

وزارة التعليم العالي و البحث العلمي
République Algérienne Démocratique et Populaire

الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique
Centre Universitaire Salhi Ahmed- Naama
Institut des sciences et technologies
Département de Mathématiques et Informatique



Mémoire de fin d'études

En vue de l'obtention du diplôme de Master
En : Mathématiques

Spécialité : Probabilités, Statistique et Application

Intitulé

La Régression Non Paramétrique Dans Un Espace Vectoriel

Présenté par :
Mebarki Oussama

Soutenu : Septembre 2022

Devant le jury composé de :

Dr. Belguerna Abderrahmane	MCA	C-Univ Naâma	Président
Dr. Laâla Zeyneb	MCB	C-Univ Naâma	Examinatrice
M. Moussaoui Fatma	MAA	C-Univ Naâma	Encadreur

Année universitaire 2021/2022

Remerciement

NOUS VOUDRONS REMERCIER MME MOUSSAOUI FATMA POUR SA DIRECTION, SES ORIENTATIONS ET SA COMPRÉHENSION, AUSSI POUR SES CONSEILS ET SES ENCOURAGEMENTS.

LES MEMBRES DE JURYS POUR AVOIR ACCEPTÉ DE LIRE ET ÉVALUER CE TRAVAIL.

Dédicace

JE DÉDIE CE MODESTE TRAVAIL À MES PARENTS QUI M'ONT SOUTENU TOUT AU LONG DE MON PARCOURS, AINSI À TOUS MES CHÈRES SŒURS, MON CHER FRÈRE, MA FAMILLE ET MES AMIS

Résumé

La régression non paramétrique est un outil statistique permettant d'écrire la relation entre une variable dépendante et une ou plusieurs variables explicatives, Nous rappelons quelques définitions et notations qui seront nécessaires pour l'élaboration des résultats établis dans ce mémoire. Nous présentons notre modèle " la fonction de la régression $r(x)$ où x réel et nous étudions la convergence presque complète et la convergence en moyenne quadratique . On en déduit, nous généralisons les mêmes résultats c-à-d " estimateur de $r(x)$, la convergence presque complète et la convergence en moyenne quadratique au cadre vectoriel

Enfin, nous avons fait une simulation sur des exemples avec le logiciel R.

Chapitre 1

Introduction général

1.1 Historique

Historiquement l'estimateur à noyau a été considéré séparément à la première fois par Nadaraya et Watson en 1964. Les deux auteurs ont adopté ces techniques d'estimation à la fonction de régression $r(x) = \mathbb{E}(Y/X = x)$. Le terme régression trouve son origine dans les travaux de Pearson et Lee (1903) qui introduisent la notion de droite de régression à propos d'observations du couple $(X, Y) = (\text{taille du père, taille du fils})$. Cette droite de régression est déterminée à l'aide d'une représentation graphique (diagram 1 de Pearson et Lee, 1903) associée aux observations $(X_i, Y_i), i = 1, \dots, n$ et qui n'est pas autre chose qu'un régressogramme proposée par Tukey en 1961 [5] a une origine intuitive immédiate : (X à valeurs dans $[0, 1]$) on divise l'intervalle $[0, 1[$ en k_n intervalles égaux et disjoints et on estime la régression par la fonction en escalier qui, sur chaque intervalle, est constante et égale à la moyenne des $Y_i, i = 1, \dots, n$ tels que X_i appartienne à cet intervalle (et nulle si aucun X_i n'appartient à l'intervalle (cette définition se généralise au cas où X prend ses valeurs dans \mathbb{R}^p)). Les propriétés asymptotique de la forme actuelle de l'estimateur à noyau de la fonction de régression se date du 1981. Ils ont été développés par Collomb (1981), dont il a démontré la convergence presque sûre. Bosq en 1987 a étudié la convergence en moyenne quadratique. Il a donné la forme explicite du biais et de la variance de l'estimateur à noyau de la fonction de régression. L'étude de la convergence en norme L_p de l'estimateur à noyau de la fonction de régression a été obtenu par Devroye en 1987.

Pour plus détails voir article de Gerard Collomb [5]

1.2 Notations et Définitions

Dans ce qui suit, nous regroupons quelques définitions ainsi que quelques outils qui seront nécessaires pour l'élaboration des résultats établis dans ce mémoire.

Définition 1

Soit la structure statistique suivant $(\Omega, \mathcal{A}, P_\theta)$, $\theta \in \Phi$, et $\Phi \in \mathbb{R}^d$.

Ω : espace fondamental.

\mathcal{A} : la tribu.

P_θ : ensemble de probabilité.

Φ : ensemble des paramètres.

Si $d < \infty$: on dit que la statistique est paramétrique.

Si $d = \infty$: on dit que la statistique est non paramétrique.

Exemple : 1

Le modèle de Bernoulli : On prend $\Omega = (0, 1)$, $\mathcal{A} = \mathcal{P}(\Omega)$: c'est l'ensemble des parties de Ω . $P_p(0) = 1 - p = q$ et $P_p(1) = p$

Donc : $(\Omega, \mathcal{A}, P_\theta) : \{\Omega = (0, 1), \mathcal{A} = \mathcal{P}(\Omega), \theta = p\}$

Définition 2

On appelle statistique fonctionnelle toute application T :

$$T : \mathcal{F} \rightarrow \Phi$$

$$F \mapsto T(F) = \theta$$

\mathcal{F} : l'ensemble des fonctions de répartition.

Φ : l'ensemble des paramètres.

T : s'appelle statistique fonctionnelle .

On dit que $\theta = f$ un paramètre fonctionnel si $T(F) = d(F) = F' = f$.

1.2.1 Le noyau d'ordre k

La définition suivante précise le noyau d'ordre k

Définition 3

Une fonction K de \mathbb{R}^p est dit noyau d'ordre k , $k \in \mathbb{N}^*$, si

$$T_{(i_1, \dots, i_p)}(K) = 0, \text{ pour tout } (i_1, \dots, i_p) \in \mathbb{N}^{*p}, \text{ vérifiant } i_j < k \text{ (} 1 \leq j < p \text{)}$$

et

$$T_j(K) \neq 0, \text{ pour tout } j \leq k$$

où

$$T_{(i_1, \dots, i_p)}(K) = \int_{\mathbb{R}^p} u_1^{i_1} \dots u_p^{i_p} K(u_1, \dots, u_p) du_1 \dots du_p, \text{ et } T_j(K) = \int_{\mathbb{R}^p} u_j^k K(u_1, \dots, u_p) du_1 \dots du_p.$$

telle que $k > 0$, c'est à dire qu'il vérifie :

$$\int t^j K(t) dt = 0, \forall j = 1, \dots, k-1 \text{ et } 0 < \left| \int t^k K(t) dt \right| < \infty.$$

1.2.2 Mode de convergence

Dans cette section, nous allons introduire différentes notions de convergence pour une suite de variables aléatoires.

Définition 4

(Convergence en Loi) Soit (X_n) une suite de variables aléatoires réelles définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. On dit que la suite (X_n) , de fonction de répartition F_n , converge en loi vers une v.a X de fonction de répartition F , si la suite $F_n(x)$ converge vers $F(x)$ en tout point x où F est continue :

$$X_n \xrightarrow{l} X, \text{ quand } n \rightarrow \infty.$$

Définition 5

(Convergence en probabilité) On dit que (X_n) converge vers X en probabilité si :

$$\forall \epsilon > 0 \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0$$

et on écrit,

$$X_n \xrightarrow{p} X, \text{ quand } n \rightarrow \infty.$$

Définition 6

(Convergence presque sûrement) On dit que la suite (X_n) de v.a.r. converge presque sûrement vers X s'il existe un élément A de la tribu \mathcal{A} tel que $P(A) = 1$ et pour tout $\omega \in A$

$$P(\omega, \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega)) = 1$$

On note

$$X_n \xrightarrow{p.s} X$$

Définition 7

(Convergence en moyenne quadratique) Soit (X_n) une suite de v.a.r. On dit qu'elle converge en moyenne quadratique vers une v.a.r. X si

$$\mathbb{E}(|X_n - X|^2) \xrightarrow{n \rightarrow +\infty} 0$$

.

Définition 8

(Convergence dans L^p) Soit (X_n) une suite de v.a.r. dans L^p . On dit qu'elle converge dans L^p vers une v.a.r. X si

$$\|X_n - X\|_p \xrightarrow{n \rightarrow +\infty} 0$$

.

Définition 9

(Convergence presque complète) On dit que la suite (X_n) de v.a.r. est convergente presque complète vers X si

$$\forall \epsilon > 0, \quad \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$$

et notée par $\lim_{n \rightarrow \infty} X_n = X$, p.co

Proposition 1

Si $\lim_{n \rightarrow \infty} X_n = X$, p.co, alors nous avons

1. $\lim_{n \rightarrow \infty} X_n = X$, p.
2. $\lim_{n \rightarrow \infty} X_n = X$, p.s.

1.2.3 Taux de convergence

1.2.3.1 Notations $O(h)$ et $o(h)$

Soit X_n variable aléatoire : Soit x et h deux séries de nombres réels. Alors, lorsque $n \rightarrow \infty$

a) $X_n = O(h) \iff \forall \epsilon > 0, \exists \delta$ et N tels que $\mathbb{P}(|\frac{X_n}{h}| > \delta) < \epsilon, \forall n > N$.

b) $X_n = o(h) \iff \forall \epsilon > 0, \lim \mathbb{P}(|\frac{X_n}{h}| > \delta) = 0$.

Le taux de convergence presque sure à 0 pour une suite des variable aléatoire est définie par :

$$\begin{aligned} X_n - X = O_{p.s.}(u_n) &\iff \mathbb{P}((X_n - X) = O_{u_n}) = 1 \\ &\iff \mathbb{P}(\exists c < \infty, \exists n, \forall m > n, |X_n - X| \leq cu_m) = 1. \end{aligned}$$

et le taux de convergence en probabilité définie par :

$$X_n - X = O_p(u_n) \iff \lim_m \limsup_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \leq mu_n) = 1$$

Maintenant on définit le taux de la convergence presque complète :

Définition 10

On dit que le taux de convergence presque complète des $(X_n)_{n \in \mathbb{N}}$ vers X est l'ordre u_n telle que $(u_n)_{n \in \mathbb{N}}$ est une suite déterministe de nombres réels positifs qui tend vers zéro si et seulement si :

$$\exists \epsilon_0 > 0, \sum_{n \in \mathbb{N}} \mathbb{P}(|X_n - X| > \epsilon_0 u_n) < \infty.$$

et on écrit : $X_n - X = O_{p.co}(u_n)$.

Cette définition du taux aléatoire introduite par Ferraty et Vieu (2006). Elle a l'avantage théorique d'impliquer les deux vitesses de convergence classiques en probabilité et presque sûre, et l'avantage pratique d'être souvent plus facile à démontrer. Dans les quinze dernières années, ce mode de convergence a été très utilisé dans des travaux concernant la statistique non-paramétrique des données fonctionnelles

Proposition 2

[1] Si $\lim_{n \rightarrow \infty} u_n = 0$, $X_n = O_{p.co}(u_n)$ et $\lim_{n \rightarrow \infty} Y_n = l_y p.co$, où l_y est une nombre réel déterministe. Alors

- i) $X_n Y_n = O_{p.co}(u_n)$.
- ii) $\frac{X_n}{Y_n} = O_{p.co}(u_n)$, avec $l_y \neq 0$.

