République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme **:**

MASTER en Mathématiques

Option : **Statistique**

Par

CHERIET Bilal

Titre :

Régression non paramètrique

Membres du Comité d'Examen :

Pr.	BRAHIMI Brahim	Professeur	UMKB	Président
Dr.	KHEIREDDINE Souraya	M.C.B	UMKB	Encadreur
Dr.	BENELMIR Imen	M.C.B	UMKB	Examinateur

Juin 2019

Dédicace

C'est avec la volonté de Dieu que je suis arrivée à ce rang pour faire ce modeste travail.

Je le dédie avec tout mon amour et mon respect, avec toute la tendresse à trés chers parents : mon père et ma mère.

*À mes frères "**Aymen** ", "**Faisal**", avec mes meilleurs vœux comme je les remercie pour leurs sacrifices, leur patience et leur encouragements.*

*À la famille « **Cheriet** » et la famille « **Lefriki** »*

*À mes chères amies :"**Saad**", "**Hamza**", "**Manel**"*

À mes chers professeurs et toutes mes amies.

Sans oublier ma promotion, dont je garderai de très bons souvenirs.

REMERCIEMENTS

Avant tout j'adresse mes remerciements à mon Dieu qui m'a donné la patience et le courage qui m'ont permis de réaliser mes souhaits dans le domaine de mon choix : les mathématiques.

Je tiens à remercier mes chers parents, mes frères que Dieu les gardent.

Encore, je remercie mon encadreur Dr. Kheireddine Souraya, pour ses conseils et ses remarques qui ont été très utiles.

Mers remerciement vont égalemment a tous les membres du Département de Mathématiques : Enseignants, Etudiants et Administrateurs.

Résumé

Dans ce mémoire nous étudions l'estimateur non paramétrique de la fonction de régression. La construction de l'estimateur est basée sur l'utilisation d'une densité $K(.)$ appelée noyau et d'un paramètre de lissage h.

Nous rappelons les propriétés asymptotiques de l'estimateur : la convergence, et la normalité asymptotique. Nous parlons aussi du choix de noyau et de paramètre de lissage.

Finalement, nous donnons des explications graphiques des résultats théoriques appliqués sur des exemples de régression linéaire et non linéaire à l'aide du logiciel R.

Table des matières

Table des figures

Introduction

La statistique fonctionnelles constitue un champ de recherches d'actualité, à la fois diversifie par ses aspects fondamentaux et par les différents domaines qu'elle recoupe : statistique non-paramétrique, statistique des opérateurs, variables et/ou modèles fonctionnels. Sans prétendre à l'exhaustivité, l'objectif de cette mémoire est de présenter quelques uns de ces aspects fonctionnels de la statistique en nous contonnant autour des modèles non paramétriques de régression.

L'estimation de la fonction de régression est un problème important dans l'analyse des données avec un large gamme d'applications en filtrage et la prévision dans les communications et le contrôle des systèmes, la reconnaissance de formes et de classification...

L'estimateur de type noyau de la régression à été largement étudié dans la littérature. Les résultats originaux par Nadaraya (1964) et Watson (1964) ont été étendues dans plusieurs journaux, et elles sont résumées par exemple dans Bosq (1998), Devroye et Györfi (1985), et Rao (1983). Citons aussi le cas de données censurées à droites : Carbonez *et al.* (1995), Kohler *et al.* (2002) et autres et le cas de données tranquées à gauche : Lemdani et Ould-Saïd (2006),...

Ce mémoire est divisé en trois chapitres :

Le premier chapitre (introdictif) est consacré à la définition de l'estimation fonctionnelle et l'estimateur à noyau de la densité, et les théorémes de convergences de variables aléatoires ..

Dans le deuxième chapitre, nous présentons l'estimation non paramétrique de la régression et les propriétés de l'éstimateur. Nous étudions ici, la normalité asymptotique de l'éstimateur et choix du noyau et de la largeur de fenêtre.

Nous terminons notre mémoire par un troisième chapitre où nous donnons des exemples par simulation qui expriment l'importance de paramètre de lissage h, la taille de l'échantillon utilisé et le noyau K... dans l'estimation non paramétrique de la fonction de régression.

Chapitre 1

Estimation fonctionnelle

Dans ce chapitre, nous donnerons quelques notions élémentaires, définitions et exemples dans l'estimation paramétrique et non paramétrique.

