République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITÉ MENTOURI - CONSTANTINE

Faculté des sciences exactes

Département de mathématiques

N° d'ordre..... :
N° Série..... :

M E M O I R E

Présenté pour obtenir le diplôme de

Magistère en Mathématiques

Option : Probabilité et Statistique

Présenté et soutenu par

M^{elle} **: BENCHOULAK Hadjer**

Thème

Bandes de confiance pour les fonctions de densité et de régression

Mémoire dirigé par : **Nahima Nemouchi**

soutenu en 2012

Devant le jury :

Président	Z. MOHDEB	-	Prof Université Mentouri Constantine.
Rapporteur :	N. NEMOUCHI	-	M.C. (A) Université Mentouri Constantine.
Examinateurs :	F. MESSACI	-	Prof. Université Mentouri Constantine.
	S. BELALOUI	-	M.C. (A) Université Mentouri Constantine.

Table des matières

0.1 Introduction

La théorie de l'estimation est une des préoccupations majeures des statisticiens. Ainsi l'estimation non paramétrique réelle a reçu un intérêt croissant tant sur le plan théorique que pratique. Cette branche de la statistique ne se résume pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon (comme cela est le cas pour la théorie de l'estimation paramétrique), elle consiste généralement à estimer à partir des observations une fonction inconnue, élément d'une certaine classe fonctionnelle, telle que la fonction de densité ou la fonction de régression à titre d'exemples.

Depuis les travaux de Rosenblatt (1956) et Parzen (1962) puis de Nadaraya- Watson (1964) portant respectivement sur les estimateurs non paramétriques des fonctions de la densité et de la régression, la méthode du noyau a été largement utilisée dans de nombreux travaux, citons quelques uns, sans prétendre être exhaustifs : Prakasa Rao (1983), Devroye et Györfi (1985), Silverman (1986), Roussas (1990), Härdle (1990), Scott (1992), Bosq et Lecoutre (1987), Wand et Jones (1995) et les références citées dans ces publications.
En s'appuyant sur l'étude du processus empirique local indéxé par certaines classes de fonctions, Deheuvels et Mason (2004) ont établi, récemment des vitesses de convergence en probabilité pour des déviations de ces estimateurs par rapport à leur espérances. Ce qui leur a permis la construction des bandes de confiance uniformes pour la densité et la régression.
l'objet central de ce mémoire est d'exposer la construction de ces dernières, en mettant en évidence que le passage des intervalles de confiance aux bandes de confiance a été possible grâce aux résultats plus forts de convergence en probabilité plutôt que de convergence en loi. Ces bandes sont données en fonction des estimateurs à noyau de Nadaraya-Watson pour la fonction de régression et de Parzen-Rosenblatt pour la densité. Afin d'atteindre ce but le présent mémoire est subdivisé en cinq chapitres.

Dans le premier chapitre nous introduisons quelques rappels sur les différentes modes de convergences et sur les inégalités exponentielles de type Bernstein qui permettent de contrôler le comportement limite des déviations des estimateurs par rapport à leur espérances. Nous évoquons, dans ce même chapitre, deux méthodes d'estimation non paramétrique de la densité : la méthode d'estimation par histogramme et la méthode d'estimation par noyau (estimateur de Parzen-Rosenblatt) sur laquelle nous nous concentrons le plus et qui peut être considérée comme une extension de l'estimateur par la méthode de l'histogramme.
Nous présentons, au chapitre 2, la méthode d'estimation par noyau de la fonction de régression (estimateur de Nadaraya-Watson), qui n'est autre qu'une généralisation de la méthode d'estimation par régressogramme étudiée brièvement.
Afin d'évaluer la qualité de l'estimation, nous traitons les propriétés asymptotiques de ces estimateurs, à savoir la convergence en moyenne quadratique et la convergence presque complète aussi bien ponctuelle qu'uniforme.

Les résultats de convergence mettent en évidence le rôle du paramètre de lissage. Pour la sélection optimale de ce paramètre , nous considérons au chapitre 3 l'approche globale pour les résultats de type convergence uniforme et l'approche ponctuelle pour les résultats de type convergence ponctuelle. Ces fenêtres minimisant les deux critères d'erreur dépendent de quantités fonctionnelles inconnues, dans le souci de parer à cette difficulté, nous nous sommes limités à étudier la méthode de plug in pour la densité et la validation croisée pour la régression.
Le chapitre 4, quant à lui, présente le coeur de ce travail de ce mémoire, c'est à dire la construction des bandes de confiance asymptotiques basées sur les estimateurs à noyau, de Nadaraya-

Watson pour la fonction de régression et de Parzen- Rosenblatt pour la densité de probabilité. Cette constuction s'appuie sur les lois uniformes du logaritme pour les estimateurs considérés, établies par Deheuvels et Mason (2004). En s'inspirant de ces travaux et en utilisant un résultat important de Härdel (1990), nous montrons que les intervalles d'estimation de la fonction de régression donnés par les bandes de confiance sont plus étroits que ceux donnés par les intervalles de confiance. Par la suite nous introduisons le travail de N. Nemouchi et Z. Mohdeb (2010) sur la constuction des estimateurs de la densité et de la régression asymptotiquement optimaux, sous l'hypothèse que les lois sous-jacentes sont gaussiennes de paramètres inconnus. Dans cet article, les estimateurs des paramètres de lissage, minimisant les critères d'erreurs, vérifient les conditions imposées par Deheuvels et Mason (2004), ceci leur a permis de déduire des bandes de confiance asymptotiques pour la densité de probabilité et pour la fonction de régression. Dans ce chapitre, nous avons aussi développé des applications sur les bandes de confiance pour la fonction de régression, introduites par Kebbabi, Messaci et Nemouchi (2010). Dans le dernier chapitre, des simulations tendent à valider les résultats obtenus par N. Nemouchi et Z. Mohdeb (2010) ainsi que par Nemouchi, Kebabi et Messaci (2010), mettant en évidence la bonne performance des bandes obtenues dès les petites tailles de l'échantillon.

Chapitre 1

Estimation de la fonction de densité

1.1 Introduction

Une méthode de modélisation d'une suite de mesures émanant de la répétition d'une expérience, est de supposer que ces mesures sont des réalisations de variables aléatoires indépendantes équi-distribuées.

La compréhension de ces mesures et la manière dont elles sont distribuées mènent à l'étude de la loi de probabilité sous-jacente.

Par exemple, en médecine pour atténuer le vertige de Ménière (découvert récemment), une association de trois types de médicaments est administrée par voie intraveineuse (un anxiolytique , un anti-vertige et un anti-émétique).

L'étude de l'assimilation de ce traitement sur n patients, revient à mesurer la concentration x_i de ces trois médicaments passée dans le sang du patient.

La modélisation de ces faits observés consiste à supposer que $x_1, x_2, ..., x_n$ sont des réalisations de n variables aléatoires indépendantes $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ ayant la même densité f.

Donc l'étude du processus d'assimilation de ce traitement dans le sang revient à connaître la densité f. Si f est complètement inconnue c'est à dire que l'on n' a aucune idée sur la forme que peut prendre cette densité, on est alors amené à l'estimer.

Dans ce cas, donner un estimateur de f, ne se résume pas a l' estimation de la moyenne et de la variance comme dans le cas gaussien. Il s'agit donc de construire une fonction (qui se trouve dans un espace de dimension infinie), c'est le cas non paramétrique. Nous allons nous intéresser à cet estimateur. Afin de faire l'étude asymptotique de ce dernier, nous énonçons d'abord les définitions, les propositions et les corollaires sur lesquels repose la suite de notre travail.

Définition 1.1.1.

On dit que la suite de variables aléatoires réelles $(\mathrm{X}_n)_{n\in\mathrm{N}}$ *converge presque complètement vers une variable aléatoire* X *lorsque* $n \to \infty$ *(et on note* $\lim_{n\to\infty} \mathrm{X}_n = \mathrm{X}$ *a.co), si et seulement si :*

$$\forall \epsilon > 0, \quad \sum_{n \in \mathrm{N}} \mathrm{P}[|\mathrm{X}_n - \mathrm{X}| > \epsilon] < \infty.$$

Définition 1.1.2.

On dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(\mathrm{X}_n)_{n \in \mathrm{N}}$ *vers* X *est d'ordre* (U_n) *(*(U_n) *étant une suite numérique déterministe), et on note* $\mathrm{X}_n = \mathrm{O}_{a.co}(\mathrm{U}_n)$*, si et seulement si :*

$$\exists \epsilon_0 > 0, \quad \sum_{n \in \mathrm{N}} \mathrm{P}[|\mathrm{X}_n - \mathrm{X}| > \epsilon_0 \mathrm{U}_n] < \infty.$$

Notons que la convergence presque complète entraîne à la fois la convergence presque sûre et la convergence en probabilité.

Proposition 1.1.1.

Si $\lim_{n \longrightarrow \infty} \mathrm{X}_n = \mathrm{X}$ *a.co, alors* X_n *converge en probabilité et presque sûrement vers X.*

preuve :

La convergence en probabilité se déduit facilement de la convergence de la série suivante

$$\sum_{n \in \mathrm{N}} \mathrm{P}[|\mathrm{X}_n - \mathrm{X}| > \epsilon] < \infty,$$

$(\mathrm{P}[|\mathrm{X}_n - \mathrm{X}| > \epsilon)$ *est le terme général d'une série convergente).*

Le lemme de Borel Contelli implique que :

$$\forall \epsilon > 0, \quad \mathrm{P}[\lim_{n \to \infty} \sup |\mathrm{X}_n - \mathrm{X}| > \epsilon] = 0.$$

De plus, $\lim_{n \to \infty} \mathrm{X}_n(\omega) \neq \mathrm{X}(\omega)$*, implique l'existence de* $\epsilon > 0$*, tel que*

$$\lim_{n \to \infty} \sup |\mathrm{X}_n(\omega) - \mathrm{X}(\omega)| > \epsilon,$$

on alors $\mathrm{P}[\lim_{n \to \infty} \mathrm{X}_n = \mathrm{X}] = 1$*, c'est à dire* $\mathrm{X}_n \longrightarrow \mathrm{X}, \quad \mathrm{P.S}.$

Outre le fait que la convergence presque complète est une convergence très forte, elle jouît des propriétés résumées ci dessous.

Proposition 1.1.2.

Soient l_x et l_y deux nombres réels déterministes et (U_n) une suite de nombres réels telle que :

$\lim_{n\to\infty} \mathrm{U}_n = 0$.

i) *Si* $\lim_{n\to\infty} \mathrm{X}_n = l_x$ *a.co et* $\lim_{n\to\infty} \mathrm{Y}_n = l_y$ *a.co, alors :*

a) $\lim_{n\to\infty}(\mathrm{X}_n + \mathrm{Y}_n) = l_x + l_y$ *a.co,*

b)$\lim_{n\to\infty}(\mathrm{X}_n \times \mathrm{Y}_n) = l_x \times l_y$ *a.co,*

c) $\lim_{n\to\infty} \frac{1}{\mathrm{X}_n} = \frac{1}{l_x}$ *a.co si* $l_x \neq 0$.

ii) *Si* $\mathrm{X}_n - l_x = \mathrm{O}_{a.co}(\mathrm{U}_n)$ *et* $\mathrm{Y}_n - l_y = \mathrm{O}_{a.co}(\mathrm{U}_n)$, *on a :*

a) $(\mathrm{X}_n + \mathrm{Y}_n) - l_x - l_y = \mathrm{O}_{a.co}(\mathrm{U}_n)$,

b) $(\mathrm{X}_n \times \mathrm{Y}_n) - l_x \times l_y = \mathrm{O}_{a.co}(\mathrm{U}_n)$,

c)$\frac{1}{\mathrm{X}_n} - \frac{1}{l_x} = \mathrm{O}_{a.co}(\mathrm{U}_n)$ *si* $l_x \neq 0$.

iii) *Si* $\mathrm{X}_n = \mathrm{O}_{a.co}(\mathrm{U}_n)$ *et* $\lim_{n\to\infty} \mathrm{Y}_n = l_y$ *a.co, alors on a :*

a) $\mathrm{X}_n \times \mathrm{Y}_n = \mathrm{O}_{a.co}(\mathrm{U}_n)$,

b) $\frac{\mathrm{X}_n}{\mathrm{Y}_n} = \mathrm{O}_{a.co}(\mathrm{U}_n)$, *si* $l_y \neq 0$.

Démonstration .

i a) *La preuve découle immédiatement de l'inégalité suivante :*

$$\mathrm{P}[|(\mathrm{X}_n + \mathrm{Y}_n) - (l_x + l_y)| > \epsilon] \leqslant \mathrm{P}[|\mathrm{X}_n - l_x| > \tfrac{\epsilon}{2}] + \mathrm{P}[|\mathrm{Y}_n - l_y| > \tfrac{\epsilon}{2}].$$

ii a) *Il suffit d'appliquer la même inégalité à* $\epsilon = \epsilon_0 \mathrm{U}_n$.

i b) *Sans perte de généralité, on pose* $l_x = 0$. *La décomposition suivante :*

$$\mathrm{X}_n \times \mathrm{Y}_n = \mathrm{X}_n(\mathrm{Y}_n - l_y) + \mathrm{X}_n \times l_y,$$

nous donne :

$$\begin{aligned} \mathrm{P}\left[|(\mathrm{X}_n \times \mathrm{Y}_n)| > \epsilon\right] &\leqslant \mathrm{P}\left[|\mathrm{Y}_n - l_y||\mathrm{X}_n| > \frac{\epsilon}{2}\right] + \mathrm{P}\left[|l_y \mathrm{X}_n| > \frac{\epsilon}{2}\right] \\ &\leqslant \mathrm{P}\left[|\mathrm{Y}_n - l_y| > \sqrt{\frac{\epsilon}{2}}\right] + \mathrm{P}\left[|\mathrm{X}_n| > \sqrt{\frac{\epsilon}{2}}\right] + \mathrm{P}\left[|\mathrm{X}_n l_y| > \frac{\epsilon}{2}\right] \\ &\leqslant \mathrm{P}\left[|\mathrm{Y}_n - l_y| > \frac{\epsilon}{2}\right] + \mathrm{P}\left[|\mathrm{X}_n| > \frac{\epsilon}{2}\right] + \mathrm{P}\left[|\mathrm{X}_n l_y| > \frac{\epsilon}{2}\right]. \end{aligned}$$

L'inégalité précédente et la convergence presque complète de X_n *et* Y_n *permettent d'écrire*

$$\sum_{n\in\mathbb{N}} \mathrm{P}[|\mathrm{X}_n \mathrm{Y}_n| > \epsilon] < \infty.$$

ii b) *Il suffit d'appliquer le résultat précédent à* $\epsilon = \epsilon_0 \mathrm{U}_n$.

i c) *La convergence presque complète de* Y_n *vers* l_y *implique l'existence de* $\delta \geq 0$ ($\delta = \frac{l_y}{2}$ *par exemple)*

tel que :

$$\sum_{n\in\mathbb{N}} \mathrm{P}[|\mathrm{Y}_n| \leqslant \delta] < \infty, \tag{1.1}$$

de plus on a :

$$\begin{aligned} \mathrm{P}\left[\frac{1}{\mathrm{Y}_n} - \frac{1}{l_y} > \epsilon\right] &= \mathrm{P}[|\mathrm{Y}_n - l_y| > \epsilon |l_y \mathrm{Y}_n|] \\ &\leqslant \mathrm{P}[|\mathrm{Y}_n - l_y| > \epsilon |l_y \mathrm{Y}_n|, |\mathrm{Y}_n| > \delta] + \mathrm{P}[|\mathrm{Y}_n \leqslant \delta|] \\ &\leqslant \mathrm{P}[|\mathrm{Y}_n - l_y| > \epsilon\delta |l_y|] + \mathrm{P}[|\mathrm{Y}_n| \leqslant \delta]. \end{aligned}$$

En utilisant la relation (1.1) et la convergence presque complète de Y_n *vers* l_y*, il vient :*

$$\forall \epsilon > 0, \quad \sum_{n\in\mathrm{N}} \mathrm{P}\left[\left| \frac{1}{\mathrm{Y}_n} - \frac{1}{l_y} \right| \geq \epsilon\right] < \infty.$$

ii c) *On procède de la même manière pour* $\epsilon = \epsilon_0 \mathrm{U}_n$.

iii a) *La définition de la convergence presque complète de* Y_n *vers* l_y *implique l'existence de* $\delta > 0$*, tel que*

$$\sum_{n\in\mathbb{N}} \mathrm{P}[|\mathrm{Y}_n| > \delta] < \infty.$$

La décomposition suivante

$$\begin{aligned} \mathrm{P}[|\mathrm{Y}_n\mathrm{X}_n| > \epsilon \mathrm{U}_n] &= \mathrm{P}[|\mathrm{Y}_n\mathrm{X}_n| > \epsilon \mathrm{U}_n, |\mathrm{Y}_n| \leqslant \delta) + \mathrm{P}(|\mathrm{X}_n\mathrm{Y}_n| > \epsilon \mathrm{U}_n, |\mathrm{Y}_n| > \delta) \\ &\leqslant \mathrm{P}(|\mathrm{X}_n| > \epsilon\delta^{-1}\mathrm{U}_n] + \mathrm{P}(|\mathrm{Y}_n| > \delta], \end{aligned}$$

associée à l' inégalité précédente et à l'hypothèse $\mathrm{X}_n = \mathrm{O}_{a.co}(\mathrm{U}_n)$, *conduit à* $\mathrm{X}_n\mathrm{Y}_n = \mathrm{O}_{a.co}(\mathrm{U}_n)$.

L'estimation d'un paramètre inconnu par un estimateur requiert que ce dernier jouisse de certaines propriétés asymptotiques, nous rappelons certaines d'entre elles ci dessous.

Définition 1.1.3.

On dit que l'estimateur $\widehat{\theta}_n(x)$ *du paramètre* $\theta(x)$

1) est faiblement consistant si

$$\forall x \in \mathbb{R} \quad \widehat{\theta}_n(x) \xrightarrow{\mathrm{P}} \theta(x), \qquad quand\ n \to \infty.$$

2) est faiblement et uniformément consistant si

$$\sup_{x\in\mathrm{R}} |\widehat{\theta}_n(x) - \theta(x)| \xrightarrow{\mathrm{P}} 0, \qquad quand \quad n \to \infty.$$

3) est fortement consistant si

$$\forall x \in \mathbb{R}, \qquad \widehat{\theta}_n(x) \xrightarrow{\mathrm{P.S}} \theta(x), \qquad quand\ n \to \infty.$$

4) est fortement et uniformément consistant si

$$\sup_{x\in\mathrm{R}} |\widehat{\theta}_n(x) - \theta(x)| \xrightarrow{\mathrm{P.S}} 0, \qquad quand\ n \to \infty.$$

5) est asymptotiquement sans biais si

$$\forall x \in \mathbb{R}, \qquad \lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n(x)) = \theta(x).$$

6) est asymptotiquement et uniformément sans biais si

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \mathrm{E}(\widehat{\theta}_n(x) - \theta(x)) = 0.$$

7) est convergent en moyenne quadratique si

$$\lim_{n\to\infty} \mathrm{E}(\widehat{\theta}_n(x) - \theta(x))^2 = 0.$$

Nous allons donner deux versions des inégalités exponentielles de type Bernstein qui nous serons utiles pour l'établissement des résultats que nous avons choisi de reprendre. Nous supposons que $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$ *est une suite de variables aléatoires réelles, indépendantes et centrées.*

Corollaire 1.1.1.

a) Si pour tout $m \geq 2$*, il existe un réel* C_m *strictement positif et une constante a positive, tels que :*

$$\mathrm{E}|\mathrm{X}_1^m| \leqslant \mathrm{C}_m a^{2(m-1)},$$

alors on a

$$\forall \epsilon > 0, \qquad \mathrm{P}\left[\left| \sum_{i=1}^{n} \mathrm{X}_i \right| > \epsilon n\right] \leqslant 2\exp\left\{\frac{-\epsilon^2 n}{2a^2(1+\epsilon)}\right\}.$$

b)Supposons que les $(\mathrm{X}_i)_{1\leqslant i\leqslant n}$ *dépendent de* n $(\mathrm{X}_i = \mathrm{X}_{i,n})$*.*

Si pour tout $m \geq 2$*, il existe un réel* C_m *strictement positif et une suite* (a_n) *de réels positifs, tels que :*

$$\mathrm{E}|\mathrm{X}_1^m| \leqslant \mathrm{C}_m a_n^{2(m-1)},$$

et si

$\mathrm{U}_n = n^{-1}a_n^2 \log n$*, vérifie* $\lim_{n\to\infty} \mathrm{U}_n = 0$*, alors on a*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{X}_i = \mathrm{O}_{a.co}(\sqrt{\mathrm{U}_n}).$$

Tandisque que ce résultat s'applique à des variables dont on a majoré les moments d'ordre m, le corollaire suivant est donné pour des variables identiquement distribuées et bornées.

Corollaire 1.1.2.

a) S'il existe une constante positive $\mathrm{M} < \infty$*, telle que :*

$$|\mathrm{X}_1| \leqslant \mathrm{M},$$

alors on a :

$$\forall \epsilon \geq 0, \qquad \mathrm{P}\left[\left| \sum_{i=1}^{n} \mathrm{X}_i \right| > \epsilon n\right] \leqslant 2 \exp\left\{\frac{-\epsilon^2 n}{2\sigma^2(1 + \frac{\mathrm{M}\epsilon}{\sigma^2})}\right\},$$

où

$$\sigma^2 = \mathrm{E}\mathrm{X}_1^2.$$

b) Supposons que les $(\mathrm{X}_i)_{1 \leqslant i \leqslant n}$ *dépendent de* n *et que* $\sigma_n^2 = \mathrm{E}\mathrm{X}_i^2$*, s'il existe* $\mathrm{M} = \mathrm{M}_n < \infty$ *telle que :*

$$|\mathrm{X}_1| \leqslant \mathrm{M},$$

et

$$\frac{\mathrm{M}}{\sigma_n^2} \leqslant \mathrm{C} < \infty$$

et si

$\mathrm{U}_n = n^{-1}\sigma_n^2 \log n$*, vérifie* $\lim_{n\to\infty} \mathrm{U}_n = 0$*,*

alors on a

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{X}_i = \mathrm{O}_{a.co}(\sqrt{\mathrm{U}_n}).$$

Les démonstrations de ces corollaires sont basées sur la proposition suivante dont la preuve peut être trouvée dans Uspensky (1937).

Proposition 1.1.3.