1.3 Outils

L'inégalité exponentielle de Hoeffding est utilisée dans la démonstration de problème de convergence presque complète et elle donnée ci-dessous :

Lemme 1

[3] Soit $\Delta_1, \dots, \Delta_n$ des variables aléatoires centrées, indépendantes et de même loi, telle qu'il existe deux réels positifs δ_1 et δ_2 vérifiant :

$$|\Delta_1| \leq \delta_1 \text{ et } \mathbb{E}|\Delta_1|^2 \leq \delta_2.$$

Alors, pour tout $\epsilon \in]0, \frac{\delta_1}{\delta_2}[$, on a :

$$\mathbb{P} \left(n^{-1} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon \right) \leq 2 \exp \left(\frac{-n\epsilon^2}{4\delta_2} \right)$$

1.4 Plan de travail

Dans ce mémoire, nous étudierons l'estimation non paramétrique de la fonction de régression par la méthode du noyau dans un espace vectoriel et nous concentrons sur deux modes de convergence : la convergence presque complète et la convergence en moyenne

quadratique.

Ce manuscrit est présenté en quatre chapitre :

Ce chapitre introductif destiné premièrement à un historique sur l'estimation de la fonction de régression par la méthode du noyau et nous présentons également, un nombre important de définitions et notations permettant de clarifier le vocabulaire utilisé dans la suite.

Dans la deuxième chapitre, nous étudions l'estimation de la fonction de régression dans le cas réelle. Nous présentons la convergence presque complète et la convergence en moyenne quadratique dans le cas où les variables sont indépendantes et identiquement distribuées.

Dans le troisième chapitre, on généralise les résultats du chapitre précédent lorsque la variable explicative est vectorielle.

Dans le chapitre quatre, nous donnons des exemples par simulation dans le deux cas : cas où la variable explicative est réelle et le cas vectorielle.

Nous terminons ce travail par un conclusion générale.

Chapitre 2

Régression Non-Paramétrique Réelle

Dans ce chapitre on essayera estimer la fonction de la régression notée $r(x)$ où x est un réel. On introduit notre modèle ainsi son estimateur dans la première section. Nous étudions la convergence presque complète dans la deuxième section, et dans la dernière section nous traitons la convergence en moyenne quadratique.

2.1 Le modèle non-paramétrique

Définition 11

Soient X et Y deux variables aléatoires sur (Ω, A, P) . La régression non paramétrique est une fonction notée r vérifie

$$Y = r(x) + \epsilon$$

où ϵ est une variable aléatoire réelle centrée et indépendante de X .

Définition 12

La variable aléatoire X s'appelle variable explicative et la variable aléatoire Y s'appelle variable réponse.

Remarque 1

La fonction de régression nous permet de prévoir la variable Y sachant la variable explicative X . Cette fonction est explicitement donnée par

$$r(x) = E(Y/X = x) \tag{2.1}$$

Définition 13

Soit $\{(X_i, Y_i), i = 1, \dots, n\}$ un n -échantillon de (X, Y) . On notera f la densité par rapport à la mesure de Lebesgue sur \mathbb{R} de la variable explicative X . L'estimateur à noyau de la fonction r est défini par

$$\hat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \quad (2.2)$$

K est une fonction de \mathbb{R} dans \mathbb{R} , h_n est un paramètre réel strictement positif.

2.2 Convergence presque complète

(Sous les hypothèses)

(H1) la fonction r (resp f) est de C^k

(H2) $f(x) > 0$

(H3) $\lim_{n \rightarrow \infty} h_n = 0, \lim_{n \rightarrow \infty} \frac{nh_n}{\log n} = \infty$

(H4) K est bornée intégrable à support compact vérifiant $\int K(t)dt = 1, \int t^j K(t)dt = 0, \forall j = 1, \dots, k - 1$ et $0 < \int t^k K(t)dt < \infty$

(H5) La variable réponse est telle que : $|Y| < M < \infty$

Théorème 1

(vitesse de convergence presque complète ponctuelle sous condition de dérivabilité)

Considérons (H1) avec $k > 0$ et supposons que les conditions (H2) et (H5) soient réalisées, On a

$$\hat{r}(x) - r(x) = O(h_n^k) + O\left(\sqrt{\frac{\log n}{nh_n}}\right) p.c.o \quad (2.3)$$

Démonstration :

Nous avons

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}$$

où $\hat{g}(x) = \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)$ et $\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)$

On considère la décomposition suivante :

$$\begin{aligned}
\widehat{r}(x) - r(x) &= \frac{1}{\widehat{f}(x)}(\widehat{g}(x) - g(x) + g(x) - r(x)\widehat{f}(x)) \\
&= \frac{1}{\widehat{f}(x)}(\widehat{g}(x) - g(x)) + \frac{r(x)}{\widehat{f}(x)}(f(x) - \widehat{f}(x)) \\
&= \frac{1}{\widehat{f}(x)}(\widehat{g}(x) - \mathbb{E}\widehat{g}(x) + \mathbb{E}\widehat{g}(x) - g(x)) + \frac{r(x)}{\widehat{f}(x)}(f(x) - \mathbb{E}\widehat{f}(x) + \mathbb{E}\widehat{f}(x) - \widehat{f}(x))
\end{aligned}$$

Lemme 2

$$\mathbb{E}\widehat{f}(x) - f(x) = O(h_n^k) \text{ p.co} \quad (2.4)$$

$$\mathbb{E}\widehat{g}(x) - g(x) = O(h_n^k) \text{ p.co} \quad (2.5)$$

Démonstration du Lemme 2

Pour la première équation

$$\begin{aligned}
\mathbb{E}(\widehat{f}(x)) &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \frac{1}{nh_n} \mathbb{E}\left[\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\
&= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right)\right] \\
&= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) f(X_1) dX_1
\end{aligned}$$

le changement des variables $T = \frac{x - X_1}{h_n}$, $X_1 = x - Th_n \implies dX_1 = -h_n dT$ donc

$$\mathbb{E}(\widehat{f}(x)) = \int K(T) f(x - Th_n) dT$$

on utilise le développement de Taylor

$$f(x - Th_n) = f(x) - h_n T f^{(1)}(x) + \frac{h_n^2 T^2}{2} f^{(2)}(x) + \dots + \frac{h_n^k T^k}{k} f^{(k)}(x) + o(h_n^k)$$

d'où

$$\begin{aligned}\mathbb{E}(\widehat{f}(x)) &= \int K(T)[f(x) - h_n T f^{(1)}(x) + \frac{h_n^2}{2} T^2 f^{(2)}(x) + \dots + \frac{h_n^k}{k!} T^k f^{(k)}(x) + o(h_n^k)] dT \\ \mathbb{E}(\widehat{f}(x)) &= f(x) \int K(T) dT - h_n f^{(1)}(x) \int T K(T) dT + \frac{h_n^2}{2} f^{(2)}(x) \int T^2 K(T) dT \\ &\quad + \dots + \frac{h_n^k}{k!} f^{(k)}(x) \int T^k K(T) dT + o(h_n^k) \\ \mathbb{E}(\widehat{f}(x)) &= f(x) + \frac{h_n^k}{k!} f^{(k)}(x) \int T^k K(T) dT \dots + o(h_n^k)\end{aligned}$$

Pour la deuxième équation

$$\begin{aligned}\mathbb{E}(\widehat{g}(x)) &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\ &= \frac{1}{nh_n} \mathbb{E}\left[\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right) Y_1\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right) \mathbb{E}(Y_1 / X = X_1)\right] \\ &= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) f(X_1) r(X_1) dX_1 \\ &= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) g(X_1) dX_1\end{aligned}$$

le changement des variables $T = \frac{x - X_1}{h_n}$, $X_1 = x - Th_n \implies dX_1 = -h_n dT$ donc

$$\mathbb{E}(\widehat{g}(x)) = \int K(T) g(x - Th_n) dT$$

on utilise le développement de Taylor

$$g(x - Th_n) = g(x) - h_n T g^{(1)}(x) + \frac{h_n^2}{2} T^2 g^{(2)}(x) + \dots + \frac{h_n^k}{k} T^k g^{(k)}(x) + o(h_n^k)$$

d'où

$$\mathbb{E}(\widehat{g}(x)) = \int K(T)[g(x) - h_n T g^{(1)}(x) + \frac{h_n^2}{2} T^2 g^{(2)}(x) + \dots + \frac{h_n^k}{k!} T^k g^{(k)}(x) + o(h_n^k)] dT$$

$$\begin{aligned} \mathbb{E}(\widehat{g}(x)) &= g(x) \int K(T)dT - h_n g^{(1)}(x) \int TK(T)dT + \frac{h_n^2}{2} g^{(2)}(x) \int T^2 K(T)dT + \\ &\quad \dots + \frac{h_n^k}{k!} g^{(k)}(x) \int T^k K(T)dT + o(h_n^k) \\ \mathbb{E}(\widehat{g}(x)) &= g(x) + \frac{h_n^k}{k!} g^{(k)}(x) \int T^k K(T)dT \dots + o(h_n^k) \quad \blacksquare \end{aligned}$$

Lemme 3

$$\widehat{g}(x) - \mathbb{E}\widehat{g}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right) p.co \quad (2.6)$$

$$\widehat{f}(x) - \mathbb{E}\widehat{f}(x) = O\left(\sqrt{\frac{\log n}{nh}}\right) p.co \quad (2.7)$$

Démonstration du Lemme 3

Pour la première équation

$$\begin{aligned} \widehat{g}(x) - \mathbb{E}\widehat{g}(x) &= \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right) - \mathbb{E} \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h}\right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} (Y_i K\left(\frac{x - X_i}{h}\right) - \mathbb{E} Y_i K\left(\frac{x - X_i}{h}\right)) \end{aligned}$$

On pose $\Delta_i = \frac{1}{h} [Y_i K\left(\frac{x - X_i}{h}\right) - \mathbb{E} Y_i K\left(\frac{x - X_i}{h}\right)]$