1.1 Estimation paramétrique et non paramétrique

Premièrement, nous appelons modèle statistique, le triplet $(\mathbb{E},\mathcal{A},\mathbb{P})$ où $\mathbb{E}$ est l'éspace des observations (par exemples des réels), $\mathcal{A}$ une tribu sur $\mathbb{E}$ et $\mathbb{P}$ une famille de probabilité sur $(\mathbb{E},\mathcal{A})$:
Soit $X : \Omega \to \mathbb{E}$ une application mesurable. On peut toujours écrire $\mathbb{P}$ par $(\mathbb{P}_\theta, \theta \in \Theta)$.
Soit h une application de $\mathbb{P}$ dans Θ'. Estimer $h(P)$ c'est essayer de l'évaluer au vu de l'observation d'un échantillon de la variable aléatoire X qui est à valeurs dans $\mathbb{E}$. Donc, le paramètre à estimer est l'application

$$h : P \to \Theta' \quad \text{ou} \quad \Theta \to \Theta'$$
$$\theta \to h\left(P_\theta\right)$$

Un éstimateur de h est une fonction $h_n : x \to h_n(X_1, ..., X_n)$ mesurable par rapport à l'observation $(X_1, ..., X_n)$.

Définition 1.1.1 ***Estimation paramétrique :*** *Si l'on sait à priori que h appartient à une famille paramétrée $\{h(x,\theta), \theta \in \Theta\}$ où $\Theta \subset \mathbb{R}^s$ et $h(.,.)$ est une fonction connue, on parle alors d'éstimation paramétrique, car éstimer h revient à éstimer le paramétr fini dimensionnel θ.*

Définition 1.1.2 ***Estimation non paramétrique :*** *Par contre, si l'on sait seulement que h appartient à $\mathbb{P}$ ensemble des lois de probabilités qui est un éspace de dimension infinie, alors on dit que l'on fait de l'éstimation non paramétrique ou de l'éstimation fonctionnelle*

Dans ce qui suit, on suppose que l'on a observé un échantillon $X_1, X_2, ..., X_n$ à valeurs dans $\mathbb{R}^s$ muni de sa tribu borélienne $\mathcal{B}$. De plus, on suppose que les $\{X_i, i = 1, ..., n\}$ sont indépendantes et identiquement distribuées $(i.i.d)$ $\mu \in P_0$ une famille de loi sur $(\mathbb{R}^s; \mathcal{B})$.

i) La densité de probabilité : Si P_0 est une famille de loi dominée par une loi λ,donc elle admet (théorème de Radon-Nykadim) une densité $f = \frac{d\mu}{d\lambda}$ c'est un paramètre dans L^1. Si $\frac{d\mu}{d\lambda}$ admet une version bornée (respectivement continue et bornée) alors on peut la considérer comme un paramétre dans L^2(respectivement dans C_b $(\mathbb{R}^s)$).

Enfin, si f_μ est différentiable, on définit de nouveaux paramétres fonctionnels : les dérivées partielles de f_μ :

ii) La fonction de répartition : C'est la fonction définie par

$$F_\mu(x_1, ..., x_s) = \mu\left(\prod_{i=1}^{s}]-\infty; x_i]\right) \qquad , (x_1, ..., x_s) \in \mathbb{R}^s.$$

iii) La fonction des quantiles : Pour $s = 1$, la fonction quantile d'ordre p définie par

$$F_\mu^{-1}(p) = Q(p) = \inf\{t \in \mathbb{R}; F_\mu(t) \geq p\} \qquad 0 < p < 1.$$

F_μ^{-1} est un paramètre à valeur dans l'éspace de fonctions réelles définies sur $]0; 1[$monotones non décroissantes et continues à gauche.

v) La fonction caractéristique : Elle est définie par

$$\hat{\mu}(t) = E_{\mu}\left[\exp\left\{i\left\langle t, x\right\rangle\right\}\right] \quad \text{où} \quad t, x \in \mathbb{R}^s.$$

$\hat{\mu}$ est un paramètre dans $C_b\ (\mathbb{R}^s)$.

iv) Le paramètre de régression : Supposons que l'on observe un échantillon$\{(X_i, Y_i), \quad i = 1, ..., n\}$ d'un couple (X, Y) à valeurs dans $\mathbb{R}^{s_1} \times \mathbb{R}^{s_2}$ est soit μ_Y^x , $x \in \mathbb{R}^{s_1}$une famille de versions des lois conditionnelles de Y sachant $X = x$: Toute fonction de la forme $r : x \to r\left(\mu_Y^x\right)$ est un paramètre de régression. Les plus usuals sont :

1) L'éspérance conditionnelle (qui est la fonction de régression),

2) La densité conditionnelle,

3) Le mode conditionnel,

4) La fonction de répartition conditionnelle,

5) Le quantile conditinnel.

