence

The purpose of this chapter is to study the asymptotic behaviour of the recursive kernel density estimator. This estimator was introduced and investigated by [2] for independent and α -mixing sequences. In this work, we are interested in η -weak dependence, which is different from the notion of α -mixing. We provide the variance and the mean squared error of this estimator. The asymptotic normality is also discussed. A simulation study for two η -dependent models which are not necessarily α -mixing shows the advantage in time computation to consider the recursive kernel estimation rather than the Parzen-Rosenblatt one.

5.1 Introduction

Let $(X_i)_{i \in \mathbb{Z}}$ be a stationary sequence of η -weakly dependent real-valued random variables with unknown density f . Studying the properties of nonparametric estimators under different assumptions of dependence has received considerable attention. In this work, we use the generalized weak dependence studied by [15] and we consider the recursive version of a kernel estimator of the density f introduced by [2] and given by

$$f_n^\ell(x) = \frac{1}{\sum_{i=1}^n h_i^{1-\ell}} \sum_{i=1}^n \frac{1}{h_i^\ell} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R} \quad (1.1)$$

where $\ell \in [0, 1]$ is a smoothing parameter, this latter is considered here to obtain a more general estimator, which includes recursive estimators defined by [10] or by [47], h_n is a given positive sequence decreasing to 0, and K is a bounded Lipschitzian density function, convergent to 0 at infinity.

We note that the recursive form (1.1) allows us to update the estimation whenever additional observations are obtained which is not the case with Parzen-Rozenblatt estimator. Indeed the estimator (1.1) satisfies

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

with $K_{n+1}(x - X_{n+1}) = \frac{1}{h_{n+1}^\ell} \sum_{j=1}^{n+1} h_j^{1-\ell} K\left(\frac{x - X_{n+1}}{h_{n+1}}\right)$.

From a practical point of view, this iterative procedure provides an important saving in computational time and memory in many situations such as control problems. In particular, the use of recursive estimates is appropriate in time series analysis and that is why our study is performed with dependent observations. In this context there are always good reasons for studying the behaviour of recursive estimators when the independence is not fulfilled. Kernel estimator has been studied under different assumptions of dependence from first papers of Rosenblatt, under mixing until weak dependence works, we cite from others [15], [16], [1], [17], and [4] for more generalized approaches of weak dependence. It is motivated by the fact that there are models which are not mixing, so this frame of dependence contains a lot of models, like linear and bilinear models, and several classes of the models commonly used in the econometrics illustrating phenomena like fading memory in the finance. Many of the previous results concerning asymptotic properties of recursive kernel estimators are established under the assumption that the $(X_t)_t$ are independent or α -mixing. We refer, for instance, to [48], [10] and to [47], where the quadratic mean convergence and strong consistency are studied under assumption of independence. [2] generalized all these versions using different values of the parameter $\ell \in [0, 1]$. The mean square convergence and the asymptotic normality were studied by [31] under strong mixing. For the fixed values $\ell = \frac{1}{2}$ and $\ell = 1$, the asymptotic normality under negative association was treated by [29]. Tran [44] obtained the uniform convergence of this recursive estimator for $\ell = 1$ under α -mixing. Our investigation deals with studying the convergence and the asymptotic normality of the recursive estimator (1.1) under η -weak dependence extending in this way both the works of [15] to a non-recursive frame and those of [2] concerning indepen-

dence and strong mixing for the estimator (1.1).

5.2 Assumptions

Among others, Wu ([49]) proposed some dependence measures that provide a natural framework for a limit theory for stationary processes. In this paper, we assume that the sequence $(X_t)_{t \in \mathbb{Z}}$ is η -weakly dependent (see [9]). In order to state our results, the following assumptions are required on the kernel K and the bandwidth h_n .

- (H.1) $nh_n^{1+\epsilon} \rightarrow \infty$, as $n \rightarrow \infty$, for some $\epsilon > 0$.
- (H.2) $\beta_{n,r} = \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^r \rightarrow \beta_r < \infty$, as $n \rightarrow \infty$, $r \leq 3$.
- (K.1) $\int vK(v)dv = 0$, $\int v^2K(v)dv < \infty$.
- (K.2) $\int K^2(v)dv < \infty$.
- (F.1) f has bounded and continuous derivatives of order 2.

Remarks 1. Assumptions (K.1), (K.2), and (F.1) are classical in kernel estimation, while assumption (H.1) is appropriate in the dependence context, other assumptions about the bandwidth will be added as needed.

In [47], assumption (H.2) is used to study the behaviour of the bias of the recursive estimator

$$f_n^{DW}(x) = \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \frac{1}{\sqrt{h_i}} K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}.$$

In [46] also, a similar assumption is used to investigate the mean squared error of a recursive local polynomial estimator under α -mixing condition. In fact this assumption is appropriate in the recursivity, because in this case, the stationarity of X_i does not lead to the stationarity of $\frac{1}{h_i^2} K\left(\frac{x - X_i}{h_i}\right)$, so it is used to facilitate calculus.

In particular, if $h_n = n^{-\nu}$, $0 < \nu < 1$, we have

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

If $\nu r \geq 1$, then $\beta_r = \infty$.

If $\nu r < 1$, then $\sum_{i=1}^n i^{-\nu r} \sim \frac{n^{1-\nu r}}{1-\nu r}$ then $\beta_r = \frac{1}{1-\nu r}$.

5.3 Results

We are now in position to state the main results of this paper.

5.3.1 Convergence in mean squared error

Let f_{X_s} and $f_{(X_s, X_t)}$ denote respectively the marginal and the joint densities of (X_s, X_t) . We assume through this section, that

$$(F.2) \sup_{|s-t| \geq 1} \|g_{s,t}\|_\infty < \infty, \text{ with } g_{s,t} = f_{(X_s, X_t)} - f_{X_s} \otimes f_{X_t}.$$

Variance

The following proposition gives the asymptotic behavior of the variance of the recursive estimator (1.1) for η -weakly dependent sequences.

Proposition 8. *Suppose that the stationary sequence (X_n) is η -weakly dependent. Suppose also that the assumptions (H.2), (K.2) and (F.2) are satisfied. Define for some $\ell \in [0, 1]$,*

$$U_{n,\ell}(N) = \frac{1}{n} \sum_{k=N+1}^{n-1} \sum_{p=1}^n \left(\frac{h_{k+p}}{h_n} \right)^{-2\ell-1} \eta_k.$$

One has the following

(i) *If $\lim_{N \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} N^{2+\nu} U_{n,\ell}(N) = 0$, for some $\nu > 0$, then*

$$\text{Var} f_n^\ell(x) = \frac{\beta_{1-2\ell}}{nh_n \beta_{1-\ell}^2} f(x) \int K^2(x) dx + o\left(\frac{1}{nh_n}\right),$$

as $n \rightarrow \infty$.

(ii) *If $\lim_{N \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} N^2 U_{n,\ell}(N) < \infty$, then*

$$\text{Var} f_n^\ell(x) = \frac{\beta_{1-2\ell}}{nh_n \beta_{1-\ell}^2} f(x) \int K^2(x) dx + O\left(\frac{1}{nh_n}\right),$$

as $n \rightarrow \infty$.

Remarks 2. – In the special case when $\ell = 1$ and $h_i = h_n$, we get the classic Parzen-Rosenblatt kernel density estimator given by,

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad x \in \mathbb{R} \tag{3.2}$$

and the variance of the recursive estimator obtained in Proposition 8 coincides with the result of [16] under η -dependence in a non-recursive frame.

Note that in this case, $\lim_{n \rightarrow \infty} U_{n,\ell}(R) = U(R) := \sum_{k>R} \eta_k$.

– The estimator (1.1) generalizes other recursive estimators, indeed, substituting respectively the values $\ell = 0$, $\ell = 1$, we get the recursive estimators of Deheuvels and Wolverton (see [10] and [48]), and our result on the variance is similar to the result of [10] under independence.

– Amiri [2] derived the variance of the estimator (1.1) under independence and α -mixing conditions. So our results are extended to more general notion of weak dependence.

– Different expressions of the estimator (1.1) depending on $\ell \in [0, 1]$ allow a comparison with Parzen-Rosenblatt estimator in term of variance, we have

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

With the usual selection of the bandwidth $h_n = n^{-\frac{1}{5}}$. The quantity $\frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} = \frac{(4+\ell)}{5(4+2\ell)} < 1$, $\forall \ell \in [0, 1]$, hence the variance is reduced with the recursive estimation. Comparison between recursive kernel estimators with different values of $\ell \in [0, 1]$ is also possible, note that $\frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2}$ increases with ℓ , the optimal choice is then for $\ell = 0$.

Mean squared error

There are various ways to quantify the performance of the estimator (1.1). We will focus here on the mean squared error MSE and its two components namely bias and variance. We have

$$\begin{aligned} MSE(n, h) &= \mathbb{E}(f_n^\ell(x) - f(x))^2 \\ &= Bias^2 + \text{Var}(f_n^\ell(x)). \end{aligned}$$

From Proposition 8 above, we can infer the following.

(Recall that $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$.)

Corollary 1. *If assumptions (H.2), (K.1), (K.2), (F.1) and (F.2) hold then*

$$MSE(n, h_n) \sim \frac{h_n^4 \beta_{3-\ell}^2}{4\beta_{1-\ell}^2} \left(f^{(2)}(x) \int x^2 K(x) dx \right)^2 + \frac{\beta_{1-2\ell}}{nh_n \beta_{1-\ell}^2} f(x) \int K^2(x) dx,$$

as n tends to infinity and the optimal bandwidth is given by

$$h_{n,opt} = \left(\frac{\beta_{1-2\ell} f(x) \int K^2(x) dx}{\beta_{3-\ell}^2 (f^{(2)}(x) \int x^2 K(x) dx)^2} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Remarks 3. – Large values of the bandwidth give good estimates with small variance, whereas to obtain small bias one must choose small bandwidth. In order to achieve good rates of convergence, both components bias and variance were then considered in the expression of MSE above, and the optimal bandwidth satisfies the compromise between the bias and the variance.