Comme K et Y sont bornés alors

$$|\Delta_i| \leq \frac{C}{h}.$$

Alors il suffit de calculer

$$\mathbb{V}(\Delta_i) = \mathbb{V}\left(\frac{1}{h} Y_i K\left(\frac{x - X_i}{h}\right)\right)$$

et

$$\mathbb{E}(\Delta_i^2) \leq \mathbb{E}(\mu_i^2)$$

où $\mu_i = \frac{1}{h}[Y_i K(\frac{x-X_i}{h})]$, alors,

$$\begin{aligned}\mathbb{E}(\mu_i^2) &= \mathbb{E}\left(\frac{1}{h^2}Y^2 K^2\left(\frac{x-X_i}{h}\right)\right) \\ &= \mathbb{E}\left(\frac{1}{h^2}\mathbb{E}(Y^2/X_1)K^2\left(\frac{x-X_i}{h}\right)\right) \\ &= \frac{1}{h^2} \int \phi(X_1)f(X_1)K^2\left(\frac{x-X_1}{h}\right)du\end{aligned}$$

où $\phi(X_1) = \mathbb{E}(Y^2/X = X_1)$

On considère le même changement des variables $z = \frac{(x-X_1)}{h}$ donc $dz = \frac{-du}{h}$ pour démontrer que :

$$\mathbb{E}(\mu_i^2) = \frac{1}{h} \int \phi(x-zh)f(x-zh)K^2(z)dz$$

Puisque la fonction ϕ est borné, la densité f continue et le noyau K est à support compact, alors, il existe une constante C telle que :

$$\mathbb{E}(\mu_i^2) \leq \frac{C}{h}$$

La preuve de ce lemme est basée sur l'inégalité d'Hoeffding , pour laquelle alors, pour tout $\forall \epsilon > 0$ et pour tout δ_2 on a :

$$\mathbb{P}\left[\frac{1}{n} \left| \sum_{i=1}^n \Delta_i \right| > \epsilon\right] \leq 2 \exp\left(\frac{-n\epsilon^2}{4\delta_2}\right)$$

En prenant

$$\epsilon = \epsilon_0 \sqrt{\frac{\log n}{nh}}$$

pour tout $\epsilon_0 > 0$ à

$$\begin{aligned}P[|\hat{g}(x) - \mathbb{E}\hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}}] &\leq 2 \exp\left(\frac{-n\epsilon_0^2 h \log n}{4nhC}\right) \\ &\leq 2 \exp\left(\frac{-\epsilon_0^2 \log n}{4C}\right) \\ &\leq 2n^{-\frac{\epsilon_0^2}{4C}}\end{aligned}$$

Donc

$$\sum_{i=1}^n P[|\hat{g}(x) - \mathbb{E}\hat{g}(x)| > \epsilon_0 \sqrt{\frac{\log n}{nh}}] \leq \sum_{i=1}^n 2n^{-\frac{\epsilon_0^2}{4C}}$$

Il suffit de choisir $\frac{\epsilon_0^2}{4C} > 1$ pour que la série converge d'après la série de Riemann. Pour $\hat{f}(x)$ la démonstration est similaire à celle qui précède de $\hat{f}(x) - \mathbb{E}\hat{f}(x) = O(\sqrt{\frac{\log n}{nh}})$

■

Lemme 4

Sous les hypothèses du lemme précédent

$$\exists \delta > 0, \text{ tel que } \sum_{i=1}^n P[|\hat{f}(x)| < \delta] < \infty \quad (2.8)$$

Démonstration du Lemme 4

le Lemme 3 entraîne en particulier la convergence presque complète de $\hat{f}(x)$ vers $f(x)$.

Ainsi, pour tout $\epsilon > 0$ on a

$$\sum_{i=1}^n \mathbb{P}[|\hat{f}(x) - f(x)| > \epsilon] < \infty$$

On remarque que

$$\hat{f}(x) \leq \frac{f(x)}{2} \implies |\hat{f}(x) - f(x)| \geq \frac{f(x)}{2}$$

On peut écrire, alors

$$\mathbb{P}[|\hat{f}(x)| \leq \frac{f(x)}{2}] \leq \mathbb{P}[|\hat{f}(x) - f(x)| > \frac{f(x)}{2}]$$

Donc

$$\sum_{i=1}^n \mathbb{P}[|\hat{f}(x)| \leq \frac{f(x)}{2}] \leq \sum_{i=1}^n \mathbb{P}[|\hat{f}(x) - f(x)| > \frac{f(x)}{2}] < \infty$$

Il suffit maintenant de prendre $\delta = \frac{f(x)}{2}$ ■

2.3 Convergence en moyenne quadratique

Théorème 2

Sous les conditions suivantes :

(H1) Les fonctions r et f sont 2 -fois continûment différentiables au voisinage de x .

(H2) La densité f de la variable explicative est strictement positive au point x .

(H3) Le paramètre de lissage est tel que :

$$\lim_{n \rightarrow +\infty} h_n = 0 \text{ et } \lim_{n \rightarrow +\infty} nh_n = \infty$$

(H4) Le noyau K est d'ordre 2, borné et intégrable.

(H5) La variable réponse est telle que : $|Y| < M < \infty$

on a

$$\mathbb{E}[\hat{r}(x) - r(x)]^2 = B^2(x)h_n^4 + V(x)\frac{1}{nh_n} + o(h_n^4) + o\left(\frac{1}{nh_n}\right) \quad (2.9)$$

où

$$B(x) = \frac{\int t^2 K(t) dt}{2} \left(\frac{g^{(2)}(x) - r(x)f^{(2)}(x)}{f(x)} \right) \text{ et } V(x) = \int K^2(t) dt \frac{(\phi(x) - r^2(x))}{f(x)}$$

Démonstration La démonstration se fait par un calcul séparé des deux parties : partie biais et partie variance. En effet, l'erreur quadratique peut être exprimé

$$\mathbb{E}[\hat{r}(x) - r(x)]^2 = [\mathbb{E}(\hat{r}(x)) - r(x)]^2 + V[\hat{r}(x)]$$

On considère le développement usuel de $\frac{1}{z}$:

$$\frac{1}{z} = 1 - (z - 1) + \dots + (-1)^p (z - 1)^p + (-1)^{p+1} \frac{(z - 1)^{p+1}}{z}, \forall z \neq 0, \forall p \in \mathbb{N}^*$$

Une application de ce développement pour $z = \frac{\hat{f}(x)}{\mathbb{E}(\hat{f}(x))}$ et $p = 1$ nous permet d'écrire

$$\hat{r}(x) = \frac{\hat{g}(x)}{\mathbb{E}(\hat{f}(x))} \left[1 - \frac{\hat{f}(x) - \mathbb{E}(\hat{f}(x))}{\mathbb{E}(\hat{f}(x))} \right] + \frac{(\hat{f}(x) - \mathbb{E}(\hat{f}(x)))^2}{(\mathbb{E}(\hat{f}(x)))^2} \hat{r}(x) \quad (2.10)$$

Ceci implique que,

$$\mathbb{E}\hat{r}(x) = \frac{\mathbb{E}\hat{g}(x)}{\mathbb{E}\hat{f}(x)} - \frac{A_1}{(\mathbb{E}\hat{f}(x))^2} + \frac{A_2}{(\mathbb{E}\hat{f}(x))^2} \quad (2.11)$$

où

$$A_1 = \mathbb{E}\widehat{g}(x)(\widehat{f}(x) - \mathbb{E}\widehat{f}(x)) \text{ et } A_2 = \mathbb{E}[(\widehat{f}(x) - \mathbb{E}\widehat{f}(x))^2\widehat{r}(x)]$$

Pour calculer le terme A_1 on utilise l'indépendance puis l'équidistribution des couples (X_i, Y_i) donc :

$$\begin{aligned} A_1 &= \text{Cov}(\widehat{g}(x), \widehat{f}(x)) = \frac{1}{(nh)^2} \sum_{i=1}^n \text{Cov}(Y_i K(\frac{x - X_i}{h}), K(\frac{x - X_i}{h})) \\ &= \frac{1}{h} \mathbb{E}[Y \frac{1}{h} K^2(\frac{x - X}{h})] - \frac{1}{n} \mathbb{E}[Y \frac{1}{h} K(\frac{x - X}{h})] \mathbb{E}[\frac{1}{h} K(\frac{x - X}{h})] \end{aligned}$$

$$A_1 = \frac{1}{nh} (g(x) \int K^2(t) dt) - \frac{1}{n} (g(x) - o(1))(f(x) - o(1)) \quad (2.12)$$

$$= \frac{1}{nh} g(x) \int K^2 dt + o(\frac{1}{nh}) \quad (2.13)$$

Concernant maintenant le terme A_2 , puisque Y est bornée et puisque K est positif, on peut borner sûrement $\widehat{r}(x)$ par une constante C , et arriver ainsi à

$$|A_2| \leq C \mathbb{V}(\widehat{f}(x))$$

donc

$$A_2 = O(\frac{1}{nh}) \quad (2.14)$$

$$\mathbb{E}[\widehat{r}(x)] - r(x) = ((\mathbb{E}\widehat{f}(x))^{-1} \mathbb{E}\widehat{g}(x) - r(x)) - (\mathbb{E}\widehat{f}(x))^{-2} \text{Cov}(\widehat{g}(x), \widehat{f}(x)) \quad (2.15)$$

$$+ ((\mathbb{E}\widehat{f}(x))^{-2} \mathbb{E}(\widehat{f}(x) - \mathbb{E}\widehat{f}(x))^2 \widehat{r}(x)) \quad (2.16)$$

Comme la variable Y est borné, on peut trouver une constante $C > 0$ telle que $\widehat{r}(x) \leq C$.

D'où,

$$\begin{aligned}\mathbb{E}[\widehat{r}(x)] - r(x) &= ((\mathbb{E}\widehat{f}(x))^{-1}\mathbb{E}\widehat{g}(x) - r(x)) - (\mathbb{E}\widehat{f}(x))^{-2}\text{Cov}(\widehat{g}(x), \widehat{f}(x)) \\ &\quad + ((\mathbb{E}\widehat{f}(x))^{-2}V(\widehat{f}(x)))^2O(1)\end{aligned}$$

Pour la partie dispersion nous inspirons de techniques de Sarda et Vieu (2000) et de Bosq et Lecoutre (1987) et en vertu de l'expression 2.10, nous déduisons que

$$\mathbb{V}([\widehat{r}(x)]) = (\mathbb{E}\widehat{f}(x))^{-2}\mathbb{V}[\widehat{g}(x)] - 2(\mathbb{E}\widehat{f}(x))^{-3}(\mathbb{E}\widehat{g}(x))\text{Cov}(\widehat{g}(x), \widehat{f}(x)) \quad (2.17)$$

$$+ (\mathbb{E}\widehat{f}(x))^{-4}(\mathbb{E}\widehat{g}(x))^2\mathbb{V}(\widehat{f}(x)) + o\left(\frac{1}{nh_n}\right) \quad (2.18)$$

Finalement, le Théorème 2 est une conséquence des lemmes suivants :

Lemme 5

Sous les conditions du Théorème 2 on a

$$\mathbb{E}(\widehat{f}(x)) = f(x) + h_n^2 \frac{f^{(2)}(x)}{2!} \int t^2 K(t) dt + o(h_n^2) \quad (2.19)$$

$$\mathbb{E}(\widehat{g}(x)) = g(x) + h_n^2 \frac{g^{(2)}(x)}{2!} \int t^2 K(t) dt + o(h_n^2) \quad (2.20)$$

Démonstration du Lemme 5

$$\begin{aligned}\mathbb{E}[\widehat{f}(x)] &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{nh_n} \mathbb{E}\left[\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right)\right] \\ &= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) f(X_1) dX_1\end{aligned}$$

le changement des variables $t = \frac{x - X_1}{h_n}$ donc

$$\mathbb{E}[\widehat{f}(x)] = \int K(t) f(x - th_n) dt$$

on utilise le développement de Taylor

$$f(x - th_n) = f(x) + th_n f^{(1)}(x) + t^2 \frac{h_n^2}{2} f^{(2)}(x) + o(h_n^2)$$

d'où

$$\mathbb{E}[\widehat{f}(x)] = f(x) + \frac{h_n^2}{2} \int t^2 K(t) f^{(2)}(x) dt + o(h_n^2)$$