1.2 Estimation non-paramétrique d'une densité

1.2.1 L'estimateur de la fonction de densité

Supposons que nous observons n variables aléatoires indépendantes et identiquement distribués X_1,$\cdots$, $X_n$de densité de probabilité par rapport à la mesure de Lebesgue une fonction inconnue f de $\mathbb{R}$ dans $[0, +\infty[$. L'objectif de notre étude est la construction d'un estimateur de f , c'est-à-dire une fonction $\hat{f}_n(x) = f_n(x, X_1, ..., X_n)$ mesurable par rapport à la tribu engendrée par $(X_1, ..., X_n)$.

Notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 et considérons la fonction de répartition empirique

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Xi \leq x\}}\,; \ \ \forall x \in \mathbb{R}$$

La loi forte des grands nombres permet d'affirmer que F_n est un estimateur de F. Il est même possible d'obtenir des intervalles de confiance et de tester l'adéquation des données à différentes lois. Néanmoins, il n'est pas évident d'utiliser $\hat{F}_n$ pour estimer f .
Une des premières idées intuitives est de considérer pour $h > 0$ petit

$$\hat{f}_n(x) = \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} = \frac{1}{2nh}\sum_{i=1}^{n} \mathbf{1}_{\{-h \leq Xi - x \leq h\}}$$

L'idée la plus naturelle est d'estimer f en un point x et de voir ce qui ce passe au voisinage de x. Une des premières idées intuitives est de considérer pour $h > 0$ petit

$$\begin{aligned} f(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h} \\ &= \lim_{h \to 0} \frac{F(x+h) - f(x)}{2h} + \frac{F(x) - F(x-h)}{2h}. \end{aligned}$$

En remplaçant alors F par F_n, on obtient

$$\hat{f}_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}. \tag{1.1}$$

qui peut aussi s'écrire sous la forme

$$\begin{aligned} \hat{f}_{n,X}(x) &= \frac{1}{2h}\frac{1}{n}\sum_{i=1}^{n} \left\{ \mathbf{1}_{\{X_i \leq x+h\}} - \mathbf{1}_{\{X_i \leq x-h\}} \right\} \\ &= \frac{1}{2nh}\sum_{i=1}^{n} \mathbf{1}_{\{x-h \leq X_i \leq x+h\}} \\ &= \frac{1}{2nh}\sum_{i=1}^{n} \mathbf{1}_{\left\{-1 \leq \frac{X_i - x}{h} \leq 1\right\}} \\ &=: \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right). \end{aligned}$$

Cet estimateur, appelé estimateur de Rosenblatt (1956), c'est le premier exemple d'estimateur à noyau construit à l'aide d'une fonction appelée noyau : $K(t) := \frac{1}{2}\mathbf{1}_{\{-1 \leq t \leq 1\}}$.
Définissions maintenant plus généralement la notion d'estimateur à noyau :

Définition 1.2.1 *Soit* $K : \mathrm{IR} \longrightarrow \mathrm{IR}$ *une fonction intégrable, positive et telle que* $\int K(u)\,du = 1$, K *est appelée noyau. Soit* $h := h_n > 0$ *un paramètre de lissage (fenêtre) qui dépend de la taille de l'échantillon* n. *Pour toute* $n \in \mathrm{IN}^*$, *l'estimateur à noyau de la densité de la v.a.* X *au point* x, *noté* $\hat{f}_{n,X}(x)$ *est donné par :*

$$\hat{f}_{n,X}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{1.2}$$

Voici quelques exemples de noyaux classiques :

- Noyau Triangulaire : $K(t) = (1 - |t|)\,\mathbf{1}_{\{|t|\leq 1\}}$,
- Noyau Biweight : $K(t) = \frac{15}{16}\left(1 - t^2\right)^2 \mathbf{1}_{\{|t|\leq 1\}}$,
- Noyau Gaussien : $K(t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-t^2/2\right\}, \ \ t \in \mathbb{R}$,
- Noyau d'Epanechnikov : $K(t) = \frac{3}{4}\left(1 - t^2\right)\,\mathbf{1}_{\{|t|\leq 1\}}$,

La figure (Fig.**1.1**) si-après présente l'allure des quatre noyaux cités.