Si

$$\forall m \geq 2,\ |\mathrm{E}(\mathrm{X}_i^m)| \leqslant \left(\frac{m!}{2}\right)(a_i)^2 b^{m-2},$$

alors

$$\forall \epsilon \geq 0, \qquad \mathrm{P}\left[\sum_{i=1}^{n} \left| \mathrm{X}_i \right| > \epsilon \mathrm{A}_n\right] \leqslant 2 \exp\left\{\frac{-\epsilon^2}{2(1 + \frac{b\epsilon}{\mathrm{A}_n})}\right\},$$

où

$(a_i)_{1 \leqslant i \leqslant n}$ *sont des réels positifs,* $b \in \mathbb{R}^+$ *et* $\mathrm{A}_n^2 = a_1^2 + a_2^2 + a_3^2 + ... + a_n^2$*.*

La démonstration du corollaire (1.1.1)

a) *En remplaçant* $b = a^2$ *et* $\mathrm{A}_n = a\sqrt{n}$ *dans la proposition précédente, on aboutit à a) .*

b) *En posant* $\epsilon = \epsilon_0\sqrt{\mathrm{U}_n}$ *dans* a) *et comme* U_n *tend vers zéro , pour une certaine constante C' on a :*

$$\begin{aligned} \mathrm{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n}\mathrm{X}_i\right| > \epsilon_0\mathrm{U}_n\right] &\leqslant 2\exp\left\{\frac{-\epsilon_0^2\log n}{2\left(1+\epsilon_0\sqrt{\mathrm{U}_n}\right)}\right\} \\ &\leqslant 2n^{-\mathrm{C}'}\epsilon_0^2 . \end{aligned}$$

D' où, pour un choix convenable de ϵ_0 *on déduit que*

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i = \mathrm{O}_{a.co}(\sqrt{\mathrm{U}_n}).$$

La démonstration du corollaire (1.1.2)

a) *En appliquant la proposition (1.1.3) à* $a_i^2 = \sigma^2$, $\mathrm{A}_n^2 = n\sigma^2$ *et b=M on aboutit à a).*

b) *Comme* $\frac{\mathrm{MU}_n}{\sigma_n^2}$ *tend vers zéro, il suffit de reprendre le résultat a) pour* $\epsilon = \epsilon_0\sqrt{\mathrm{U}_n}$ *, on arrive donc à l'existence d'une constante C' telle que :*

$$\begin{aligned} \mathrm{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n}\mathrm{X}_i\right| > \epsilon_0\mathrm{U}_n\right] &\leqslant 2\exp\left\{\frac{-\epsilon_0^2\log n}{2\left(1+\epsilon_0\sqrt{\frac{\mathrm{MU}_n}{\sigma_n^2}}\right)}\right\} \\ &\leqslant 2n^{-\mathrm{C}'}\epsilon_0^2 . \end{aligned}$$

Pour ϵ_0 *bien choisi (assez grand) le terme de droite est le terme général d'une série convergente. Ainsi s'achève la preuve de ce corollaire.*

1.2 Estimation de la fonction de répartition

Soit $\mathrm{X}_{1,n} < \mathrm{X}_{2,n} < ... < \mathrm{X}_{n,n}$ *la statistique d'ordre associée à* $\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n$*, une suite de variables aléatoires indépendantes identiquement distribuées de même loi que* X *de fonction de répartition* $\mathrm{F}(x) = \mathrm{P}(\mathrm{X}_1 \leqslant x)$*, et de densité de probabilité (par rapport à la mesure de Lebesgue)*

f. F et f sont complètement inconnues (cadre non paramétrique).
Avant d'étudier la construction et les propriétés de l' estimateur à noyau de f (une fonction mesurable par rapport à la tribu engendrée par $(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n)$*), nous donnons brièvement un aperçu sur l' estimateur de la fonction de répartition.*
Un estimateur sans biais classique de la fonction de répartition $\mathrm{F}(x)$ *est la fonction de répartition empirique définie par :*

$$\begin{aligned}
\mathrm{F}_n(x) &= \frac{1}{n}\sum_{i=1}^{n} 1_{\{\mathrm{X}_i \leqslant x\}} \\
&= \frac{1}{n}\sum_{i=1}^{n} 1_{\{\mathrm{X}_{i,n} \leqslant x\}} \\
&= \begin{cases} 0 & si\ x < \mathrm{X}_{1,n} \\ \frac{k}{n} & si\ \mathrm{X}_{k,n} \leqslant x < \mathrm{X}_{k+1,n} \quad k = 1, ..., n-1. \\ 1 & si\ x \geq \mathrm{X}_{n,n}\,. \end{cases}
\end{aligned}$$

La loi forte des grands nombres nous montre que c'est un estimateur fortement consistant c'est à dire :

$$\forall x \in \mathbb{R}, \qquad \mathrm{F}_n(x) \to \mathrm{F}(x)$$

Le théorème de Glivenko-Cantelli permet d'améliorer ce résultat puisqu'il donne la convergence uniforme, il s'énonce

$$\forall x \in \mathbb{R}, \qquad \sup_{x\in\mathbb{R}} |\mathrm{F}_n(x) - \mathrm{F}(x)| \xrightarrow{\text{P.S}} 0, \quad quand\ n \to \infty$$

De plus le théorème central limite permet d'obtenir la normalité asymptotique de cet estimateur, résultat amélioré par le théorème de Donsker qui montre la convergence de F_n *en tant que processus vers le processus de Wiener.*
Nous remarquons aussi qu'à partir de cet estimateur nous pouvons définir un estimateur du $q^{ème}$ *quantile (ou quantile d'ordre q) à savoir*

$$\mathrm{F}^{-1}(q) = \inf_{x\in\mathbb{R}}\{x : \mathrm{F}(x) \geq q\}, \qquad 0 < q < 1$$

qui peut être estimé par

$$\mathrm{F}_n^{-1}(q) = \inf_{x\in\mathbb{R}}\{x : \mathrm{F}_n(x) \geq q\}.$$

1.3 Estimation de la densité par histogramme

L'histogramme est l' estimateur de la densité non paramétrique le plus ancien et le plus simple, il est introduit par John Graunt au ***XVII***^{e} *siècle, il est considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini.*

Cette méthode consiste à estimer f en un point x par la proportion des variables aléatoires $(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_n)$, qui se trouvent dans un intervalle de longueur un paramètre de lissage h_n et qui contient x. Elle est donc basée sur le choix d'un point d'origine a_0 et d'une partition $\mathrm{B}_k = ([a_k, a_{k+1}[)_{k=1,...,p}$ en P intervalles du support de X . Si nous notons n_k le nombre de variables dans la classe $[a_k, a_{k+1}[$ et $h_n = a_{k+1} - a_k$, l'estimateur de f sur $[a_k, a_{k+1}[$ du type histogramme est :

$$\hat{f}_n(x, h_n) = \frac{n_k}{nh} = \frac{\sum_{i=1}^{n} \mathrm{I}_{[a_k^{(\mathrm{X}_i)}, a_{k+1}[}}{nh} \qquad pour\ x \in [a_k, a_{k+1}[.$$

Les propriétés asymptotiques de cet estimateur ont été détaillées dans le livre de Bosq et Lecoutre (1987) et le livre de Simonoff (1996).

Pour tout $x \in [a_k, a_{k+1}]$, l'étude asymptotique de l'erreur quadratique moyenne de f_n et l'erreur quadratique moyenne intégrée def_n ont été établies par Lecoutre (1982).
Nous avons pour tout $x \in [a_k, a_{k+1}]$, pour tout $k \in (1, ..., p)$ et si la densité f vérifie certaines hypothèses de régularité alors les deux critères d'erreurs cités précédemment sont donnés respectivement par les expressions suivantes :

$$\mathrm{E}(\hat{f}_n(x, h_n) - \mathrm{E}f(x))^2 = \frac{f(x)}{nh} + \frac{f'(x)}{4}(h - 2(x - a_k))^2 + o(n^{-1}) + o(h^3)$$

et

$$\int_{\mathbb{R}} \mathrm{E}(\hat{f}_n(x, h_n) - \mathrm{E}f(x))^2 = \frac{1}{nh} + \frac{h^2 \int_{\mathbb{R}} (f'(x))^2 dx}{12} + o(n^{-1}) + o(h^3).$$

Ces deux quantités tendent vers zéro, quand h tend vers zéro et nh tend vers l'infini.
L'erreur quadratique moyenne intégrée permet de déterminer le paramètre de lissage optimal de l'histogramme, cette valeur minimise aussi ce critère. Elle s'écrit :

$$h_{(opt)} = \left(\frac{6}{\int_{\mathbb{R}} (f'(x))^2 dx} \right)^{1/3} n^{-1/3}.$$

Geoffrey (1974)a étudié la convergence uniforme et presque complète de cet estimateur.

Nous constatons que l' histogramme a de bonnes propriétés statistiques, par contre il n'est robuste ni pour le choix du paramètre de lissage h, ni pour celui de a_0. Le deuxième désavantage est sa discontinuité qui ne peut pas s' adapter au cas où f, la densité à estimer, vérifie certaines hypothèses de régularité.

Afin de résoudre ce problème, l' estimateur de Parzen Rosenblatt a été introduit, il généralise intuitivement la méthode d' estimation par histogramme, et il est très utilisé en estimation non paramétrique.

1.4 L'estimateur à noyau

Du fait que

$$f(x) \simeq \frac{\mathrm{F}(x+h) - \mathrm{F}(x-h)}{2h} \qquad \textit{pour } h_n \textit{ petit.}$$

Rosenblatt (1956) *a donné un estimateur de* f*, en remplaçant* F *par son estimateur* F_n*.*

D'où

$$f_n(x, h_n) = \frac{\mathrm{F}_n(x+h) - \mathrm{F}_n(x-h)}{2h},$$

où F_n *est la fonction de répartition empirique. Cet estimateur peut encore s'écrire :*

$$\begin{aligned} f_n(x, h_n) &= \sum_{i=1}^{n} \frac{1_{\{x-h<\mathrm{X}_i \leqslant x+h\}}}{2nh} \\ &= \frac{1}{2hn} \sum_{i=1}^{n} \mathrm{I}_{\{-1<\frac{x-\mathrm{X}_i}{h} \leqslant 1\}} \\ &= \frac{1}{2hn} \sum_{i=1}^{n} \mathrm{K}_0\left(\frac{x-\mathrm{X}_i}{h}\right) \end{aligned} \tag{1.2}$$

avec

$$\mathrm{K}_0(u) = \frac{1}{2}\, 1_{\{-1<u\leqslant 1\}}.$$

Dans ce même article, Rosenblatt a mesuré la qualité de cet estimateur, en calculant son biais et sa variance, donnés respectivement par :

$$\begin{aligned} \mathrm{E}f_n(x, h_n) - f(x) &= \frac{1}{2h}\mathrm{E}(\mathrm{F}_n(x+h) - \mathrm{F}_n(x-h)) - f(x) \\ &= \frac{1}{2h}(\mathrm{F}(x+h) - \mathrm{F}(x-h)) - f(x). \end{aligned} \tag{1.3}$$

et par :

$$\begin{aligned} \mathrm{V}ar[f_n(x, h_n)] &= \frac{1}{4nh_n^2}\left[\mathrm{F}(x+h_n)(1-\mathrm{F}(x+h_n)) + \mathrm{F}(x-h_n)(1-\mathrm{F}(x-h_n))\right] \\ &- \frac{1}{4nh_n^2}\left[2\mathrm{F}\left(inf((x-h_n),(x+h_n))\right) + 2\mathrm{F}(x+h_n)\mathrm{F}(x-h_n)\right]. \end{aligned}$$

Nous remarquons que si $h_n \to 0$ *et* $nh_n \to \infty$ *quand* $n \to \infty$*, on a :*

$$\lim_{n\to\infty} \mathrm{E}[f_n(x, h_n)] = f(x)$$

et

$$\lim_{n\to\infty} \mathrm{V}ar[f_n(x, h_n)] = 0.$$

$f_n(x, h_n)$ *est un estimateur consistant. Nous remarquons qu'il n'a pas le problème du choix d'origine* a_0 *comme le cas de l' histogramme mais il présente l'inconvénient d'être discontinu aux points* $\mathrm{X}_i \pm h$.
Ainsi une généralisation de cet estimateur a été introduite par Parzen (1962) en posant

$$f_n(x, h_n) = \frac{1}{nh_n} \sum_{i=1}^{n} \mathrm{K}\left(\frac{\boldsymbol{X}_i - x}{h_n}\right), \tag{1.4}$$

où (h_n) *est une suite de réels strictement positifs, tendant vers zéro quand* $n \to \infty$*, (appelée fenêtre) et* K *est une fonction mesurable définie de* $\mathbb{R} \longrightarrow \mathbb{R}$*, appelée noyau.*

Exemple de noyaux K *les plus utilisés dans l'estimation de la densité :*

- *Noyau rectangulaire :*

$$\mathrm{K}_1(x) = \begin{cases} \frac{1}{2}, & si\ |x| \leqslant 1\,; \\ \\ 0, & si\ |x| > 1. \end{cases}$$

- *Noyau triangulaire :*

$$\mathrm{K}_2(x) = \begin{cases} 1 - |x|, & si\ |x| \leqslant 1\,; \\ \\ 0, & si\ |x| > 1. \end{cases}$$

- *Noyau d'Epanechnikov ou parabolique :*

$$\mathrm{K}_3(x) = \begin{cases} \frac{3}{4}(1 - x^2), & si\ x \in [-1, 1]\,; \\ \\ 0, & sinon. \end{cases}$$

- *Noyau quadratique :*

$$\mathrm{K}_4(x) = \begin{cases} \frac{15}{16}(1 - x^2)^2, & si\ x \in [-1, 1]\,; \\ \\ 0, & sinon\ . \end{cases}$$

- *Noyau cubique :*

$$\mathrm{K}_5(x) = \begin{cases} \frac{35}{32}(1-x^2)^3, & si\ x \in [-1,1]\,; \\ \\ 0, & sinon. \end{cases}$$

- *Noyau gaussien :*

$$\mathrm{K}_6(x) = \frac{1}{\sqrt{2\Pi}} \exp\left(\frac{-1}{2}x^2\right), \quad x \in \mathrm{R}.$$

- *Noyau sinus :*

$$\mathrm{K}_7(x) = \begin{cases} \frac{1}{2\pi}\left(\frac{\sin(\frac{x}{2})}{\frac{x}{2}}\right)^2, & si\ x \neq 0\,; \\ \\ \frac{1}{2\pi}, & si\ x = 0. \end{cases}$$

- *Noyau cosinus :*

$$\mathrm{K}_8(x) = \begin{cases} \frac{\pi}{4}\cos\left(\frac{\pi x}{2}\right), & si\ -1 \leqslant x \leqslant 1\,; \\ \\ 0, & sinon. \end{cases}$$

- *Noyau de Silverman :*

$$\mathrm{K}_9(x) = \frac{1}{2}\exp\left(-|x|/\sqrt{2}\right)\sin\left(|x|/\sqrt{2} + \frac{\pi}{4}\right), \quad x \in \mathbb{R}$$

L'estimateur de Parzen - Rosenblatt *a connu un très grand succès parmi les estimateurs non paramétriques, ceci est dû à sa simplicité et sa convergence vers la densité f pour tous les modes (convergence dans* L_1*, presque sûre, en probabilité en moyenne quadratique et presque complète) et il nous laisse aussi le choix sur le noyau* K *.*

1.4.1 Propriétés de l'estimateur

Le pilier des premiers résultats de la convergence de cet estimateur est le théorème de Bochner (1955)*, rappelé ci dessous :*

Théorème 1.4.1. *(Bochner)*

Soit $K : (\mathbb{R}^m, \beta^m) \to (\mathbb{R}, \beta)$ une fonction mesurable, où β^p est la tribu borélienne de $\mathbb{R}^p$, vérifiant :

$$\exists\, \mathrm{M} \quad (constante)\ telle\ que, \quad \forall z \in \mathbb{R}^m, |\mathrm{K}(z)| \leqslant \mathrm{M},$$

$$\int_{\mathbb{R}^m} |\mathrm{K}(z)|dz < \infty,$$

et

$$\|z\|^m |\mathrm{K}(z)| \longrightarrow 0 \qquad quand\ \|z\| \longrightarrow \infty.$$

Par ailleurs, soit

$g : (\mathbb{R}^m, \beta^m) \longrightarrow (\mathbb{R}, \beta)$ *une fonction telle que*

$$\int_{\mathbb{R}^m} |g(z)|dz < \infty,$$

Si g *est continue, et si* $0 < h_n \longrightarrow 0$, *quand* $n \to \infty$ *alors :*

$$\lim_{n\to\infty} \frac{1}{h_n^m} \int_{\mathrm{R}^m} \mathrm{K}\left(\frac{z}{h_n}\right) g(x-z)dz = g(x) \int_{\mathrm{R}^m} \mathrm{K}(z)dz. \tag{1.5}$$

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

1.4.2 La consistance de l'estimateur

L'estimateur à noyau de la densité dépend de deux paramètres la fenêtre h *et le noyau* K. *Le noyau* K *établit l'aspect du voisinage de* x *et* h *contrôle la taille de ce voisinage, donc* h *est le paramètre prédominant pour avoir de bonnes propriétés asymptotiques, néanmoins le noyau* K *ne doit pas être négligé, comme le montre le travail de Parzen (1962) cité ci dessous sur la consistance de cet estimateur. Cette dernière est obtenue, en se basant sur l'étude asymptotique du biais, de la variance et de la décomposition suivante :*

$$\mathrm{E}[f_n(x,h_n) - f(x)]^2 = \mathrm{V}ar[f_n(x,h_n)] + [\mathrm{B}iais\{f_n(x,h_n)\}]^2.$$

Dans la suite, nous supposons que K *est un noyau vérifiant les conditions suivantes.*

(K.1) *K est bornée, c'est à dire* $\sup_{x\in\mathbb{R}} |\mathrm{K}(x)| < \infty$,

(K.2) $\lim |x|\mathrm{K}(x) = 0$, *quand* $|x| \to \infty$,

(K.3) $\mathrm{K} \in \mathrm{L}_1(\mathbb{R})$, *c'est à dire* $\int_{\mathbb{R}} |\mathrm{K}(x)|dx < \infty$,

(K.4) $\int_{\mathbb{R}} \mathrm{K}(x)dx = 1$.

1) Etude du biais :

Proposition 1.4.1.

Sous les hypothèses [(K.1), (K.2), (K.3) et (K.4)], et si f est continue alors

$$\forall x \in \mathbb{R} \ \lim_{n\to\infty} \mathrm{E}\left[f_n(x, h_n)\right] = f(x). \tag{1.6}$$

En effet :

$$\begin{aligned} \mathrm{E}(f_n(x, h_n)) &= \mathrm{E}\left[\frac{1}{nh_n}\sum_{i=1}^{n} \mathrm{K}\left(\frac{x - \mathrm{X}_i}{h_n}\right)\right] \\ &= \frac{1}{h_n}\mathrm{E}\left[\mathrm{K}\left(\frac{x - \mathrm{X}_i}{h_n}\right)\right] \\ &= \frac{1}{h_n}\int_{\mathrm{R}} \mathrm{K}\left(\frac{x - t}{h_n}\right) f(t)dt. \end{aligned}$$

En posant $x - t = z$, on arrive à :

$$\mathrm{E}\left[f_n(x, h_n)\right] = \frac{1}{h_n}\int_{\mathrm{R}} \mathrm{K}\left(\frac{z}{h_n}\right) f(x - z)dz.$$

Comme K et f vérifient les conditions du théorème de Bochner, et $\lim_{n\to\infty} h_n = 0, \quad n \to \infty$, *on a alors*

$$\lim_{n\to\infty} \frac{1}{h_n}\int_{\mathbb{R}} \mathrm{K}\left(\frac{x - t}{h_n}\right) f(t)dt = f(x)\int_{\mathbb{R}} \mathrm{K}(z)dz.$$

d'où

$$\lim_{n\to\infty} \mathrm{E}\left[f_n(x, h_n)\right] = f(x).$$

Nous constatons que le biais de l'estimateur converge vers zéro quand la fenêtre tend vers zéro, de plus vu son expression, on constate qu'il ne dépend pas du nombre des variables, il dépend surtout du noyau K.

2) Etude de la variance de $f_n(x, h_n)$

Proposition 1.4.2.

Sous les conditions[(K.1), (K.2), (K.3) et (K.4)] et si f est continue en tout point x de $\mathbb{R}$, alors :

$$\lim_{n\to\infty} \mathrm{V}ar[f_n(x, h_n)] = 0$$

En effet

$$\begin{aligned}
\mathrm{V}ar[f_n(x, h_n)] &= \mathrm{E}[f_n(x, h_n)]^2 - [\mathrm{E}f_n(x, h_n)]^2 \\
&\leqslant \mathrm{E}[f_n(x, h_n)]^2 \\
&\leqslant \frac{1}{n}\mathrm{E}\left[\frac{1}{h_n}\mathrm{K}\left(\frac{x - \mathrm{X}_i}{h_n}\right)\right]^2 \\
&\leqslant \frac{1}{nh_n^2}\int \mathrm{K}^2\left(\frac{x-t}{h_n}\right) f(t)dt \\
&\leqslant \frac{1}{nh_n^2}\int \mathrm{K}^2\left(\frac{z}{h_n}\right) f(x-z)dz.
\end{aligned}$$

Remarquons que (K.1) et (K.3) impliquent que le noyau est de carré intégrable et les hypothèses sur h_n, K *et f assurent que :*

$$\frac{1}{nh_n^2}\int \mathrm{K}^2\left(\frac{z}{h_n}\right) f(x-z)dz \sim \frac{1}{nh_n} f(x)\int \mathrm{K}^2(z)dz,$$

d'où

$$\lim_{n\to\infty} \mathrm{V}ar[f_n(x, h_n)] = 0, \qquad \textit{quand } nh_n \to \infty.$$

Ces deux propositions impliquent la convergence en moyenne quadratique et donc, à fortiori, la consistance de l'estimateur.

1.4.3 Consistance forte

Dans le même article, Parzen *a établi la normalité asymptotique, ainsi que la convergence uniforme en probabilité. Son travail est un outil important et a été largement développé par plusieurs chercheurs (*Devroye *et* Györfi *(1985),* Silverman *(1986),* Izenman *(1991),* Scott*(1992)). Pour une étude plus détaillée, voir par exemple (*Bosq *et* Lecoutre *(1987) et* Tsybakov*(2009). En (1976)* Nadaraya *a énoncé le théorème suivant sur la consistance forte de l' estimateur.*

Théorème 1.4.2.

Si K(.) est à variation bornée et si pour tout $\gamma > 0$, la série $\sum_{i=1}^{\infty} \exp(-\gamma n h_n^2)$ converge, alors

$$\upsilon_n = \sup_{x \in \mathrm{R}} |f_n(x, h_n) - f(x)| \xrightarrow{\mathrm{P}} 0, \qquad \textit{quand } n \longrightarrow \infty.$$

si et seulement si la densité f est uniformément continue.

La démonstration de ce théorème est basée sur la définition et les lemmes suivants.

Définition 1.4.1.

Soit μ une application de $\mathbb{R}^+ \to \mathbb{R}$. μ est dite à variation bornée si

$$\forall t > 0 \qquad \mathrm{T}_\mu(t) = \sup \left\{ \sum_{i=1}^{n-1} |\mu(t_{i+1}) - \mu(t_i)| \right\} < \infty,$$

où sup *est pris sur toutes les subdivisions $([t_j, t_{j+1}[)_{0 \leqslant i \leqslant n-1}$ de $[0, t]$.*

Lemme 1.4.1.

$\forall n > 0,\ \forall \lambda > 0,\ \exists \mathrm{C} > 0$ *tel que :*

$$p \left[\sup_{-\infty < x < \infty} |\mathrm{F}_n(x) - \mathrm{F}(x)| > \lambda \right] \leqslant \mathrm{C} \exp(-\alpha n \lambda)^2,$$

où F *et* F_n *désignent respectivement la fonction de répartition et la distribution empirique.*

Lemme 1.4.2.

Pour toute fonction g, on a :

$$\lim_{n \to \infty} \sup_{x \in \mathrm{R}} |f_n(x, h_n) - g(x)| \xrightarrow{\mathrm{P}} 0,$$

si seulment si

$$\lim_{n \to \infty} \sup_{x \in \mathrm{R}} |\mathrm{E} f_n(x, h_n) - g(x)| = 0.$$

Lemme 1.4.3.