Pour la deuxième équation

$$\begin{aligned} \mathbb{E}[\widehat{g}(x)] &= \mathbb{E}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\ &= \frac{1}{nh_n} \mathbb{E}\left[\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) Y_i\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right) Y_1\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right) \mathbb{E}[Y_1 | X = X_1]\right] \\ &= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) f(X_1) r(X_1) dX_1 \\ &= \frac{1}{h_n} \int K\left(\frac{x - X_1}{h_n}\right) g(X_1) dX_1 \end{aligned}$$

le changement des variables $t = \frac{x - X_1}{h_n}$ donc

$$\mathbb{E}[\widehat{g}(x)] = \int K(t) g(x - th_n) dt$$

on utilise le développement de Taylor

$$g(x - th_n) = g(x) + th_n g^{(1)}(x) + t^2 \frac{h_n^2}{2} g^{(2)}(x) + o(h_n^2)$$

Ainsi

$$\mathbb{E}[\widehat{g}(x)] = g(x) + \frac{h_n^2}{2} \int t^2 K(t) g^{(2)}(x) dt + o(h_n^2)$$

Corollaire 1

Sous les conditions du Théorème 2 on a

$$(\mathbb{E}\widehat{f}(x))^{-1}\mathbb{E}[\widehat{g}(x)] - r(x) = B(x)h_n^2 + o(h_n^2) \quad (2.21)$$

Démonstration du Corollaire 1 On a

$$\begin{aligned} & (\mathbb{E}\widehat{f}(x))^{-1}\mathbb{E}[\widehat{g}(x)] - r(x) \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2)}{f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)} - r(x) \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2) - r(x)[f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)]}{f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)} \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) + o(h_n^2) - r(x)[f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x) + o(h_n^2)]}{f(x) + O(h_n^2)} \\ &= \frac{g(x) + \frac{h_n^2}{2} \int t^2 K(t)g^{(2)}(x) - r(x)[f(x) + \frac{h_n^2}{2} \int t^2 K(t)f^{(2)}(x)]}{f(x)} + O(h_n^2) \\ &= \frac{h_n^2}{2} \int t^2 K(t) \left(\frac{g^{(2)}(x) - r(x)f^{(2)}(x)}{f(x)} \right) + O(h_n^2) \end{aligned}$$

Lemme 6

Sous les conditions du Théorème 2 on a

$$\mathbb{V}[\widehat{g}(x)] = \frac{1}{nh} f(x)\phi(x) \int K^2(t)dt + O\left(\frac{1}{nh_n}\right) \quad (2.22)$$

avec $\phi(x) = \mathbb{E}[Y^2|X = x]$

Démonstration du Lemme 6 on a

$$\begin{aligned} \mathbb{V}[\widehat{g}(x)] &= \mathbb{V}\left[\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)Y_i\right] \\ &= \frac{1}{n^2 h_n^2} \mathbb{V}\left[\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)Y_i\right] \\ &= \frac{1}{nh_n^2} [\mathbb{E}[K^2\left(\frac{x - X_1}{h_n}\right)Y_1^2] - (\mathbb{E}[K\left(\frac{x - X_1}{h_n}\right)Y_1])^2] \end{aligned}$$

Il est clair que

$$\frac{1}{h_n^2} (\mathbb{E}[K(\frac{x - X_1}{h_n})Y_1])^2 = (\mathbb{E}[\widehat{g}(x)])^2 = O(h^4)$$

D'où

$$\begin{aligned} \mathbb{V}[\widehat{g}(x)] &= \frac{1}{nh_n^2} \mathbb{E}[K^2(\frac{x - X_1}{h_n})\mathbb{E}[Y_1^2/X = X_1]] + O(\frac{1}{nh_n}) \\ &= \frac{1}{nh_n^2} \mathbb{E}[K^2(\frac{x - X_1}{h_n})\mathbb{E}[Y_1^2/X = X_i]] + O(\frac{1}{nh_n}) \\ &= \frac{1}{nh_n^2} \int K^2(\frac{x - X_1}{h_n})f(X_1)\phi(X_1)dX_1 + O(\frac{1}{nh_n}) \end{aligned}$$

le changement de variable $t = \frac{x - X_1}{h_n}$ et on utilise la continuité de la fonction $f(x)\phi(x)$

pour arriver à

$$\mathbb{V}[\widehat{g}(x)] = \frac{1}{nh_n} f(x)\phi(x) \int K^2(t)dt + o(\frac{1}{nh_n})$$

Lemme 7

Sous les conditions du Théorème 2 on a

$$Cov(\widehat{g}(x), \widehat{f}(x)) = \frac{1}{nh_n} g(x) \int K^2(t)dt + o(\frac{1}{nh_n}) \quad (2.23)$$

Démonstration du Lemme 7 On a

$$\begin{aligned} Cov(\widehat{g}(x), \widehat{f}(x)) &= Cov(\frac{1}{nh_n} \sum_{i=1}^n K(\frac{x - X_i}{h_n})Y_i, \frac{1}{nh_n} \sum_{i=1}^n K(\frac{x - X_j}{h_n})) \\ &= \frac{1}{n^2 h_n^2} Cov(\sum_{i=1}^n K(\frac{x - X_i}{h_n})Y_i, \frac{1}{nh_n} \sum_{i=1}^n K(\frac{x - X_j}{h_n})) \\ &= \frac{1}{nh_n^2} [\mathbb{E}(K^2(\frac{x - X_1}{h_n})Y_1) \\ &\quad - \mathbb{E}(K(\frac{x - X_1}{h_n})Y_1)\mathbb{E}(K(\frac{x - X_1}{h_n}))] \end{aligned}$$

En vertu du calcul précédent, (voir le lemme 2) on peut conclure

$$\frac{1}{h_n^2} [\mathbb{E}(K(\frac{x - X_1}{h_n})Y_1)\mathbb{E}(K(\frac{x - X_1}{h_n}))] = O(1)$$

Ainsi,

$$\begin{aligned}
\text{Cov}(\widehat{g}(x), \widehat{f}(x)) &= \frac{1}{h_n} [\mathbb{E}(K^2(\frac{x - X_1}{h_n})\mathbb{E}[Y_1/X = X_1]) + O(\frac{1}{nh_n})] \\
&= \frac{1}{nh_n^2} [\mathbb{E}(K^2(\frac{x - X_1}{h_n})\mathbb{E}[Y_1/X = X_i]) + O(\frac{1}{nh_n})] \\
&= \frac{1}{nh_n} \int K^2(\frac{x - X_1}{h_n})g(X_1)dX_1 + O(\frac{1}{nh_n})
\end{aligned}$$

On considère le changement des variables usuelle $t = \frac{x - X_1}{h_n}$ et on utilise la continuité de la fonction $g(x)$ pour arriver à

$$\text{Cov}(\widehat{g}(x), \widehat{f}(x)) = \frac{1}{nh_n}g(x) \int K^2(t)dt + o(\frac{1}{nh_n})$$

Lemme 8

Sous les hypothèses (H1)-(H2) et (H4) on a,

$$V[\widehat{f}(x)] = \frac{1}{nh_n}f(x) \int K^2(t)dt + o(\frac{1}{nh_n}) \quad (2.24)$$

Démonstration du Lemme 8 on a

$$\begin{aligned}
\mathbb{V}[\widehat{f}(x)] &= \mathbb{V}[\frac{1}{nh_n} \sum_{i=1}^n K(\frac{x - X_i}{h_n})] \\
&= \frac{1}{n^2h_n^2} \mathbb{V}[\sum_{i=1}^n K(\frac{x - X_i}{h_n})] \\
&= \frac{1}{nh_n^2} [\mathbb{E}[K^2(\frac{x - X_1}{h_n})] - (\mathbb{E}[K(\frac{x - X_1}{h_n})])^2]
\end{aligned}$$

Il est clair que

$$\frac{1}{h_n^2} (\mathbb{E}[K(\frac{x - X_1}{h_n})])^2 = (\mathbb{E}[\widehat{f}(x)])^2 = O(1)$$

D'où

$$\begin{aligned}\mathbb{V}[\widehat{f}(x)] &= \frac{1}{nh_n^2} \mathbb{E}\left[K^2\left(\frac{x - X_1}{h_n}\right)\right] + o\left(\frac{1}{nh_n}\right) \\ &= \frac{1}{nh_n} \int K^2\left(\frac{x - X_1}{h_n}\right) f(X_1) dX_1 + o\left(\frac{1}{nh_n}\right)\end{aligned}$$

Le changement des variables $t = \frac{x - X_1}{h_n}$ et si on utilise la continuité de la fonction $f(x)$ pour arriver à

$$\mathbb{V}[\widehat{f}(x)] = \frac{1}{nh_n} f(x) \int K^2(t) dt + o\left(\frac{1}{nh_n}\right)$$

Chapitre 3

Régression Non-Paramétrique

Vectorielle

Dans ce chapitre, nous généralisons les résultats obtenus dans le chapitre précédent au cadre vectoriel.

3.1 Le modèle non-paramétrique

Soit le modèle de régression vectorielle suivant :

$$Y_i = r(X_i) + \epsilon_i \quad (3.1)$$

Où ;

Y_i : la variable aléatoire à expliquer (réel).

X_i : la variable aléatoire explicative à valeur sur \mathbb{R}^p

p un entier strictement positif.

ϵ_i : variable aléatoire centrée indépendante de X .

L'estimateur à noyau pour la fonction r est :

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}, x \in \mathbb{R}^p \quad (3.2)$$

où $\hat{g}(x) = \frac{1}{nh^p} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)$ et $\hat{f}(x) = \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$

3.2 Convergence presque complète

Hypothèses : Le modèle r est renforcé par les conditions :

— (H1) r et f sont k fois continument dérivable autour de x .