Code R .

```
K1=function(t){(1-abs(t))*ifelse(abs(t)<=1,1,0)}
K2=function(t){(15/16)*((1-t^2)^2)*ifelse(abs(t)<=1,1,0)}
K3=function(t){dnorm(t)}
K4=function(t){ifelse(abs(t)<1,(3/4)*(1-t^2),0)}
op=par(mfrow=c(2,2))
curve(K1(x),-1,1,ylab="K(x)",main="Triangulaire")
curve(K2(x),-1,1,ylab="K(x)",main="Biweight")
curve(K3(x),-4,4,ylab="K(x)",main="gaussien")
curve(K4(x),-1,1,ylab="K(x)",main="Epanechikov")
par(op)
```

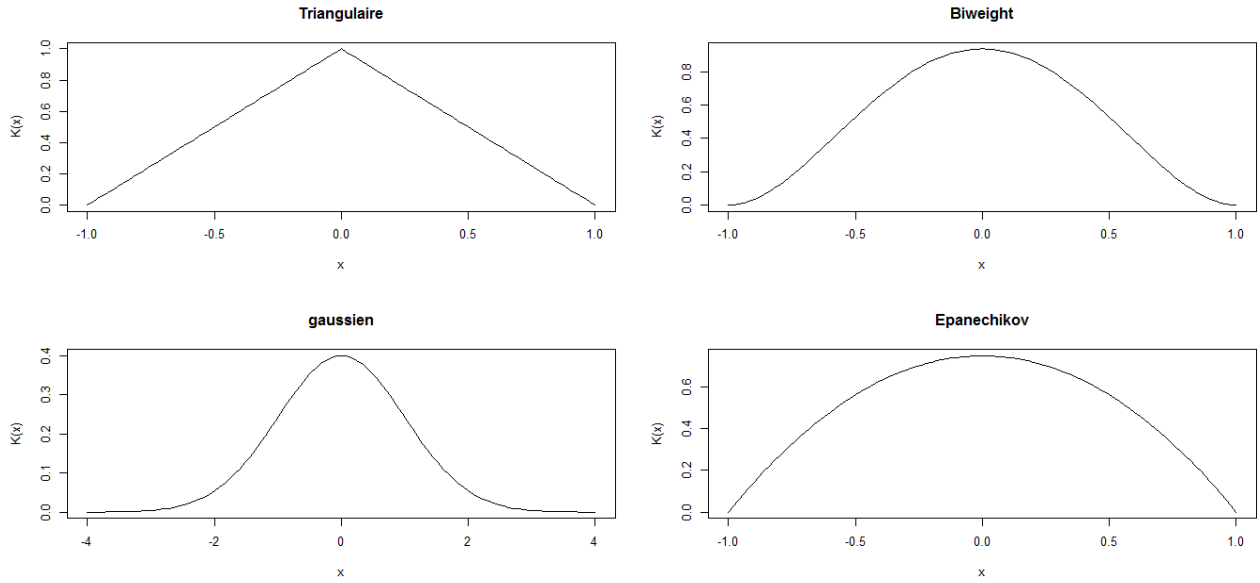

FIG. 1.1 – Allures des noyaux : Triangulaire, Biweight, Gaussien et Epanechnikov.

Définition 1.2.2 *Notons par* $*$ *le produit de convolution i.e :*

$$f * g(x) = \int f(y)\, g(x-y)\, dy = \int g(y)\, f(x-y)\, dy.$$

Lemme 1.2.1 (Bochner) *1) Soit* K *un noyau de Parzen -Rosenblatt et* $f \in L^1$ *alors en tout point* x *de continuité de* f *on a*

$$\lim_{h\to 0} (f * K_h)(x) = f(x)$$

2) Soit maintenant K *un noyau quelconque ; si* $f \in L^1$ *est uniformément continue, alors*

$$\lim_{h\to 0} \sup_{x\in \mathrm{IR}} |f * K_h(x) - f(x)| = 0.$$

1.3 Théorèmes de convergences de variables aléatoires

1.3.1 Convergence dans L^p

Définition 1.3.1 *Soit (X_n) une suite de variable aléatoire.réelle. dans $L^p(\Omega, \mathcal{A}, \mathbb{P})$, $0 < p < \infty$. On dit que X_n converge vers X dans L^p si*

$$\|X_n - X\|_P \underset{n\to+\infty}{\longrightarrow} 0$$

Proposition 1.3.1 *Soit p et q des réels tels que : $1 \leq p < q$. Si (X_n)converge dans L^q vers X, alors la convergence a également lieu dans L^p.*

$$\|X_n - X\|_P \leq \|X_n - X\|_q$$

Corollaire 1.3.1 *Si on a :*

$$X_n \xrightarrow{L^2} X$$

alors

$$X_n \xrightarrow{L^1} X.$$

1.3.2 Convergence en loi

Définition 1.3.2 *Soit (X_n)et X des vecteurs aléatoires à valeurs dans l'espace probabilisable $(\mathbb{R}^P, \mathcal{B}_{\mathbb{R}^P})$. On dit que la suite (X_n) converge en loi vers X si, pour toute fonction h de $\mathbb{R}^p$ vers $\mathbb{R}$, continue et bornée, on a*

$$\lim_{n\to+\infty} Eh(X_n) = Eh(X)$$

On note $X_n \xrightarrow{\mathcal{L}} X$ et on dit aussi parfois que la loi de X_n converge vers celle de X.