– The parameter ℓ can be considered as an other smoothing parameter, although its effect on the variance-bias balance is very much smaller than the the bandwidth sequence $(h_n)_n$ effect. Thus, the selection of an optimal value of ℓ is not so important and this parameter is considered here to obtain a more general estimator, which includes recursive estimators defined by [10], [47].

5.3.2 Asymptotic normality

In a recursive frame, arguing as in [16], we employ the classic technique of proving a central limit theorem (see also [31]) for weakly dependent random variables by decomposing the deviation $f_n^\ell(x) - \mathbb{E}(f_n^\ell(x))$ as a sum based on big blocks separated by small ones. The contribution of small blocks is negligible and the big blocks are shown to be approximately independent. The Lindeberg condition gives then the following result which is devoted to the asymptotic normality of the recursive kernel density estimator (1.1) under the more generalized notion of η -weak dependence.

For this purpose, we impose the following additional assumptions.

(D.1) $\eta_k = O(\exp(-\lambda k))$, for some $\lambda > 0$.

(H.3) If u_n, v_n are two sequences satisfying $u_n \sim v_n$ then $h_{u_n} \sim h_{v_n}$.

Theorem 6. *Suppose that the stationary sequence (X_n) is η -weakly dependent. Assume that assumptions (D.1), (H.1) and (H.3) hold. Then the sequence*

$$\left(\sqrt{nh_n} \left(f_n^\ell(x) - E f_n^\ell(x) \right) \right)_n$$

converges in distribution as $n \rightarrow \infty$, to a centered normal law $\mathcal{N}(0, \sigma^2)$ with variance $\sigma^2 = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(x) dx$.

Remarks 4. The conclusion of Theorem 6 is similar to the result of [16] for Parzen-Rosenblatt estimator under Assumption (H.1) in the context of the weak dependence with the arithmetic decay of the dependence coefficients $\theta_n = O(n^{-12-a})$, with $a > 0$.

For the case of α -mixing sequences, [38] proved the CLT for the Parzen-Rosenblatt estimator under the condition $\alpha_n = O(n^{-a})$, $a > 0$.

[31] obtained the CLT of the recursive estimator (1.1) in the particular cases $\ell = \frac{1}{2}$, $\ell = 1$ and for stationary α -mixing sequences satisfying

$\left(\frac{n}{h_n}\right)^{\frac{1}{2}} \sum_{i \geq q_n} [\alpha(i)]^{1-\epsilon} \rightarrow 0$, for some $q_n \rightarrow \infty$, with the following assumption on the bandwidth $nh_n^{3-2\epsilon} \rightarrow \infty$, as $n \rightarrow \infty$, for some $0 < \epsilon < 1$.

Amiri ([2]) generalized recently Masry's works. He showed the CLT of the recursive estimator (1.1) for different values of $\ell \in [0, 1]$ under the independence and the α -mixing condition with a geometric coefficient $\alpha_n = O(\exp(-an))$, $a > 0$ under the assumption $\frac{nh_n^3}{(\ln n)^\epsilon} \rightarrow \infty$, as $n \rightarrow \infty$, for some $\epsilon > 0$.

5.4 Application

This section is aimed to discuss the behavior of the estimator (1.1) based on some η -weakly dependent processes. Our results allow to derive the *MSE* and the asymptotic normality in various cases. To apply these results, we check the conditions of Proposition 8, Corollary1 and Theorem6 on such models.

5.4.1 Bernoulli shift

Next corollary gives the *MSE* of the recursive estimator based on Bernoulli shift. It can be obtained with two kind of decaying of the dependence coefficients : arithmetic or geometric.

Corollary 2. Let $X_n = \sum_{-\infty}^{+\infty} a_i \epsilon_{n-i}$ be the Bernoulli shift process with $a_i = O(i^{-A})$, for some $A > 5$. Suppose that $h_n = cn^{-\frac{1}{5}}$, for $c > 0$ and that assumptions (K.1) and (K.2) hold. Then for $\ell \in [0, 1]$,

$$\lim_{n \rightarrow \infty} n^{\frac{4}{5}} MSE(n, cn^{-\frac{1}{5}}) = \left(\frac{(4 + \ell) f^{(2)}(x) \int v^2 K(v) dv}{2 + \ell} \right)^2 + \frac{(4 + \ell)^2}{20 + 10\ell} f(x) \int K^2(x) dx.$$

Similar result can be applied to the Bernoulli shift with geometric coefficients.

The following corollary is an immediate consequence of Theorem 1.1. We focus on the asymptotic normality of Bernoulli shift with geometric coefficients.

Corollary 3. Let $X_n = \sum_{-\infty}^{+\infty} a_i \epsilon_{n-i}$ be the Bernoulli shift process with $a_i = O(\exp(-\lambda i))$, for some $\lambda > 0$. If $h_n = cn^{-\frac{1}{5}}$, $c > 0$, then the sequence

$$\left(\sqrt{n^{\frac{4}{5}}} \left(f_n^\ell(x) - \mathbb{E} f_n^\ell(x) \right) \right)_n$$

converges in distribution as $n \rightarrow \infty$, to a centered normal law $\mathcal{N}(0, \sigma^2)$ with variance $\sigma^2 = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(x) dx$.

5.4.2 Bilinear process

To apply our results on an other η -dependent model, we choose the bilinear process defined in (2.2) with the coefficients

$$b_1 = a, b_j = 0, \forall j > 1, a_j = 0, \forall j \geq 1. \quad (4.3)$$

Corollary 4. Let $X_n = aX_{n-1} + b + \epsilon_n$ be a bilinear process. Suppose that $h_n = cn^{-\frac{1}{5}}$, for $c > 0$ and that assumptions (K.1) and (K.2) hold. Then for $\ell \in [0, 1]$,

$$\lim_{n \rightarrow \infty} n^{\frac{4}{5}} MSE(n, cn^{-\frac{1}{5}}) = \left(\frac{(4 + \ell) f^{(2)}(x) \int v^2 K(v) dv}{2 + \ell} \right)^2 + \frac{(4 + \ell)^2}{20 + 10\ell} f(x) \int K^2(x) dx.$$

Corollary 5. Let $X_n = aX_{n-1} + b + \epsilon_n$ be a bilinear process. If $h_n = cn^{-\frac{1}{5}}$, $c > 0$, then the sequence

$$\left(\sqrt{n^{\frac{4}{5}}} \left(f_n^\ell(x) - \mathbb{E} f_n^\ell(x) \right) \right)_n$$

converges in distribution as $n \rightarrow \infty$, to a centered normal law $\mathcal{N}(0, \sigma^2)$ with variance $\sigma^2 = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(x) dx$.

5.5 Simulation

The purpose of this simulation study is to compare the recursive estimator (1.1) with the classical Parzen-Rosenblatt estimator under the assumption of η -dependence. The example investigated is the Bernoulli shift given by

$$X_i = \frac{1}{2}(X_{i-1} + \epsilon_i), i = \overline{1, n},$$

with Bernoulli innovations $\mathbb{P}(\epsilon_i = 0) = \mathbb{P}(\epsilon_i = 1) = \frac{1}{2}$.

This can be written in the form $X_i = H(\epsilon_i, \epsilon_{i-1}, \dots)$ with $H(x) = \sum_{k \geq 0} 2^{-(k+1)} x_k$.

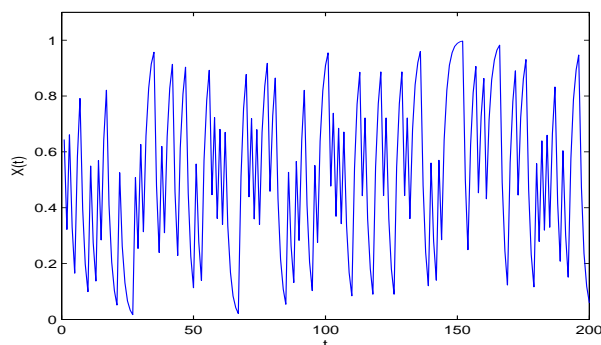


FIGURE 5.1 – Bernoulli shift, $n = 200$

This model (figure 5.1) has stationary uniform distribution on the interval $[0, 1]$, but it satisfies no α -mixing condition. It is however η -weakly dependent if

$$\delta_p = \mathbb{E}(|H(\epsilon_{i-j}, j \in \mathbb{Z}) - H(\epsilon_{i-j} I_{|j| < p})|) \longrightarrow 0 \quad \text{as } p \rightarrow \infty,$$

with $\eta_p = 2\delta_{\lfloor \frac{p}{2} \rfloor}$ (see [9], page 41). This condition is satisfied for this model, in fact

$$\begin{aligned} \mathbb{E}(|H(\epsilon_{i-j}, j \in \mathbb{Z}) - H(\epsilon_{i-j} I_{|j| < p})|) &= \mathbb{E}(|H(\epsilon_{i-j} I_{|j| \geq p})|) \\ &= \mathbb{E}\left(\sum_{j \geq p} 2^{-(j+1)} \epsilon_{i-j}\right) \\ &= \frac{1}{2} \sum_{j \geq p} 2^{-(j+1)} \longrightarrow 0 \quad \text{as } p \rightarrow \infty. \end{aligned}$$

In figures (5.2), (5.3), (5.4), and (5.5), the density function is estimated by two kernel estimators, the recursive estimator (1.1) with different values of

$\ell \in [0, 1]$ and Parzen-Rosenblatt estimator. We choose Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)\mathbb{I}_{|u| \leq 1}$ and the bandwidth $h_n = n^{-\frac{1}{5}}$, with samples of size $n = 300$ and $n = 500$. In order to show the advantage of the recursive

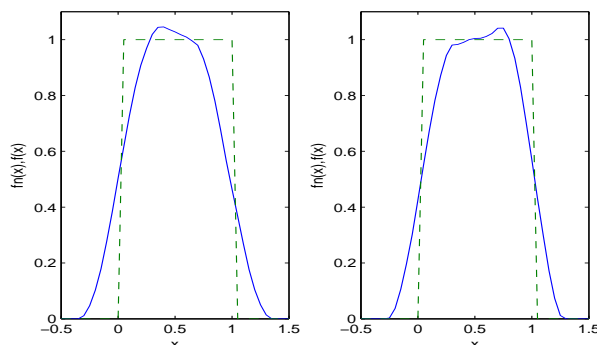


FIGURE 5.2 – (—) Rosenblatt kernel estimators with $n = 300, 500$, (- - -) real density.