— (H2) x étant un point fixé de \mathbb{R}^p

$$f(x) > 0$$

(f est la densité de x)

— (H3)

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} \frac{nh^p}{\log(n)} = \infty$$

Où h est le paramètre de lissage

— (H4) K est borné et intégrable et à support compact, pour tout p -uplet d'entier positifs (i_1, i_2, \dots, i_p)

— (H5) K vérifie :

$$T(i_1, i_2, \dots, i_p) = \int_{\mathbb{R}^p} u_1^{i_1} \dots u_p^{i_p} K(u_1, \dots, u_p) du_1, \dots, du_p$$

$\forall j \in 1, \dots, p$ on a :

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

et

$$0 < \left| \int t^k K(t) dt \right| < +\infty$$

— (H6)

$$|Y| < M < \infty$$

Théorème 3

Sous les hypothèses précédentes on arrive à

$$\widehat{r}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log(n)}{nh^p}}\right) \text{ en}(p.co) \quad (3.3)$$

Preuve. [Preuve de théorème] La preuve est très similaire du celle du théorème 3 en nous contenant d'insister sur les parties de la démonstration pour lesquelles l'aspect multidimensionnel. Dans cette preuve c désigne une constante, et $\int_{\mathbb{R}^p} K(u)du = 1$ et $rf = g$. Le résultat final sera prouvé dès que seront vérifiées les 5 propriétés suivant :

$$\mathbf{E}(\widehat{g}(x)) - g(x) = O(h^k) \text{ en}(p.co) \quad (3.4)$$

$$\mathbf{E}(\widehat{f}(x)) - f(x) = O(h^k) \text{ en}(p.co) \quad (3.5)$$

$$\mathbf{E}(\widehat{g}(x)) - \widehat{g}(x) = O\left(\sqrt{\frac{\log(n)}{nh^p}}\right) \text{ en}(p.co) \quad (3.6)$$

$$\mathbf{E}(\widehat{f}(x)) - \widehat{f}(x) = O\left(\sqrt{\frac{\log(n)}{nh^p}}\right) \text{ en}(p.co) \quad (3.7)$$

$$\exists \epsilon > 0, \sum_{i=1}^n \mathbb{P}(\widehat{f}(x) \leq \epsilon) < \infty \text{ en}(p.co) \quad (3.8)$$

$$\widehat{r}(x) = \frac{\widehat{g}(x)}{\widehat{f}(x)}$$

Et $\widehat{g}(x) = \frac{1}{nh^p} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)$ et $\widehat{f}(x) = \frac{1}{nh^p} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ On a :

$$\mathbb{E}(\widehat{g}(x)) = \frac{1}{h^p} \mathbb{E}\left[Y K\left(\frac{x-X}{h}\right)\right] \quad (3.9)$$

En conditionnant par rapport à X on arrive à :

$$\mathbb{E}(\widehat{g}(x)) = \frac{1}{h^p} \int_{-\infty}^{+\infty} r(u) K\left(\frac{x-u}{h}\right) f(u) du, z = \frac{x-u}{h} \quad (3.10)$$

$$\mathbb{E}(\widehat{g}(x)) = \int_{-\infty}^{+\infty} g(x-zh) K(z) dz \quad (3.11)$$

$$g(x-zh) - g(x) = \sum_{j=1}^k \frac{(-h)^j}{j!} \sum_{i_1+\dots+i_p=j} (T_k(i_1, \dots, i_p) \left[\frac{\partial^j g}{\partial x_1^{i_1} \dots \partial x_p^{i_p}} \right](z) + O(h^j)) \quad (3.12)$$

On utilise la condition (5), on trouve :

$$\mathbb{E}(\widehat{g}(x)) - g(x) = \frac{(-h)^k}{k!} \sum_{j=1}^k \left[\frac{\partial^k g}{\partial x_j^k} \right](x) T_k(j) + O(h^k) \quad (3.13)$$

ceci achève la preuve.

$$\mathbb{E}(\widehat{g}(x)) - g(x) = O(h^k)$$

Pour montrer 3.5 on suit les mêmes étapes que 3.4 on arrive à :

$$\mathbb{E}(\widehat{f}(x)) - f(x) = \frac{(-h)^k}{k!} \sum_{j=1}^k \left[\frac{\partial^k f}{\partial x_j^k} \right](x) T_k(j) + O(h^k) \quad (3.14)$$

ceci achève la preuve.

$$\mathbb{E}(\widehat{f}(x)) - f(x) = O(h^k)$$

Pour 3.6 on a besoin du lemme 1

si on applique ce lemme aux variables :

$$\Delta_i = \frac{1}{h^p} (Y_i K\left(\frac{x-X_i}{h}\right) - \mathbb{E} Y_i K\left(\frac{x-X_i}{h}\right))$$

$$|\Delta_i| \leq \frac{c}{h^p} \quad (3.15)$$

car K et Y sont bornés $\mathbb{V}(\Delta_i) = \mathbb{E}(\Delta_i^2) - (\mathbb{E}(\Delta_i))^2$ et $\mathbb{E}(\Delta_i) = 0$

on a

$$\mathbb{E}(\mu_i^2) \leq \frac{c}{h^p} \quad (3.16)$$

on applique maintenant le lemme :

$$\mathbb{P}[|\mathbb{E}(\widehat{g}(x)) - g(x)| > \epsilon] \leq 2 \exp\left(-\frac{n\epsilon^2 h^p}{4c}\right) \quad (3.17)$$

si on pose :

$$\epsilon = \epsilon_0 \sqrt{\frac{\log(n)}{nh^p}}$$

on arrive finalement :

$$\mathbb{P}[|\mathbb{E}(\widehat{g}(x)) - g(x)| > \epsilon_0 \sqrt{\frac{\log(n)}{nh^p}}] \leq 2n^{-c\epsilon_0^2} \quad (3.18)$$

on peut choisir $\epsilon_0 > \frac{1}{\sqrt{c}}$

D'où le résultat.

Pour 3.7 il suffit de reprendre les calculs précédents de 3.6

$$\mathbb{P}[|\mathbb{E}(\widehat{f}(x)) - f(x)| > \epsilon_0 \sqrt{\frac{\log(n)}{nh^p}}] \leq 2n^{-c\epsilon_0^2} \quad (3.19)$$

Pour 3.8 il suffit de poser $\epsilon = \frac{f(x)}{2}$ pour que

$$\widehat{f} \neq 0 \text{ en } (p.c.o)$$

Des résultats précédents on arrive à

$$\widehat{r}(x) - r(x) = O(h^k) + O\left(\sqrt{\frac{\log(n)}{nh^p}}\right)$$

3.3 Convergence en moyenne quadratique

Théorème 4

Sous les conditions suivantes :

- (H1) r et f sont k fois continuellement dérivables autour de x
- (H2) x étant un point fixé de \mathbb{R}^p

$$f(x) > 0$$

- (H3)

$$\lim_{n \rightarrow \infty} h = 0 \text{ et } \lim_{n \rightarrow \infty} nh^p = \infty$$

- (H4)

K est borné intégrable positif symétrique et à support compact

- (H5)

$$\phi(u) = \mathbb{E}(Y^2/X = u) \text{ est continue au point } x$$

- (H6)

$$|Y| < M < \infty$$

Théorème 5

Convergence en moyenne quadratique ponctuelle multi-variée sous condition de dérivabilité. considérons le modèle (H1) avec $k = 2$ et supposons que les conditions (H2), (H3), (H4), (H5) et (H6) soient réalisées.

On a

$$\mathbb{E}[\widehat{r}(x) - r(x)]^2 = B^2(x)h^4 + V(x)\frac{1}{nh^p} + o(h^4 + \frac{1}{nh^p}) \quad (3.20)$$

où

$$B(x) = \frac{1}{2f(x)} \sum_{j=1}^p T_K(j) \left(\left[\frac{\partial^k g}{\partial x_j^K} \right](x) - \left[r(x) \frac{\partial^k f}{\partial x_j^K} \right](x) \right) \quad (3.21)$$

$$\mathbb{V} = \int_{\mathbb{R}^p} K^2(t) dt \frac{(\phi(x) - r^2(x))}{f(x)} \quad (3.22)$$

Preuve du Théorème 5 On considère comme toujours que K est unitaire (i.e K est d'intégrale 1). le résultat énoncé va découler des 2 résultats suivants :

$$\mathbb{E}\widehat{r}(x) - r(x) = B(x)h^2 + o(h^2) \quad (3.23)$$

et

$$\mathbb{V}\widehat{r}(x) = \mathbb{V}(x) \frac{1}{nh^p} + o\left(\frac{1}{nh^p}\right) \quad (3.24)$$

Preuve de 3.23 concernant ce terme de biais notons que le résultat 2.11 reste valable dans cadre multidimensionnel en reprenant les preuves de 2.12 et 2.14 et en les adaptant à notre cadre multivarié on arrive à

$$A_1 = \frac{1}{nh^p} g(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right) \quad (3.25)$$

et

$$A_2 = O\left(\frac{1}{nh^p}\right) \quad (3.26)$$

Enfin 3.23 découle directement de 2.11, 3.13, 3.13, 3.25 et 3.26. Preuve de 3.24 la première chose consiste à établir , en suivant les même que pour démontrer 2.17 que l'on a

$$\mathbb{V}(\widehat{r}(x)) = \frac{\mathbb{V}(\widehat{g}(x))}{(\mathbb{E}\widehat{f}(x))^2} - 4 \frac{\mathbb{E}\widehat{g}(x) \text{Cov}(\widehat{g}(x), \widehat{f}(x))}{(\mathbb{E}\widehat{f}(x))^3} + 3 \mathbb{V}(\widehat{g}(x)) \frac{(\mathbb{E}\widehat{f}(x))^2}{(\mathbb{E}\widehat{f}(x))^4} + o\left(\frac{1}{nh^p}\right) \quad (3.27)$$

Ensuite ,il suffit de reprendre les preuves de 3.6 et 3.7 et d'utiliser la condition (H5) qui va permettre de préciser les constantes Ainsi on arrive à :

$$\mathbb{V}(\widehat{g}(x)) = \frac{1}{h^p} f(x) \phi(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right) \quad (3.28)$$

$$\mathbb{V}(\widehat{f}(x)) = \frac{1}{h^p} g(x) \phi(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right) \quad (3.29)$$

quant à la Covariance , elle se calcule de manière analogue et nous avons

$$Cov(\widehat{g}(x), \widehat{f}(x)) = \frac{1}{h^p} g(x) \int_{\mathbb{R}^p} K^2(t) dt + o\left(\frac{1}{nh^p}\right) \quad (3.30)$$

Chapitre 4

Application sous R

Soient un échantillon $(X_i, Y_i), i = 1, \dots, n$ et $Y_i = r(X_i) + \epsilon_i$ avec ϵ_i : v.a centré indépendant de X_i et

$$r(x) = \mathbb{E}(Y/X = x) = \frac{\int y f(x, y) dy}{f_X(x)} \quad (4.1)$$

est la fonction de regression, où $f_X(x)$ est la densité de la variable X . Nous avons vu que $r(x)$ est estimé par la quantité :

$$\hat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)} = \frac{\Phi_{n,X}(x)}{f_{n,X}(x)} \quad (4.2)$$

Il dépend de la taille de l'échantillon n , du noyau K et de la fenêtre h

Dans la suite de ce chapitre, On peut écrire notre modèle sous forme :

$$y = r(x) + \epsilon,$$

et nous étudions les deux cas :

- Régression linéaire : $y = 1 + 2x + \epsilon$
- Régression non linéaire : $y = \exp(x) + \epsilon$

Dans chaque modèle :

- 1) X est de loi normale centré de variance $\sigma^2 = 0.2$. 2) ϵ un terme d'erreur de loi $N(0; 1)$.
- 3) Etudier les cas suivants : a) Paramètre de lissage h fixe et n varié, noyau normal (noyau

à support non compact)

b) Paramètre de lissage h fixe et n varié, noyau d'epanechnikov (noyau à support compact)

c) n fixe, h varié et noyau normal

d) n fixe, h varié et noyau d'epanechnikov

4.1 Régression linéaire

Soit le modèle linéaire

$$y = 1 + 2x + \epsilon$$

4.1.1 Paramètre de lissage h fixe et n varié, noyau normal (noyau à support non compact)

Dans ce cas, on choisit le paramètre de lissage $h = n^{-\frac{1}{5}}$ (fixé), n varié ($n = 50; 100; 500$) et K est un noyau gaussien : $K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$