Si

$$\lim_{n \to \infty} \sup_{x \in \mathbb{R}} |f_n(x, h_n) - g(x)| = 0 \qquad \mathrm{P.S}$$

alors g est uniformément continue.

L'étude de la convergence, presque complète des estimateurs à noyau dans le cadre de la variable aléatoire fonctionnelle introduite par Ferraty *et* Vieu *(2000), s'inspire énormément du cas réel exposé ci dessous. Cette convergence implique la convergence en probabilité et la convergence presque sûre mais n'implique pas la convergence en moyenne quadratique, comme le montre l'exemple suivant :*
Soit l'espace de probabilité $([0,1], \mathrm{B}([0,1]), \lambda)$ *où* $\mathrm{B}([0,1]$ *désigne la tribu borélienne de* $[0,1]$ *et* λ *est la restriction de la mesure de Lebesgue à cet ensemble. La suite* $f_n = n1_{[0,\frac{1}{n^2}]}$ *converge presque complètement vers* 0 *mais ne converge pas en moyenne quadratique et inversemnet la suite* $g_n = 1_{[0,\frac{1}{n}]}$ *converge en moyenne quadratique vers* 0 *mais ne converge pas presque complètement.*

1.4.4 La convergence presque complète

Maintenant, nous introduisons les hypothèses de base permettant de donner un théorème général sur la convergence presque complète.

- ***f est continue au voisinage de x, un point fixé de*** $\mathbb{R}$. (1.7)

- ***Le paramètre de lissage*** h_n ***est tel que***

$$\lim_{n\to\infty} h_n = 0 \; et \; \lim_{n\to\infty} \frac{\log n}{n h_n} = 0. \tag{1.8}$$

- ***Le noyau*** K ***est tel que***

K *est d'ordre k au sens de Gasser c'est à dire :*

$$\int t^j \mathrm{K}(t)dt = 0 \qquad \forall j = 1, 2....k-1 \quad et \quad 0 < \left| \int t^k \mathrm{K}(t)dt \right| < \infty \tag{1.9}$$

et

(K.5) K *est borné, intégrable et à support compact.*

Théorème 1.4.3.

Si les conditions (K.5), (1.7) et (1.8) sont vérifiées alors :

$$\lim_{n\to\infty} f_n(x, h_n) = f(x), \qquad a.co. \tag{1.10}$$

Démonstration .

La démonstration de ce théorème est basée sur la décomposition suivante :

$$f_n(x,h_n) - f(x) = (f_n(x,h_n) - \mathrm{E}[f_n(x,h_n)]) - (f(x) - \mathrm{E}[f_n(x,h_n)]). \tag{1.11}$$

Le résultat du théorème découle alors des deux lemmes suivants.

Lemme 1.4.4.

Si les conditions (K.5), (1.7) et (1.8) sont vérifiées on a :

$$\lim_{n\to\infty} \mathrm{E}[f_n(x,h_n)] = f(x). \tag{1.12}$$

Preuve .
Nous avons :

$$\mathrm{E}[f_n(x,h_n)] = \frac{1}{h_n}\int_{-\infty}^{+\infty} \mathrm{K}\left(\frac{x-t}{h_n}\right) f(t)dt,$$

posons $z = \frac{x-t}{h_n}$

$$\mathrm{E}[f_n(x,h_n)] = \int_{\mathrm{R}} \mathrm{K}(z) f(x - zh_n)dz.$$

La continuité uniforme de f *sur le support compact de* K *entraine*

$$f(x - zh_n) \to f(x), \; uniformément\, en\, z.$$

D'où

$$\lim_{n\to\infty} \mathrm{E}[f_n(x,h_n)] = f(x).$$

Lemme 1.4.5.

Sous les hypothèses (K.5) , (1.7) et (1.8) on a :

$$f_n(x,h_n) - \mathrm{E}[f_n(x,h_n)] = \mathrm{O}_{a.co}\left(\sqrt{\frac{\log n}{nh_n}}\right) \tag{1.13}$$

Preuve .
Nous avons,

$$\begin{aligned} f_n(x,h_n) - \mathrm{E}[f_n(x,h_n)] &= \frac{1}{n}\sum_{i=1}^{n} h_n^{-1}\left[\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right) - \mathrm{EK}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \Gamma_i, \end{aligned}$$

où

$$\Gamma_i = h_n^{-1}\left[\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right) - \mathrm{EK}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right].$$

En utilisant l'hypothèse (K.5) on a :

$$|\Gamma_i| < \frac{c}{h_n}.$$

D'autre part le changement de variable $z = \frac{x-t}{h_n}$ *, nous donne*

$$\begin{aligned} h^{-1}\mathrm{E}\left[h^{-1}\mathrm{K}^2\left(\frac{\mathrm{X}-x}{h_n}\right)\right] &= h^{-2}\int \mathrm{K}^2\left(\frac{x-t}{h_n}\right) f(t)dt \\ &= h^{-1}\int \mathrm{K}^2(z) f(x-zh_n)dz. \end{aligned}$$

Comme K *est bornée et* f *est continue sur le support compact de* K*, on a l'existence d'une constante* C *telle que :*

$$\mathrm{E}\Gamma_i^2 < \frac{\mathrm{C}}{h_n}.$$

On obtient alors, en appliquant le corollaire (1.1.2) de l'inégalité exponentielle de type Bernstein*,*

$$f_n(x,h_n) - \mathrm{E}f_n(x,h_n) = \mathrm{O}_{a.co}\left(\sqrt{\frac{\log n}{nh_n}}\right). \tag{1.14}$$

Ce résultat est plus fort que le résultat demandé.

En remplaçant l'hypothèse (1.7) par :

- ***f est k fois continûment dérivable autour du point x.*** (1.15)

On obtient une vitesse de convergence presque complète ponctuelle de l'estimateur à noyau.

Théorème 1.4.4.

Sous les conditions ((K.5) , (1.8) et (1.15)on a :

$$f_n(x,h_n) - f(x) = \mathrm{O}(h_n^k) + \mathrm{O}\left(\sqrt{\frac{\log n}{nh_n}}\right) \qquad a.co \tag{1.16}$$

Preuve .
En reprenant la décomposition de la preuve précédente, le résultat du théorème sera établi par les lemmes précédent et suivant :

Lemme 1.4.6.

Sous les conditions (1.8), (1.9) et (1.15) on a :

$$\mathrm{E}f_n(x,h_n) - f(x) = \mathrm{O}(h_n^k) \tag{1.17}$$

Preuve .
On a :

$$\mathrm{E}f_n(x,h_n) = \int \mathrm{K}\left(\frac{x-t}{h_n}\right) f(t)dt.$$

En posant $z = \frac{x-t}{h_n}$ *on obtient :*

$$\mathrm{E}f_n(x,h_n) = \int \mathrm{K}(z) f(x-zh)dz.$$

La condition (1.15) nous permet de développer f *au voisinage de* x*.*

$$f(x-zh_n) = f(x) + \sum_{i=1}^{k-1} \frac{(-1)^i (zh_n)^i}{i!} f^i(x) + \frac{(-1)^k (zh_n)^k}{k!} f^k(\theta_z)$$

où θ_z *entre* x *et* $x - zh_n$*.*

D'autre part la condition (1.9) sur l'ordre de K *au sens de Gasser et Müller nous donne :*

$$\mathrm{E}f_n(x,h_n) = f(x) + \frac{(-1)^k h_n^k}{k!} \int z^k \mathrm{K}(z) f^k(\theta_z)dz.$$

La compacité du support de K *et la condition de (1.15) impliquent la convergence uniforme en* z *de* $f^k(z)$ *vers* $f^k(x)$*, d'où :*

$$\mathrm{E}f_n(x,h_n) - f(x) = (-1)^k h^k \int z^k \mathrm{K}(z) \frac{f^k(x)}{k!} dz + \mathrm{O}(h_n^k).$$

Cette relation permet d'achever la preuve du théorème.
Le résultat du théorème précédent peut être établi uniformément. Il suffit de conserver toutes les hypothèses et de donner une autre version de l'hypothèse (1.15).

- ***qui consiste à choisir un compact S de*** $\mathbb{R}$ ***sur lequel f est*** k ***fois continûment dérivable*** , (1.18)

  et de rajouter l'hypothèse de type Lipschitz *sur le noyau* K*,*

- *il existe* $\mathrm{C} < \infty$*,* $\quad \forall x \in \mathrm{S}, \qquad \forall y \in \mathrm{S}$,

$$|\mathrm{K}(x) - \mathrm{K}(y)| \leqslant \mathrm{C}|x-y|. \tag{1.19}$$

Nous avons aussi besoin de la condition suivante sur le paramètre de lissage,

- *il existe* $\xi < \infty$, *tel que ,*

$$\lim_{n\to\infty} n^{2\xi-1}h_n = +\infty \tag{1.20}$$

Théorème 1.4.5.

Considérons le modèle (1.18), sous les hypothèses (K.5), (1.8), (1.9), (1.19),(1.20), on a

$$\sup_{x\in \mathrm{S}} |f_n(x,h_n) - f(x)| = \mathrm{O}(h_n^k) + \mathrm{O}\left(\sqrt{\frac{\log n}{nh_n}}\right), \qquad a.co \tag{1.21}$$

Preuve .
La démonstration de ce résultat utilise une décomposition similaire à celle du théorème (1.4.3).

$$\sup_{x\in \mathrm{S}} |f_n(x,h_n) - f(x)| \leqslant \sup_{x\in \mathrm{S}} |f_n(x,h_n) - \mathrm{E}f_n(x,h_n)| + \sup_{x\in \mathrm{S}} |\mathrm{E}f_n(x,h_n) - f(x)|. \tag{1.22}$$

Nous remarquons que grâce à l'hypothèse (1.18) et à la compacité du support du noyau K *les étapes de la démonstration utilisée dans le lemme (1.4.6) peuvent être faites uniformément en* $x \in \mathrm{S}$.

D'où

$$\sup_{x\in \mathrm{S}} |\mathrm{E}f_n(x,h_n) - f(x)| = \mathrm{O}(h_n^k). \tag{1.23}$$

Il reste à montrer l'égalité suivante :

$$\sup_{x\in \mathrm{S}} |f_n(x,h_n) - \mathrm{E}f_n(x,h_n)| = \mathrm{O}_{ac.o}\left(\sqrt{\frac{\log n}{nh_n}}\right). \tag{1.24}$$

S *est un compact de* $\mathbb{R}$*, il existe donc un recouvrement fini de S tel que :*

$$\mathrm{S} \subset \cup_{k=1}^{z_n} \mathrm{S}_k,$$

où

$$\mathrm{S}_k =]t_k - l_n, t_k + l_n[,$$

avec

$$l_n = n^{-2\xi},\ l_n = \mathrm{C}z_n^{-1}$$

et

$$t_x = \arg \min_{t\in\{t_1,t_2\ldots t_{z_n}\}} |x - t|.$$

On a :

$$\begin{aligned} \sup_{x\in \mathrm{S}} |\mathrm{E}f_n(x,h_n) - f_n(x,h_n)| &\leqslant \sup_{x\in \mathrm{S}} |f_n(x,h_n) - f_n(t_x,h_n)| \\ &+ \sup_{x\in \mathrm{S}} |f_n(t_x,h_n) - \mathrm{E}f_n(t_x,h_n)| \\ &+ \sup_{x\in \mathrm{S}} |\mathrm{E}f_n(x,h_n) - \mathrm{E}f_n(t_x,h_n)|. \end{aligned}$$

Comme K *est* Lipschitzien*, l'hypothèse (1.20) nous donne*

$$\begin{aligned}\sup_{x\in \mathrm{S}}|f_n(x,h_n)-f_n(t_x,h_n)| &\leqslant \sup_{x\in \mathrm{S}}\frac{1}{nh_n}\sum_{i=1}^{n}\left|\,\mathrm{K}\left(\frac{\mathrm{X}_i-x}{h_n}\right)-\mathrm{K}\left(\frac{\mathrm{X}_i-t_x}{h_n}\right)\right| \\ &\leqslant \sup_{x\in \mathrm{S}}\frac{\mathrm{C}}{h_n^2}|x-t_x| \\ &\leqslant \frac{l_n\mathrm{C}}{h_n^2} \\ &\leqslant \frac{\mathrm{C}}{(h_n n^{\xi})^2}=o\left(\frac{\log n}{nh_n}\right),\end{aligned}$$

et de manière évidente on a l'existence d'une constante C *telle que :*

$$\sup_{x\in \mathrm{S}}|\mathrm{E}f_n(x,h_n)-\mathrm{E}f_n(t_x,h_n)| \leqslant \frac{\mathrm{C}}{(n^{\xi}h_n)^2}.$$

En ce qui concerne le terme $\sup_{x\in \mathrm{S}}|f_n(t_x,h_n)-\mathrm{E}f_n(t_x,h_n)|$ *on a pout tout* $\epsilon>0$,

$$\begin{aligned}\mathrm{P}\left[\sup_{x\in \mathrm{S}}|f_n(t_x,h_n)-\mathrm{E}f_n(t_x,h_n)|>\epsilon\right] &= \mathrm{P}\left[\max_{k=1,\dots z_n}|f_n(t_k,h_n)-\mathrm{E}f_n(t_k,h_n)|>\epsilon\right] \\ &\leqslant z_n\mathrm{P}\left[\frac{1}{n}\sum_{i=1}^{n}\left|\,\mathrm{U}_i-\mathrm{EU}_i\,\right|>\epsilon\right]. \qquad (1.25)\end{aligned}$$

où $\mathrm{U}_i=\frac{1}{h_n}\mathrm{K}\left(\frac{\mathrm{X}_i-t_k}{h_n}\right)$.

En suivant les mêmes étapes que la preuve du lemme (1.4.5), nous pouvons alors majorer aisément U_i *et* EU_i^2*, comme suit*

$$\mathrm{U}_i\leqslant\frac{\mathrm{C}}{h_n} \quad et \quad \mathrm{EU}_i^2\leqslant\frac{\mathrm{C}}{h_n},$$

Nous somme en position d'appliquer l'inégalité de type exponentielle de Bernstein*, énoncer dans le corollaire (1.1.2), cette inégalité associée à (1.25) donnent directement,*

$$\begin{aligned}\mathrm{P}\left[\sup_{x\in \mathrm{S}}\left|\,f_n(t_x,h_n)-\mathrm{E}f_n(t_x,h_n)\,\right|>\epsilon\right] &\leqslant z_n\exp(-cn\epsilon^2h_n) \\ &\leqslant n^{2\xi}\exp(-cn\epsilon^2h_n). \qquad (1.26)\end{aligned}$$

En posant, $\epsilon = \epsilon_0\sqrt{\frac{\log n}{nh_n}}$*, on obtient*

$$\sum \mathrm{P}\left[\sup_{x\in \mathrm{S}} \left| f_n(t_x, h_n) - \mathrm{E}f_n(t_x, h_n) \right| > \epsilon_0\sqrt{\frac{\log n}{nh_n}}\right] < \infty, \tag{1.27}$$

pour ϵ_0 *choisi suffisamment grand.*

Remarque 1.4.1. *Nous pouvons définir l'estimateur de Parzen-Rosenblatt dans le cadre multivarié. Lorsque la variable* X *est à valeurs dans* $\mathbb{R}^p$ *l'estimateur s'écrit ,*

$$f_n(x, h_n) = \frac{1}{nh_n^p}\sum_{i=1}^{n} \mathrm{K}\left(\frac{\boldsymbol{X}_i - x}{h_n}\right), \qquad x \in \mathbb{R}^p \tag{1.28}$$

où $\mathrm{K} : \mathbb{R}^p \to \mathbb{R}$ *est une fonction multivarié définie comme le produit de noyaux univariés* K_j*, tels que* $\mathrm{K}(u) = \mathrm{K}(u_1, ..., u_p) = \prod_{j=1}^{p} \mathrm{K}_j(u_j), \;\; u \in \mathbb{R}^p$.

Chapitre 2

Estimation de la fonction de régression

2.1 Introduction

Le modèle de la régression est l'un des modèle les plus fréquemment rencontrés en statistique paramétrique et non paramétrique.

Soient $(\mathrm{X}_1, \mathrm{Y}_1), (\mathrm{X}_2, \mathrm{Y}_2), ..., (\mathrm{X}_n, \mathrm{Y}_n)$ *des couples de variables aléatoires indépendantes et de même loi que* (X, Y).

Dans le modèle de régression non paramétrique on suppose l' existence d'une fonction "r*" qui exprime la valeur moyenne de la variable à expliquer* Y *en fonction de la variable explicative* X*, c'est à dire :*

$$\mathrm{Y} = r(\mathrm{X}) + \epsilon,$$

où ϵ *est une variable centrée et indépendante de* X.
Dans notre travail, nous nous intéressons au modèle où les données $\{\mathrm{X}_i,\ 1 \leqslant i \leqslant n\}$ *sont strictement aléatoires et non dégénérées et nous suppoosons aucune hypothèse paramétrique sur la loi du couple (X,Y) ; d'une manière plus explicite, (X,Y) est à valeurs dans* $\mathbb{R}^2$*, il admet une densité jointe* $f(x, y)$ *sur* $\mathbb{R}^2$ *et une densité marginale* $f(x) > 0$ *(par rapport à la mesure de Lebesgue sur* $\mathbb{R}$*).*
La variable Y *est supposée intégrable, c'est à dire* $\mathrm{E}(|\mathrm{Y}|) < \infty$*, on peut alors définir la fonction régression r(x) par :*

$$r(x) = \mathrm{E}(\mathrm{Y}/\mathrm{X} = x) = \frac{\int_{\mathbb{R}} yf(x, y)dy}{\int_{\mathbb{R}} f(x, y)dy} = \frac{\int_{\mathbb{R}} yf(x, y)dy}{f(x)}.$$

r(x) est la fonction qui réalise la meilleure approximation de Y sachant X=x au sens des moindres carrés. Dans ce problème, l' estimation de r(x) est de type non paramétrique. L' estimateur de la fonction r(x) que nous considérons est l'estimateur à noyau introduit par Nadaraya-Watson, ce dernier appartient à la classe des estimateurs linéaires qui regroupe aussi les estimateurs par fonction splines, par projection ou séries orthogonales et par ondelettes.

2.2 La construction de l'estimateur de la fonction de régression

Dans cette section nous présentons la construction de l'estimateur à noyau de Nadaraya-Watson, dont l'idée remonte à Tukey (1961) où il introduit un estimateur à noyau de type, régressogramme de la fonction de régression défini par :

$$r_n(x,h_n)=\frac{\sum_{i=1}^{n}\mathrm{Y}_i 1_{[t_k,t_{k+1}[}(\mathrm{X}_i)}{\sum_{i=1}^{n}1_{[t_k,t_{k+1}[}(\mathrm{X}_i)} \qquad \textit{pour } x\in[t_k,t_{k+1}[,$$

où $[t_k,t_{k+1}[$, $k\in\mathbb{N}$ *est une partition du support de* X.

Bosq (1969) *fût le premier à donner une étude des propriétés statistiques de cet estimateur, il a montré la convergence uniforme presque sûre du régressogramme sur un intervalle* $[a,b]$*, quand*

$$h_n=\mathrm{O}(n^{-\alpha}), \qquad 0<\alpha<1.$$

Sabry (1978) *a obtenu la convergence uniforme presque sûre sur* $[0, \sqrt{\frac{logn}{loglogn}}]$*. En* (1982) Lecoutre *a procédé à l'extension de tous ces résultats à* $\mathbb{R}$ *; et pour une étude plus approfondie de cet estimateur (la convergence du biais et de la variance) on se réfère au livre de* Bosq *et* Lecoutre (1987).
De façon analogue à l'histogramme, afin d'éviter le problème du positionnement des bornes des intervalles de la partition, un autre estimateur a été construit, comme suit :

$$\forall x, \qquad r_n(x,h_n)=\frac{\sum \mathrm{Y}_i 1_{[x-h^{(\mathrm{X}_i)};x+h[}}{\sum 1_{[x-h_;^{(\mathrm{X}_i)}x+h[}}, \tag{2.1}$$

où h_n *un paramètre réel strictement positif.*
Le désavantage de ce nouveau estimateur est sa discontinuité, sa généralisation a été introduite et étudiée par Nadaraya (1964) *et* Walson (1964). *Ce dernier a vu son domaine d' application croître de plus en plus et beaucoup de résultats importants ont été obtenus en ex U.R.S.S aussi bien qu'ailleurs (cf Nadaraya (1989), Wand et Jones (1995), Watson(1964)).*
Une des multiples façons de construire l'estimateur de Nadaraya-Watson que nous abordons est d'utiliser les estimateurs de $f(x,y)$ *et de* $f(x)$*. Soient* $\mathrm{J}(x,y)$ *une densité de probabilité sur* $\mathbb{R}^2$*, posons* $\mathrm{J}_1(x)=\int_{-\infty}^{\infty}\mathrm{J}(x,y)dy$.
Si $h_n\to 0$*, quand* $n\to\infty$ *, alors*

$$f_n(x,y)=\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{J}\left(\frac{x-\mathrm{X}_i}{h_n},\frac{y-\mathrm{Y}_i}{h_n}\right),$$

$$g_n(x,h_n)=\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right),$$

peuvent être considérés respectivement, comme les estimateurs de $f(x,y)$ *et de* $f(x)$ *et*

$$\begin{aligned} r_n(x,h_n) &= \frac{\int_{-\infty}^{\infty} y f_n(x,y)dy}{g_n(x)} \\ &= \frac{h_n \sum_{i=1}^n m\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right)} + \frac{\sum_{i=1}^n \mathrm{Y}_i \mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right)}, \end{aligned} \tag{2.2}$$

où

$$m(x) = \int_{\mathrm{R}} y\mathrm{J}(x,y)dy,$$

peut être considéré comme, un estimateur de r(x). Pour un choix de

$$\mathrm{J}(x,y) = \mathrm{J}_1(x)\mathrm{L}(y),$$

avec

$$\int y\mathrm{L}(y)dy = 0,$$

alors $m(x) = 0$ *et l' estimateur défini dans la relation (2.2) se réduit à*

$$r_n(x,h_n) = \frac{\sum_{i=1}^n \mathrm{Y}_i \mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{J}_1\left(\frac{x-\mathrm{X}_i}{h_n}\right)}. \tag{2.3}$$

Donc l' estimateur à noyau de la régression est donné par :

$$r_n(x,h_n) = \begin{cases} \frac{\sum_{i=1}^n \mathrm{Y}_i \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)} = \frac{\Phi_n(x,h_n)}{f_n(x,h_n)}, & si \sum_{i=1}^n \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right) \neq 0\,; \\ \\ \frac{1}{n}\sum_{i=1}^n \mathrm{Y}_i, & si \sum_{i=1}^n \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right) = 0. \end{cases}$$

où

$$\Phi_n(x,h_n) = \frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right).$$

La construction de l' estimateur à noyau de Nadaraya-Watson dépend de deux paramètres, le paramètre de lissage h dont le choix est crucial pour obtenir de bonnes propriétés asymptotiques (citées ci-dessous) et le noyau K dont on ne peut pas négliger le rôle pour la réduction du biais. D'une manière analogue aux propriétés asymptotiques de l' estimateur de Parzen Rosenblatt, nous étudions dans cette partie deux modes de convergence, la convergence en moyenne quadratique et la convergence presque complète.
En plus des conditions (K.1-K.4) sur le noyau K, *nous avons besoin des hypothèses suivantes.*
(K.6) $\int_{\mathbb{R}} u\mathrm{K}(u)du = 0,$

(K.7) $\int_{\mathbb{R}} u^2\mathrm{K}(u)du < \infty.$

2.3 Les propriétés de l'estimateur de la fonction de régression

2.3.1 La consistance

En vu de la décomposition suivante :

$$\mathrm{E}(r_n(x,h_n)-r(x))^2=\mathrm{V}ar(r_n(x,h_n))+(\mathrm{E}r_n(x,h_n)-r(x))^2.$$

L'étude asymptotique du biais et de la variance de l' estimateur de Nadaraya-Watson détermine les conditions suffisantes à la consistance de cet estimateur.