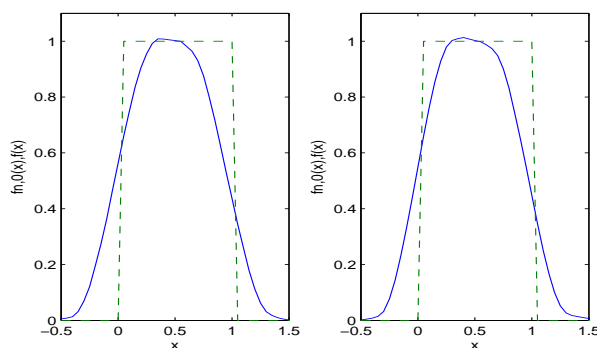


FIGURE 5.3 – (—) Recursive kernel estimators with $\ell = 0, n = 300, 500$, (- - -) real density.

estimation, we use the following recursive algorithm for updating $f_n^\ell(x)$:

1. Observe $X_i, i = \overline{1, n}$.
2. Choose h_i, K .
3. Compute $f_n^\ell(x) = S_n^{-1}V_n$ with $S_n = \sum_{i=1}^n h_i^{1-\ell}, V_n = \sum_{i=1}^n K_i$ and $K_i = \frac{1}{h_i^\ell} K\left(\frac{x-X_i}{h_i}\right)$.

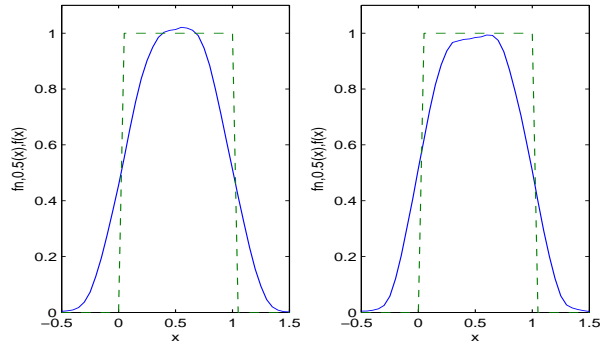


FIGURE 5.4 – (—) Recursive kernel estimators with $\ell = \frac{1}{2}$, $n = 300, 500$, (- -) real density.

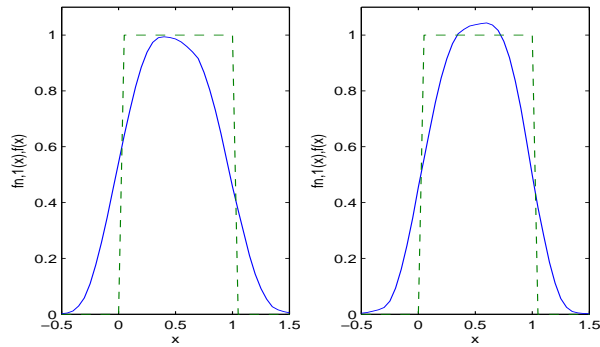


FIGURE 5.5 – (—) Recursive kernel estimators with $\ell = 1$, $n = 300, 500$, (- -) real density.

4. Save $f_n^\ell(x)$.
5. Observe X_{n+1} .
6. Compute h_{n+1} , $K_{n+1}^\ell = S_{n+1}^{-1}K_{n+1}$
7. Compute $f_{n+1}^\ell(x) = S_{n+1}^{-1}S_n f_n^\ell(x) + K_{n+1}^\ell$.

Next we compare this recursive algorithm with Parzen-Rosenblatt one. The time of computation in seconds is given for different sample sizes $n = 100$, $n = 200$, $n = 300$ and $n = 500$ in the following table.

n	$f_n(x)$	$f_n^\ell(x), \ell = 1$	$f_n^\ell(x), \ell = 0.5$	$f_n^\ell(x), \ell = 0$
100	0.24	0.18	0.14	0.15
200	0.59	0.47	0.41	0.42
300	1.34	0.86	0.9	0.81
500	3.32	2.09	1.95	2.15

Table 1 : Comparison of the time computation of the recursive estimator and Rosenblatt one.

The time of computation is clearly smaller with the recursive estimator than with Rosenblatt one.

Moreover, it is interesting to note that the variance is also smaller with the recursive estimator, in Figure 5.6, we compare the variance of the recursive estimator and Rosenblatt one in term of the sample size from 40 to 200 using naturally the smoothing parameter $h_n = n^{-\frac{1}{5}}$. Indeed, Figure 5.6 reveals that the variance is reduced with the recursive estimator even for moderate sample size.

Bilinear processes

We discuss now the bilinear process defined in (2.2) by the non-negative coefficients,

$$b_1 = a, b_j = 0, \forall j > 1, a_j = 0, \forall j \geq 1, \tag{5.4}$$

where the innovations $\epsilon_n \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and we get the model $X_n = aX_{n-1} + b + \epsilon_n$.

This model is η -weak dependent if $0 < a < 1$.

We can see that the stationary law of this process is Gaussian with mean $\frac{b}{1-a}$ and variance $\frac{a^2\sigma_\epsilon^2}{1-a^2}$.

Note that according to the choice (5.4), there exists an integer $N = 2$ such that $a_i = b_i = 0$, for $i \geq N$, then the coefficient of dependence is given by $\eta_k = O(\exp(-\lambda k))$, for $\lambda > 0$, assumption (D.1) is then fulfilled.

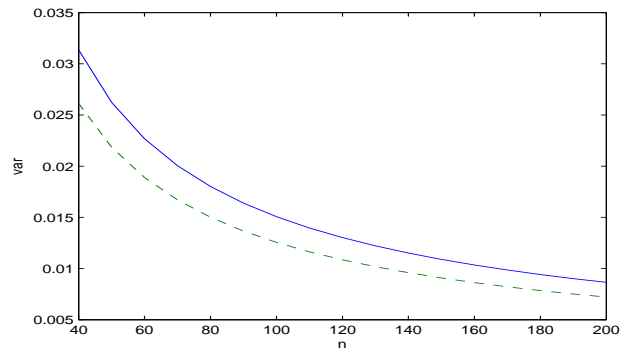


FIGURE 5.6 – Comparison of variance of the recursive estimator (- - -) with Rosenblatt one (—) in term of the sample size using the bandwidth $h_n = n^{-\frac{1}{5}}$.

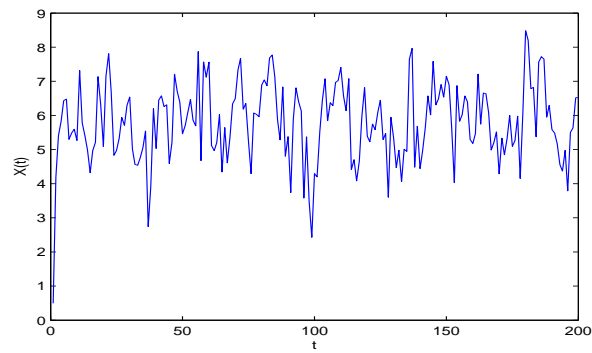


FIGURE 5.7 – Bilinear process, $n = 200$.

Since the density of the innovations is Gaussian, we can check recursively that the density f satisfies assumption (F.1), the joint densities $f_{(X_s, X_t)}$ of our model are bounded and assumption (F.2) is also satisfied. (See [36]).

Now, we simulate this bilinear process with $a = 0.5$, $b = 3$ and $\sigma_\epsilon = 1$.

From Proposition 3, it is η -weakly dependent since $\lambda = 0.5 < 1$.

The recursive estimator with different values of the parameter $\ell \in [0, 1]$, $f_{n,0}(x)$, $f_{n,0.5}(x)$, $f_{n,1}(x)$ and Rosenblatt one $f_n(x)$ are given in the next figures. The time of computation and the variance are also reduced with this process.

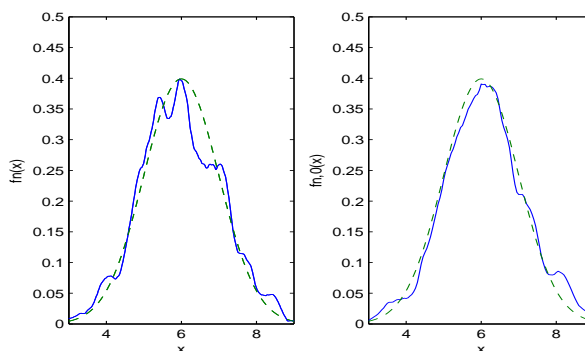


FIGURE 5.8 – (—) Rosenblatt estimator (right) and recursive estimator with $\ell = 0$ (left) $n = 500$, (- - -) real density of the bilinear process.

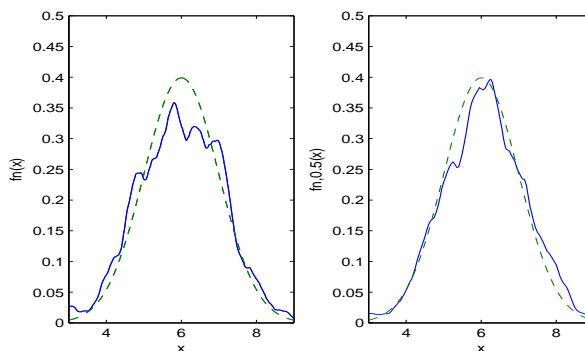


FIGURE 5.9 – (—) Rosenblatt estimator (right) and recursive estimator with $\ell = \frac{1}{2}$ (left) $n = 500$, (- - -) real density of the bilinear process.