Code R

```
n=50 # taille de l
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y=1+2*X+E # Mod
```

```
# Noyau Normale K(t) c
```

```
K=function(t)(1/sqrt(2*pi))*exp(-0.5 * t^2)
```

```
# param
```

```
h = nm.2
```

```
# Initiation
```

```
s=100 # taille de l
```

```
a=min(X) # borne inf
```

```
b=max(X) # borne sup
```

```

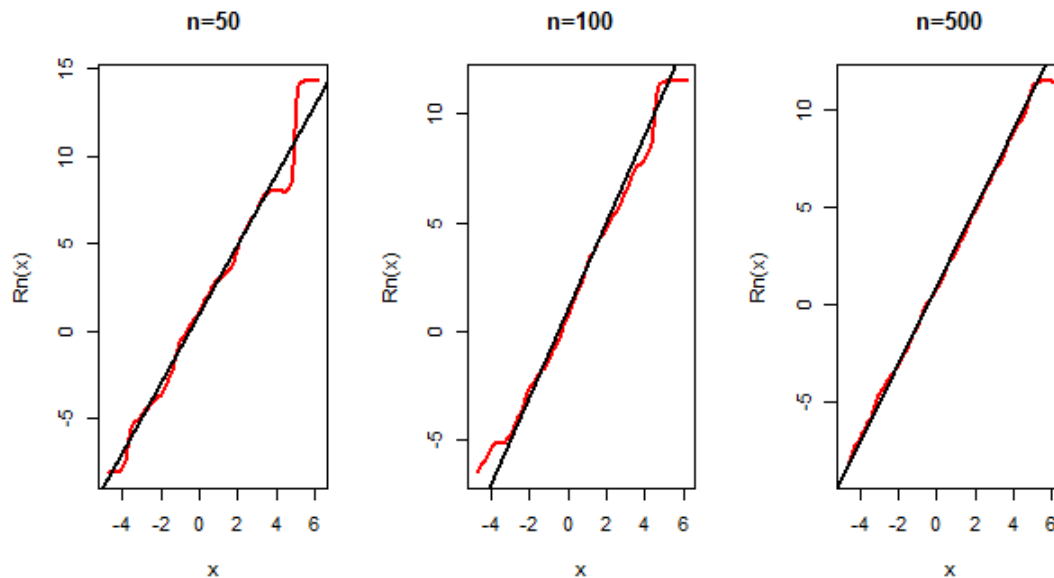
x=seq(a,b,length=s) # Intervalle [a,b]
V=numeric(n)
fn=numeric(s)
# Densit
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
# Fonction Hn(.)
W=numeric(n)
Hn=numeric(s)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn # R
op=par(mfrow=c(1,3))
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)
### Pour n =100###
n=100
X=rnorm(n,0,2)
E=rnorm(n)
Y=1+2*X+E
h = n-0.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)

```

```

for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=100",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)
#### Pour n =500####
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=1+2*X+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=500",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)

```



4.1.2 n fixe, h varié et noyau normal

Dans cette section, nous prenons le paramètre de lissage dans l'intervalle $]0; 1[$ et avec des tests graphique en va déterminer le paramètre h optimal (au sens graphique). On fixe la taille de l'échantillon $n = 250$ et le noyau K est normal, l'estimation obtenue avec les valeurs de h varié de $0 : 1$ à $0 : 9$ sont données dans la figure. Il est clair que la valeur de h optimale est de $h = 0 : 7$ (ligne 3, colonne 1)

Code R

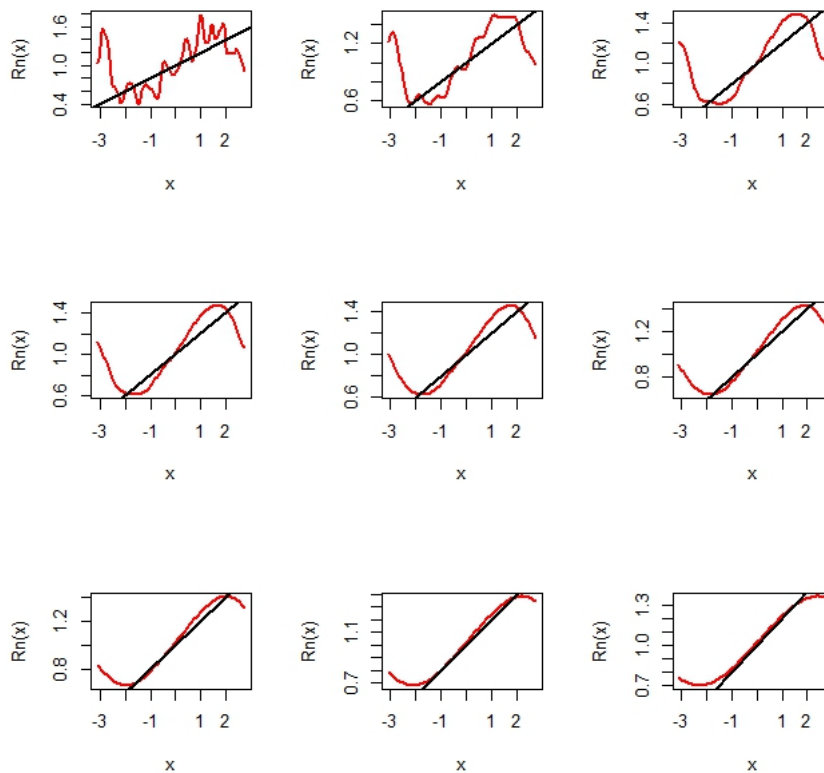
```
n=250 # taille de l'échantillon
X=rnorm(n,0,2)
E=rnorm(n)
Y=1+2*X+E
# Noyau Normale K(t) c'est une densité
K=function(t)(1/sqrt(2*pi))*exp(-0.5 * t^2)
# paramètre de lissage h
h=seq(.1,.9,length=9)
# Initiation
s=100 # taille de l'intervalle [a,b]
a=min(X) # borne inf
```

```

b=max(X) ‡ borne sup
x=seq(a,b,length=s) ‡ Intervalle [a,b]
V=array(dim=c(n,s,9))
fn=array(dim=c(s,9))
W=array(dim=c(n,s,9))
Hn=array(dim=c(s,9))
‡ density fn(x)
for(k in 1 :9)
for(j in 1 :s)
for(i in 1 :n) V[i,j,k]=K((x[j]-X[i])/h[k])
fn[j,k]=sum(V[,j,k])/(n*h[k])
‡ fonction Hn(x)
for(k in 1 :9)
for(j in 1 :s)
for(i in 1 :n) W[i,j,k]=K((x[j]-X[i])/h[k])*Y[i]
Hn[j,k]=sum(W[,j,k])/(n*h[k])
Rn=array(dim=c(s,9))
for(k in 1 :9) Rn[,k]=Hn[,k]/fn[,k]
‡ Graphes
x11() ‡ nouvelle fenetre graphique
op=par(mfrow=c(3,3))
for(k in 1 :9)
plot(x,Rn[,k],xlab="x", ylab="Rn(x)", main=" ",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)

par(op)

```



4.1.3 Paramètre de lissage h fixe et n varié, noyau d'épanechnikov (noyau à support compact)

Dans ce cas, on choisit le paramètre de lissage $h = n^{-\frac{1}{5}}$ (fixé), n varié ($n = 50; 100; 500$) et K est un noyau d'épanechnikov : $K(t) = \frac{3}{4}(1 - t^2)\mathbb{1}_{|t| \leq 1}$

Code R

```
n=50 # taille de l
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y=1+2*X+E # Mod
```

```
# Noyau D'EPANECHNIKOV K(t) c
```

```
K = function(t)ifelse(abs(t) < 1, (3/4) * (1 - t^2), 0)
```

```
# param
```

```
h = n^-0.2
```

```
# Initiation
```

```

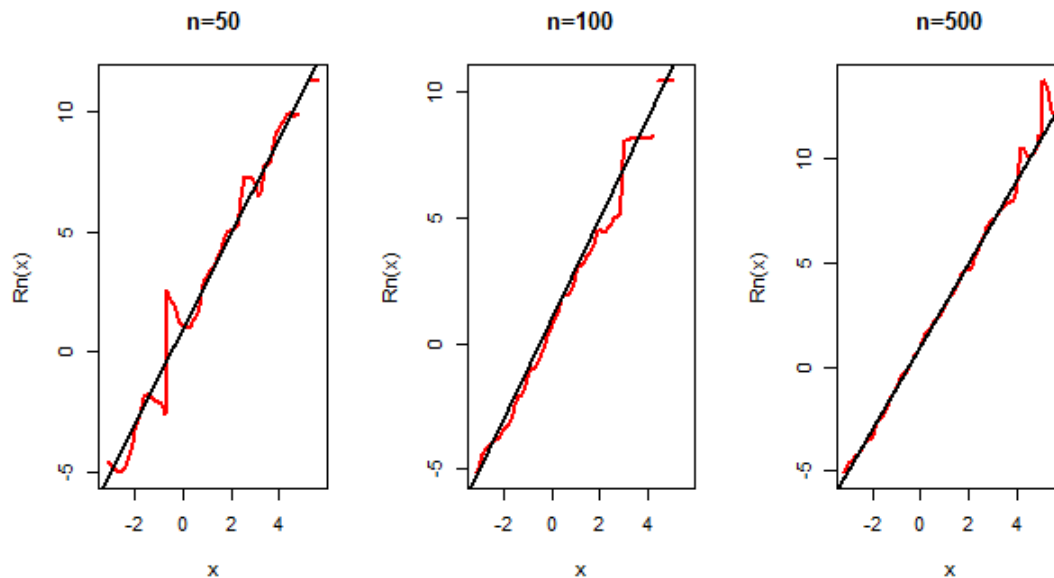
s=100 # taille de l
a=min(X) # borne inf
b=max(X) # borne sup
x=seq(a,b,length=s) # Intervalle [a,b]
V=numeric(n)
fn=numeric(s)
# Densit
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
# Fonction Hn(.)
W=numeric(n)
Hn=numeric(s)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn # R
op=par(mfrow=c(1,3))
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)
### Pour n =100###
n=100
X=rnorm(n,0,2)
E=rnorm(n)
Y=1+2*X+E
h = n-0.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)

```

```

fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=100",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)
#### Pour n =500####
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=1+2*X+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=500",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)

```



4.1.4 n fixe, h varié et noyau d'épanechnikov

Dans cette section, nous prenons le paramètre de lissage dans l'intervalle $]0; 1[$ et avec des tests graphique en va déterminer le paramètre h optimal (au sens graphique). On fixe la taille de l'échantillon $n = 250$ et le noyau K d'épanechnikov, l'estimation obtenue avec les valeurs de h varié de $0 : 1$ à $0 : 9$ sont données dans la figure. Il est clair que la valeur de h optimale est de $h = 0 : 7$ (ligne 3, colonne 1)