1) Etude asymptotique de la variance

Proposition 2.3.1.

Sous les hypothèses de la proposition (1.4.2) et si $\mathrm{EY}^2<\infty$*, alors en chaque point de continuité des fonctions r(x), f(x) et* $\sigma^2(x)=\mathrm{V}ar(\mathrm{Y}/\mathrm{X}=x)$
On a

$$\mathrm{V}ar[r_n(x,h_n)]=\frac{1}{nh_n}\left\{\frac{\sigma^2(x)}{f(x)}\int_{\mathbb{R}}\mathrm{K}^2(u)du\right\}(o(1)+1).$$

où $\quad f(x)>0.$

Preuve
Soit la fonction $\psi(x)=\int_{\mathbb{R}}y^2f(x,y)dy$*, en se basant sur le lemme de Bochner on a*

$$\begin{aligned}
var(\phi_n(x,h_n)) &= \frac{1}{nh_n^2}\left\{\mathrm{E}\left[\mathrm{Y}^2\mathrm{K}^2\left(\frac{x-\mathrm{X}}{h_n}\right)\right]-\left[\mathrm{EYK}\left(\frac{x-\mathrm{X}}{h_n}\right)\right]^2\right\}.\\
&= \frac{1}{nh_n}\left\{\int_{\mathbb{R}}\mathrm{K}^2(u)\psi(x-h_nu)du-h_n\left(\int_{\mathbb{R}}\mathrm{K}(u)f(x-uh_n)r(x-h_nu)du\right)^2\right\}\\
&= \frac{1}{nh_n}\psi(x)\int_{\mathbb{R}}\mathrm{K}^2(u)du(1+o(1)),
\end{aligned}$$

$$\mathrm{E}\left[\{f_n(x,h_n)-\mathrm{E}f_n(x,h_n)\}\{\phi_n(x,h_n)-\mathrm{E}(\phi_n(x,h_n))\}\right]=\frac{1}{nh_n}\phi(x)\int_{\mathbb{R}}\mathrm{K}^2(u)du(1+o(1))$$

et

$$varf_n(x,h_n)=\frac{1}{nh_n}f(x)\int_{\mathbb{R}}\mathrm{K}^2(u)du(1+o(1)).$$

Posons

$$\mathrm{B}_n(x) = \begin{pmatrix} f_n(x,h_n) \\ \\ \phi_n(x,h_n) \end{pmatrix}$$

et

$$\mathrm{A}(x) = \begin{pmatrix} \frac{-r(x)}{[f(x)]^2}, & \frac{1}{f(x)} \end{pmatrix}.$$

La matrice de variance covariance de $\mathrm{B}_n(x)$ *est alors donnée par l'expression suivante*

$$\Sigma := \frac{1}{nh_n} \begin{pmatrix} f(x) & \phi(x) \\ \\ \phi(x) & \psi(x) \end{pmatrix} \int_{\mathbb{R}} \mathrm{K}^2(u)du \ (1+o(1)).$$

En remarquant, que

$$\begin{aligned} varr_n(x,h_n) &= \mathrm{A}\Sigma\mathrm{A}^t \\ &= \frac{1}{nh_n} \begin{pmatrix} \frac{\psi(x)}{[f(x)]^2} - \frac{(\phi(x))^2}{[f(x)]^3} \end{pmatrix} \int_{\mathbb{R}} \mathrm{K}^2(u)du \ (1+o(1)), \end{aligned}$$

où A^t *désigne la transposée de* A*, on obtient alors*

$$varr_n(x,h_n) = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{R}} \mathrm{K}^2(u)du \right\} \ (o(1)+1).$$

2) Etude asymptotique du biais

L'étude asymptotique du biais repose sur la proposition suivante.

Proposition 2.3.2.

Sous les hypothèses de la proposition (2.3.1) et

a) Si $|\mathrm{Y}| \leqslant \mathrm{C}_1 < \infty$ *P.S et si* $nh_n \to \infty$*, quand* $n \to \infty$*, alors*

$$\mathrm{E}r_n(x,h_n) = \frac{\mathrm{E}[\Phi_n(x,h_n)]}{\mathrm{E}[f_n(x,h_n)]} + \mathrm{O}\left(\frac{1}{nh_n}\right).$$

b) Si $\mathrm{EY}^2 < \infty$*,* $nh_n^2 \to \infty$*, quand* $n \to \infty$*,alors :*

$$\mathrm{E}r_n(x,h_n) = \frac{\mathrm{E}[\Phi_n(x,h_n)]}{\mathrm{E}[f_n(x,h_n)]} + \mathrm{O}\left(\frac{1}{\sqrt{n}h_n}\right).$$

Preuve :

En multipliant les deux expressions de l'identité suivante

$$\frac{1}{f_n(x,h_n)} = \frac{1}{\mathrm{E}f_n(x,h_n)} - \frac{f_n(x,h_n) - \mathrm{E}(f_n(x,h_n))}{[\mathrm{E}f_n(x,h_n)]^2} + \frac{[f_n(x,h_n) - \mathrm{E}f_n(x,h_n)]^2}{f_n(x,h_n)[\mathrm{E}f_n(x,h_n)]^2},$$

par $\Phi_n(x)$ *et en passant à l'espérance nous obtenons*

$$\begin{aligned} \mathrm{E}r_n(x,h_n) &= \frac{\mathrm{E}\phi_n(x)}{\mathrm{E}f_n(x,h_n)} - (\mathrm{E}f_n(x,h_n))^{-2}\mathrm{E}[(\phi_n(x,h_n) \\ &- \mathrm{E}[\phi_n(x,h_n)])(f_n(x,h_n) - \mathrm{E}f_n(x,h_n))] \\ &+ \mathrm{E}\{(f_n(x,h_n)^{-1}(\mathrm{E}f_n(x,h_n))^{-2}\phi_n(x,h_n)[f_n(x,h_n) - \mathrm{E}f_n(x,h_n)]^2\} \\ &= \frac{\mathrm{E}\phi_n(x,h_n)}{\mathrm{E}f_n(x,h_n)} + [c_n^1(x) + c_n^2(x)](\mathrm{E}f_n(x,h_n))^{-2}, \end{aligned} \tag{2.4}$$

où

$$c_n^{(1)}(x) = \mathrm{E}[(\Phi_n(x,h_n) - \mathrm{E}[\Phi_n(x,h_n)])(f_n(x,h_n) - \mathrm{E}[f_n(x,h_n)])],$$

$$c_n^{(2)}(x) = \mathrm{E}\{(f_n(x,h_n)^{-1}\Phi_n(x,h_n)[f_n(x,h_n) - \mathrm{E}f_n(x,h_n)]^2\}.$$

Comme $var[f_n(x,h_n)] \sim \frac{1}{nh_n}f(x)\int_{\mathbb{R}}\mathrm{K}^2(t)dt$ *(cf la preuve de la proposition (1.4.2) et*

$var[\Phi_n(x,h_n)] \sim \frac{1}{nh_n}\int_{\mathbb{R}} y^2 f(x,y)dy \int_{\mathbb{R}}\mathrm{K}^2(t)dt$ *(cf la proposition précédente), alors*

$$|c_n^1(x)| \leqslant (var\phi_n(x,h_n))^{\frac{1}{2}}(varf_n(x,h_n))^{\frac{1}{2}} = \mathrm{O}\left(\frac{1}{nh_n}\right). \tag{2.5}$$

Nous constatons aussi que l'hypothèse a) implique l'inégalité suivante

$$|c_n^2(x)| \leqslant \mathrm{C}_1 var(f_n(x,h_n)) \sim \frac{1}{nh_n}f(x)\int_{\mathbb{R}}\mathrm{K}^2(t)dt = \mathrm{O}\left(\frac{1}{nh_n}\right) \tag{2.6}$$

En combinant les relations (2.4), (2.5), (2.6) nous obtenons le résultat a).
Pour montrer le cas b), il suffit de remarquer que la relation (2.5) est toujours valable mais la relation (2.6) devient

$$\begin{aligned} |c_n^2(x)| &\leqslant \mathrm{E}\{\max_{1\leqslant i\leqslant n}|\mathrm{Y}_i|\}\mathrm{E}[f_n(x,h_n) - \mathrm{E}f_n(x,h_n)]^2, \\ &\leqslant \left[\sum_{i=1}^n \mathrm{EY}_i^2\right]^{\frac{1}{2}}\left[\mathrm{E}\left[f_n(x,h_n) - \mathrm{E}\left[f_n(x,h_n)\right]\right]^4\right]^{\frac{1}{2}} \\ &= \sqrt{n}(\mathrm{EY}^2)^{\frac{1}{2}}\mathrm{O}\left(\frac{1}{nh_n}\right) \\ &= \mathrm{O}\left(\frac{1}{\sqrt{n}h_n}\right). \end{aligned} \tag{2.7}$$

Les relations, (2.5), (2.6) et (2.7) donnent le résultat b).

Maintenant nous sommes en mesure d'énoncer le résultat suivant.

Proposition 2.3.3.

Si les conditions (K.4), (K.6) et (K.7) sont vérifiées et si $f(.)$ et $r(.)$ sont de classe $\mathrm{C}^2(\mathbb{R})$ et si $|\mathrm{Y}|$ est borné.
Alors :

$$\mathrm{E}(r_n(x,h_n)) - r(x) = \frac{h_n^2}{2}\left\{\left\{r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right\}\int_{\mathbb{R}} u^2\mathrm{K}(u)du\right\}(1+o(1)). \qquad (2.8)$$

Remarque 2.3.1.
1)- Les conditions (K.4), (K.6) et (K.7) peuvent être remplacées par le noyau K est d'ordre 2 au sens de Gasser et Müller.
2)-$o(1)$ dans la relation (2.8) est égale à $\mathrm{O}(h) + \mathrm{O}((nh)^{-1})$.

Preuve
On a :

$$\begin{aligned}\frac{\mathrm{E}(\Phi_n(x,h_n))}{\mathrm{E}(f_n(x,h_n))} - r(x) &= \left[\mathrm{EK}\left(\frac{x-\mathrm{X}}{h_n}\right)\right]^{-1}\left\{\int_{\mathbb{R}}\frac{1}{h_n}\mathrm{K}\left(\frac{x-t}{h_n}\right)\phi(t)dt - r(x)\int_{\mathbb{R}}\frac{1}{h_n}\mathrm{K}\left(\frac{x-t}{h_n}\right)f(t)dt\right\} \\ &= \left\{(f(x))^{-1}\left\{\frac{h_n^2}{2}\phi''(x) - \frac{h_n^2}{2}r(x)f''(x)\right\}\int_{\mathbb{R}} u^2\mathrm{K}(u)du + \phi(x) - r(x)f(x)\right\}(1+o(1\end{aligned}$$

Comme $\phi(x) = r(x)f(x)$. L'équation précédente peut s'écrire :

$$\mathrm{E}r_n(x,h_n) - r(x) = \left\{\frac{h_n^2}{2}\left\{r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right\}\int_{\mathbb{R}} u^2\mathrm{K}(u)du\right\}(1+o(1)). \qquad (2.9)$$

D'où

$$\lim_{n\to\infty}\mathrm{E}r_n(x,h_n) = r(x).$$

2.3.2 Convergence presque complète

En se basant sur la preuve donnée dans Ferraty *et* Vieu (2003)*, nous traitons dans ce paragraphe la convergence presque complète de l'estimateur à noyau de la fonction de régression. Nous gardons quelques conditions précédentes, auxquelles nous rajoutons les hypothèses suivantes.*

- ***f, r sont des fonctions continues au voisinage de*** x***, un point fixé de*** $\mathbb{R}$**.** (2.10)

La densité f *et la variable* Y *sont telles que*

$$\boldsymbol{f > 0} \tag{2.11}$$

et

$$|\mathrm{Y}| < \mathrm{M} < \infty, \tag{2.12}$$

où M *est une constante réelle positive.*

Théorème 2.3.1.

Sous les hypothèses (1.8)*, (2.10), (2.11),* (2.12)*,(K.4) et (K.5), on a :*

$$\lim_{n\to\infty} r_n(x, h_n) = r(x). \qquad a.co$$

Démonstration .

La Démonstration de ce théorème est basée sur la décomposition suivante :

$$\begin{aligned} r_n(x,h_n) - r(x) &= \frac{1}{f_n(x,h_n)}\left[(\phi_n(x,h_n) - \mathrm{E}\phi_n(x,h_n)) + (\mathrm{E}\phi_n(x,h_n) - \phi(x))\right] \\ &+ \left[(f(x) - \mathrm{E}f_n(x,h_n)) + (\mathrm{E}f_n(x,h_n) - f_n(x,h_n))\right]\frac{r(x)}{f_n(x,h_n)}. \end{aligned}$$

où $\phi(x) = r(x)f(x)$*. Le résultat énoncé découle des lemmes suivants :*

Lemme 2.3.1.

Sous les hypothèses (2.10), (2.12), (K.4) et (K.5) on a :

$$\lim_{n\to\infty} \mathrm{E}\phi_n(x, h_n) = \phi(x).$$

Preuve .
Nous avons :

$$\begin{aligned} \mathrm{E}\phi_n(x,h_n) &= \mathrm{E}\left[\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{Y}_i\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right] \\ &= \frac{1}{h_n}\mathrm{E}\left[\mathrm{YK}\left(\frac{x-\mathrm{X}}{h_n}\right)\right]. \end{aligned}$$

Le conditionnement par rapport à $\mathrm{X} = x$ *nous donne :*

$$\mathrm{E}\phi_n(x,h_n) = \frac{1}{h_n}\mathrm{E}\left[r(x)\mathrm{K}\left(\frac{x-\mathrm{X}}{h_n}\right)\right];$$

où

$$\begin{aligned}\mathrm{E}\phi_n(x,h_n) &= \frac{1}{h_n}\int r(t)\mathrm{K}\left(\frac{x-t}{h_n}\right)f(t)dt \\ &= \frac{1}{h_n}\int \phi(t)\mathrm{K}\left(\frac{x-t}{h_n}\right)dt.\end{aligned}$$

En posant $u=\frac{x-t}{h_n}$, *on obtient :*

$$\mathrm{E}g(x) = \int \phi(x-uh_n)\mathrm{K}(u)du.$$

Comme K *est à support compact, la continuité uniforme de* ϕ *et la condition (K.4), nous donnent :*

$$\lim_{n\to\infty}\mathrm{E}\phi_n(x,h_n)=\phi(x).$$

Lemme 2.3.2.

Sous les hypothèses (2.10), (2.12), (K.5) et (K.4) on a :

$$\lim_{n\to\infty}(\mathrm{E}\phi_n(x,h_n)-\phi_n(x,h_n))=0 \qquad a.co$$

Preuve .
On a :

$$\phi_n(x,h_n)-\mathrm{E}\phi_n(x,h_n)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{Z}_i,$$

où

$$\mathrm{Z}_i=\frac{1}{h}\left[\mathrm{Y}_i\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)-\mathrm{E}\left(\mathrm{Y}_i\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right)\right].$$

De plus, la condition (2.12) et (K.5) nous donnent :

$$|\mathrm{Z}_i|\leqslant\frac{\mathrm{C}}{h_n}.$$

Occupons nous maintenant du moment d'ordre 2 de Z_i,

$$\mathrm{EZ}_i^2 = var\left(\frac{1}{h_n}\mathrm{Y}_i\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right) \leqslant \mathrm{E}\Gamma_i^2,$$

où

$$\Gamma_i = \frac{1}{h_n}\mathrm{Y}_i\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right).$$

En utilisant le conditionnement par rapport à la variable X*, on obtient*

$$\begin{aligned}\mathrm{E}\Gamma_i^2 &= h_n^{-2}\mathrm{E}\left[\Phi(\mathrm{X})\mathrm{K}^2\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right] \\ &= h_n^{-2}\int \Phi(t)\mathrm{K}^2\left(\frac{x-t}{h_n}\right)f(t)dt,\end{aligned}$$

où

$$\Phi(x) = \mathrm{E}[\mathrm{Y}^2/\mathrm{X}=x].$$

Le changement de variable $z = \frac{x-t}{h_n}$*, nous permet d'écrire :*

$$\mathrm{E}\Gamma_i^2 = \frac{1}{h_n}\int \Phi(x-zh_n)f(x-zh_n)\mathrm{K}^2(z)dz\ .$$

la continuité de f *sur le support compact* K*, les conditions (2.12) et (K.4) impliquent :*

$$\mathrm{E}\Gamma_i^2 \leqslant \frac{\mathrm{C}}{h_n}.$$

Les conditions du corollaire (1.1.2) étant satisfaites, nous déduisons :

$$\frac{1}{n}\sum_i \mathrm{Z}_i = \mathrm{O}_{a.co}\left(\sqrt{\frac{\log n}{nh_n}}\right).$$

La convergence presque complète de la densité établie dans le paragraphe précédent assure les convergences suivantes

$$\lim_{n\to\infty}\mathrm{E}[f_n(x,h_n)] = f(x),$$

$$\lim_{n\to\infty}\mathrm{E}[f_n(x,h_n)] - f(x) = 0 \qquad a.co.$$

Il ne nous reste qu'à prouver le lemme suivant pour finaliser la démonstration de ce théorème.

Lemme 2.3.3.

Sous les conditions (1.8), (2.10), (K.4) et (K.5) on a

$$\exists \delta > 0, \qquad \sum_{n=1} \mathrm{P}[f_n(x, h_n) \leqslant \delta] < \infty. \tag{2.13}$$

Le théorème (1.4.3), entraîne la convergence presque complète de $f_n(x, h_n)$ vers $f(x)$, c'est à dire

$$\forall \epsilon, \qquad \sum_{n=1} \mathrm{P}\left[|f_n(x, h_n) - f(x)| > \epsilon\right] < \infty.$$

De plus on a

$$|f_n(x, h_n)| \leqslant \frac{f(x)}{2} \Longrightarrow |f_n(x, h_n) - f(x)| > \frac{f(x)}{2}.$$

D'où

$$\mathrm{P}\left[|f_n(x, h_n) \leqslant \frac{f(x)}{2}|\right] \leqslant \mathrm{P}\left[|f_n(x, h_n) - f(x)| > \frac{f(x)}{2}\right].$$

Comme $f(x) > 0$, en posant $\delta = \epsilon = \frac{f(x)}{2}$, on arrive au résultat.

Le résultat du théorème est obtenu comme une conséquence directe des lemmes précédents. Pour l'étude de la vitesse de convergence presque complète ponctuelle, nous essayons de considérer le modèle suivant, en gardant toutes les hypothèses précédentes.

r et f sont k fois continûment dérivables autour de x. (2.14)

Théorème 2.3.2.

Considérons le modèle (2.14) avec $k > 0$ et supposons que les hypothèses (1.8), (1.9), (2.12), (2.11), (K.5) soient réalisées, alors on a

$$|r_n(x, h_n) - r(x)| = \mathrm{O}(h_n^k) + \mathrm{O}\left(\sqrt{\frac{\log n}{n h_n}}\right) \qquad a.co.$$

En utilisant la décomposition précédente, le résultat énoncé se base sur le lemme suivant :

Lemme 2.3.4.
Sous les hypothèses (1.9), (2.14) et (K.5)

$$\mathrm{E}\phi_n(x, h_n) - \phi(x) = \mathrm{O}(h_n^k). \tag{2.15}$$

Preuve .
L'expression de $\mathrm{E}\phi_n(x, h_n)$ est analogue à la précédente. En effet, on a :

$$\mathrm{E}\phi_n(x, h_n) = \int \phi(x - z h_n)\mathrm{K}(z)dz.$$

Le modèle (2.14), nous permet de développer ϕ au voisinage de x, ceci nous permet d'écrire :

$$\phi(x - h_n z) = \phi(x) + \sum_{j=1}^{k-1} \frac{(-1)^j (z h_n)^j}{j!} \phi^j(x) + \frac{(-1)^k (z h_n)^k}{k!} \phi^k(\theta_z),$$

où θ_z est entre x et $x - z h_n$.