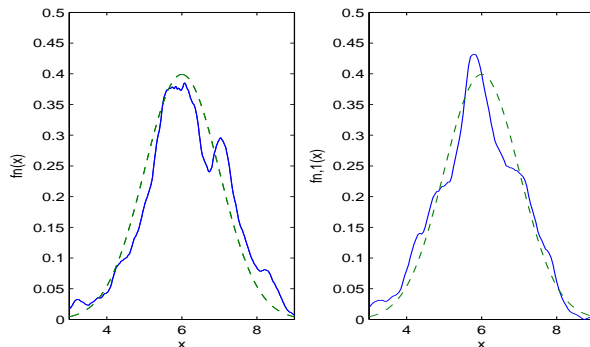


FIGURE 5.10 – (—) Rosenblatt estimator (right) and recursive estimator with $\ell = 1$ (left), $n = 500$, (- -) real density of the bilinear process.

All these simulations allow to conclude that the recursive estimator (1.1) has a clear advantage compared to Parzen-Rosenblatt estimator.

5.6 Proofs

In this section, the next lemma will be needed to justify some convergences, it is a direct consequence of the lemma of Toeplitz. (See [2]).

Lemma 3. *Let ω_n be a sequence of real numbers converging to ω , for $r \leq 3$. under assumptions (H.1), (H.2), we have*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n} \right)^r \omega_i = \beta_r \omega.$$

Proof of Proposition 8. Let $\hat{f}_n^\ell(x)$ be the recursive kernel estimator of $f(x)$ based on the iid variables distributed as X_1 . For convenience, define the covariance quantity

$$A_{i,j} = \text{Cov} \left(h_i^{-\ell} K \left(\frac{x - X_i}{h_i} \right), h_j^{-\ell} K \left(\frac{x - X_j}{h_j} \right) \right). \quad (6.5)$$

We have,

$$\text{Var}(f_n^\ell(x)) = \text{Var}(\hat{f}_n^\ell(x)) + 2 \left(\sum_{i=1}^n h_i^{1-\ell} \right)^{-2} \left(\sum_{i>j} A_{i,j} \right). \quad (6.6)$$

Lemma 4. (Amiri B. 2010)

Under assumptions (H.2), (K.2), we get

$$\lim_{n \rightarrow \infty} nh_n \text{Var}(\hat{f}_n^\ell(x)) = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(x) dx.$$

Proof of Lemma4

$$\begin{aligned} \text{Var}(\hat{f}_n^\ell(x)) &= \left(\sum_{i=1}^n h_i^{(1-\ell)} \right)^{-2} \sum_{i=1}^n \text{Var}\left(\frac{1}{h_i^\ell} K\left(\frac{x - X_i}{h_i}\right)\right) \\ &= \left(\sum_{i=1}^n h_i^{(1-\ell)} \right)^{-2} \sum_{i=1}^n \frac{1}{h_i^{2\ell}} \left(\int K^2\left(\frac{x-u}{h_i}\right) f(u) du - \left(\int K\left(\frac{x-u}{h_i}\right) f(u) du \right)^2 \right) \\ &:= I_1 - I_2. \end{aligned}$$

For the first term I_1 , we have

$$I_1 = \frac{\beta_{n,1-\ell}^{-2}}{n^2 h_n} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^{(1-2\ell)} \int \frac{1}{h_i} K^2\left(\frac{x-u}{h_i}\right) f(u) du,$$

applying now Lemma 3, we get

$$\lim_{n \rightarrow \infty} nh_n I_1 = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(u) du.$$

For the second term I_2

$$\begin{aligned} I_2 &= \left(\sum_{i=1}^n h_i^{(1-\ell)} \right)^{-2} \sum_{i=1}^n \frac{1}{h_i^{2\ell}} \left(\int K\left(\frac{x-u}{h_i}\right) f(u) du \right)^2 \\ &= \left(\sum_{i=1}^n \frac{1}{h_i^{(1-\ell)}} \right)^2 \sum_{i=1}^n h_i^{2(1-\ell)} \left(\int \frac{1}{h_i} K\left(\frac{x-u}{h_i}\right) f(u) du \right)^2, \\ &= \frac{1}{n \beta_{n,1-\ell}^2} \frac{1}{n} \sum_{i=1}^n \left(\frac{h_i}{h_n}\right)^{2(1-\ell)} \left(\int \frac{1}{h_i} K\left(\frac{x-u}{h_i}\right) f(u) du \right)^2, \end{aligned}$$

Then Lemma 3 gives

$$\lim_{n \rightarrow \infty} nh_n I_2 = 0.$$

□.

We have now to control the covariance term. We have,

$$\begin{aligned} \Delta_n(x) &= |\text{Var}(f_n^\ell(x)) - \text{Var}(\hat{f}_n^\ell(x))| \\ &\leq 2\left(\sum_{i=1}^n h_i^{1-\ell}\right)^{-2} \left(\sum_{i>j} |A_{i,j}| \mathbb{I}_{\{1 \leq i-j \leq R\}} \right. \\ &\quad \left. + \sum_{i>j} |A_{i,j}| \mathbb{I}_{\{R+1 \leq i-j \leq n-1\}} \right), \end{aligned} \quad (6.7)$$

for some arbitrary R to be chosen later.

The proof of the proposition will follow from the next lemma.

Lemma 5. *Suppose that the sequence (X_n) satisfies the η -weak dependence condition. If the assumptions (H.2) and (F.2) are satisfied, then for any $R > 0$,*

$$|\Delta_n(x)| \leq \frac{R}{n} + \frac{U_{n,\ell}(R)}{nh_n^3},$$

$$\text{recall that } U_{n,\ell}(R) = \frac{1}{n} \sum_{k=R+1}^{n-1} \sum_{p=1}^n \left(\frac{h_{k+p}}{h_n}\right)^{-2\ell-1} \eta_k.$$

Proof of Lemma 5.

Set $k = i - j$, $p = j$. From the first term of (6.7), (using the same arguments as in [2], page 34), we get since $h_{k+p} \leq h_p$,

$$\begin{aligned} &\sum_{k=1}^R \sum_{p=1}^n |A_{k+p,p}| \\ &= \sum_{k=1}^R \sum_{p=1}^n h_{k+p}^{-\ell} h_p^{-\ell} \left| \mathbb{E} \left(K\left(\frac{x - X_{k+p}}{h_{k+p}}\right) K\left(\frac{x - X_p}{h_p}\right) \right) - \mathbb{E} \left(K\left(\frac{x - X_{k+p}}{h_{k+p}}\right) \right) \mathbb{E} \left(K\left(\frac{x - X_p}{h_p}\right) \right) \right| \\ &= \sum_{k=1}^R \sum_{p=1}^n h_{k+p}^{-\ell} h_p^{-\ell} \left| \int \int K\left(\frac{x-u}{h_{k+p}}\right) K\left(\frac{x-v}{h_p}\right) g_{k+p,p}(u,v) du dv \right| \\ &\leq \sup_{|k| \geq 1} \|g_{p+k,p}\|_\infty \sum_{k=1}^R \sum_{p=1}^n h_p^{2(1-\ell)}. \end{aligned}$$

We obtain,

$$2\left(\sum_{i=1}^n h_i^{1-\ell}\right)^{-2} \sum_{k=1}^R \sum_{p=1}^n |A_{k+p,p}| \leq 2 \sup_{k \geq 1} \|g_{k+p,p}\|_\infty \frac{R\beta_{n,2(1-\ell)}}{n\beta_{n,(1-\ell)}^2}. \quad (6.8)$$

By using the η -dependence condition, we control the second part of (6.7) (recall that h_n is decreasing),

$$\begin{aligned} & 2\left(\sum_{i=1}^n h_i^{1-\ell}\right)^{-2} \sum_{k=R+1}^{n-1} \sum_{p=1}^n \left| \text{Cov}\left(h_{k+p}^{-\ell} K\left(\frac{x - X_{k+p}}{h_{k+p}}\right), h_p^{-\ell} K\left(\frac{x - X_p}{h_p}\right)\right) \right| \\ & \leq 2\left(\sum_{i=1}^n h_i^{1-\ell}\right)^{-2} \sum_{k=R+1}^{n-1} \sum_{p=1}^n h_{k+p}^{-\ell} h_p^{-\ell} \left(h_{k+p}^{-1} \|K\|_\infty \text{Lip}(K) + h_p^{-1} \|K\|_\infty \text{Lip}(K) \right) \eta_k \\ & \leq 2\|K\|_\infty \text{Lip}(K) \left(\sum_{i=1}^n h_i^{1-\ell}\right)^{-2} \sum_{k=R+1}^{n-1} \sum_{p=1}^n h_{k+p}^{-2\ell-1} \eta_k \\ & \leq n^{-2} h_n^{-2(1-\ell)} h_n^{-2\ell-1} 2\|K\|_\infty \text{Lip}(K) \beta_{n,(1-\ell)}^{-2} \sum_{k=R+1}^{n-1} \sum_{p=1}^n \left(\frac{h_{k+p}}{h_n}\right)^{-2\ell-1} \eta_k \\ & \leq 2n^{-1} \beta_{n,(1-\ell)}^{-2} h_n^{-3} \|K\|_\infty \text{Lip}(K) U_{n,\ell}(R), \end{aligned} \quad (6.9)$$

where $U_{n,\ell}(R) = \frac{1}{n} \sum_{k=R+1}^{n-1} \sum_{p=1}^n \left(\frac{h_{k+p}}{h_n}\right)^{-2\ell-1} \eta_k$.

From (6.8) and (6.9), we obtain

$$\Delta_n(x) \leq \sup_{k \geq 1} \|g_{k+p,p}\|_\infty \frac{R\beta_{n,2(1-\ell)}}{n\beta_{n,(1-\ell)}^2} + \frac{2\|K\|_\infty \text{Lip}(K) U_{n,\ell}(R)}{nh_n^3 \beta_{n,(1-\ell)}^2}.$$

The assumptions (H.2) and (F.2) give the result and lemma 1 is then proved.

We now come back to the proof of the proposition. Since $\lim_{R \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} R^{2+\nu} U_{n,\ell}(R) = 0$, if we choose $R = \lfloor h_n^{-\frac{3}{3+\nu}} \rfloor$, with $\nu > 0$, we obtain

$$\Delta_n(x) \leq \frac{1}{nh_n} \frac{\sup_{k \geq 1} \|g_{k+p,p}\| \beta_{n,2(1-\ell)}}{h_n^{\frac{-\nu}{3+\nu}} \beta_{n,(1-\ell)}^2} + \frac{1}{nh_n} \frac{4\|K\|_\infty \text{Lip}(K)}{h_n^{\frac{-\nu}{3+\nu}} \beta_{n,(1-\ell)}^2}$$

Consequently, assumptions (H.2) and (F.2) give

$$\Delta_n(x) = o\left(\frac{1}{nh_n}\right). \quad (6.10)$$

Part (i) of Proposition 8 is then proved.