Proposition 1.3.2 *Soit (X_n)et X des v.a.r. de fonction de répartition (F_n) et F respectivement. La suite (X_n) converge en loi vers X si, et seulement si,*

$$\lim_{n\to+\infty} F_n(x) = F(x);$$

en tout point x où F est continue.

Exemple 1.3.1 *On considère la suite (X_n) de v.a.r. telle que, pour tout n, la v.a.r. X_n ait pour loi*

$$P\left(X_n = 2 + \frac{1}{n}\right) = 1$$

i.e. la loi de X_n est la dirac en $1 + \frac{2}{n}(P_{X_n} = \delta_{2+\frac{1}{n}})$. En raison de la convergence de la suite $(2 + \frac{1}{n})$ vers 2, on a :

$$\forall x > 2, \quad \exists n_0 \ : \forall n > n_0, 2 + \frac{1}{n} < x$$
$$\forall x > 2, \quad \exists n_0 \ : \forall n > n_0, F_n(x) = P(X_n \leq x) = 1.$$

Par ailleurs, pour tout $x \leq 2$, on a :

$$F_n(x) = P(X_n \leq 2) = 0 :$$

Définissons alors X la v.a.r. de loi δ_2. Sa fonction de répartition est alors :

$$F_X(x) = \begin{cases} 0 & si \quad x < 2 \\ 1 & si \quad x \geq 2 \end{cases}$$

On remarque que la fonction F_X est continue sur $\mathbb{R} \setminus \{2\}$ et que, sur cet ensemble, on a :

$$\lim_{n\to+\infty} F_n(x) = F(x).$$

Ainsi, d'après la proposition précédente, on a la convergence de X_n vers X. Il est intéressant de noter que la convergence des fonctions de répartition n'a pas lieu au point de discontinuité de F puisque l'on a, pour tout n,

$$F_n(2) = 0 \neq F(2) = 1.$$

Théorème 1.3.1 *Soit (X_n) et X des vecteurs aléatoires de $\mathbb{R}^p$, absolument continus de densité $(f_{n,X})$ et f par rapport à la mesure de Lebesgue dans $\mathbb{R}^p$. Si on a, λ_p-presque-partout,*

$$\lim_{n\to+\infty} f_{n,X} = f$$

alors

$$X_n \xrightarrow{\mathcal{L}} X$$

Théorème 1.3.2 (Théorème de Paul Lévy)

1. Si (X_n) est une suite de variables aléatoires dans $\mathbb{R}^p$ convergeant en loi vers une variable aléatoire X dans $\mathbb{R}^p$, alors la suite (φ_{Xn}) des fonctions caractéristiques associée à la suite (X_n) converge en tout point vers la fonction caractéristique φ_X de X, i.e.
$$X_n \xrightarrow{\mathcal{L}} X \Rightarrow \forall x \in \mathbb{R}^p, \varphi_{Xn}(x) \longrightarrow \varphi_X(x)$$

2. Soit (X_n) est une suite de variables aléatoires dans $\mathbb{R}^p$. Si la suite (φ_{Xn}) de ses fonctions caractéristiques converge simplement vers une fonction φ continue en 0, alors φ est la fonction caractéristique d'une variable aléatoire X et X_n converge en loi vers X, i.e.

$$\varphi_{Xn}(x) \longrightarrow \varphi_X(x) \quad , \forall x \in \mathbb{R}^p \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{L}} X \ ,$$

Théorème 1.3.3 ***(Théorème de Cramer-Wold)*** *Soit (X_n) et X des vecteurs aléatoires dans $\mathbb{R}^p$. On a alors l'équivalence suivante :*

$$X_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \forall u \in \mathbb{R}^p \ : u'X_n \xrightarrow{\mathcal{L}} u'X \ ,$$