La condition (1.9) sur K, *implique :*

$$\mathrm{E}\phi_n(x, h_n) = \phi(x) + (-1)^k h^k \frac{\int z^k \mathrm{K}(z) \phi^k(\theta_z) dz}{k!}.$$

La convergence uniforme de $\phi^k(\theta_z)$ vers $\phi^k(x)$ (assurée par le modèle (2.14)) et la condition (K.5), nous donnent

$$\mathrm{E}\phi_n(x, h_n) - \phi(x) = \mathrm{O}(h_n^k).$$

Les lemmes (1.4.5), (2.3.2), (2.3.3) et (2.3.4) nous assurent que :

$$\mathrm{E}f_n(x, h_n) - f_n(x, h_n) = \mathrm{O}\left(\sqrt{\frac{\log n}{n h_n}}\right), \qquad a.co$$

$$\mathrm{E}g_n(x, h_n) - g_n(x, h_n) = \mathrm{O}\left(\sqrt{\frac{\log n}{n h_n}}\right), \qquad a.co$$

et

$$\exists \delta > 0 \qquad \sum_{n=1} \mathrm{P}[f_n(x, h_n) \leqslant \delta] < \infty.$$

En combinant tous les résultats cités précédemment, on arrive au résultat cherché.
Nous essayons maintenant d'établir une vitesse convergence presque complète uniforme. Il suffit d'une part de considérer un compact S *de* $\mathbb{R}$ *tel que la condition (2.14) soit remplacée par le modèle suivant*

$$\textbf{\textit{r et f sont k continûment dérivables sur S.}} \tag{2.16}$$

et de supposer d'autre part l'existence de $\theta > 0$, tel que

$$\inf_{x \in \mathrm{S}} f(x) > \theta. \tag{2.17}$$

Nous gardons toutes les conditions citées précédemment, auxquelles nous rajoutons la condition Lipschitzienne*(1.19) sur le noyau* K *et la condition (1.20) sur le paramètre de lissage h.*

Théorème 2.3.3.
Soient les modèles (2.16), (2.17) avec k>0 et les conditions (K.5), (2.12), (1.8), (1.9), (1.19) et (1.20), on a

$$\sup_{x\in \mathrm{S}} |r_n(x,h_n) - r(x)| = \mathrm{O}(h_n^k) + \mathrm{O}\left(\sqrt{\frac{\log n}{nh_n}}\right), \qquad a.co$$

Preuve .
Une décomposition similaire au cas ponctuel, nous permet d'écrire

$$\begin{aligned}
\sup_{x\in \mathrm{S}} |r_n(x,h_n) - r(x)| &\leqslant \frac{\sup_{x\in \mathrm{S}} |\phi_n(x,h_n) - \phi(x)|}{\inf_{x\in \mathrm{S}} |f_n(x,h_n)|} + \sup_{x\in \mathrm{S}} |f(x) - f_n(x,h_n)| \frac{\sup_{x\in \mathrm{S}} |r(x)|}{\inf_{x\in \mathrm{S}} |f_n(x,h_n)|} \\
&\leqslant \frac{\sup_{x\in \mathrm{S}} |\phi_n(x,h_n) - \mathrm{E}(\phi_n(x,h_n))|}{\inf_{x\in \mathrm{S}} |f_n(x,h_n)|} + \frac{\sup_{x\in \mathrm{S}} |\mathrm{E}(\phi_n(x,h_n)) - \phi(x)|}{\inf_{x\in \mathrm{S}} |f_n(x,h_n)|} \\
&+ \left\{ \sup_{x\in \mathrm{S}} |f(x) - \mathrm{E}(f_n(x,h_n))| + \sup_{x\in \mathrm{S}} |\mathrm{E}(f_n(x,h_n)) - f_n(x,h_n)| \right\} \frac{\sup_{x\in \mathrm{S}} |r(x)|}{\inf_{x\in \mathrm{S}} |f_n(x,h_n)|}
\end{aligned}$$

Les approximations en $\mathrm{O}(h_n^k)$ *traitées précédemment peuvent se généraliser via (K.5) et (2.14) comme suit*

$$\sup_{x\in \mathrm{S}} |\mathrm{E}g_n(x,h_n) - g(x)| = \mathrm{O}(h_n^k)$$

et

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

Comme r est borné, la preuve de ce théorème s' achèvera à partir des lemmes suivants :

Lemme 2.3.5. *Sous les hypothèses (1.8), (2.16), (1.19) et (1.20) on a*

$$\sup_{x\in \mathrm{S}} |\mathrm{E}\phi_n(x,h_n) - \phi_n(x,h_n)| = \mathrm{O}\left(\sqrt{\frac{\log n}{nh_n}}\right), \qquad a.co$$

Preuve .
S *est un compact de* $\mathbb{R}$*, il existe un recouvrement fini de* S *tel que :*

$$\mathrm{S} \subset \cup_{k=1}^{z_n} \mathrm{S}_k$$

où

$$\mathrm{S}_k =]t_k - l_n, t_k + l_n[\ \ et \ \ l_n = n^{-\beta}$$

Posons

$$t_x = \arg \min_{t\in\{t_1,t_2\ldots,t_{z_n}\}} |x - t|$$

avec

$$l_n = n^{-2\xi}, \ l_n = \mathrm{C}z_n^{-1}.$$

On a

$$\sup_{y\in \mathrm{S}} |\mathrm{E}\phi_n(x,h_n) - \phi_n(x,h_n)| \leqslant \mathrm{A}_1 + \mathrm{A}_2 + \mathrm{A}_3;$$

où

$$\mathrm{A}_1 = \sup_{y\in \mathrm{S}} |\phi_n(x,h_n) - \phi_n(t_x,h_n)|\,,$$

$$\mathrm{A}_2 = \sup_{y\in \mathrm{S}} |\phi_n(t_x,h_n) - \mathrm{E}\phi_n(t_x,h_n)|\,,$$

$$\mathrm{A}_3 = \sup_{y\in \mathrm{S}} |\mathrm{E}(\phi_n(t_x,h_n)) - \mathrm{E}\phi_n(x,h_n)|\,.$$

Concernant le terme A_1*, comme le noyau* K *est lipschitzien et la variable* Y *est bornée, on a*

$$\begin{aligned} |\phi_n(t_x,h_n) - \phi_n(x,h_n)| &= \frac{1}{nh_n}\sum_{i=1}^{n} |\mathrm{Y}_i| \left| \left[\mathrm{K}\left(\frac{t_x - \mathrm{X}_i}{h_n}\right) - \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)\right] \right| \\ &\leqslant \frac{\mathrm{C}}{h_n}\frac{|t_x - x_i|}{h_n^2} \\ &= \frac{\mathrm{C}l_n}{h_n^2}. \end{aligned}$$

La condition (1.20) implique :

$$\mathrm{A}_1 = o\left(\frac{\log n}{nh_n}\right).$$

Une manière de démonstration analogue à la précédente, nous permet d'écrire

$$\mathrm{A}_3 = o\left(\frac{\log n}{nh_n}\right).$$

Pour ce qui concerne le terme A_2*, on a :* $\forall \epsilon > 0$

$$\begin{aligned} \mathrm{P}\left[\sup_{y\in \mathrm{S}} |\phi_n(t_x,h_n) - \mathrm{E}\phi_n(t_x,h_n)| > \epsilon\right] &= \mathrm{P}\left[\max_{\{j=1,\ldots,z_n\}} |\phi_n(t_j,h_n) - \mathrm{E}\phi_n(t_j,h_n)| > \epsilon\right] \\ &\leqslant z_n \mathrm{P}\left[|\phi_n(t_j,h_n) - \mathrm{E}\phi_n(t_j,h_n)| > \epsilon\right] \\ &\leqslant z_n \mathrm{P}\left[\frac{1}{nh_n}\sum_{i=1}^{n} |\mathrm{U}_i - \mathrm{EU}_i| > \epsilon\right]. \end{aligned}$$

où

$$\mathrm{U}_i = \mathrm{Y}_i \mathrm{K}\left(\frac{\mathrm{X}_i - t_j}{h_n}\right).$$

Il suffit de trouver des majorants pour U_i *et* EU_i^2*, pour pouvoir appliquer le corollaire (1.1.2). D'après la démonstration du lemme (1.4.5), on a .*

$$|\mathrm{U}_i| \leqslant \frac{\mathrm{C}}{h_n} \qquad et \qquad \mathrm{EU}_i^2 \leqslant \frac{\mathrm{C}}{h_n}$$

Maintenant nous sommes en mesure d'appliquer le corollaire (1.1.2) :

$$\forall \epsilon > 0, \qquad \mathrm{P}\left[\sup_{y\in \mathrm{S}} |\phi_n(t_x, h_n) - \mathrm{E}\phi_n(t_x, h_n)| > \epsilon\right] \leqslant n^{2\xi} \exp\left(-\mathrm{C} n\epsilon^2 h\right).$$

Comme

$$\lim_{n\to\infty} \sqrt{\frac{\log(n)}{nh_n}} = 0$$

en posant

$$\epsilon = \epsilon_0 \sqrt{\frac{\log(n)}{nh_n}},$$

on obtient alors, pour un choix adéquat de ϵ_0

$$\forall \epsilon > 0, \qquad \sum_n \mathrm{P}\left[\sup_{x\in \mathrm{S}} |\phi_n(t_x, h_n) - \mathrm{E}\phi_n(t_x, h_n)| > \epsilon_0 \sqrt{\frac{\log(n)}{nh_n}}\right] < \infty.$$

Pour achever la preuve du théorème, énonçons le dernier lemme suivant.

Lemme 2.3.6.
Sous les conditions (K.5), (1.8), (1.9), (2.16), (2.17) on a :

$$\exists \delta > 0\ , \qquad \sum_{n=1}^{\infty} \mathrm{P}[\inf_{x\in \mathrm{S}} |f_n(x, h_n)| \leqslant \delta] < \infty. \tag{2.18}$$

Preuve .
D'après le théorème (1.4.5) on a

$$\forall \epsilon > 0, \ \sum_{n=1}^{\infty} \mathrm{P}[\sup_{x\in \mathrm{S}} |f_n(x, h_n) - f(x)| > \epsilon] < \infty\,.$$

Si

$$\inf_{x\in \mathrm{S}} |f_n(x, h_n)| \leqslant \frac{\theta}{2},$$

alors

$$\sup_{x\in \mathrm{S}} |f_n(x, h_n) - f(x)| > \frac{\theta}{2}.$$

où θ *vérifie (2.17).*
Pour $\epsilon = \frac{\theta}{2}$*, la convergence uniforme complète* $f_n(x, h_n)$ *vers* $f(x)$ *nous donne*

$$\sum_{n=1}^{\infty} \mathrm{P}\left[\inf_{x\in \mathrm{S}} |f_n(x, h_n)| \leqslant \frac{\theta}{2}\right] < +\infty.$$

Chapitre 3

Le choix du paramètre de lissage .

Les vitesses de convergence données dans les chapitres précédents dépendent de deux paramètres : la fonction de noyau K *dont l'efficacité est peu influente et le paramètre de lissage* h_n*, dont le choix est crucial aussi bien pour l'approche ponctuelle que pour la globale que nous exposons ci après.*
Il est utile de rappeler les hypothèses, précédemment introduites, et qui sont nécessaires pour la suite de notre travail.

- $f(.)$ *et* $r(.)$ *sont de classes* $\mathrm{C}^2(\mathbb{R})$*,*
- $\lim_{n\to\infty} h_n = 0, \qquad$ *quand* $n \to \infty$,
- $|\mathrm{Y}|$ *est bornée.*
- K *satisfait les propriétés suivantes : (K.1)- (K.4) et (K.6)-(K.7).*

3.1 Choix du paramètre de lissage

3.1.1 Etude du critère d'erreur quadratique moyenne de $r_n(x, h_n)$

L'erreur quadratique moyenne MSE (mean square error) est une mesure permettant d'évaluer la similarité de r_n *par rapport à la fonction de régression inconnue* r*, au point* x *donné de* $\mathbb{R}$*.*

Notre but est de minimiser

$$\mathrm{MSE}(r_n(x, h_n)) = \mathrm{E}(r_n(x, h_n) - r(x))^2.$$

Le développement de cette expression faite précédemment, nous donne

$$\mathrm{MSE}(r_n(x, h_n)) = varr_n(x, h_n) + (biais(r_n(x, h_n))^2.$$

Nous constatons d'une part que les expressions du biais de $r_n(x,h_n)$ et de la variance de $r_n(x,h_n)$ (voir les propositions (2.3.3), (2.3.1)) permettent de conclure qu'une grande valeur de h_n donne une augmentation du biais et une diminution de la variance (estimation fortement biaisée) et qu'un faible paramètre h_n, donne une diminution du biais et une augmentation de la variance (phénomène de sous lissage).

D'autre part, sous les hypothèses de ces mêmes propositions, nous obtenons

$$\mathrm{MSE}(r_n(x,h_n)) = \frac{h_n^4}{4}\left\{\left[r''(x)+2r'(x)\frac{f'(x)}{f(x)}\right][u^2\mathrm{K}(u)]+o(1)\right\}^2+\frac{1}{nh_n}\left\{\frac{\sigma^2(x)}{f(x)}[\mathrm{K}^2]\right\}(1+o(1)),$$

où $[u^p\mathrm{K}^q(u)] = \int t^p\mathrm{K}^q(t)dt$. Pour trouver donc un compromis entre le biais et la variance nous minimisons par rapport à h_n l'expression de l'erreur quadratique moyenne asymptotique AMSE (asymptotique mean squared error) donnée par

$$\mathrm{AMSE}[r_n(x,h_n)] = \frac{h_n^4}{4}\left\{r''(x)+2r'(x)\frac{f'(x)}{f(x)}\right\}^2[u^2\mathrm{K}]^2+\frac{1}{nh_n}\times\frac{\sigma^2(x)}{f(x)}[\mathrm{K}^2(u)].$$

Comme AMSE est une fonction convexe. La fênetre $h^{\mathrm{MSE}}_{opt(r_n(x,h_n))} = \arg\min_h[\mathrm{AMSE}(r_n(x,h_n))]$ est solution de l'équation suivante

$$\frac{\partial}{\partial h_n}\left[\frac{h_n^4}{4}\left\{r''(x)+2r'(x)\frac{f'(x)}{f(x)}\right\}^2[u^2\mathrm{K}]^2+\frac{1}{nh_n}\times\frac{\sigma^2(x)}{f(x)}(\mathrm{K}^2(u))\right]=0.$$

lorsque $\left[r''(x)+2r'(x)\frac{f'(x)}{f(x)}\right]^2[u^2\mathrm{K}]\neq 0,$

d'où

$$h^{\mathrm{MSE}}_{opt(r_n(x,h_n))} = n^{\frac{-1}{5}}\left\{\frac{\frac{\sigma^2(x)}{f(x)}[\mathrm{K}^2(x)]}{\left\{\left[r''(x)+2r'(x)\frac{f'(x)}{f(x)}\right][t^2\mathrm{K}]\right\}^2}\right\}^{\frac{1}{5}}.$$

On s'intéresse maintenant à l'approche globale pour la sélection du paramètre h_n, pour cela on introduit le critère d'erreur quadratique intégrée moyenne ou MISE (mean integrated squared error) de $r_n(x,h_n)$.

$$\mathrm{MISE}[r_n(x,h_n)] \;=\; \mathrm{E}\left[\int_{\mathbb{R}}(r_n(x,h_n)-r(x))^2dx\right],$$

En appliquant le théorème de Fubini, on a

$$\mathrm{MISE}[r_n(x,h_n)] = \left[\int_{\mathbb{R}} \mathrm{E}(r_n(x,h_n) - r(x))^2 dx\right],$$

Sous les mêmes hypothèses que les propositions (2.3.3) et (2.3.1), on a

$$\mathrm{AMISE}r_n(x,h_n) = \frac{h_n^4}{4}\int\left\{r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right\}^2 dx[u^2\mathrm{K}] + \frac{1}{nh_n}\int\frac{\sigma(x)^2}{f(x)}dx[\mathrm{K}^2(u)].$$

La fênetre $h^{\mathrm{MISE}}_{opt(r_n(x,h_n))}$ *minimisant l'AMISE du critère global est :*

$$h^{\mathrm{MISE}}_{opt(r_n(x,h_n))} = n^{\frac{-1}{5}}\left\{\frac{\int\frac{\sigma_I^2(x)}{f(x)}[\mathrm{K}^2]dx}{\int\left\{r''(x) + 2r'(x)\frac{f'(x)}{f(x)}\right\}^2 dx[t^2\mathrm{K}]}\right\}^{\frac{1}{5}}.$$

Un travail similaire se fait pour le choix optimum du paramètre de lissage dans le cas de l' estimateur de Parzen- Rosemblatt, nous obtenons :

$$h^{\mathrm{MSE}}_{opt(f_n(x,h_n))} = n^{\frac{-1}{5}}\left\{\frac{f(x)[\mathrm{K}^2]}{(f''(x))^2[t^2\mathrm{K}]^2}\right\}^{\frac{1}{5}}, \tag{3.1}$$

$$h^{\mathrm{MISE}}_{opt(f_n(x,h_n))} = n^{\frac{-1}{5}}\left\{\frac{[\mathrm{K}^2]}{[t^2\mathrm{K}]^2\int_{\mathrm{R}}(f''(x))^2 dx}\right\}^{\frac{1}{5}}. \tag{3.2}$$

quand $f''(x) \neq 0$.

En insérant $h^{\mathrm{MISE}}_{opt(f_n(x,h_n))}$ *dans* $\mathrm{MSE}[f_n(x,h_n)]$ *on peut montrer que le taux de convergence est d'ordre* $n^{-4/5}$*, il est plus faible que celui de l'histogramme dont l'ordre est égale à* $n^{-2/3}$.

Nous notons que l'expression de h_n *optimal, minimisant asymptotiquement les quatre critères d'erreurs à la forme* $\mathrm{C}n^{-1/5}$ *où la constante C est en fonction de la distribution et de termes aléatoires inconnues . Pour parer à cette difficulté, il existe deux famille de méthodes célèbres.*
La famille des méthodes de validation croisées (cross validation) et la famille des méthodes Plug in. Nous nous limitons à une méthode de Plug in pour la densité et une méthode de la validation croisée pour la régression.

3.2 Quelques méthodes d'optimisation de h_n

3.2.1 La méthode Plug in (ré-injection)

L'idée de cette méthode consiste à estimer la partie inconnue dans l'équation (3.2) c'est à dire la deuxième dérivée de f, par un estimateur consistant (ceci revient à estimer la seconde dérivée de f(x) avant f(x)).

Cette estimation nécessite un choix d' un nouveau noyau (il peut être égale à K) et d'un nouveau paramètre de lissage. Ce dernier est choisi d'une façon optimale par rapport à AMISE et à ce nouveau noyau.

La nouvelle équation de h_{opt} alors identique à l' équation (3.2), mais dont le dénominateur apparaît la quatrième dérivée de f(x) .

Il est clair que ce processus ne s' arrête pas, pour remédier à ce problème, Sheather, Jones et Marron (1996) ont proposé un nombre d'estimations successives égale à deux, leurs méthodes sont alors décrites jusqu'à la sixième dérivée.

Rule of thumb (La méthode du pouce)

Cette méthode consiste à remplacer f dans $h^{\mathrm{MSE}}_{opt(f_n(x,h_n))}$ par une densité normale d'écart type σ^2 inconnue à estimer.
Dans ce cas, la fenêtre donnée par la méthode du pouce, s'écrit

$$h_{\mathrm{RT}} = \left\{ \frac{8\sqrt{\pi}[\mathrm{K}^2]}{3[t^2\mathrm{K}^2]^2} \right\}^{\frac{1}{5}} n^{-1/5}\widehat{\sigma},$$

où

$$\widehat{\sigma} = \min\left\{ \hat{\mathrm{S}}, \frac{\mathrm{IQ}}{1,349} \right\},$$

$\hat{\mathrm{S}}$ *est l'estimateur de l'écart type et* IQ *est l'estimateur de l'écart interquartile.*

Notons que pour un noyau Gaussien *(respectivement* Epanechnikov *et* cubique*), les calculs nous donnent.*

$$h_{\mathrm{RT}} = \left\{ \frac{8\sqrt{\pi}}{3[t^2\mathrm{K}^2]^2} \right\}^{\frac{1}{5}} n^{-1/5}\widehat{\sigma} = 1,06\ \widehat{\sigma}n^{-1/5} \qquad (respect \quad 2,34\,\widehat{\sigma}n^{-1/5}\ et\ 2,78\,\widehat{\sigma}n^{-1/5}).$$

3.2.2 La méthode de validation croisée pour la régression

La méthode de la validation croisée, concernant l'estimation non paramétrique de la régression, a été étudiée par Hall (1984)*,* Härdle *et Marron* (1985)*,* Härdle *et Kelly (1987). L'idée*

de la validation croisée consiste à minimiser par rapport à h_n la distance $d_1(r_n, r)$ définie par

$$d_1(r_n, r) = \int_{\mathbb{R}} \left(r_n(x, h_n) - r(x)\right)^2 \omega(x) f(x) dx,$$

où $\omega(.)$ est une fonction de poids arbitraire (voir Härdle et Kelly (1987) *).*
On a

$$d_1(r_n, r) = \int_{\mathbb{R}} r_n(x, h_n)^2 \omega(x) f(x) dx - 2 \int_{\mathbb{R}} r_n(x, h_n) r(x) \omega(x) f(x) dx + \int_{\mathbb{R}} r(x)^2 \omega(x) f(x) dx.$$

Comme le dernier terme ne dépend pas de h_n, la minimisation $d_1(r_n, r)$ revient à optimiser :

$$\int_{\mathbb{R}} r_n(x, h_n)^2 \omega(x) f(x) dx - 2 \int_{\mathbb{R}} r_n(x, h_n) r(x) \omega(x) f(x) dx. \tag{3.3}$$

Nous constatons que la fonction de régression $r(.)$ et la densité $f(.)$ sont inconnues de plus

$$\int_{\mathbb{R}} r_n(x, h_n) r(x) \omega(x) f(x) dx = \mathrm{E}(r_n(x, h_n) \mathrm{Y} \omega(\mathrm{X})),$$

d'où un estimateur du deuxième terme de l'équation (3.3) est :

$$\frac{1}{n} \sum_{i=1}^{n} r_i(\mathrm{X}_i) \mathrm{Y}_i \omega(\mathrm{X}_i),$$

avec :

$$r_i(\mathrm{X}_i) = \frac{\sum_{j \neq i}^{n} \mathrm{Y}_j \mathrm{K}\left(\frac{x - \mathrm{X}_j}{h_n}\right)}{\sum_{j \neq i}^{n} \mathrm{K}\left(\frac{x - \mathrm{X}_j}{h_n}\right)}.$$

D'une manière similaire le premier terme de l'équation (3.3) est estimé par :

$$\frac{1}{n} \sum_{i=1}^{n} \left\{ r_i^2(\mathrm{X}_i) \omega(\mathrm{X}_i) \right\}.$$

Maintenant, nous sommes en mesure de minimiser la quantité suivante

$$\frac{1}{n} \sum_{i=1}^{n} r_i^2(\mathrm{X}_i) \omega(\mathrm{X}_i) - \frac{2}{n} \sum_{i=1}^{n} r_i(\mathrm{X}_i) \mathrm{Y}_i \omega(\mathrm{X}_i),$$

sans contraintes.

En plus nous remarquons que :

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} r_i^2(\mathrm{X}_i)\omega(\mathrm{X}_i) - \frac{2}{n}\sum_{i=1}^{n} r_i(\mathrm{X}_i)\mathrm{Y}_i\omega(\mathrm{X}_i) &= \frac{1}{n}\sum_{i=1}^{n}\left\{r_i(\mathrm{X}_i) - \mathrm{Y}_i\right\}^2\omega(\mathrm{X}_i) \\ &- \frac{1}{n}\sum_{i=1}^{n}\mathrm{Y}_i^2\omega(\mathrm{X}_i).\end{aligned}$$

Comme le dernier terme ne dépend pas de h_n, le critère à optimiser est

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\mathrm{Y}_i - r_i(\mathrm{X}_i)\right\}^2\omega(\mathrm{X}_i).$$

Chapitre 4

Les bandes de confiance

4.1 Introduction

L' approximation presque sûre du processus empirique uniforme par une suite de ponts browniens introduites par Komlós, Major et Tusnády ([K.M.T]) (1975) est une source importante pour démontrer une loi du logarithme itéré pour l' estimateur à noyau de Parzen Rosenblatt de la densité (voir Stute (1982). Ces résultats sont raffinés par Deheuvels et Mason (1992). En s'appuyant sur la théorie des processus empiriques uniformes, Deheuvels et Mason (2004) ont établi les lois uniformes du logarithme des estimateurs de Nadaraya -Watson et de Parzen-Rosenblatt.
Cette étude leur a permis de construire des bandes de confiance asymptotiques uniformes basées sur la consistance uniforme des estimateurs sur des intervalles fermés de $\mathbb{R}$.
Soient $(\mathrm{X}_1, \mathrm{Y}_1), (\mathrm{X}_2, \mathrm{Y}_2)...(\mathrm{X}_n, \mathrm{Y}_n)$ *des couples aléatoires à valeurs dans* $\mathbb{R}^2$ *et de même loi que (X ,Y) de densité jointe sur* $\mathbb{R}^2$, $f(x, y)$ *et* $f(x)$ *la densité marginale de X par rapport à la mesure de Lebesgue sur* $\mathbb{R}$ *. Nous adoptons les hypothèses et les notations suivantes :*

- $\mathrm{I} = [a, b]$ *et* $\mathrm{J} = [a', b']$ *dénotent deux intervalles fixés de* $\mathbb{R}$ *tels que*

$$-\infty < a' < a < b < b' < +\infty$$

- *On note /I/=b-a la mesure de Lebesgue de I.*

En plus de la ***(K.4)*** *le noyau* K *est tel que :*

- ***(K.8)*** *K est à variation bornée et continue à droite sur* $\mathbb{R}$,

- ***(K.9)*** *K est à support compact sur*$[-\alpha/2, \alpha/2]$ *, pour un certain* $0 < \alpha < \infty$.