The second part is also proved, in the same manner, by choosing $R = \lfloor h_n^{-1} \rfloor$.

Proof of Corollary 1.

The bias is not affected by the dependence. It is analytical, in fact, we have

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

using the Taylor expansion, the assumptions (F.1), (H.2), (K.1)

$$\begin{aligned} &h_n^{-2} (\mathbb{E}(f_n^\ell(x)) - f(x)) \\ &= \frac{1}{2} \frac{\beta_{n,3-\ell}}{\beta_{n,1-\ell}} \int v^2 K(v) f''(x - \theta h_i v) dv, \quad 0 < \theta < 1. \end{aligned}$$

The dominated convergence theorem leads to

$$\lim_{n \rightarrow \infty} \int v^2 K(v) f''(x - \theta h_i v) dv = f''(x) \int v^2 K(v) dv, \quad (6.11)$$

Now applying Lemma 3

$$\lim_{n \rightarrow \infty} h_n^{-2} (\mathbb{E}(f_n^\ell(x)) - f(x)) = \frac{\beta_{3-\ell}}{\beta_{1-\ell}} f''(x) \int v^2 K(v) dv. \quad (6.12)$$

The above approximations for bias and variance lead to the expression of the *MSE*.

Of central importance is the way in which *MSE* changes as a function of h_n , for small values of h_n the second term of *MSE* becomes larger but as h_n gets larger so the first term increases. There is an optimal bandwidth $h_{n,opt}$ which minimizes *MSE* given by

$$h_{n,opt} = \arg \min_{h_n} MSE(h_n) = \left(\frac{\beta_{1-\ell}^2 f(x) \int K^2(x) dx}{\beta_{3-\ell}^2 (f^{(2)}(x) \int x^2 K(x) dx)^2} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Using assumptions (H.2), (K.1), (K.2) and (F.1), we get

$$h_{n,opt} = cn^{-\frac{1}{5}}, \quad c < \infty.$$

Proof of Corollary 2.

For the Bernoulli shift with arithmetic coefficients $a_i = O(i^{-A})$, $A > 5$, the behavior of the coefficient η_k depends on the coefficients a_i . Indeed, in [15], it is proved that

$$\eta_k \leq \sum_{i \geq k} a_i.$$

Using the arithmetic decay above, we get $\eta_k = O(k^{1-A})$, with $A > 5$. The conditions of Proposition 8 are satisfied, in fact for some $0 < \nu < 1$,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} N^{2+\nu} U_{n,\ell}(N) &\leq \limsup_{N \rightarrow \infty} \frac{N^{2+\nu}}{A-1} \sum_{k \geq N+1} k^{1-A} \\ &= \lim_{N \rightarrow \infty} \frac{N^{4+\nu-A}}{(A-1)(A-2)} = 0. \end{aligned}$$

The proof of Corollary 2 is obvious from Corollary 1. The choice $h_n = n^{-\frac{1}{5}}$ gives $\lim_{n \rightarrow \infty} \frac{1}{n} \beta_{n,r} = \frac{5}{5-r} < \infty$, satisfying then assumption (H.2). We deduce that,

$$\lim_{n \rightarrow \infty} n^{\frac{4}{5}} \mathbb{E}(f_n^\ell(x) - f(x))^2 = \left(\frac{(4+\ell)f^{(2)}(x) \int v^2 K(v) dv}{2+\ell} \right)^2 + \frac{(4+\ell)^2}{20+10\ell} f(x) \int K^2(x) dx.$$

Proof of Corollary 3.

Arguing as in Corollary 2, the decay of the coefficients $a_i = O(\exp(-\lambda i))$, ($\lambda > 0$) ensures that $\eta_k = O(\exp(-\lambda k))$, assumption (D.1) of Theorem 6 is then satisfied.

Now if we take $h_n = n^{-\frac{1}{5}}$, assumptions (H.1) and (H.3) are satisfied and the proof of Corollary 3 is then completed.

Proof of Corollary 4.

Assumptions of Proposition 8 are satisfied, in fact, the joint densities $f_{(X_s, X_t)}$ of our model are bounded (see [36]) and assumption (F.2) is then checked. Similarly to the proof of Corollary 2, we get the result.

Proof of Corollary 5.

According to the choice (5.4) in our model, it exists an integer $N = 2$ such that

$$a_i = b_i = 0, \text{ for } i \geq N,$$

then the coefficient of dependence $\eta_k = O(\exp(-\lambda k))$, for some $\lambda > 0$ which satisfies exactly the assumption (D.1), the result is a direct application of Theorem 6.

Proof of Theorem 6. To prove the central limit theorem for weakly dependent sequences, we refer to the standard techniques used for i.i.d. sequences.

For this, we consider two integer-valued sequences $(p_n)_{n \in \mathbb{N}}$, $(q_n)_{n \in \mathbb{N}}$ such that

$$\lim_{n \rightarrow \infty} p_n = \infty, \quad \lim_{n \rightarrow \infty} q_n = \infty, \quad \text{and } p_n = o(n), \quad q_n = o(p_n).$$

Define the Bernstein blocks, for $m = 1, \dots, r_n$ and $r_n = \lfloor \frac{n}{p_n + q_n} \rfloor$ by

$$I_m = [(m-1)(p_n + q_n) + 1, (m-1)(p_n + q_n) + p_n] \cap \mathbb{N},$$

$$I'_m = [(m-1)(p_n + q_n) + p_n + 1, (m-1)(p_n + q_n) + p_n + q_n] \cap \mathbb{N}.$$

We now consider the following decomposition

$$S_n = \sqrt{nh_n} \left(f_n^\ell(x) - E f_n^\ell(x) \right) = S_{n1} + S_{n2} + S_{n3},$$

where

$$\begin{aligned} S_{n1} &= \sum_{m=1}^{r_n} T_m, \quad \text{with, } T_m = \sum_{j \in I_m} \psi_{n,j}, \\ S_{n2} &= \sum_{m=1}^{r_n} T'_m, \quad \text{with, } T'_m = \sum_{j \in I'_m} \psi_{n,j}, \\ S_{n3} &= \sum_{j=N+1}^n \psi_{n,j}, \quad N := r_n(p_n + q_n), \end{aligned}$$

and

$$\psi_{n,j} = \left(\frac{h_n^{2\ell-1}}{n} \right)^{\frac{1}{2}} \frac{h_j^{-\ell}}{\beta_{n,(1-\ell)}} \left(K\left(\frac{x - X_j}{h_j}\right) - \mathbb{E}K\left(\frac{x - X_j}{h_j}\right) \right).$$

The required asymptotic normality follows once the following statements are proved :

(i) S_n and S_{n1} have the same limiting behavior.

(ii) $\lim_{n \rightarrow \infty} \left| \mathbb{E} \exp(itS_{n1}) - \prod_{m=1}^{r_n} \mathbb{E} \exp(it(T_m)) \right| = 0.$

$$(iii) \quad \forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \sum_{m=1}^{r_n} \frac{1}{\text{Var}(S_{n1})} \mathbb{E} \left(T_m^2 \mathbb{I}_{\{|T_m| > \epsilon \sqrt{\text{Var}(S_{n1})}\}} \right) = 0.$$

For the proof of the statement (i), we write,

$$\begin{aligned} & |\mathbb{E}(\exp itS_n) - \mathbb{E}(\exp itS_{n1})| \\ & \leq \mathbb{E}(|\exp itS_n - \exp itS_{n1}| \mathbb{I}_{|S_n - S_{n1}| < \epsilon}) + E(|\exp itS_n - \exp itS_{n1}| \mathbb{I}_{|S_n - S_{n1}| \geq \epsilon}) \\ & \leq \epsilon |t| \mathbb{E}(\mathbb{I}_{|S_n - S_{n1}| < \epsilon}) + 2\mathbb{E}(\mathbb{I}_{|S_n - S_{n1}| \geq \epsilon}) \\ & \leq \epsilon |t| + 2\mathbb{P}(|S_n - S_{n1}| \geq \epsilon) \\ & \leq \epsilon |t| + \frac{2\text{Var}(S_n - S_{n1})}{\epsilon^2}. \end{aligned} \tag{6.13}$$

We have,

$$\text{Var}(S_n - S_{n1}) = \text{Var}(S_{n2}) + \text{Var}(S_{n3}) + 2\text{Cov}(S_{n2}, S_{n3}). \tag{6.14}$$

For the first term,

$$\begin{aligned} \text{Var}(S_{n2}) &= \sum_{m=1}^{r_n} \text{Var}(T'_m) + 2 \sum_{1 \leq i < j \leq r_n} \text{Cov}(T'_i, T'_j) \\ &= \sum_{m=1}^{r_n} \sum_{i \in I'_m} \text{Var}(\psi_{n,i}) + 2 \sum_{m=1}^{r_n} \sum_{i < j, i, j \in I'_m} \text{Cov}(\psi_{ni}, \psi_{nj}) + 2 \sum_{1 \leq i < j \leq r_n} \sum_{s \in I'_i} \sum_{t \in I'_j} \text{Cov}(\psi_{ns}, \psi_{nt}) \\ &= E_{1,n} + E_{2,n} + E_{3,n}. \end{aligned}$$

Because h_n is decreasing, we obtain the following control for E_1

$$E_{1,n} = \frac{h_n^{2\ell-1}}{n\beta_{n,1-\ell}^2} \sum_{m=1}^{r_n} \sum_{i \in I'_m} \text{Var} \left(h_i^{-\ell} K \left(\frac{x - X_i}{h_i} \right) \right) \leq \frac{\|K\|_\infty^2 r_n q_n}{nh_n \beta_{n,(1-\ell)}^2}.$$