Preuve. *Supposons en premier lieu que X_n converge en loi vers X. La fonction g de $\mathbb{R}^p$ vers $\mathbb{R}$ définie par $g(x) = u\prime x$, pour u dans $\mathbb{R}^p$, est une forme linéaire. Elle est donc continue. Ainsi, d'après le théorème de Slutsky, on a la convergence*

$$u'X_n \xrightarrow{\mathcal{L}} u'X$$

Réciproquement, supposons que pour tout u dans $\mathbb{R}^p$, on ait

$$u'X_n \xrightarrow{\mathcal{L}} u'$$

Le théorème de Paul Lévy, nous donne alors la convergence

$$\varphi_{u'Xn}(t) \longrightarrow \varphi_{u'X}(t)$$

pour tout t dans $\mathbb{R}$. Celle-ci prise en $t = 1$, nous donne :

$$\varphi_{u'Xn}(1) = \varphi_{Xn}(u) \longrightarrow \varphi_X(u) = \varphi_{u'X}(1)$$

dont on tire, en utilisant la réciproque du théorème de Paul Lévy, la convergence $X_n \xrightarrow{\mathcal{L}} X$

■

1.3.3 Convergence en probabilité

a) Cas des variables aléatoires réelles

Définition 1.3.3 *On dit que la suite (X_n) de v.a.r. converge en probabilité vers la variable aléatoire X, si*

$$\forall \varepsilon > 0,\ P(|X_n - X| > \varepsilon) \to 0;\ \textit{quand}\ n \to +\infty$$

On note $X_n \xrightarrow{P} X$.

Remarquons qu'il est équivalent de dire

$$\lim_{n\to+\infty} P(|X_n - X| > \varepsilon) = 0$$

et

$$\lim_{n\to+\infty} P(|X_n - X| \leq \varepsilon) = 1$$

Le théorème suivant nous donne une condition suffisante pour avoir la convergence en probabilité vers une constante.

Théorème 1.3.4 *Soit (X_n) une suite de v.a.r. dans L^2. Si on a*

$$\lim_{n\to+\infty} EX_n = a \quad et \quad \lim_{n\to+\infty} VarX_n = 0$$

alors

$$X_n \xrightarrow{p} a$$

Preuve. Grâce à l'inégalité de Bienaymé-Tchebychev, on peut écrire, pour tout $\varepsilon > 0$

$$P(|X_n - a| > \varepsilon) \leq \frac{E(X_n - a)^2}{\varepsilon^2}$$

Or, on a déjà vu que

$$E(X_n - a)^2 = Var\ X_n + (E\ X_n - a)^2$$

D'où :

$$\forall \varepsilon > 0,\ P(|X_n - a| > \varepsilon) \leq \frac{Var\ X_n + (E\ X_n - a)^2}{\varepsilon^2}$$

et en utilisant les deux hypothèses, on a bien :

$$\forall \varepsilon > 0,\ \lim_{n \longrightarrow +\infty} P(|X_n - a| > \varepsilon) = 0$$

et donc $X_n \xrightarrow{P} a$. ■

Proposition 1.3.3 *La convergence dans L^1 entraîne celle en probabilité.*

Preuve. Remarquons que l'on a :

$$\begin{aligned} \|X_n - X\|_{L^1} &= E\,|X_n - X_P| \\ &= \int_{|X_n - X_P| > \varepsilon} |X_n - X_P|\, dP + \int_{|X_n - X_P| \leq \varepsilon} |X_n - X_P|\, dP \\ &\geq \int_{|X_n - X_P| > \varepsilon} |X_n - X_P|\, dP \geq \varepsilon P\left(|X_n - X_P| > \varepsilon\right). \end{aligned}$$

La convergence de (X_n) vers X dans L_1 entraîne alors que, pour tout ε strictement positif, on a :

$$P\left(|X_n - X_P| > \varepsilon\right) \underset{n \to +\infty}{\longrightarrow} 0$$

■

Théorème 1.3.5 ***(Théorème de Slutsky)*** *Soit (X_n) et X des v.a.r. Si (X_n) converge en probabilité vers la v.a. X et si g est une application continue de $\mathbb{R}$ vers $\mathbb{R}$, alors*

$$g(X_n) \xrightarrow{P} g(X)$$

b) Cas des vecteurs aléatoires

Définition 1.3.4 *Soit (X_n) et X des vecteurs aléatoires à valeurs dans $\mathbb{R}^p$.*

On dit que (X_n) converge en probabilité vers le vecteur aléatoire X si ses p composantes $X_{n,1}, ..., X_{n,p}$ convergent en probabilité vers les composantes $X_1, ..., X_p$ de X.
Le théorème suivant permet de donner une définition équivalente à cette convergence en probabilité .