Nous supposons certaines conditions sur la distribution du couple (X , Y) à savoir :

- ***(F.1)*** $f(x, y)$ *est continue sur J*$\times\mathbb{R}$ *;*

- ***(F.2)*** $f(x)$ *est continue et strictement positive sur J ;*

• ***(F.3)*** $\mathrm{Y}1_{\{\mathrm{X}\in\mathrm{J}\}}$ *est bornée ;*

Si ***(F.3)*** *n'est pas vérifiée, nous supposons la condition suivante :*

• ***(F.4)*** $\beta_{\mathrm{M}(\psi)} = \sup_{x\in\mathrm{J}} \mathrm{E}(\mathrm{M}(|\psi(\mathrm{Y})|)/\mathrm{X} = x) < \infty$ *;*

où ψ *est une fonction réelle mesurable et bornée sur chaque compact de* $\mathbb{R}$ *et* $\mathrm{M}(x) = x^p$ *pour un certain* $p > 2$.

Dans ce paragraphe on s'intéresse au traitement plus général de la fonction de régression conditionnelle c' est à dire

$$r_\psi(x) \quad = \quad \mathrm{E}(\psi(\mathrm{Y})/\mathrm{X} = x).$$

Ceci permet de déduire l'étude des fonctionnelles de la densité conditionnelle de Y *sachant* $\mathrm{X} = x$.

(par exemple la fonction de répartition conditionnelle de Y *sachant* $\mathrm{X} = x$ *définie par,* $\mathrm{E}(\mathrm{Y}/\mathrm{X} = x) = \mathrm{E}[1_{]-\infty<\mathrm{Y}<y]}/\mathrm{X} = x]$.

Nous remarquons que les conditions (F.1)-(F.3) impliquent que la fonction régression de $\psi(x)$ *sachant* $\mathrm{X} = x$ *et la variance de* $\psi(x)$ *sachant* $\mathrm{X} = x$ *sont bien définies et elles sont égales à :*

$$\begin{aligned} r_\psi(x) &= \mathrm{E}(\psi(\mathrm{Y})/\mathrm{X} = x) \\ &= \frac{1}{f_{\mathrm{X}}(x)} \int_{\mathbb{R}} \psi(y) f(x,y) dy \\ &= \frac{m_\psi(x)}{f(x)} \end{aligned}$$

où

$$m_\psi(x) = \int_{\mathbb{R}} \psi(y) f(x,y) dy$$

et

$$\begin{aligned} \sigma_\psi^2 &= \mathrm{V}ar(\Psi(\mathrm{Y})/\mathrm{X} = x) \\ &= \frac{1}{f(x)} \int_{\mathbb{R}} (\psi(y) - r_\psi(x))^2 f(x,y) dy. \end{aligned}$$

Pour la réalisation de leurs travaux de 2004, Deheuvels et Mason ont choisi une fenêtre aléatoire $\mathrm{H}_n(x)$ *et une fonction aléatoire* $\Theta_n(x)$*, telles que :*

(B.1) $\mathrm{P}\left(c_1 h_n \leqslant \inf_{x\in\mathrm{I}} \mathrm{H}_n(x) \leqslant \sup_{x\in\mathrm{I}} \mathrm{H}_n(x) \leqslant c_2 h_n\right) \to 1$*, quand* $n \to \infty$ *;*

(B.2) $\forall \epsilon > 0,\ \mathrm{P}\left(\sup_{x\in\mathrm{I}} \left| \frac{\mathrm{H}_n(x)}{h_n} - \mathrm{C}(x) \right| \geqslant \epsilon\right) \to 0$ *quand* $n \to \infty$;

(Θ.1) $\forall \epsilon > 0, \qquad \mathrm{P}\left(\sup_{x\in \mathrm{I}} \left| \frac{\Theta_n(x)}{\Theta(x)} - 1 \right| \geqslant \epsilon \right) \to 0$, *quand* $n \to \infty$;

où $\mathrm{H}_n(x) = h_n \mathrm{C}_n(x)$, $\Theta(x)$ *et* $\mathrm{C}(x)$ *sont des fonctions spécifiées positives et continues sur I,* c_1 *et* c_2 *sont des constantes strictement positives et* h_n *est un paramètre réel positif vérifiant les conditions suivantes :*

(H.1) $h_n \to 0$, *quand* $n \to \infty$;

(H.2) $\frac{nh_n}{logn} \to \infty$, *quand* $n \to \infty$;

(H.3) $nh_n \left\{ \frac{\log n}{\mathrm{M}^{inv}(n)^2} \right\} \to \infty$, *quand* $n \to \infty$,*øù* $\mathrm{M}^{inv}(n) = n^{1/p}$

Posons

$$r_{\psi;n}(x;h_n) = \begin{cases} \frac{\sum_{i=1}^n \psi(\mathrm{Y}_i)\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}, & si\ f_n(x,h_n) \neq 0\ ; \\ \\ \frac{1}{n}\sum_{i=1}^n \psi(\mathrm{Y}_i), & si\ f_n(x,h_n) = 0. \end{cases}$$

$$\sigma^2_{\psi;n}(x;n) = \begin{cases} \frac{\sum_{i=1}^n (\psi(\mathrm{Y}_i) - r_{\psi;n}(x;h_n))^2 \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}{\sum_{i=1}^n \mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}, & si\ f_n(x,h_n) \neq 0, \\ \\ \frac{1}{n}\sum_{i=1}^n (\psi(\mathrm{Y}_i) - \frac{1}{n}\sum_{j=1}^n \psi(\mathrm{Y}_j))^2, & si\ f_n(x,h_n) = 0. \end{cases}$$

$$\hat{\mathrm{E}}r_{\psi;n}(x;h_n) = \begin{cases} \frac{\mathrm{E}(\psi(\mathrm{Y})\mathrm{K}\left(\frac{x-\mathrm{X}}{h_n}\right))}{\mathrm{E}(\mathrm{K}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}, & si\ \mathrm{E}f_n(x;h_n) \neq 0\ ; \\ \\ \mathrm{E}(\psi(\mathrm{Y})), & si\ \mathrm{E}f_n(x;h_n) = 0. \end{cases}$$

Nous pouvons alors énoncer le résultat suivant.

Théorème 4.1.1. *(Deheuvels et Mason)*
Sous les conditions (F.1-3), (K.8-9), (H.1-3), (K.4), (B.1) et (Θ_1), *on a pour* $n \to \infty$

$$\sup_{x\in \mathrm{I}} \pm \left\{ \frac{n\mathrm{H}_n(x)}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/\mathrm{H}_n(x))} \right\}^{\frac{1}{2}} \Theta_n(x) \left\{ r_{\psi;n}(x;\mathrm{H}_n(x)) - \widehat{\mathrm{E}}r_{\Psi;n}(x;\mathrm{H}_n(x)) \right\} \xrightarrow{\mathrm{P}} \tag{4.1}$$

$$\left[\sup_{x\in \mathrm{I}} \left\{ \frac{\Theta^2(x)\sigma^2_\psi(x)}{f(x)} \right\} \int_{\mathrm{R}} \mathrm{K}^2(t)dt \right]^{\frac{1}{2}}.$$

Si de plus, la condition(B.2) est vérifiée alors,

$$\left\{ \frac{nh_n}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n)} \right\}^{\frac{1}{2}} \sup_{x\in \mathrm{I}} \pm \Theta_n(x) \left\{ r_{\psi;n}(x;\mathrm{H}_n(x)) - \widehat{\mathrm{E}}r_{\psi;n}(x;\mathrm{H}_n(x)) \right\} \xrightarrow{\mathrm{P}} \tag{4.2}$$

$$\left[\sup_{x\in \mathrm{I}}\left\{\frac{\Theta^2(x)\sigma_\psi^2(x)}{\mathrm{C}(x)f(x)}\right\}\int_{\mathrm{R}}\mathrm{K}^2(t)dt\right]^{\frac{1}{2}}.$$

où $\log_{\theta,k}(|\mathrm{I}|/h_n)=\log(\theta\vee u[\mathrm{K}^2])$ *avec* $\theta=7$ *(voir la remarque 4.1.3).*

Donnons maintenant les résultats concernant l'estimateur, à noyau , de la densité.

Théorème 4.1.2. *(Deheuvels et Mason(2004))*

Sous les conditions (F.1-2), (K.8-9), (H.1-3), (B.1) et (Θ_1)*, on a*

$$\sup_{x\in \mathrm{I}}\pm\left\{\frac{n\mathrm{H}_n(x)}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/\mathrm{H}_n(x))}\right\}^{\frac{1}{2}}\Theta_n(x)\left\{f_n(x;\mathrm{H}_n(x))-\mathrm{E}f_n(x;\mathrm{H}_n(x))\right\}\xrightarrow{\mathrm{P}} \tag{4.3}$$

$$\left[\sup_{x\in \mathrm{I}}\left\{\Theta^2(x)f(x)\right\}\int_{\mathrm{R}}\mathrm{K}^2(t)dt\right]^{\frac{1}{2}}.$$

Si de plus la condition (B.2)est vérifiée, il vient

$$\left\{\frac{h_n}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n)}\right\}^{\frac{1}{2}}\sup_{x\in \mathrm{I}}\pm\Theta_n(x)\left\{f_n(x;\mathrm{H}_n(x))-\mathrm{E}f_n(x;\mathrm{H}_n(x))\right\}\xrightarrow{\mathrm{P}} \tag{4.4}$$

$$\left[\sup_{x\in \mathrm{I}}\left\{\frac{\Theta^2(x)f_{\mathrm{X}}(x)}{\mathrm{C}(x)}\right\}\int_{\mathrm{R}}\mathrm{K}^2(t)dt\right]^{\frac{1}{2}}.$$

Remarque 4.1.1.
L'étude des déviations $\left\{r_{\psi;n}(x;\mathrm{H}_n(x))-\widehat{\mathrm{E}}r_{\psi;n}(x;\mathrm{H}_n(x))\right\}$ *et* $\left\{f_n(x;\mathrm{H}_n(x))-\mathrm{E}f_n(x;\mathrm{H}_n(x))\right\}$ *permettent de déterminer respectivement les bandes de confiance de* $r(x)$ *et de* $f(x)$*. Notons que pour un choix convenable de la fonction aléatoire*
$\mathrm{L}_n(x)=\mathrm{L}_n(\mathrm{X}_1,\mathrm{X}_2...,\mathrm{X}_n;x)$ *; les relations du théorème (4.1.2) peuvent s'écrire sous la forme :*

$$\sup_{x\in\mathbb{I}}\pm\left\{\frac{1}{\mathrm{L}_n(x)}\right\}\left\{r_n(x;\mathrm{H}_n(x))-\widehat{\mathrm{E}}r_n(x;\mathrm{H}_n(x))\right\}\xrightarrow{\mathrm{P}}1, \tag{4.5}$$

et si de plus $\mathrm{H}_n(x)$ *est choisi de telle façon que le biais de* $r_n(x;\mathrm{H}_n(x))$ *soit négligeable c'est à dire :*

$$\sup_{x\in\mathbb{I}}\pm\left\{\frac{1}{\mathrm{L}_n(x)}\right\}\left\{\widehat{\mathrm{E}}r_n(x;\mathrm{H}_n(x)-r(x)\right\}\xrightarrow{\mathrm{P}}0. \tag{4.6}$$

alors, ces deux relations combinées donnent

$\forall\epsilon>0$

$$\lim_{n\to\infty}\mathrm{P}\left(r(x)\in\left[r_n(x;\widehat{\mathrm{H}}_n(x))-(1+\epsilon)\mathrm{L}_n(x),\ r_n(x;\widehat{\mathrm{H}}_n(x))+(1+\epsilon)\mathrm{L}_n(x)\right]\right)=1,\forall x\in\mathbb{I} \tag{4.7}$$

et

$$\lim_{n\to\infty} \mathrm{P}\left(r(x) \in \left[r_n(x;\widehat{\mathrm{H}}_n(x)) - (1-\epsilon)\mathrm{L}_n(x),\, r_n(x;\widehat{\mathrm{H}}_n(x)) + (1-\epsilon)\mathrm{L}_n(x)\right]\right) = 0, \forall x \in \mathbb{I}. \quad (4.8)$$

Comme les relations (4.7), (4.8) sont vérifiées pour tout $\epsilon > 0$, on a alors pour tout $x \in \mathrm{I}$

$$[r_n(x;\widehat{\mathrm{H}}_n(x)) - \mathrm{L}_n(x), r_n(x;\widehat{\mathrm{H}}_n(x)) - \mathrm{L}_n(x)],$$

constituent des bandes de confiance optimales simultanées asymptotiques pour r(x). D'une manière analogue nous obtenons des bandes de confiance pour f(x)

Remarque 4.1.2.

Un résultat important de Härdle *(1990), montre que si f(x) et r(x) sont deux fois continûment dérivables sur $\mathbb{R}$, f(x)>0 et si la condition **(F.3)** est vérifiée, alors pour $h_n = n^{-\alpha}$, $\alpha \in [1/5, 1[$, on a la normalité asymptotique de l' estimateur de Nadaraya-Watson c'est à dire :*

$$\sqrt{nh}[r_n(x;h_n(x)) - r(x)] \xrightarrow{\mathcal{L}} (\mathrm{D}(x), v^2(x)),$$

avec :

$$\mathrm{D}(x) = \left\{r''(x) + 2r'(x)\frac{f'(x)}{f(x)}[u^2\mathrm{K}]\right\}$$

et

$$v^2(x) = \frac{\sigma^2(x)}{f(x)}[\mathrm{K}^2].$$

Comme le terme du biais de $r_n(x,h_n)$ est négligeable par rapport au terme de la variance, l'intervalle de confiance au seuil de α de la fonction régression, est donné par

$$\left[r_n(x;h_n) - z_{1-\alpha/2}\sqrt{\frac{\sigma^2(x)[\mathrm{K}^2]}{nh_nf(x)}}, r_n(x;h_n) + z_{1-\alpha/2}\sqrt{\frac{\sigma^2(x)[\mathrm{K}^2]}{nh_nf(x)}}\right],$$

où $z_{1-\alpha/2}$ est le $(1-\alpha/2)$ quantile de la loi normale centré réduite. Notons que cette méthode donne un intervalle de confiance à $(1-\alpha/2)\%$ des valeurs de r(x) et non pas toutes les valeurs de r(x), par exemple pour un point $x_0 \in \mathbb{R}$ fixé, on a

$$\mathrm{P}\left(r(x_0) \in \left[r_n(x_0;h_n) - 1.96\sqrt{\frac{\sigma^2(x_0)[\mathrm{K}^2]}{nh_nf(x_0)}},\, r_n(x_0;h_n) + 1.96\sqrt{\frac{\sigma^2(x_0)[\mathrm{K}^2]}{nh_nf(x)}}\right]\right) \approx 95\%$$

quand $n \to \infty$, où 1.96 est la valeur du quantile d'ordre 2.5% de la loi normale centrée réduite. En conséquence des remarques (4.1.1), (4.1.2) les intervalles d'estimation de r(x) donnés par les bandes de confiance, sont plus étroites que ceux donnés par les intervalles de confiance.

Remarque 4.1.3.

Supposons que f(x) est la densité d'une loi normale centrée réduite et soit K le noyau d' Epanechnikov, alors la fenêtre optimale donnée par la méthode du pouce (traitée dans le chapitre précédent) est égale à $h_{\mathrm{RT}} = 2.34n^{-1/5}$, *si* $\mathbb{I} = [-2, 2]$ *ona*

$$\theta = 7, \qquad \frac{|\mathbb{I}|}{h_{\mathrm{RT}}}\left\{\int_{\mathbb{R}} \mathrm{K}^2(t)dt\right\} \geqslant 7 \Rightarrow h_{\mathrm{RT}} \leqslant \frac{12}{35} \Rightarrow n \geqslant 2697,$$

pour

$$\theta = 15, \qquad \frac{|\mathbb{I}|}{h_{\mathrm{RT}}}\left\{\int_{\mathbb{R}} \mathrm{K}^2(t)dt\right\} \geqslant 15 \Rightarrow h_{\mathrm{RT}} \leqslant \frac{4}{25} \Rightarrow n \geqslant 121839.$$

Une telle taille d'échantillon est rarement rencontrée, nous optons donc pour le choix de $\theta = 7$.

N.Nemouchi et Z. Mohdeb ont appliqué les résultats de Deheuvels et Mason (2004) à loi normale de moyenne et d'écart type inconnus. Pour un choix approprié de l' estimation, de la fenêtre, ils ont obtenu des bandes de confiance, basées sur les estimateurs à noyau de type Nadaraya Watson pour la fonction de régression et de type Parzen-Rosenblatt pour la densité de probabilité . Soient $(\mathrm{X}_1, \mathrm{Y}_1), (\mathrm{X}_2, \mathrm{Y}_2), ...$ *n variables aléatoires indépendantes et identiquement distribuées de même loi que le couple gaussien (X , Y). Nous résumons, ci après leurs résultats. X et Y suivent la loi normale de moyennes respectives* $\mu_{\mathrm{X}}, \mu_{\mathrm{Y}}$, *d' écarts type respectifs* $\sigma_{\mathrm{X}} > 0$, $\sigma_{\mathrm{Y}} > 0$ *et de coefficient de corrélation* ρ *inconnus. Nous utilisons les mêmes notations et les mêmes hypothèses que dans Deheuvels et Mason (2004) .*
Dans ce modèle, pour tout $x \in \mathbb{I}$ *la fonction de régression et la variance conditionnelle de* Y *sachant* $\mathrm{X} = x$ *sont bien définies et données par :*

$$r(x) = \mathrm{E}(\mathrm{Y}/\mathrm{X} = x) = \int_{\mathbb{R}} y\frac{f_{\mathrm{X,Y}}(x, y)}{f_{\mathrm{X}}(x)}dx = \mu_{\mathrm{X}}y + \rho\frac{\sigma_{\mathrm{Y}}}{\sigma_{\mathrm{X}}}(x - \mu_{\mathrm{X}}),$$

et

$$v^2(x) = var(\mathrm{Y}/\mathrm{X} = x) = (1 - \rho^2)\sigma^2_{\mathrm{Y}}.$$

Dans ces expressions, $f(x)$ *désigne la densité marginale de la variable X et* $f(x, y)$ *désigne la densité jointe de (X,Y).*
Sous les hypothèses :

$$\lim_{n\to\infty} nh_n = \infty \quad et \quad \lim_{n\to\infty} nh_n = 0$$

le facteur de centrage

$$\widehat{\mathrm{E}}r_n(x; h_n) = \begin{cases} \frac{\left[\mathrm{EYK}\left(\frac{x-\mathrm{X}}{h_n}\right)\right]}{\mathrm{EK}\left(\frac{x-\mathrm{X}_i}{h_n}\right)}, & si\ \mathrm{EK}\left(\frac{x-\mathrm{X}}{h_n}\right) \neq 0\,; \\ \mathrm{E}(\mathrm{Y}), & si\ \mathrm{EK}\left(\frac{x-\mathrm{X}}{h_n}\right) = 0. \end{cases}$$

vérifie :

$$\lim_{n\to\infty}\left\{\mathrm{E}[r_n(x;\mathrm{H}_n(x))-\widehat{\mathrm{E}}r_n(x;h_n)]\right\}=0.$$

Nous avons vu dans les chapitres précédents que l'obtention de propriétés asymptotiques des estimateurs à noyau requiert d'imposer des conditions au paramètre de lissage. Ceci nous amène au problème du choix optimum de $\mathrm{H}_{n,i}(x)$ *i=1,2, pour l'erreur quadratique moyenne de* $f_n(x,h_n)$ *et de* $r_n(x,h_n)$*. Plus précisément , nous cherchons à construire des estimateurs des ces paramètres optimums en fonction des estimateurs empiriques* $\widehat{\mu}_{\mathrm{X}},\widehat{\mu}_{\mathrm{Y}},\widehat{\sigma}_{\mathrm{X}},\widehat{\sigma}_{\mathrm{Y}},\widehat{\rho}$ *tels que*

$$\frac{\widehat{\mathrm{H}}_{n,i}(x)}{\mathrm{H}_{n,i}}\xrightarrow{\mathrm{P}}1$$

$n\to\infty$ *pour* $i=1,2$.
Nous déduisons alors des bandes de confiance asymptotiques pour $f(x)$ *respectivement pour r(x) en fonction de* $f_n(x;\widehat{\mathrm{H}}_{n,1}(x))$ *(respectivement de* $r_n(x;\widehat{\mathrm{H}}_{n,2}(x))$*).*

4.2 Résultats :

4.2.1 Evaluation de l'erreur quadratique moyenne de $f_n(x,h_n)$

D'après le chapitre 3 la valeur h_n *qui minimise l'erreur quadratique moyenne de* $f_n(x,h_n)$ *est :*

$$\begin{aligned}\mathrm{H}_{n,1}(x) &= n^{\frac{-1}{5}}\left\{\frac{f(x)[\mathrm{K}^2]}{(f''(x))^2[t^2\mathrm{K}]^2}\right\}^{\frac{1}{5}}\\ &= n^{\frac{-1}{5}}\sigma_{\mathrm{X}}\left(\frac{\sqrt{2\pi}[\mathrm{K}^2]\exp\frac{1}{2}\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2}{[t^2\mathrm{K}]^2\left[-1+\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2\right]^2}\right)^{\frac{1}{5}}\end{aligned}\tag{4.9}$$

Puisque dans l' expression précédente μ_{X} *et* σ_{X} *sont inconnus, nous les remplaçons par leurs estimateurs empiriques donnés par*

$$\widehat{\mu}_{\mathrm{X}}=\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i$$

et

$$\widehat{\sigma}^2_{\mathrm{X}}=\frac{1}{n}\sum_{i=1}^{n}(\mathrm{X}_i-\widehat{\mu}_{\mathrm{X}})^2,$$

Ce qui conduit à estimer $\mathrm{H}_{n,1}(x)$ *par :*

$$\widehat{\mathrm{H}}_{n,1}(x) = n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}}\left(\frac{\sqrt{2\pi}[\mathrm{K}^2]\exp\frac{1}{2}\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2}{[t^2\mathrm{K}]^2\left[-1+\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2\right]^2}\right)^{\frac{1}{5}}. \tag{4.10}$$

Posons

$$\Theta_{n,1}(x) = \sqrt{\frac{\widehat{\sigma}_{\mathrm{X}}}{[\mathrm{K}^2]}}\left(\frac{\sqrt{(2\pi)^3}\sqrt{[\mathrm{K}^2]}exp\frac{3}{2}\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2}{[t^2\mathrm{K}]\left|-1+\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2\right|}\right)^{\frac{1}{5}}. \tag{4.11}$$

Remarquons que $\Theta_{n,1}(x)$, $\widehat{h}_{n,1}(x) = \frac{\widehat{\mathrm{H}}_{n,1}}{\widehat{\sigma}_{\mathrm{X}}}$ *et* $h_n(x) = \frac{\mathrm{H}_{n,1}(x)}{\sigma_{\mathrm{X}}}$ *vérifient les conditions suivantes de Deheuvels et Mason (2004) c'est à dire :*

(B.1) $\forall\epsilon > 0$, *pour* $n \to \infty$

$$\mathrm{P}\left(\inf_{x\in\mathrm{I}} h_{n,1}(x)) - \epsilon n^{-1/5} \leqslant \inf_{x\in\mathrm{I}} \widehat{h}_{n,1}(x)) \leqslant \sup_{x\in\mathrm{I}} \widehat{h}_{n,1}(x) \leqslant \sup_{x\in\mathrm{I}} h_{n,1}(x) + \epsilon n^{-1/5}\right) \xrightarrow{\mathrm{P}} 1$$