The last bound, together with assumption (H.2), gives $E_{1,n} = O(\frac{q_n}{p_n h_n})$. To control the term $E_{2,n}$, we use again assumption (H.2),

$$|E_{2,n}| \leq \sum_{m=1}^{r_n} \sum_{i, j \in I'_m, i < j} |\mathbb{E}(\psi_{ni} \psi_{nj})| \leq \frac{\|K\|_\infty^2 r_n q_n^2}{nh_n \beta_{n,(1-\ell)}^2} = O\left(\frac{q_n^2}{p_n h_n}\right).$$

For the third term

$$\begin{aligned} E_{3,n} &= 2 \sum_{1 \leq i < j \leq r_n} \text{Cov} \left(\sum_{s \in I'_i} \psi_{ns}, \sum_{t \in I'_j} \psi_{nt} \right) \\ &= \frac{2h_n^{2\ell-1}}{n\beta_{n,(1-\ell)}^2} \sum_{1 \leq i < j \leq r_n} \text{Cov} \left(\sum_{s \in I'_i} \xi_s, \sum_{t \in I'_j} \xi_t \right), \end{aligned}$$

where

$$\xi_s := h_s^{-\ell} \left(K\left(\frac{x - X_s}{h_s}\right) - \mathbb{E}K\left(\frac{x - X_s}{h_s}\right) \right). \quad (6.15)$$

Let $k = j - i$, $1 \leq k \leq r_n - 1$. Using the η -weak dependence, we get

$$\begin{aligned} E_{3,n} &\leq \frac{2h_n^{2\ell-1}}{n\beta_{n,(1-\ell)}^2} \sum_{k=1}^{r_n-1} \sum_{j=1}^{r_n} \sum_{s \in I'_j} \sum_{t \in I'_j} h_s^{-\ell} h_t^{-\ell} (h_s^{-1} \|K\|_{\infty} \text{Lip}(K) + h_t^{-1} \|K\|_{\infty} \text{Lip}(K)) \eta_{k(p_n+q_n)}. \end{aligned}$$

We deduce since $h_s^{-\ell} < h_t^{-\ell}$ and using assumption (D.1),

$$\begin{aligned} E_{3,n} &\leq \frac{4h_n^{-2} \|K\|_{\infty} \text{Lip}(K)}{n\beta_{n,(1-\ell)}^2} \sum_{k=1}^{r_n-1} \sum_{j=1}^{r_n} \sum_{s \in I'_j} \sum_{t \in I'_j} \left(\frac{h_t}{h_n}\right)^{-2\ell-1} \exp(-\lambda k q_n). \\ &\leq \frac{4h_n^{-2} \|K\|_{\infty} \text{Lip}(K) (q_n - 1)^2 r_n \exp(-\lambda q_n) (1 - \exp(-\lambda q_n (r_n + 1)))}{n\beta_{n,(1-\ell)}^2 (1 - \exp(-\lambda q_n))} = O\left(\frac{q_n^2}{p_n h_n^2}\right). \end{aligned}$$

From the previous bounds of $E_{1,n}$, $E_{2,n}$, $E_{3,n}$, we should choose two sequences p_n and q_n in such a way that,

$$\begin{cases} \lim_{n \rightarrow \infty} \frac{q_n}{p_n h_n} = 0, \\ \lim_{n \rightarrow \infty} \frac{q_n^2}{p_n h_n} = 0, \\ \lim_{n \rightarrow \infty} \frac{q_n^2}{p_n h_n^2} = 0. \end{cases} \quad (6.16)$$

If so we obtain,

$$\lim_{n \rightarrow \infty} \text{Var}(S_{n2}) = 0.$$

Our purpose now is to control $\text{Var}(S_{n3})$. For this we use the same arguments as in [2] (page 39). We have,

$$\text{Var}(S_{n3}) = \sum_{j=N+1}^n \text{Var}(\psi_{n,j}) + 2 \sum_{N+1 \leq i < j \leq n} \text{Cov}(\psi_{n,i}, \psi_{n,j}) = I_{1,n} + I_{2,n}. \quad (6.17)$$

First consider the term $I_{1,n}$,

$$\sum_{j=1}^N \text{Var}(\psi_{n,j}) + I_{1,n} = \sum_{j=1}^n \text{Var}(\psi_{n,j}) \sim nh_n \text{Var}(f_n^\ell(x)), \quad (6.18)$$

as n tends to infinity. To explain the last equivalence, we use (6.10) and we obtain,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{nh_n \text{Var}(f_n^\ell(x))}{\sum_{j=1}^n \text{Var}(\psi_{n,j})} \\ &= \lim_{n \rightarrow \infty} \frac{nh_n (\sum_{i=1}^n h_i^{1-\ell})^{-2} \left(\sum_{i=1}^n h_i^{-2\ell} \text{Var}(K(\frac{x-X_i}{h_i})) + 2 \sum_{1 \leq i < j \leq n} A_{i,j} \right)}{nh_n (\sum_{j=1}^n h_j^{1-\ell})^{-2} \sum_{j=1}^n h_j^{-2\ell} \text{Var}(K(\frac{x-X_j}{h_j}))} = 1, \end{aligned}$$

recall that $A_{i,j}$ is already defined in (6.5).

In the same way, since $N \sim n$, under assumption (H.3) and using Proposition 8, we prove that

$$\sum_{j=1}^N \text{Var}(\psi_{n,j}) \sim \left(\frac{nh_n}{Nh_N} \right) Nh_N \text{Var} f_N^\ell(x) \longrightarrow \sigma^2 \quad \text{as } n \rightarrow \infty,$$

recall also that $\sigma^2 = \frac{\beta_{1-2\ell}}{\beta_{1-\ell}^2} f(x) \int K^2(x) dx$.

Consequently we obtain then from (6.18),

$$\lim_{n \rightarrow \infty} I_{1,n} = 0.$$

Now we turn our attention to the term $I_{2,n}$. We have,

$$I_{2,n} \leq \frac{nh_n}{(\sum_{i=1}^n h_i^{1-\ell})^2} \sum_{N+1 \leq i < j \leq n} |\text{Cov}(\xi_i, \xi_j)|,$$

where ξ_i is defined in the relation (6.15).
From (6.10), since $N \sim n$, we get

$$\lim_{n \rightarrow \infty} I_{2,n} = 0.$$

Equality (6.17) yields then $\lim_{n \rightarrow \infty} \text{Var}(S_{n3}) = 0$.

Now by Schwartz inequality we obtain

$$\lim_{n \rightarrow \infty} \text{Cov}(S_{n2}, S_{n3}) = 0.$$

Part (i) is then deduced from (6.13).

The purpose of Part (ii) is to prove that the sequence $(T_m)_m$ is asymptotically independent. For this, we need, arguing as in [16], the following lemma.

Lemma 6. *If the stationary sequence $(X_n)_n$ is η -weakly dependent then*

$$\left| \mathbb{E}(\exp(it \sum_{m=1}^{r_n} T_m)) - \prod_{m=1}^{r_n} \mathbb{E}(\exp(itT_m)) \right| \leq 2Lip(K)|t| \frac{r_n(r_n+1)p_n\eta_{q_n}}{\beta_{n,1-\ell}\sqrt{nh_n^3}}$$

Proof of Lemma 6. We have,

$$\begin{aligned} & \left| \mathbb{E}(\exp(it \sum_{m=1}^{r_n} T_m)) - \prod_{m=1}^{r_n} \mathbb{E}(\exp(itT_m)) \right| \tag{6.19} \\ & \leq \left| \text{Cov}(\exp(it \sum_{m=1}^{r_n-1} T_m), \exp(itT_{r_n})) \right| + \left| \mathbb{E}(\exp(it \sum_{m=1}^{r_n-1} T_m)) - \prod_{m=1}^{r_n-1} \mathbb{E}(\exp(itT_m)) \right|. \end{aligned}$$

Let us note that $\exp(it \sum_{m=1}^{r_n-1} T_m)$ can be written as a function f depending on the arguments

$$(X_1, \dots, X_{p_n}, X_{p_n+q_n+1} + \dots + X_{2p_n+q_n}, \dots, X_{(r_n-2)(p_n+q_n)+1}, \dots, X_{(r_n-2)(p_n+q_n)+p_n}).$$

In the same way, the quantity $\exp(itT_{r_n})$ is a function g depending on

$$(X_{(r_n-1)(p_n+q_n)+1}, \dots, X_{(r_n-1)(p_n+q_n)+p_n}),$$

such that

$$\begin{aligned} & f(X_1, \dots, X_{p_n}, X_{p_n+q_n+1} + \dots + X_{2p_n+q_n}, \dots, X_{(r_n-2)(p_n+q_n)+1}, \dots, X_{(r_n-2)(p_n+q_n)+p_n}) \\ & = \exp(it \sum_{m=1}^{r_n-1} T_m) \end{aligned}$$

and that

$$g(X_{(r_n-1)(p_n+q_n)+1}, \dots, X_{(r_n-1)(p_n+q_n)+p_n}) = \exp(itT_{r_n}).$$

We can check that

$$\text{Lip}(g) \leq 2|t| \left(\frac{\text{Lip}(K)}{\beta_{n,1-\ell}} \right) \left(\frac{h_n^{2\ell-1}}{n} \right)^{\frac{1}{2}} \sum_{j \in I_{r_n}} h_j^{-\ell-1} \leq \frac{2|t| \text{Lip}(K) p_n}{\beta_{n,1-\ell} \sqrt{nh_n^3}}.$$

and that

$$\text{Lip}(f) \leq \frac{2|t| \text{Lip}(K)}{\beta_{n,1-\ell}} \left(\frac{h_n^{2\ell-1}}{n} \right)^{\frac{1}{2}} \sum_{m=1}^{r_n-1} \sum_{j \in I_m} h_j^{-\ell-1} \leq \frac{2|t| r_n p_n \text{Lip}(K)}{\beta_{n,1-\ell} \sqrt{nh_n^3}}.$$

The η -weak dependence condition gives then

$$\left| \text{Cov} \left(\exp(it \sum_{m=1}^{r_n-1} T_m), \exp(itT_{r_n}) \right) \right| \leq \frac{2p_n \text{Lip}K |t|}{\beta_{n,1-\ell}} \left(\frac{r_n + 1}{\sqrt{nh_n^3}} \right) \eta_{q_n}.$$

Iterating the formula (6.19) and the last inequality, we obtain

$$\left| \mathbb{E}(\exp(it \sum_{m=1}^{r_n-1} T_m)) - \prod_{m=1}^{r_n} \mathbb{E}(\exp(itT_m)) \right| \leq \frac{2r_n p_n \text{Lip}K |t|}{\beta_{n,1-\ell}} \left(\frac{r_n + 1}{\sqrt{nh_n^3}} \right) \eta_{q_n}.$$

Lemma 6 is then proved.