Théorème 1.3.6 *Soit $\|.\|$ une norme quelconque dans $\mathbb{R}^p$ et (X_n) et X des vecteurs aléatoires dans $\mathbb{R}^p$. La suite (X_n) converge en probabilité vers la v.a. X si, et seulement si,*

$$\|X_n - X\| \xrightarrow{p} 0, \quad quand \quad n \to +\infty$$

Proposition 1.3.4 *Considérons des suites (X_n) et (Y_n) de v.a.r. Si on a les convergences :*

$$\begin{aligned} X_n &\xrightarrow{p} X \\ et \quad Y_n &\xrightarrow{p} Y \end{aligned}$$

et si g est une fonction continue de $\mathbb{R}^2$ dans $\mathbb{R}$, alors

$$g(X_n; Y_n) \xrightarrow{p} g(X; Y).$$

Corollaire 1.3.2 *Soit (X_n) et (Y_n) des suites de v.a.r. Si on a les convergences :*

$$\begin{aligned} X_n &\xrightarrow{p} X \\ et \quad Y_n &\xrightarrow{p} Y \end{aligned}$$

alors

$$X_n + \lambda Y_n \xrightarrow{p} X + \lambda Y$$

pour tout $\lambda \in \mathbb{R}$, et

$$X_n.Y_n \xrightarrow{p} X.Y$$

1.3.4 Convergence presque sûre

Définition 1.3.5 *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*

$$\forall \omega \in A : \lim_{n \to +\infty} X_n(\omega) = X(\omega)$$

On note

$$X_n \overset{p.s}{\to} X$$

Théorème 1.3.7 *La suite de v.a.r. (X_n) converge presque sûrement vers X si la suite de v.a.r. (Y_m) définie par :*

$$Y_m = \sup_{n \geq m} |X_n - X|$$

alors (Y_m) converge en probabilité vers 0.

Proposition 1.3.5 *Si, pour tout ε strictement positif, la série de terme général $P[|X_n| > \varepsilon]$ est convergente, i.e.*

$$\forall \varepsilon > 0 \quad \sum_n P[|X_n| > \varepsilon] < +\infty$$

alors (X_n) converge presque sûrement vers zéro.

Chapitre 2

Estimation non paramétrique de la fonction de régression

Dans ce chapitre, nous donnerons la définition de l'estimateur non paramétrique de la régression par la méthode du noyau. Nous étudions, les propriétés asymptotiques de l'estimateur et le choix du noyau et le paramètre de lissage.

2.1 L'estimateur non paramétrique de régression

On suppose que l'on a observé un échantillon $\{(X_i, Y_i)\,;\ i = 1, ..., n\}$ et on veut expliquer la variable aléatoire Y_i par X_i. De plus, on suppose que le modèle d'où proviennent les données a la forme

$$Y_i = r(X_i) + \varepsilon_i.$$

où ε_i est l'aléatoire centré et indépendante de X_i et r est une application mesurable réelle. La fonction de régression $r\,(.) = E[Y/X = .]$,apportant de l'information sur la relation de dépendance inconnue de Y et X ; un problème important est l'estimation de r à partir de l'observation de n copies (X_i, Y_i) , $i = 1, ..., n$ qui suivent la meme loi que (X, Y) .
Supposons que (X, Y) a une densité $f : (x, y) \to f\,(x, y)$ sur $\mathbb{R}^2$ et que $f_X : x \to f_X\,(x) =$

$\int f(x,y)\,dy > 0$ (densité de X).

Alors

$$\forall x \in \mathbb{R}, r(x) = E[Y/X = x] = \frac{\int y f(x,y) dy}{f_X(x)}$$

$$f_{n,X}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{X_i - x_i}{h_n}\right)$$

puis on considére l'estimateur de la régression

$$\forall x \in \mathbb{R}, \quad r_n(x) = \frac{\int y f(x,y) dy}{f_X(x)} \mathbf{1}_{f_X(x)\neq 0.} \tag{2.1}$$

Définition 2.1.1 *Si K est un noyau d'ordre 1, l'estimateur défini par (2.1) vérifie*

$$\forall x \in \mathbb{R}, \quad r_n(x) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{X_i - x}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right)}$$
$$= \frac{\sum_{i=1}^{n} Y_i K_{h_n}(X_i - x)}{\sum_{i=1}^{n} K_{h_n}(X_i - x)}$$

où $K_{h_n}(.) = K(./h_n)$.,donc l'estimareur à noyau de la régression est donné par :

$$r_n(x) = \frac{\sum_{i=1}^{n} Y_i K_{h_n}(X_i - x)}{\sum_{i=1}^{n} K_{h_n}(X_i - x)} = \frac{\Phi_{n,X(x)}}{f_{n,X(x)}},$$

C'est l'estimateur à noyau introduit par Nadaraya-Watson (Nadaraya, 1964 et Watson, 1964).

La construction de cet estimateur dépend de deux paramètres, le paramètre de lissage h dont le choix est crucial pour obtenir de bonnes propriétés asymptotiques et la noyau K dont on ne peut pas négliger le rôle pour la réduction du biais.