(B.2) $\forall\epsilon > 0$, *pour* $n \to \infty$

$$\mathrm{P}\left(\sup_{x\in\mathrm{I}}\left|\frac{\widehat{h}_{n,1}(x)}{h_n} - \frac{h_{n,1}(x)}{h_n}\right| \geqslant \epsilon\right) \to 0$$

(Θ*.1)* $\forall\epsilon > 0$, *pour* $n \to \infty$

$$\mathrm{P}\left(\sup_{x\in\mathrm{I}}\left|\frac{\Theta_{n,1}(x)}{\Theta_1(x)} - 1\right| > \epsilon\right) \to 0$$

où

$$\Theta_1(x) = \sqrt{\frac{\sigma_{\mathrm{X}}}{[\mathrm{K}^2]}}\left(\frac{\sqrt{(2\pi)^3}\sqrt{[\mathrm{K}^2]}exp\frac{3}{2}\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2}{[t^2\mathrm{K}]\left|-1+\left(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}}\right)^2\right|}\right)^{\frac{1}{5}}.$$

4.2.2 Evaluation de l' erreur quadratique intégrée.

L'erreur quadratique moyenne intégré de $f_n(x, h_n)$ *atteint son minimum au point :*

$$h_{n,2} = n^{\frac{-1}{5}}\left\{\frac{[\mathrm{K}^2]}{[t^2\mathrm{K}]^2\int_{\mathrm{R}}(f''(x))^2dx}\right\}^{\frac{1}{5}}.$$

Posons

$$\widehat{h}_{n,2} = n^{\frac{-1}{5}} \widehat{\sigma}_{\mathrm{X}} \left\{ \frac{8\sqrt{\pi}[\mathrm{K}^2]}{3[t^2\mathrm{K}]^2} \right\}^{\frac{1}{5}} \tag{4.12}$$

et

$$\Theta_{n,2}(x) = \left\{ \frac{[\mathrm{K}^2]}{\sqrt{2\pi}\widehat{\sigma}_{\mathrm{X}}} \exp \frac{-1}{2} \left(\frac{x - \widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}} \right)^2 \right\}. \tag{4.13}$$

Nous remarquons que $\widehat{h}_{n,2}$ vérifie la condition de Deheuvels et Mason (2004) c' est à dire : $\forall \epsilon > 0$,

$$\mathrm{P}\left(h_{n,2} - \epsilon\ n^{-1/5} < \widehat{h}_{n,2} < h_{n,2} + \epsilon\ n^{-1/5} \right) \longrightarrow 1,$$

quand $n \to \infty$. De plus, nous avons pour tout $\epsilon > 0$,

$$\mathrm{P}\left(\sup_{x \in \mathrm{I}} \left| \frac{\Theta_{n,2}(x)}{\Theta_2(x)} - 1 \right| > \epsilon \right) \to 0,$$

quand $n \to \infty$, avec

$$\Theta_2(x) = \frac{1}{\sqrt{f(x)[\mathrm{K}^2]}}.$$

4.2.3 Évaluation de l' erreur quadratique moyenne

En utilisant toujours la méthode Plug in, le paramètre optimal

$$\begin{aligned} \mathrm{H}_{n,3}(x) &= n^{\frac{-1}{5}} \left(\frac{\frac{v^2(x)}{f(x)}[\mathrm{K}^2]}{\left[r''(x) + 2r'(x)\frac{f'(x)}{f(x)} \right][t^2\mathrm{K}]} \right)^{\frac{1}{5}} \\ &= n^{\frac{-1}{5}} \sigma_{\mathrm{X}} \left(\frac{(1-\rho^2)\sigma_{\mathrm{X}}^2\sqrt{2\pi} \exp \frac{1}{2}(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}})^2[\mathrm{K}^2]}{[t^2\mathrm{K}]^2 4\rho^2(x-\mu_{\mathrm{X}})^2} \right)^{\frac{1}{5}} \end{aligned}$$

de l'erreur quadratique moyenne de $r_n(x, h_n)$ est estimée par :

$$\widehat{\mathrm{H}}_{n,3}(x) = n^{\frac{-1}{5}} \widehat{\sigma}_{\mathrm{X}} \left\{ \frac{(1-\widehat{\rho}^2)\widehat{\sigma_{\mathrm{X}}}^2\sqrt{2\pi} \exp \frac{1}{2}(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}})^2[\mathrm{K}^2]}{4[t^2\mathrm{K}]^2\widehat{\rho}^2(x-\widehat{\mu}_{\mathrm{X}})^2} \right\}^{\frac{1}{5}}, \tag{4.14}$$

avec

$$\widehat{\rho} = \frac{1}{n} \frac{\sum_{i=1}^{n} (\mathrm{X}_i - \widehat{\mu}_{\mathrm{X}})(\mathrm{Y}_i - \widehat{\mu}_{\mathrm{Y}})}{\widehat{\sigma}_{\mathrm{X}}\widehat{\sigma}_{\mathrm{Y}}}$$

où

$$\widehat{\mu}_{\mathrm{Y}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Y}_i \text{ et } \widehat{\sigma}^2_{\mathrm{Y}} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Y}_i - \widehat{\mu}_{\mathrm{Y}})^2$$

Posons

$$\Theta_{n,3}(x) = \left\{ \frac{(1-\widehat{\rho}^2)^{-2}|\widehat{\rho}(x-\widehat{\mu})|^{-1}\exp -(\frac{x-\widehat{\mu}}{\widehat{\sigma}_{\mathrm{X}}})^2}{4\pi\widehat{\sigma}_{\mathrm{X}}^{(\frac{3}{2})}\widehat{\sigma}^5_{\mathrm{Y}}[t^2\mathrm{K}]} \right\}^{\frac{1}{5}}. \tag{4.15}$$

De manière analogue , nous montrons que $\mathrm{H}_{n,3}(x), \widehat{\mathrm{H}}_{n,3}(x), \Theta_{n,3}(x), \Theta_3(x)$ *vérifient les conditions* $((\mathrm{B}.1),(\mathrm{B}.2),(\Theta_1)$ *de Deheuvels et Mason (2004), avec*

$$\Theta_3(x) = \left\{ \frac{(1-\rho^2)^{-2}|\rho(x-\mu)|^{-1}\exp -(\frac{x-\mu}{\sigma_{\mathrm{X}}})^2}{4\pi\sigma_{\mathrm{X}}^{\frac{3}{2}}\sigma^5_{\mathrm{Y}}[t^2\mathrm{K}]} \right\}^{\frac{1}{5}} \quad . \tag{4.16}$$

Théorème 4.2.1.
Soient $\widehat{\mathrm{H}}_{n,1}(x)$ *et* $\Theta_{n,1}(x)$ *donnés par les formules (4.10) et (4.11), alors nous avons*

$$\left\{ \frac{n^{4/5}\widehat{\sigma}_{\mathrm{X}}}{2\log_{\theta,\mathrm{K}}\left(\frac{|\mathrm{I}|}{\widehat{\sigma}_{\mathrm{X}}n^{-1/5}}\right)} \right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,1}(x)\left\{ f_n(x;\widehat{\mathrm{H}}_n(x)) - f(x)\right\} \xrightarrow{\mathrm{P}} 1 \tag{4.17}$$

quand $n \to +\infty$.

Remarque 4.2.1. *Du théorème (4.2.1), nous déduisons que pour tout* $0 < \epsilon < 1$, *et pour tout* $x \in \mathrm{I}$, *nous avons :*

$$\lim_{n\to\infty} \mathrm{P}\left(f(x) \in \left[f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) - (1+\epsilon)\Delta_{n,1}(x), f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) + (1+\epsilon)\Delta_{n,1}(x)\right]\right) = 1,$$

et

$$\lim_{n\to\infty} \mathrm{P}\left(f(x) \in \left[f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) - (1-\epsilon)\Delta_{n,1}(x), f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) + (1-\epsilon)\Delta_{n,1}(x)\right]\right) = 0 \quad ,$$

où

$$\Delta_{n,1}(x) = \frac{1}{\Theta_{n,1}(x)}\left(\frac{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}})}{n^{\frac{4}{5}}\widehat{\sigma}_{\mathrm{X}}} \right)^{\frac{1}{2}}$$

Il s'ensuit alors des bandes de confiance asymptotiques suivantes pour la fonction de densité.

$$[f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) - \Delta_{n,1}(x), f_n(x;\widehat{\mathrm{H}}_{n,1}(x)) + \Delta_{n,1}(x)]$$

Théorème 4.2.2.

Les estimateurs $\widehat{h}_{n,2}$ et $\Theta_{n,2}(x)$, donnés respectivement par les formules (4.12) et (4.13) sont tels que :

$$\left\{\frac{n\widehat{h}_{n,2}}{2\log_{\theta,\mathrm{K}}\left(\frac{|\mathrm{I}|}{\widehat{\sigma}n^{-1/5}}\right)}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,2}(x)\left\{f_n(x;\widehat{h}_{n,2}) - f(x)\right\} \xrightarrow{\mathrm{P}} 1 \tag{4.18}$$

Théorème 4.2.3. *Soient $\widehat{\mathrm{H}}_{n,3}$ et $\Theta_{n,3}$ les estimateurs donnés respectivement par les formules (4.14) et (4.15), alors*

$$\left\{\frac{n^{4/5}\widehat{\sigma}_{\mathrm{X}}}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/n^{-1/5}\widehat{\sigma}_{\mathrm{X}})}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) - r(x)\right\} \xrightarrow{\mathrm{P}} 1 \quad . \tag{4.19}$$

lorsque $n \longrightarrow \infty$

Remarque 4.2.2.
Notons que pour tout $0 < \epsilon < 1$ quand $n \to \infty$, d'après le théorème(4.2.3), nous avons $\forall x \in \mathrm{I}$

$$\lim_{n\to\infty} \mathrm{P}\left(r(x) \in \left[r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) - (1+\epsilon)\Delta_{n,3}(x), r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) + (1+\epsilon)\Delta_{n,3}(x)\right]\right) = 1 \quad , \tag{4.20}$$

et

$$\lim_{n\to\infty} \mathrm{P}\left(r(x) \in \left[r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) - (1-\epsilon)\Delta_{n,3}(x), r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) + (1-\epsilon)\Delta_{n,3}(x)\right]\right) = 0 \quad . \tag{4.21}$$

où

$$\Delta_{n,3}(x) = \frac{1}{\Theta_{n,3}(x)}\left(\frac{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}})}{n^{\frac{4}{5}}\widehat{\sigma}_{\mathrm{X}}}\right)^{\frac{1}{2}} \quad .$$

Les relations (4.20) et (4.21) sont vérifiées pour tout $\epsilon > 0$, nous pouvons alors dire que les intervalles :

$$[r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) - \Delta_{n,3}(x), r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) + \Delta_{n,3}(x)] \ ;$$

Constituent des bandes de confiance optimales simultanées asymptotiques pour $r(x)$ ($\forall x \in \mathrm{I}$).

La démonstration est similaire pour les trois théorèmes, par conséquent, nous nous limitons à donner la preuve du dernier d'entre eux. En prenant $h_n = n^{-1/5}$*, les conditions (B.1), (B.2) et* Θ_1 *sont satisfaites, le théorème 4.1.1 implique alors*

$$\left\{\frac{nh_n}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n)}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;\widehat{h}_{n,3}(x)) - r_n(x;\widehat{h}_{n,3}(x))\right\} \xrightarrow{\mathrm{P}} \tag{4.22}$$

$$\left[\sup_{x\in\mathrm{I}}\left\{\frac{\Theta_3^2(x)r(x)}{\mathrm{C}(x)}\right\}\int_{\mathrm{R}}\mathrm{K}^2(t)dt\right]^{\frac{1}{2}}.$$

quand $n \to +\infty$*, où*

$$\mathrm{C}(x) = \left\{\frac{(1-\rho^2)\sigma_{\mathrm{X}}^2\sqrt{2\pi}\exp\frac{1}{2}(\frac{x-\mu_{\mathrm{X}}}{\sigma_{\mathrm{X}}})^2[k^2]}{[t^2k]^2 4\rho^2(x-\mu_{\mathrm{X}})^2}\right\}^{1/5} \quad . \tag{4.23}$$

En remplaçant $\Theta_3(x)$ *et C(x) Par leurs expressions , nous obtenons :*

$$\left\{\frac{nh_n\widehat{\sigma}_{\mathrm{X}}}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n\widehat{\sigma}_{\mathrm{X}})}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;\widehat{h}_{n,3}(x)) - r_n(x;\widehat{h}_{n,3}(x))\right\} \xrightarrow{\mathrm{P}} 1 \quad ,$$

quand $n \to +\infty$*.*
Avec ce choix du paramètre h_n*, il vient (voir Nadaraya(1989)) ,*

$$\left\{\frac{nh_n}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n)}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;h_n) - r(x)\right\} \xrightarrow{\mathrm{P}} 0$$

D'après le corollaire (2.3) de Dehewels *et* Mason (2004) *et les travaux de* Einmahl *et* Mason *(2005), il s'ensuit*

$$\left\{\frac{nh_n\widehat{\sigma}}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n)}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;\widehat{\sigma}\widehat{h}_{n,3}(x)) - r_n(x;\widehat{\sigma}\widehat{h}_{n,3})\right\} \xrightarrow{\mathrm{P}} 1 \quad ,$$

et

$$\left\{\frac{nh_n\widehat{\sigma}_{\mathrm{X}}}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/h_n\widehat{\sigma}_{\mathrm{X}})}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;\widehat{\sigma}_{\mathrm{X}}\widehat{h}_{n,3}(x)) - r(x))\right\} \xrightarrow{\mathrm{P}} 0 \quad .$$

Les deux relations précédentes donnent :

$$\left\{\frac{n^{4/5}\widehat{\sigma}_{\mathrm{X}}}{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/n^{-1/5}\widehat{\sigma}_{\mathrm{X}})}\right\}^{\frac{1}{2}} \sup_{x\in\mathrm{I}} \pm\Theta_{n,3}(x)\left\{\mathrm{E}r_n(x;\widehat{\mathrm{H}}_{n,3}(x)) - r(x)\right\} \xrightarrow{\mathrm{P}} 1 \quad .$$

quand $n\to\infty$, *où* $\widehat{\mathrm{H}}_{n,3}(x) = \widehat{\sigma}_{\mathrm{X}}\widehat{h}_{n,3}(x)$.
Nous complétons ce paragraphe, en exposant le travail de K. Kebabi, F. Messaci et N. Nemouchi (2010) sur des exemples de calcul de bandes de confiance pour la régression, pour les modèles donnés ci dessous.
Soit $(\mathrm{X}_1,\mathrm{Y}_1,\epsilon_1)$, $(\mathrm{X}_2,\mathrm{Y}_2,\epsilon_2)$,...,$(\mathrm{X}_n,\mathrm{Y}_n,\epsilon_n)$ *n vecteurs aléatoires tels que* $(\mathrm{X}_1,\mathrm{Y}_1),(\mathrm{X}_2,\mathrm{Y}_2),\ldots,(\mathrm{X}_n,\mathrm{Y}_n)$ *sont indépendants de même loi que (X, Y)et* $\epsilon_1,\epsilon_2,\ldots,\epsilon_n$ *sont de même loi que* ϵ.
Nous supposons que X suit la loi normale de moyenne μ_{X} *et de variance* σ^2_{X} *inconnus. La variable Y est liée à la variable aléatoire X par une des relations suivantes.*
a) $\mathrm{Y} = \mathrm{X}^2 + \epsilon$,
b) $\mathrm{Y} = \sin\left(\frac{3\mathrm{X}}{2}\right) + \epsilon$ *et*
c) $\mathrm{Y} = \exp(\mathrm{X} - 0.2) + \epsilon$ *;*

où ϵ *est indépendante de X et suit une loi normale non dégénérée de moyenne* μ_ϵ *et de variance* σ^2_ϵ *inconnues.*
En utilisant les mêmes techniques que le théorème (4.2.1), nous pouvons déduire que pour $0<\epsilon<1$, *et pour tout* $x\in\mathrm{I}$,

$$\lim_{n\to\infty}\mathrm{P}\left(r(x)\in\left[r_n(x;\widehat{\mathrm{H}}_n(x)) - (1+\epsilon)\Delta_n(x), r_n(x;\widehat{\mathrm{H}}_n(x)) + (1+\epsilon)\Delta_n(x)\right]\right) = 1,$$

et

$$\lim_{n\to\infty}\mathrm{P}\left(r(x)\in\left[r_n(x;\widehat{\mathrm{H}}_n(x)) - (1-\epsilon)\Delta_n(x), r_n(x;\widehat{\mathrm{H}}_n(x)) + (1-\epsilon)\Delta_n(x)\right]\right) = 0 \quad ,$$

où

$$\Delta_n(x) = \frac{1}{\Theta_n(x)}\left(\frac{2\log_{\theta,\mathrm{K}}(|\mathrm{I}|/n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}})}{n^{\frac{4}{5}}\widehat{\sigma}_{\mathrm{X}}}\right)^{\frac{1}{2}}.$$

Dans le cas a)*, on a*

$$\widehat{\mathrm{H}}_n(x) = n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}}\left(\frac{\sqrt{2\pi}\widehat{\sigma}^2_\epsilon[\mathrm{K}^2]\exp\left[\frac{-1}{2}\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2\right]}{4(\widehat{\sigma}^2_{\mathrm{X}} - 2x(x-\widehat{\mu}_{\mathrm{X}}))^2[t^2\mathrm{K}]^2}\right)^{\frac{1}{5}} \tag{4.24}$$

et

$$\Theta_n(x) = \left[\frac{\exp -2\left(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}}\right)^2}{[\mathrm{K}^2]^2[t^2\mathrm{K}]\widehat{\sigma}^5_{\mathrm{X}}\widehat{\sigma}^2_\epsilon 2(x-\widehat{\mu}_{\mathrm{X}})}\right]^{1/5} \tag{4.25}$$

***Dans le cas b)**, il vient*

$$\widehat{\mathrm{H}}_n(x) = n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}}\left(\frac{\widehat{\sigma}_\epsilon\sqrt{2\pi}\exp\left(\frac{1}{2}(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}})^2\right)[\mathrm{K}^2]}{\left(\frac{9}{4}\sin\left(\frac{3}{2}x\right)\widehat{\sigma}_{\mathrm{X}}^2 + 3(x-\widehat{\mu}_{\mathrm{X}})\cos\left(\frac{3}{2}x\right)[t^2\mathrm{K}]\right)^2}\right)^{\frac{1}{5}},$$

et

$$\Theta_n(x) = \left(\frac{\exp -(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}})^2}{[\mathrm{K}^2][t^2\mathrm{K}]\widehat{\sigma}_{\mathrm{X}}^{5/2}\widehat{\sigma}_\epsilon^4\left(\widehat{\sigma}_{\mathrm{X}}\frac{9}{4}\sin\left(\frac{3}{2}x\right) + 3(x-\widehat{\mu}_{\mathrm{X}})\cos\left(\frac{3}{2}x\right)\right)}\right)^{\frac{-1}{5}}.$$

***Dans le cas c)**, on a*

$$\widehat{\mathrm{H}}_n(x) = n^{\frac{-1}{5}}\widehat{\sigma}_{\mathrm{X}}\left[\frac{\widehat{\sigma}_\epsilon^2\sqrt{2\pi}\exp\left(\frac{1}{2}(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}})^2\right)[\mathrm{K}^2]}{[\exp(2x-0.4)(\widehat{\sigma}_{\mathrm{X}} - 2(x-\widehat{\mu}_{\mathrm{X}})[\mathrm{K}t^2]]}\right]^{\frac{1}{5}}, \tag{4.26}$$

et

$$\Theta_n(x) = \left[\frac{\exp -(\frac{x-\widehat{\mu}_{\mathrm{X}}}{\widehat{\sigma}_{\mathrm{X}}})^2}{[\mathrm{K}^2][\mathrm{K}t^2]\widehat{\sigma}_{\mathrm{X}}^{5/2}\widehat{\sigma}_\epsilon^4[\exp(x-0.2)(\widehat{\sigma}_{\mathrm{X}} - 2(x-\widehat{\mu}_{\mathrm{X}})]}\right]^{\frac{1}{5}}. \tag{4.27}$$

Chapitre 5

Simulation

En choisissant $h_n = n^{-1/5}$ *et* $\mathrm{K}(x) = \mathbb{I}_{[-1/2<(x)<1/2]}$ *, nous étudions, à travers quelques simulations les propriétés des bandes et les intervalles de confiance de la fonction de densité et de la fonction de régression pour les petits échantillons .*

Nous générons pour chacune des applications que nous proposons des échantillons de taille n=10, n=50, n=100, n=800 respectivement et nous considérons les modèles suivants.

a) *X suit une loi normale* $\mathcal{N}(0,1)$*.*

b) *Y=X+Z où X suit une loi normale* $\mathcal{N}(0,1)$ *et Z suit une loi normale* $\mathcal{N}(0,2)$*.*

c) *Y=*X^2*+ Z où X suit une loi normale* $\mathcal{N}(0,1)$ *et Z suit une loi normale* $\mathcal{N}(0,2)$*.*

d) *Y=*$\sin\left(\frac{3}{2}\mathrm{X}\right)$*+ Z où X suit une loi normale* $\mathcal{N}(0,1)$ *et Z suit une loi normale* $\mathcal{N}(0,2)$*.*

e) *Y=*$\exp\left(\mathrm{X}-0.2\right)$*+ Z où X suit une loi normale* $\mathcal{N}(0,1)$ *et Z suit une loi normale* $\mathcal{N}(0,2)$*.*

Les différents graphiques donnés aux figures $(5.1)-(5.4)$ *montrent que les fonctions à estimer ainsi que leurs estimateurs se trouvent à l'intérieur des bandes de confiance dès que la taille* $n = 10$*. De plus ces dernières se rétrécissent au fur et à mesure que le taille de l'échantillon augmente, ce qui est tout à fait prévisible. Il en ressort ainsi la bonne performance des estimateurs étudiés.*

Les figures $(5.5)-(5.8)$ *montrent que la probabilité de trouver les fonctions à estimer dans l'intervalle de confiance* $\sim 95\%$ *et les figures* $(5.9)-(5.11)$ *explique la comparaison les intervalles données par les bandes de confiance avec les intervalles de confiance, on remarque que les bandes de confiance sont plus étroites que ceux données par les intervalles de confiance.*

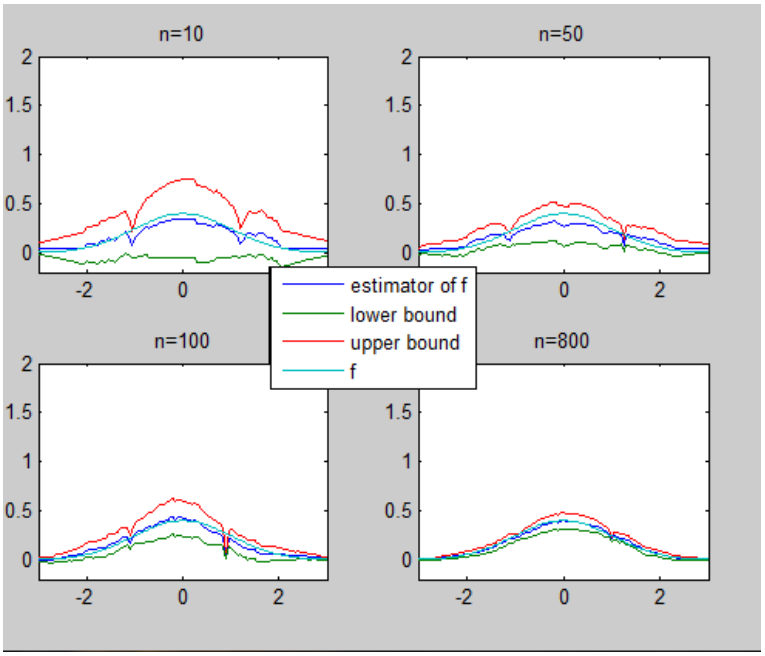

FIGURE 5.1 – les bandes de confiance de la fonction de densité dans le cas a).