To prove part (ii), we use assumptions (H.2) and (D.1). We obtain,

$$\left| \mathbb{E}(\exp(it \sum_{m=1}^{r_n} T_m)) - \prod_{m=1}^{r_n} \mathbb{E}(\exp(itT_m)) \right| = O \left(\frac{2n^2 \exp(-\lambda q_n)}{p_n \sqrt{nh_n^3}} \right).$$

To get (ii), the sequences (p_n) and (q_n) must fulfil the additional condition

$$\lim_{n \rightarrow \infty} \frac{2n^2 \exp(-\lambda q_n)}{p_n \sqrt{nh_n^3}} = 0. \tag{6.20}$$

For the point (iii), we note first that the sequence $(T_m)_m$ is bounded. Indeed,

$$|T_m| \leq \frac{\|K\|_{\infty} p_n}{\beta_{n,(1-\ell)} \sqrt{nh_n}}.$$

Consequently, using Markov inequality, we get

$$\begin{aligned}
& \sum_{1 \leq m \leq r_n} \frac{1}{\text{Var}(S_{n1})} \mathbb{E} \left(T_m^2 \mathbb{I}_{\{|T_m| > \epsilon \sqrt{\text{Var}(S_{n1})}\}} \right) \\
& \leq \frac{p_n^2 \|K\|_\infty^2}{nh_n \beta_{n,(1-\ell)}^2 \text{Var}(S_{n1})} \sum_{1 \leq m \leq r_n} P \left(|T_m| > \epsilon \sqrt{\text{Var}(S_{n1})} \right) \\
& \leq \frac{p_n^2 \|K\|_\infty^2}{nh_n \epsilon^2 \beta_{n,(1-\ell)}^2 \text{Var}(S_{n1})}.
\end{aligned}$$

Let us check that Proposition 8 leads to,

$$\lim_{n \rightarrow \infty} \text{Var}(S_{n1}) = \lim_{n \rightarrow \infty} \text{Var}(S_n) = \sigma^2.$$

In fact,

$$\text{Var}(S_n) = \text{Var}(S_{n1}) + \text{Var}(S_{n2}) + \text{Var}(S_{n3}) + 2(\text{Cov}(S_{n1}, S_{n2}) + \text{Cov}(S_{n2}, S_{n3}) + \text{Cov}(S_{n1}, S_{n3})).$$

We have already shown that

$$\lim_{n \rightarrow \infty} \text{Var}(S_{n2}) = 0, \quad \lim_{n \rightarrow \infty} \text{Var}(S_{n3}) = 0, \quad \lim_{n \rightarrow \infty} \text{Cov}(S_{n2}, S_{n3}) = 0.$$

Using Schwartz inequality again we obtain

$$\lim_{n \rightarrow \infty} \text{Cov}(S_{n1}, S_{n2}) = 0, \quad \lim_{n \rightarrow \infty} \text{Cov}(S_{n1}, S_{n3}) = 0$$

Coming back now to Lindberg (iii), we get the following bound

$$\sum_{1 \leq m \leq r_n} \frac{1}{\text{Var}(S_{n1})} \mathbb{E} \left(T_m^2 \mathbb{I}_{\{|T_m| > \epsilon \sqrt{\text{Var}(S_{n1})}\}} \right) = O\left(\frac{p_n^2}{nh_n}\right). \quad (6.21)$$

Consequently Condition (iii) is satisfied as soon as

$$\lim_{n \rightarrow \infty} \frac{p_n^2}{nh_n} = 0. \quad (6.22)$$

Finally, Theorem 6 is proved if we can construct sequences (p_n) and (q_n) fulfilling the limits (6.16), (6.20) and (6.22). Under assumption (H.1), one can take $p_n = \lfloor n^{2(\alpha+\beta)} \rfloor$, $q_n = \lfloor n^\beta \rfloor$, with $0 < \alpha + \beta < \frac{1}{4}$. \square

Perspectives de recherche

1-Etude des estimateurs de Parzen et Nadaraya dans le cas des variables fonctionnelles faiblement dépendantes

Soit $(X_i, Y_i)_{1 \leq i \leq n}$ une suite de couples aléatoires η -faiblement dépendants où Y une variable aléatoire réelle, et X est une variable aléatoire fonctionnelle c'est à dire définie sur l'espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$ ayant des valeurs sur un espace vectoriel $(H, B_{\|\cdot\|})$ muni de la semi norme $\|\cdot\|$.

Sous l'hypothèse ci-dessus, On peut toujours faire de l'inférence sur le modèle de régression

$$Y = r(X) + \epsilon.$$

L'étude pourrait commencer par répondre à la question suivante :

Si $(X_i)_{i \leq n}$ est une variable aléatoire faiblement dépendante, est ce que $(\|X_i\|)_{i \leq n}$ est faiblement dépendante aussi ?

Cela peut être vérifié si on montre que la sous tribu de \mathcal{A} engendrée par $\|X\|$ est incluse dans celle engendrée par X , ceci entraîne que le coefficient de dépendance $\eta_{\|X\|}$ de la suite $(\|X_i\|)_{i \leq n}$ soit inférieur au coefficient η_X de la suite $(X_i)_{i \leq n}$. Les propriétés de l'estimateur fonctionnel de la regression $r(x)$ donné par

$$r_n(x) = \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{\|x-X_i\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h_n}\right) \neq 0, \\ \frac{1}{n} \sum_{i=1}^n Y_i & \text{si } \sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h_n}\right) = 0. \end{cases}$$

Ce nouvel cadre d'étude conduit à de nouvelles hypothèses sur la distribution de la variable fonctionnelle et non pas sur sa densité ce qui rend l'étude

difficile par rapport au cas réel. En revanche l'estimateur fonctionnel s'impose lorsque le nombre de régresseurs est assez important devant le nombre des observations, ou dans le cas des régresseurs très corrélés.

D'après Ferraty et Vieu (2006), les résultats de l'étude de l'estimateur fonctionnel de la régression sont fortement inspirés de celui de la régression classique. Le cas α -mélangeant est traité dans (Ferraty et Vieu (2006)), donc une extension au cas plus général au sens de Doukhan et Louhichi (1999) peut être traité aussi.

2-Extension de l'étude de l'estimateur récursif de la régression dans le cas faiblement dépendant.

Comme pour la densité, Amiri (2010) définit l'estimateur récursif $r_n^\ell(x)$ de la régression donné par

$$r_n^\ell(x) = \frac{g_n^\ell(x)}{f_n^\ell(x)},$$

avec

$$g_n^\ell(x) = \frac{1}{\sum_{i=1}^n h_i^{1-\ell}} \sum_{i=1}^n \frac{1}{h_i^\ell} Y_i K\left(\frac{x - X_i}{h_i}\right), \quad x \in \mathbb{R}.$$

Il étudie les propriétés asymptotiques de cet estimateur sous l'hypothèse de l'indépendance puis sous l'hypothèse du α -mélange, dans [32], nous avons étudié des propriétés de l'estimateur récursif de la densité sous l'hypothèse de la η -dépendance, dans un futur travail, nous comptons étendre encore ces résultats pour la fonction de régression.

Un estimateur fonctionnel récursif a été aussi défini et investigué par Amiri (2011), une extension peut être aussi établie pour des variables fonctionnelles faiblement dépendantes.

Résumé

Titre :

Inférence statistique dans les modèles de régression non paramétrique

Nous avons choisi d'étudier l'inférence statistique dans des modèles de régression non paramétriques via l'étude de l'estimateur à noyau de la densité et de la régression sous différentes hypothèses. Ce travail se décline ainsi en deux parties :

Les propriétés asymptotiques des estimateurs de la densité et de la régression sont étudiées d'abord sous l'hypothèse de l'indépendance, dans la première partie de la thèse, et notamment, la convergence, la normalité et le choix de la fenêtre de lissage.

Dans la deuxième partie, nous allons vers la notion généralisée de la dépendance faible établie par Doukhan et Louhichi (1999) pour étendre les propriétés de convergence et de normalité, cela permet de traiter les séries temporelles qui s'adaptent bien avec cette nouvelle notion de dépendance faible, nous fournissons alors des exemples de modèles faiblement dépendants, nous présentons après des résultats sur un estimateur à noyau récursif sous la dépendance faible, une étude par simulation est effectuée sur des modèles faiblement dépendants.

Mots clés :

Régression non paramétrique; Estimateurs à noyau de la densité et de la régression;
Indépendance dépendance faible; convergence; normalité

ABSTRACT

Title :

Statistic inference in nonparametric regression models

We have chosen to investigate the statistic inference in nonparametric regression models by studying kernel densities and regression estimators under different assumptions .

This work is organised in two parts :

The asymptotic properties of density and regression estimators are studied first under independence assumption, particularly, convergence, normality and the choice of the smoothing window.

In the second part, we go the generalized notion of weak dependence established by Doukhan and Louhichi (1999) to extend convergence and normality properties. This allows to deal with time series, we show after, our result on a recursive kernel estimator under weak dependence, convergence in mean square error, and normality are obtained. Simulation study is done on weak dependent models.

Key Words :

non parametric regression ; kernel densities and regression estimators; Independence weak dependence; convergence; normality

ملخص

العنوان :

الإنحدار غير الوسيط الاستدلالي الإحصائي في نماذج

اخترنا أن نسلط الضوء على الاستدلالي الإحصائي في نماذج الانحدار غير الوسيط عن طريق دراسة المقدر ذو نواة لدالة الكثافة و الانحدار تحت فرضيات مختلفة.