Code R

```
n=250 # taille de l'échantillon
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y=1+2*X+E
```

```
# Noyau D'EPANECHNIKOV K(t) c'est une densité
```

```
K = function(t){ifelse(abs(t) < 1, (3/4) * (1 - t^2), 0)}
```

```
# paramètre de lissage h
```

```
h=seq(.1,.9,length=9)
```

```
# Initiation
```

```
s=100 # taille de l'intervalle [a,b]
```

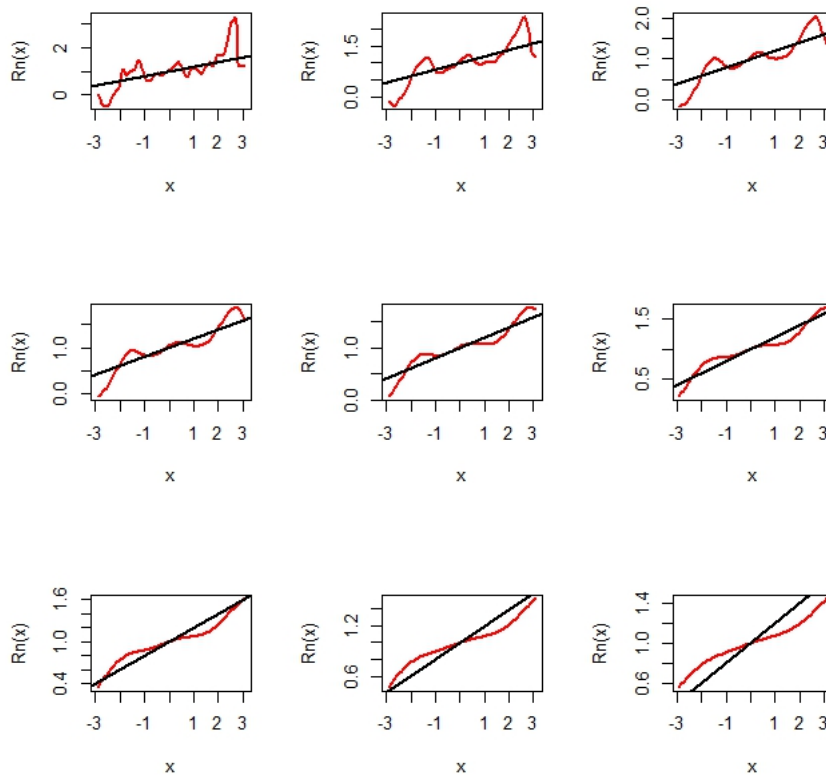
```
a=min(X) #borne inf
```

```

b=max(X) ‡ borne sup
x=seq(a,b,length=s) ‡ Intervalle [a,b]
V=array(dim=c(n,s,9))
fn=array(dim=c(s,9))
W=array(dim=c(n,s,9))
Hn=array(dim=c(s,9))
‡ density fn(x)
for(k in 1 :9)
for(j in 1 :s)
for(i in 1 :n) V[i,j,k]=K((x[j]-X[i])/h[k])
fn[j,k]=sum(V[,j,k])/(n*h[k])
‡ fonction Hn(x)
for(k in 1 :9)
for(j in 1 :s)
for(i in 1 :n) W[i,j,k]=K((x[j]-X[i])/h[k])*Y[i]
Hn[j,k]=sum(W[,j,k])/(n*h[k])
Rn=array(dim=c(s,9))
for(k in 1 :9) Rn[,k]=Hn[,k]/fn[,k]
‡ Graphes
x11() ‡ nouvelle fenetre graphique
op=par(mfrow=c(3,3))
for(k in 1 :9)
plot(x,Rn[,k],xlab="x", ylab="Rn(x)", main=" ",type='l',col=2, lwd= 2)
abline(1,.2,lwd= 2)

par(op)

```



4.2 Régression non linéaire

Soit le modèle non linéaire

$$y = \exp(x) + \epsilon$$

4.2.1 Paramètre de lissage h fixe et n varié, noyau normal (noyau à support non compact)

Dans ce cas, on choisit le paramètre de lissage $h = n^{-\frac{1}{5}}$ (fixé), n varié ($n = 50; 100; 500$) et K est un noyau gaussien : $K(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$

Code R

```
n=50 # taille de l
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y=exp(X)+E
```

```

# Noyau Normale K(t) c
K=function(t)(1/sqrt(2*pi))* exp(-0.5 * t^2)
# param
h = n^-0.2
# Initiation
s=100 # taille de l
a=min(X) #borne inf
b=max(X) # borne sup
x=seq(a,b,length=s) # Intervalle [a,b]
V=numeric(n)
fn=numeric(s)
# Densit
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
# Fonction Hn(.)
W=numeric(n)
Hn=numeric(s)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn # R
# Graphes
x11()# nouvelle fenetre graphique
op=par(mfrow=c(1,3))
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=4, lwd= 2)
lines(x,exp(x),lwd=2)
#####Pour n =100 #####
n=100

```

```

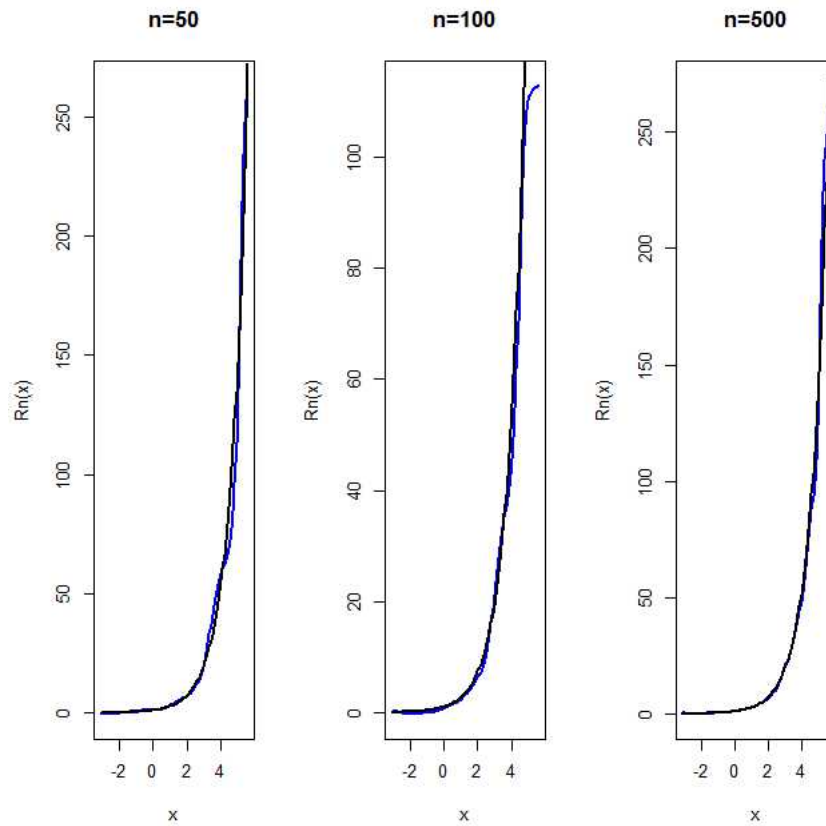
X=rnorm(n,0,2)
E=rnorm(n)
Y=exp(X)+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=100",type='l',col=4, lwd= 2)
lines(x,exp(x),lwd=2)
#####Pour n =500
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=exp(X)+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)

```

$R_n = H_n/f_n$

```
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=500",type='l',col=4, lwd= 2)
```

```
lines(x,exp(x),lwd=2)
```



4.2.2 n fixe, h varié et noyau normal

Code R

```
n=250 # taille de l'échantillon
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y = exp(x) + E
```

```
# Noyau Normale K(t) c'est une densité
```

```
K=function(t)(1/sqrt(2*pi))*exp(-0.5 * tr)
```

```
# paramètre de lissage h
```

```
h=seq(.1,.9,length=9)
```

```
‡ Initiation
```

```
s=100 ‡ taille de l'intervalle [a,b]
```

```
a=min(X) ‡ borne inf
```

```
b=max(X) ‡ borne sup
```

```
x=seq(a,b,length=s) ‡ Intervalle [a,b]
```

```
V=array(dim=c(n,s,9))
```

```
fn=array(dim=c(s,9))
```

```
W=array(dim=c(n,s,9))
```

```
Hn=array(dim=c(s,9))
```

```
‡ density fn(x)
```

```
for(k in 1 :9)
```

```
for(j in 1 :s)
```

```
for(i in 1 :n) V[i,j,k]=K((x[j]-X[i])/h[k])
```

```
fn[j,k]=sum(V[,j,k])/(n*h[k])
```

```
‡ fonction Hn(x)
```

```
for(k in 1 :9)
```

```
for(j in 1 :s)
```

```
for(i in 1 :n) W[i,j,k]=K((x[j]-X[i])/h[k])*Y[i]
```

```
Hn[j,k]=sum(W[,j,k])/(n*h[k])
```

```
Rn=array(dim=c(s,9))
```

```
for(k in 1 :9) Rn[,k]=Hn[,k]/fn[,k]
```

```
‡ Graphes
```

```
x11() ‡ nouvelle fenetre graphique
```