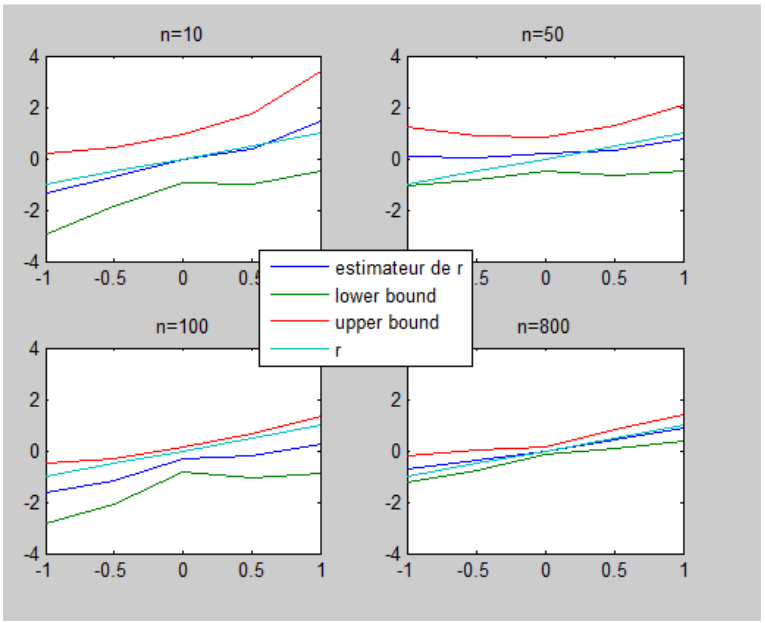

FIGURE 5.2 – les bandes de confiance de la fonction de régression dans le cas b).

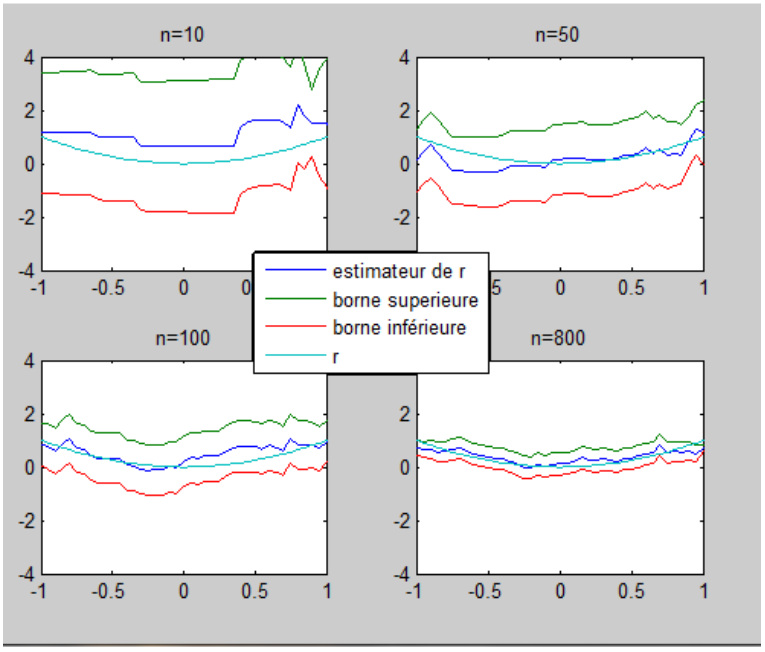

FIGURE 5.3 – les bandes de confiance de la fonction de régression dans le cas c).

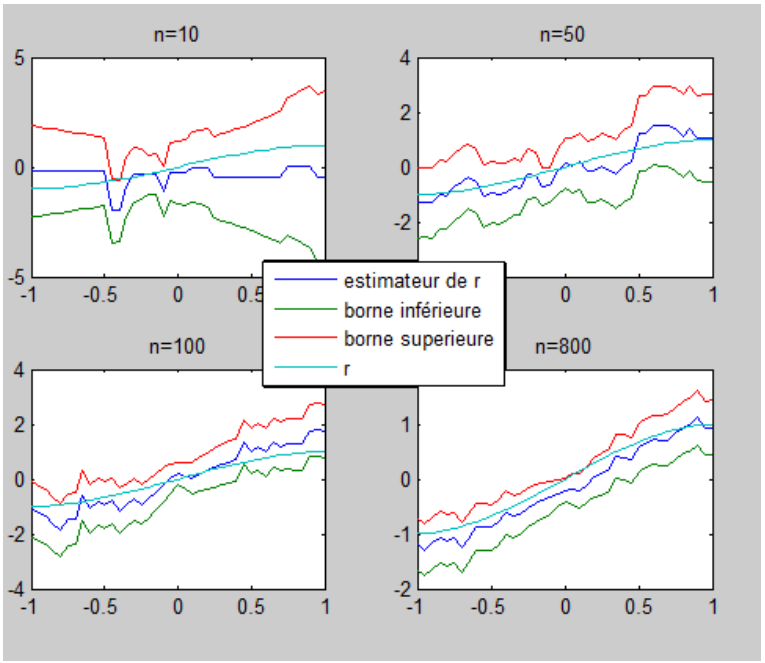

FIGURE 5.4 – les bandes de confiance de la fonction de régression dans le cas d).

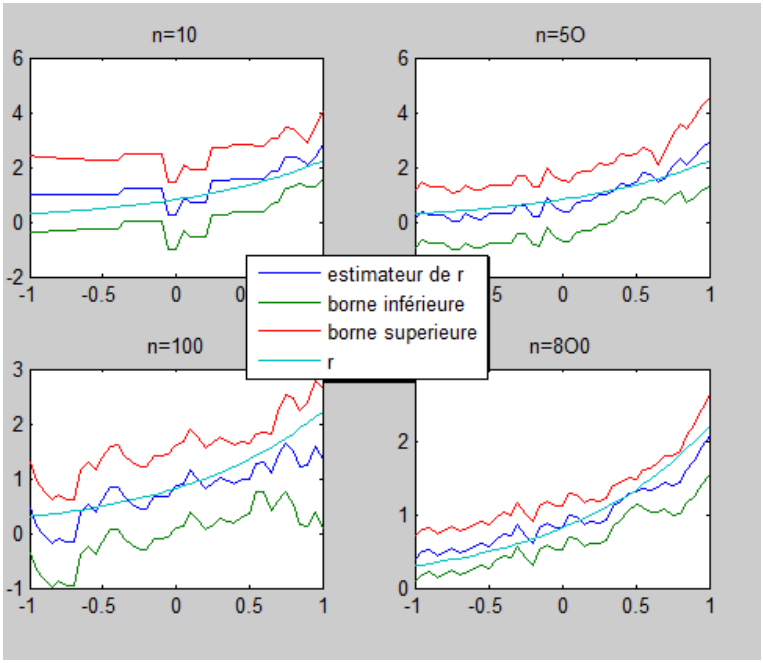

FIGURE 5.5 – les bandes de confiance de la fonction de régression dans le cas e).

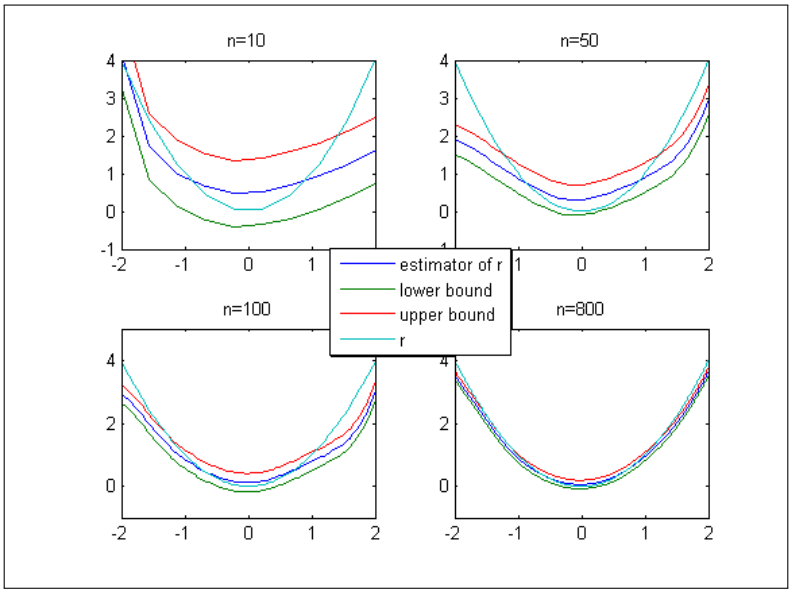

FIGURE 5.6 – Les intervalles de confiance de la fonction de régression dans le cas c).

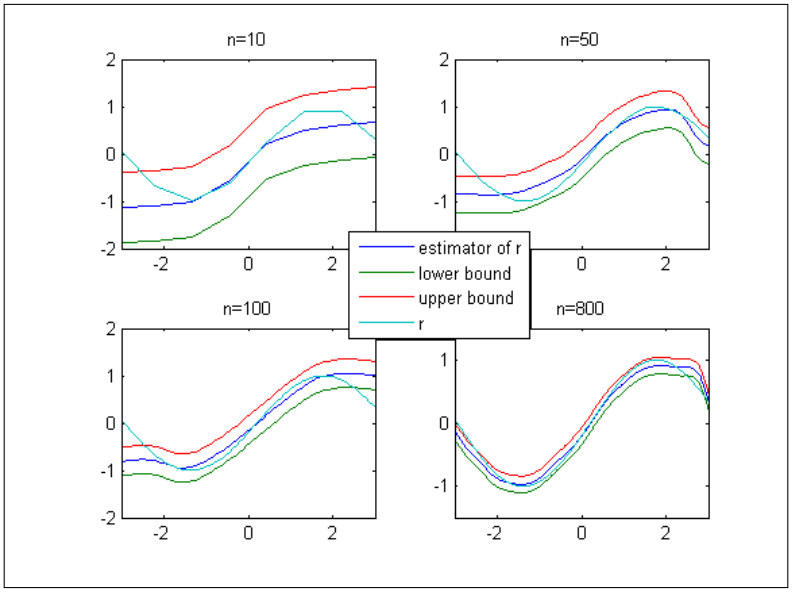

FIGURE 5.7 – Les intervalles de confiance de la fonction de régression dans le cas d).

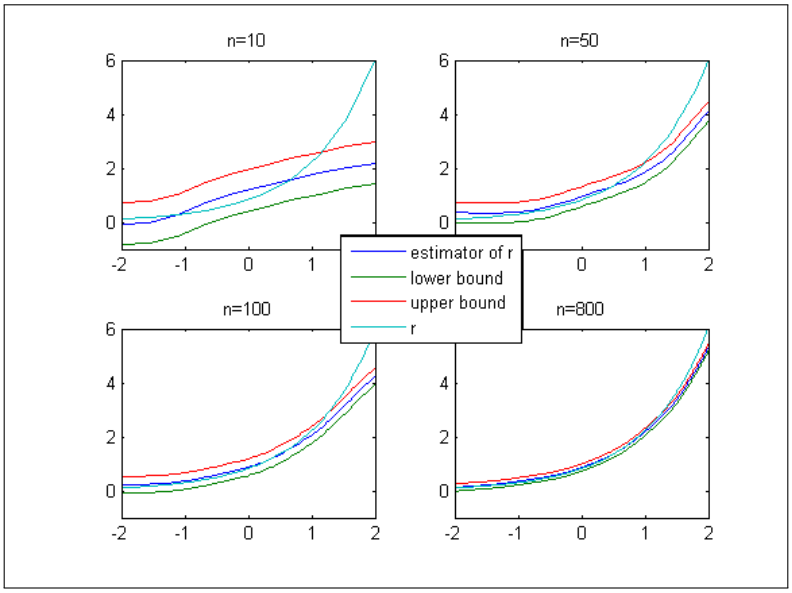

FIGURE 5.8 – Les intervalles de confiance de la fonction de régression dans le cas e).

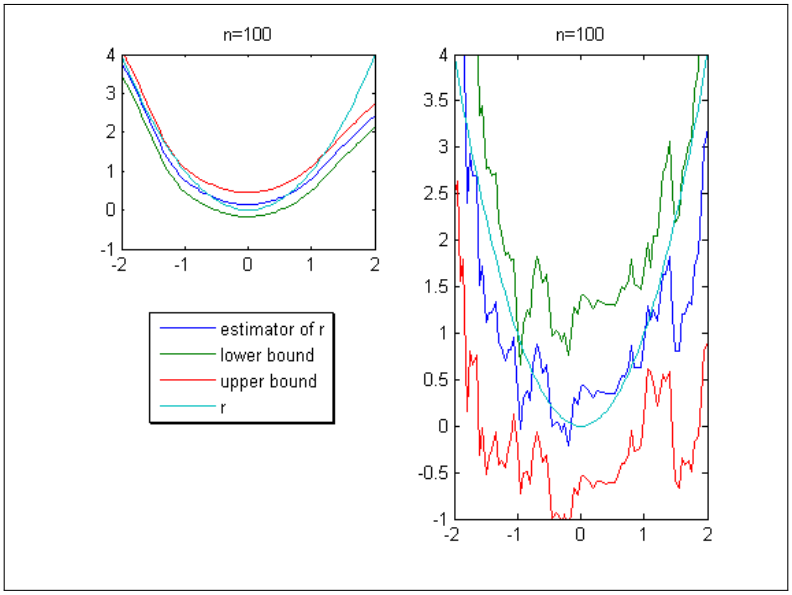

FIGURE 5.9 – Les intervalles et les bandes de confiance pour la fonction de régression dans le cas c).

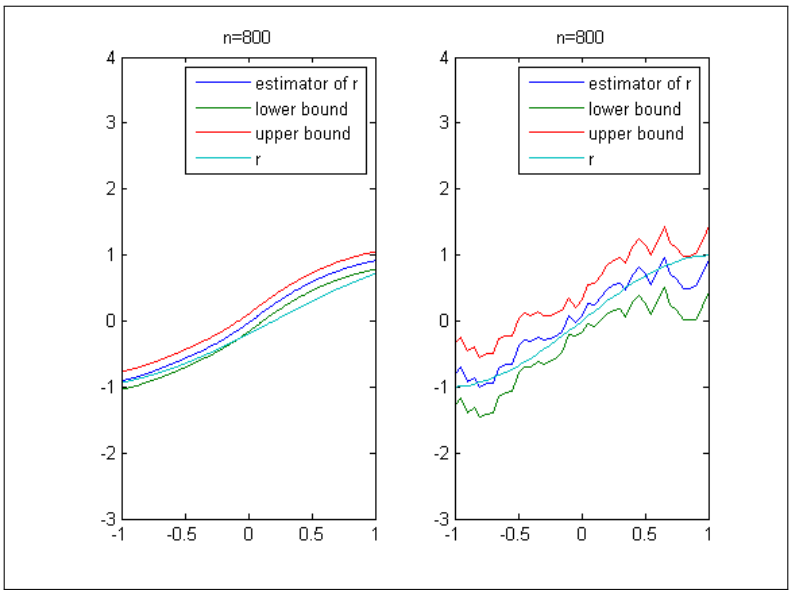

FIGURE 5.10 – Les intervalles et les bandes de confiance pour la fonction de régression dans le cas d).

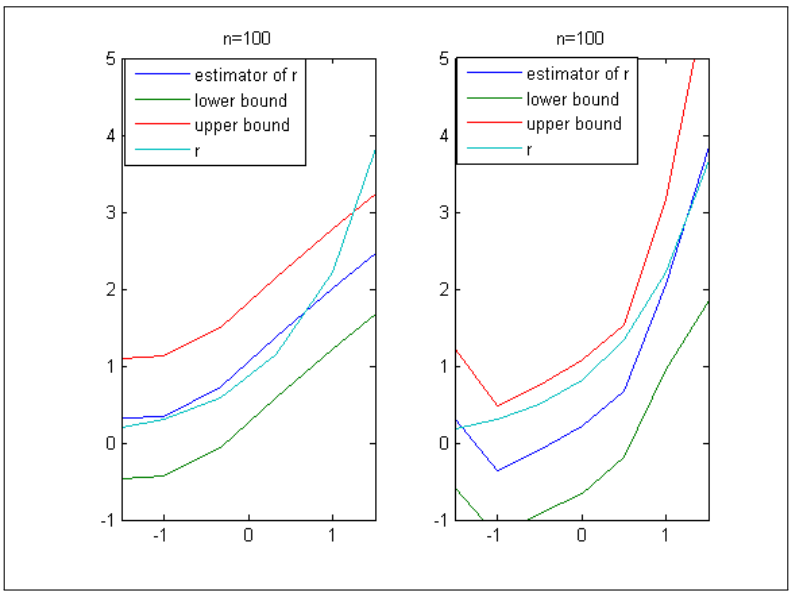

FIGURE 5.11 – Les intervalles et les bandes de confiance pour la fonction de régression dans le cas e).

Bibliographie

[1] Bochner, S. (1955). Harmonic Analysis and the Theory of probability. University of Chicago Press, Chicogo, Illinois.

[2] Bosq, D. (1969). Estimation de la densité condionnelle et de la régression. C. R. Acad. Sci. Paris Sér. A-B, 269, A661-A664.

[3] Bosq, D. and Lecoutre, J. P., (1987). Théorie de l'Estimation Fonctionnelle. Economica, Paris.

*[4] Deheuvels, P. and Mason, D. M. (1992). Functional laws of the iterated logarithm for the increments of empirical and quantile processes. Ann. Probab. **20**, 1248-1287.*

[5] Deheuvels, P. and Mason, D. M. (2004). General asymptotic confidence bands based on kernel-type function estimators. Stat. Infer. Stoc. Processes, 7.3, 225-277.

[6] Devroye, L. and Györfi, L., (1985). Nonparametric Density Estimation : The view. Wiley, New York.

[7] Einmahl, U. and Mason, D.M., (2005). Uniform in bandwidth consistency of kernel-type function estimators. Ann. Statist. ,33(3), 1380-1403.

[8] Ferraty, F. et Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. C. R . Acad. Sci., Paris. 330, No. 2, 139-142.

[9] Ferraty, F. et Vieu, P. (2002/2003). Statistique fonctionnelle : Modèles Non paramétrique de Régréssion. Cours de DEA.

[10] Geoffrey, J. (1974). Sur l'estimation d'une densité dans un espace métrique. C.R. Acad. Paris Sér. A, 278 :1449-1452.

[11] Hall, P. (1981). Laws of the iterated logarithm for nonparametric density estimators. Z. Wahrsch. Verw. Gebiete, 56, 47-61.

[12] Hall, P. (1984). Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function. Z. Wahrsch. Verw. Gebiete, 67, 175-196.

[13] Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press, Cambridge.

[14] Härdle, W. et Kelly, G. (1987). Nonparametric kernel regression estimation -optimal choice of bandwidth. Statistics, 18.1, 21-35.

[15] Härdle, W. et Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist., 13.4, 1465-1481.

[16] Izenman, A.J., (1991). Recents developments in nonparametric density estimation. J. Amer. Statist. Assoc., 86(413), 205-224.

[17] Komlós, I., Major, P., Tusnády, G. (1975). An approximation of partials sums of independent random variable and the sample distribution function. Z. Wahrsch. Verw. Gebiete, 32, 111-131.

[18] Lecoutre, J. (1982). Contribution à l'estimation non paramétrique de la régression. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France .

[19] Nadaraya, E.A. (1976). On the nonparametric estimator of Bayesian risk in the classification problem. Proc. AN. Georg SSR, 82(2), 277-280 (in Russian).

[20] Nadaraya, E. A. (1989). Nonparametric Estimation of Probability Densities and Regression Curves. Kluwer, Dordrecht.

[21] Nadaraya, E. A. (1964). On estimating Regression.Theory. Probab. Applic., 9, 141-142.

[22] Nemouchi, N. Messaci, F. Kebabi, K. (2010). Bandes de confiance de la fonction de régression pour quelques cas non linéaires. CIMA'10, Guelma.

[23] Nemouchi, N. et Mohdeb, Z.(2010). Asymptotic Confidence Bands for Density and Regression Functions in the Gaussian Case. Journal Afrika Statistika Vol. 5, No. 11, 279-287.

[24] Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076.

[25] Prakasa, Rao, B.L.S. (1983). Nonparametric Functional Estimation. Academic Press, New York.

[26] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27, 832-837.

[27] Roussas, G.(1990). Nonparametric Functional Estimation and Related Topics. NATO ASI series 355. Kluwer, Dordrecht.

[28] Sabry, H.(1978). Sur l'estimation non paramétrique des fonctions de régression. C. R. Acad. Sci. Paris. Sér. A-B, 286(20), A941-A944.

[29] Scott, D. W. (1992). Multivariate Density Estimation - Theory, Practice and Visualization. Wiley, New York.

[30] Sheather, J. Jones, M.C. and Marron, J. S. (1996). A Brief Survey of Bandwidth Selection for Density Estimation. Journal of the American Statistical Association. Vol. 91, No. 433, 401-407.

[31] Simonoff, J. S.(1996). Smoothing Methods in Statistics. Springer-Verlag.

[32] Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. Chapman Hall, London.

[33] Stute, W. (1982). A law of the iterated logarithm for kernel density estimators. Ann. Probab., 10.2, 414-422.

[34] Tsybakov, A.(2009). Introduction to Nonparametric Estimation. Springer science and business media, New York.

[35] Tukey, J. W. (1961). Curves as parameters, and touch estimation. In Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, page681-694. Univ. California Press, Berkeley, Caliv.

[36] Uspensky, J. (1935) Introduction to mathematical probability. McGraw-Hill, New York.

[37] Wand, M. P. et Jones, M. C. (1995). Kernel Smoothing. Chapman and Hall, London.
[38] Watson, G. S. (1964). Smooth Regression analysis. Sankhyà Ser. A, 26, 359-372.

Résumé :

Ce mémoire porte sur l'étude des bandes de confiance pour les fonctions de densité et de régression découlant des lois uniformes du logarithme portant sur des estimateurs à noyaux.

Sous certaines conditions sur la fenêtre aléatoire de lissage, les bandes de confiance assurent un niveau de confiance asymptotique à 100% contrairement aux intervalles de confiance. En suite nous exposons l'application de ces résultats établit à la loi gaussienne de variance et de moyenne inconnues. Enfin, nous complétons par une étude de simulation, mettant en évidence la bonne performance des bandes obtenues même pour des petites tailles de l'échantillon.

Mots clés : Bandes de confiance ; estimation non paramétrique ; fonction de densité, fonction de régression.

Abstract :

This report deals with the study of confidence bounds for density and regression functions provided from the logarithm uniform laws of kernel estimators. Under some conditions on the random smoothing window, the confidence bounds insure a level of 100% of asymptotic confidence contrary to the confidence intervals. Next, we expose the application of these results in the case of Gaussian law with unknow variance and mean . Finally, we complete by a simulation study , the good performance of bounds is obtained even for small sample sizes.

Keywords : confidence bounds ; nonparametric estimation ; density, regression function.