ينقسم هذا العمل إلى جزئين

في الجزء الأول ندرس خصائص مقدرات الكثافة و الانحدار تحت فرضية استقلالية المتغيرات و بصفة خاصة التقارب الطبيعية و اختيار وسيط التقدير

في الجزء الثاني نتجه نحو المفهوم المعمم للتعلق الضعيف المؤسس من طرف

Doukhan, Louhichi (1999)

و ذلك لتمديد الخصائص المدروسة في الجزء الأول . هذا يسمح بمعالجة السلاسل الزمنية التي تتكيف مع هذا المفهوم الجديد للتعلق الضعيف.

نقدم أمثلة لنماذج متعلقة بضعف ثم نستعرض نتائجنا حول مقدر تراجعي تحت فرضية التعلق الضعيف. في الأخير برمجت أمثلة لنماذج متعلقة بضعف بالإعلام الآلي.

الكلمات المفتاح :

إنحدار غير وسيطي -مقدرات ذات نواة لدالة الكثافة و الإنحدار -الإستقلالية الإرتباط الضعيف -
التقارب - الطبيعية

Bibliographie

- [1] Ango NZE. P., Buhlmann P., Doukhan P. (2002). Weak dependence Beyond Mixing and asymptotics for nonparametric regression. *Ann. Statist. Vol. 28, No. 2, p.325-341.*
- [2] Amiri, A.(2010). Estimateurs fonctionnels récurrents et leurs applications à la prévision. *Thèse de doctorat. Académie D'Aix-Marseille, Université D'Avignon et des pays de Vaucluse.*
- [3] Bardet J.M, Doukhan P., Léon J.R. (2008). *A functional limit theorem for η -weakly dependent processes and its applications.* Stat. Infer. Stoch. Process, **11**, p.265-280.
- [4] Bardet J.M., Doukhan, P., Lang G., and Ragache N. (2008). *Dependent Lindberg Central Limit Theorem and some applications.* ESAIM. Probability and Statistics. **12**, p. 154-172.
- [5] Blondin D. (2004). Lois limites uniformes et estimation non paramétrique de la régression . *Thèse de Doctorat. Université de Paris 6.*
- [6] Bosq, D. (1999). *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction.* Lecture Notes in Statistics. Springer-Verlag New York.
- [7] Cao, R. (2001). *Relative efficiency of local bandwidths in kernel density estimation.* Ann. Statist. **35**, p. 113-137.
- [8] Debarh M. (2007). *Some uniform results in additive regression model.* Stat. Infer. Stoch. Process, **11**, 265-280.
- [9] Dedecker, J., Doukhan, P., Gabriel, L., Léon, J. R., Louhichi, S., and Prieur, C. (2007). *Weak Dependence : With Examples and Applications, Lecture Notes in Statistics, Vol. 190.*
- [10] Deheuvels, P. (1973). *Sur l'Estimation séquentielle de la densité.* C. R. Acad. Sci., Paris, Ser. A, **276**, 1119-1121.

- [11] 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**, p.1217-1220.
- [12] Deheuvels, P., Mason D. (2004). *General asymptotic confidence bands based on kernel-type function estimators*. Stat. Inference Stoc. Processes, **7.3**, p.225- 277.
- [13] Deheuvels, P. (1977). *Estimation non paramétrique de la densité par histogrammes généralisés*, Revue de Statistique Appliquée, **XXV**, p.5-42.
- [14] Doukhan, P.(1994). *Mixing : Properties and Examples*, in : Lecture Notes in Statistics, vol. **85**, Springer-Verlag.
- [15] Doukhan, P. and Louhichi, S. (1999). *A new weak dependence condition and applications to moment inequalities*. Stochastic Processes and their Applications, **84**, p.313- 342.
- [16] Doukhan, P. and Louhichi, S. (2001). *Functional Estimation of a density under a new weak dependance condition*. Scand. J.Statist. . Vol. **28**, no **2**, p.325-341.
- [17] Doukhan, P. and Winterberger, O. (2006). *An invariance principle for weakly dependent stationary general models*. Probab. Math. Statist., **27**, 1, p.45-73.
- [18] Doukhan, P. and Neumann M. H. (2007). *Probability and moment inequalities for sums of weakly dependent random variables, with applications*. Elseiver Stochastic Processes and their Applications. **84**.
- [19] Einmahl U. and Mason, D.M (2000). *An empirical process approach to the uniform consistency of kernel type function estimators*. J. Theoretical Prob., **13**,1-37.
- [20] Einmahl U. and Mason, D.M (2005). *An empirical process approach to the uniform consistency of kernel type function estimators*. J. Theoretical Prob., **13**,1-37.
- [21] Emmanuel, R. (1989). *Théorie asymptotique des processus aléatoires faiblement dépendants*. Mathématiques et Applications Vol.31. Springer.
- [22] Ferratty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : theory and practice*. Springer, UK.
- [23] Hardle W. (1990). *Applied Nonparametric Regression* . Cambridge University Press. Cambridge.

- [24] Gangopadhyay, A.K. and Cheung, K.N. (2002). *Bayesian Approach to the choice of smoothing parameter in kernel density estimation*. Journal of nonparametric statistics, no **14**,655 -664.
- [25] Giné E., Guillou A. (2002). *Rates of strong uniform consistency for multivariate kernel density estimators*. Ann. Inst. H. Poincaré **38** p.p 907-921.
- [26] Jones, Marron, Sheather (1996). *A brief Survey of Bandwidth Selection for Density Estimation*. J. Amer. Statist. Assoc, Vol.91. No.433, p. p 401-407.
- [27] Jianqing F., Hall P., Michael A. , Prakash P.(1996). *On Local Smoothing of Nonparametric Curve Estimators*. J. Amer. Statist. Assoc, Vol.91. No. **433**, p.258-266.
- [28] Lang G., Doukhan P. (2002). *Rates in the Empirical central Limit Theorem for Stationary Processes*. Statistical Inference for Stochastic Process. No. **5**, p.199-228.
- [29] Liang, H.Y. and Baek, J. (2004). *Asymptotic normality of recursive density estimates under some dependence assumptions*, *Metrika*, **60**, p.155-166.
- [30] Linton O.B., Nielson J.B. (1995). *A kernel method of estimating structured nonparametric regression based on marginal integration*. *Biometrika*, No. **82**, p.93-100.
- [31] Masry, E. (1986). *Recursive probability density estimation for weakly dependent stationary processes*, *IEEE Trans. Inform. Theory* 32, no **2**, p.254-267.
- [32] Mezhoud, K. A., Mohdeb, Z., Louhichi, S. (2014). *Recursive kernel estimation of the density under weak dependence*. Accepted for publication in *Journal of the Korean Statistical Society*.
- [33] Nadaraya, E. A. (1965). *On nonparametric estimation of density function and regression*. *Theory. Probab. Appl*, **10**, 186-190.
- [34] Nemouchi N., Mohdeb Z. (2010). *Asymptotic Confidence Bands for Density and Regression Functions in the Gaussian Case*. *Afr. Stat. Vol. 5*, no **1** , p.279-287.
- [35] Prakasa, R. (1998). *Nonparametric Functionnal Estimation*. *Indian Statistical Institute*.

- [36] Ragache, N., Wintenberger, O. (2006). *Convergence rates for density estimators of weakly dependent time series. Dependence in probability and statistics. (Eds P.Bertail, P.Doukhan and P. Soulier). Lectures Notes in Statistics. 187. p.349-372.*
- [37] Rio, P.M. (1983). *Nonparametric estimators for time series. J. time Ser.Anal. 4, p.185-207.*
- [38] Robinson, P.M. (1983). *Nonparametric estimators for time series. J. time Ser.Anal. 4, p.185-207.*
- [39] Scott et Terrell (1987). *Biased and Unbiased Cross Validation i Density Estimation. Journal of the American Statistical Association, Vol.82. No. 400.*
- [40] Silverman B.W.(1986). *Smoothing methods in statistics. Springer Series in Statistics. New York.*
- [41] Siminoff S. Jeffrey.(1996). *Density estimation for statistics and data analysis. Chapman and Hall.*
- [42] Stone, J.C. (1982). *Optimal global rates of convergence for nonparametric regression. Ann. Statist. No. 10, p.1040-1053.*
- [43] Stone, J.C. (1985). *Additive regression and other nonparametric models. Ann. Statist. No 13, p.698-705.*
- [44] Tran, L. T. (1989). *Recursive density estimation under dependence, IEEE Trans. Inform. Theory, 35, p.1103-1108.*
- [45] Tsybakov, A. B. (2003). *Introduction à l'estimation non-paramétrique. In : SMAI, Mathématiques et Applications. Vol.41.*
- [46] Vilar-Fernandez J.M. and Vilar-Fernandez J.A. (2000). *Recursive local polynomial regression under dependence conditions. Sociedad de Estadística e Investigacion Operativa Test. Vol. 9, No.1, pp. 209-232.*
- [47] Wegman, E.J. and Davies, H. I. (1979). *Remarks on some recursive estimators of a probability density, Ann. Statist.7 (2), p.316-327.*
- [48] Wolverson, C. and Wagner, T.J. (1969). *Recursive estimates of probability densities, IEEE Trans. Syst. Cybern. 5, 307-308.*
- [49] Wu, W. (2005). *Nonlinear system theory : Another look at dependence. Proceedings of the national academy sciences of the united states,,102, pp. 14150-14154.*

-
- [50] Zhang, X. Brooks, R.D., King, M.L. (2006). *Bayesian Approach to bandwidth selection for multivariate kernel density estimation*. Computational statistics and data Analysis, no **50**, p.3009-3031.
- [51] Zougab, N. (2013). *Approche Bayésienne dans l'estimation non paramétrique de la densité de probabilité et de la courbe de régression de la moyenne*. Thèse de Doctorat. Université de Béjaia, Algérie.
- [52] N. Zougab, S. Adjabi, and C C. Kokonendji. (2012). *A Bayesian approach to bandwidth selection in univariate associate kernel estimation*. Journal of Statistical Theory and Practice. **7** (1), p.p. 8-23.