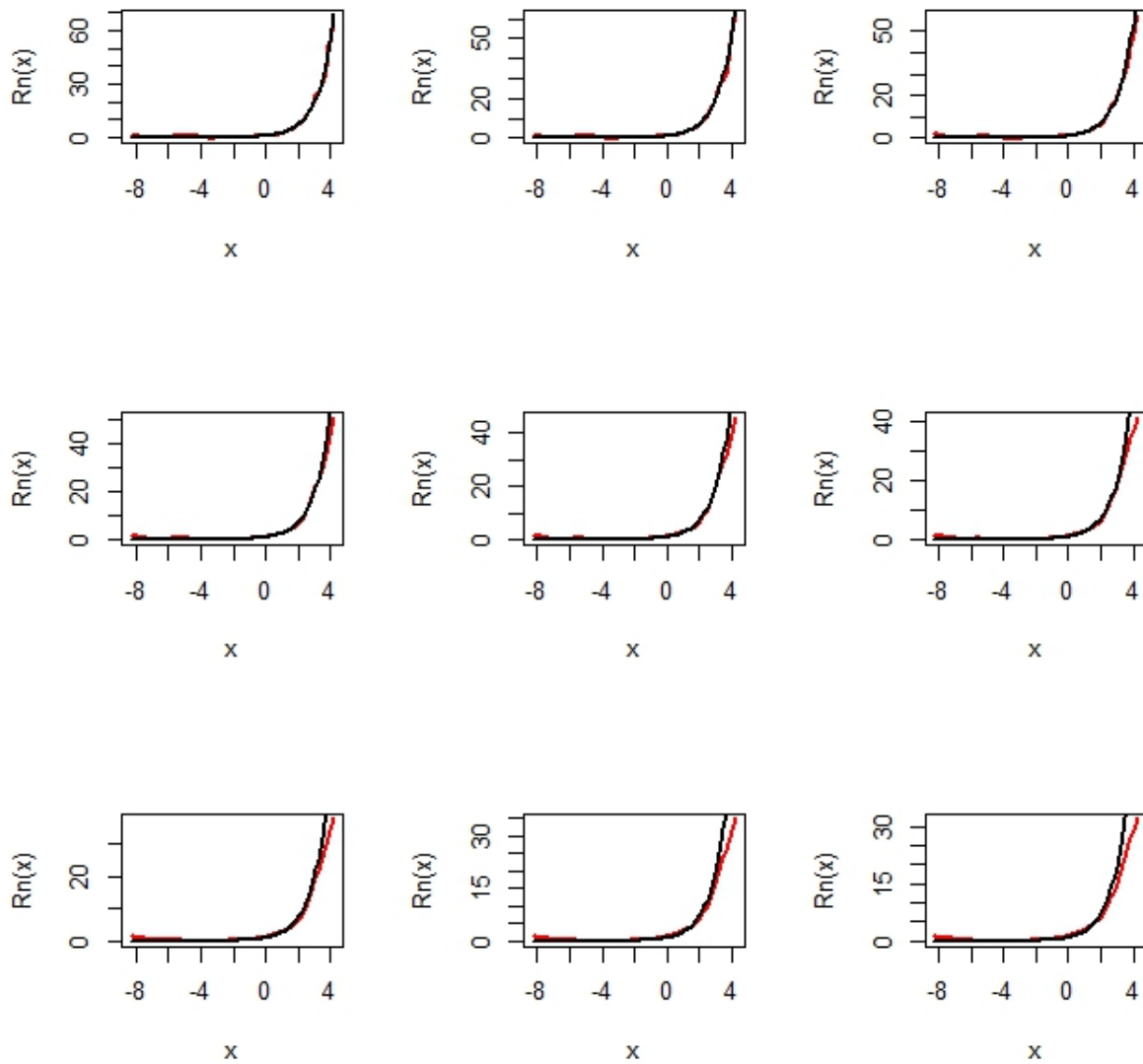
```
op=par(mfrow=c(3,3))
```

```
for(k in 1 :9)
```

```
plot(x,Rn[,k],xlab="x", ylab="Rn(x)", main=" ",type='l',col=2, lwd= 2)
```

```
lines(x,exp(x),lwd= 2)
```

```
par(op)
```



4.2.3 Paramètre de lissage h fixe et n varié, noyau d'épanechnikov (noyau à support compact)

Dans ce cas, on choisit le paramètre de lissage $h = n^{-\frac{1}{5}}$ (fixé), n varié ($n = 50; 100; 500$) et K est un noyau d'épanechnikov : $K(t) = \frac{3}{4}(1 - t^2)\mathbb{1}_{|t| \leq 1}$

Code R

```
n=50 # taille de l
```

```

X=rnorm(n,0,2)
E=rnorm(n)
Y=exp(X)+E # Mod
# Noyau D'EPANECHNIKOV K(t) c
K = function(t)ifelse(abs(t) < 1, (3/4) * (1 - t^2), 0)
# param
h = n^-0.2
# Initiation
s=100 # taille de l
a=min(X) # borne inf
b=max(X) # borne sup
x=seq(a,b,length=s) # Intervalle [a,b]
V=numeric(n)
fn=numeric(s)
# Densit
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
# Fonction Hn(.)
W=numeric(n)
Hn=numeric(s)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn # R
op=par(mfrow=c(1,3))
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=2, lwd= 2)
abline(1,2,lwd=2)
# # # Pour n =100# # #

```

```

n=100
X=rnorm(n,0,2)
E=rnorm(n)
Y=exp(X)+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]
Hn[j]=sum(W)/(n*h)
Rn =Hn/fn
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=100",type='l',col=2, lwd= 2)
lines(x,exp(x),lwd=2)
#### Pour n =500####
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=exp(X)+E
h = n-.2
V=numeric(n)
for(j in 1 :s)
for(i in 1 :n) V[i]=K((x[j]-X[i])/h)
fn[j]=sum(V)/(n*h)
W=numeric(n)
for(j in 1 :s)
for(i in 1 :n) W[i]=K((x[j]-X[i])/h)*Y[i]

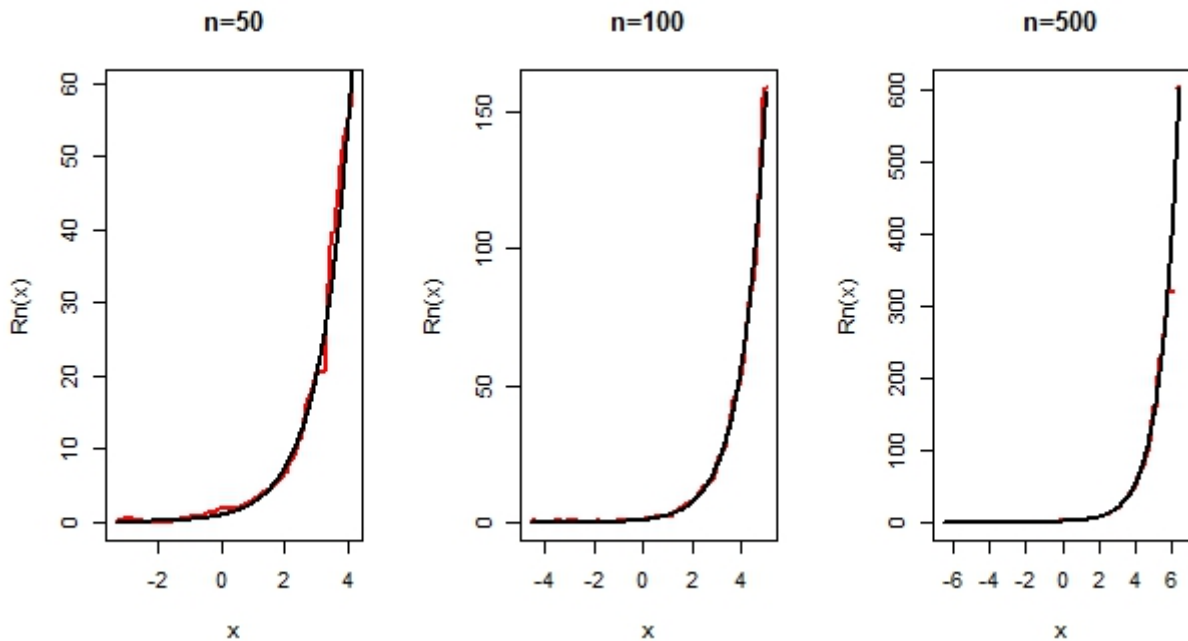
```

```
Hn[j]=sum(W)/(n*h)
```

```
Rn =Hn/fn
```

```
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=500",type='l',col=2, lwd= 2)
```

```
lines(x,exp(x),lwd=2)
```



4.2.4 n fixe, h varié et noyau d'épanechnikov

Code R

```
n=250 # taille de l'échantillon
```

```
X=rnorm(n,0,2)
```

```
E=rnorm(n)
```

```
Y = exp(x) + E
```

```
# Noyau D'EPANECHNIKOV K(t) c'est une densité
```

```
K = function(t)ifelse(abs(t) < 1, (3/4) * (1 - t^2), 0)
```

```
# paramètre de lissage h
```

```
h=seq(.1,.9,length=9)
```

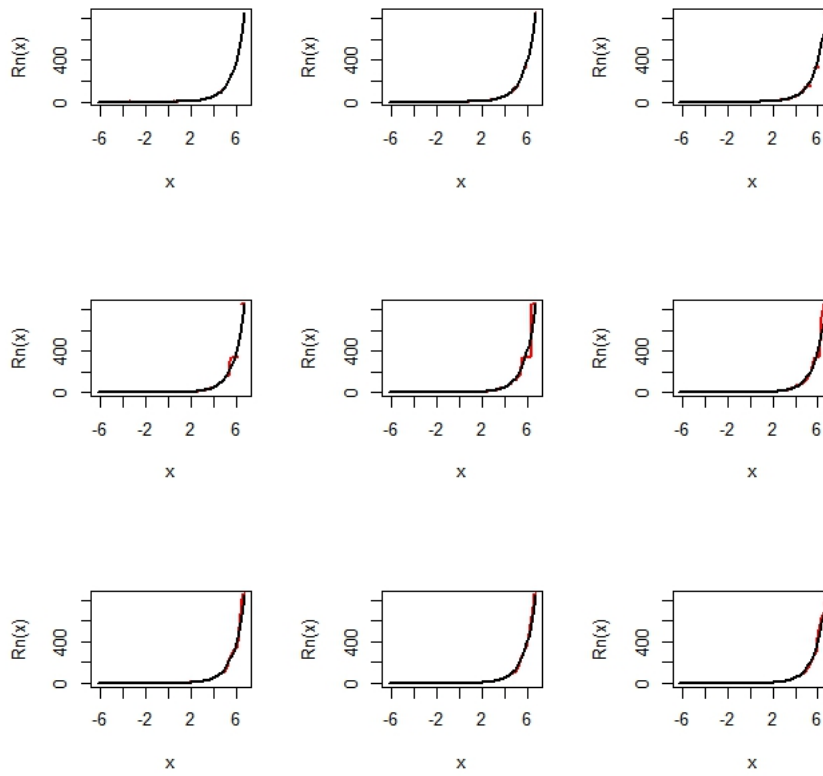
```
# Initiation
```

```

s=100 # taille de l'intervalle [a,b]
a=min(X) # borne inf
b=max(X) # borne sup
x=seq(a,b,length=s) # Intervalle [a,b]
V=array(dim=c(n,s,9))
fn=array(dim=c(s,9))
W=array(dim=c(n,s,9))
Hn=array(dim=c(s,9))
# density fn(x)
for(k in 1 :9)
  for(j in 1 :s)
    for(i in 1 :n) V[i,j,k]=K((x[j]-X[i])/h[k])
  fn[j,k]=sum(V[,j,k])/(n*h[k])
# fonction Hn(x)
for(k in 1 :9)
  for(j in 1 :s)
    for(i in 1 :n) W[i,j,k]=K((x[j]-X[i])/h[k])*Y[i]
  Hn[j,k]=sum(W[,j,k])/(n*h[k])
Rn=array(dim=c(s,9))
for(k in 1 :9) Rn[,k]=Hn[,k]/fn[,k]
# Graphes
x11() # nouvelle fenetre graphique
op=par(mfrow=c(3,3))
for(k in 1 :9)
  plot(x,Rn[,k],xlab="x", ylab="Rn(x)", main=" ",type='l',col=2, lwd= 2)
  lines(x,exp(x),lwd=2)

par(op)

```



Bibliographie

- [1] Ferraty, F and Vieu, P. Non parametric functional data analysis theory and practice. *Springer-Verlag*. (2006).
- [2] Hérdele, W. Applied Nonparametric Regression. *Universität de Berlin*. 1990.
- [3] Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of American Statistical Association*. 58, 13-30, (1963)
- [4] Bosq, D Nonparametric statistics for stochastic Processes *Université Pierre et Marie Curie , Institut de statistique*.1998.
- [5] Gerard, C. Estimation Non-paramétrique de la Regression : Revue Bibliographique. *International Statistical Review*, 49 (1981), 75-93 Longman Group Limited/Printed in Great Britain.