2.2 Les propriétés de l'estimateur

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.

nous supposons que K est un noyau vérifant les conditions suivantes :

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

(**H.2**) $\lim |x| K(x) \to 0$, quand $|x| \to +\infty$.

(**H.3**) $K \in L_1(\mathbb{R})$, c'est à dire $\int_{\mathbb{R}} |K(x)| dx < +\infty$

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

(**H.5**) $\int_{\mathbb{R}} xK(x)du = 0$

(**H.6**) $\int_{\mathbb{R}} x^2 K(x) dx < +\infty$

(**H.7**) K est bornée, intégrable et à support compact.

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.

2.2.1 Etude asymptotique du biais et de la variance

Etude asymptotique du biais

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

Proposition 2.2.1 *Sous les hypothèse de la proposition (2.3) et*

a) Si $|Y| \leq C_1 < \infty$ *P.S et si* $nh_n \to \infty$, *quand* $n \to \infty$, *alors :*

$$E[r_n(x)] = \frac{E[\Phi_{n,X}(x)]}{E[f_{n,X}(x)]} + O\left(\frac{1}{nh_n}\right)$$

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

$$E[r_n(x)] = \frac{E[\Phi_{n,X}(x)]}{E[f_{n,X}(x)]} + O\left(\frac{1}{\sqrt{n}h_n}\right)$$

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

Proposition 2.2.2 *Si les condition* $(\mathbf{H}.4), (\mathbf{H}.5)$ *et* $(\mathbf{H}.6)$ *sont vérifiées et si* $f_X(.)$ *et* $r(.)$ *sont le classe* $C^2(\mathbb{R})$ *et si* $|Y|$ *est borné.*

Alors

$$E\left[r_n\left(x\right)\right]-r\left(x\right)]=\frac{h_n^2}{2}\left\{\left\{r''\left(x\right)+2r'\left(x\right)\frac{f_X'(x)}{f_X(x)}\right\}\int_{\mathbb{R}}u^2K\left(u\right)du\right\}\left(1+o\left(1\right)\right) \qquad (2.2)$$

Remarque 2.2.1 *1)- Les conditions* $(\mathbf{H}.4), (\mathbf{H}.5)$ *et* $(\mathbf{H}.6)$ *peuvent être remplacées par le noyau* K *est d'ordre* 2 *au sens de Gasser et Müller.*

2)- dans la relation (2.2) est égale à $O(h)+O\left((nh)^{-1}\right)$.

Preuve.

$$E[r_n(x)-r(x)]=\left[EK\left(\frac{x-X}{h_n}\right)\right]^{-1}\left\{\int_{\mathbb{R}}\frac{1}{h_n}K\left(\frac{x-t}{h_n}\right)\Phi\left(t\right)dt-r\left(x\right)\int_{\mathbb{R}}\frac{1}{h_n}K\left(\frac{x-t}{h_n}\right)f\left(t\right)dt\right\}$$

$$=\left\{\left(f\left(x\right)\right)^{-1}\left\{\frac{h_n^2}{2}\Phi''\left(x\right)-\frac{h_n^2}{2}r\left(x\right)f''\left(x\right)\right\}\int_{\mathbb{R}}u^2K\left(u\right)du+\Phi\left(x\right)-r\left(x\right)f\left(x\right)\right\}\left(1+o\left(1\right)\right)$$

comme $\Phi(x)=r(x)f(x)$. *L'équation précédente peut s'écrire :*

$$E[r_n(x)-r(x)]=\left\{\frac{h_n^2}{2}\left\{r''\left(x\right)+2r'\left(x\right)\frac{f_X'(x)}{f_X(x)}\right\}\int_{\mathbb{R}}u^2K\left(u\right)du\right\}\left(1+o\left(1\right)\right)$$

D'où

$$\lim_{n\to\infty}E[r_n(x)]=r(x)$$

■

Etude asymptotique de la variance

Proposition 2.2.3 *Sous* $EY^2 < \infty$, *alors en chaque point de continuité des fonctions* $r(x), f_X(x)$ *et* $\sigma^2(x) = Var(Y/X = x)$ *on a*

$$Var\left[r_n(x)\right] = \frac{1}{nh_n}\left\{\frac{\sigma^2(x)}{f_X(x)}\int K^2(u)\,du\right\}(o(1)+1) \tag{2.3}$$

où $f_X(x) > 0$.

Preuve. *Soit la fonction* $\psi(x) = \int y^2 f(x;y)dy$,*en se basant sur le lemme de* ***Bochner*** *1.2.1on a*

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

$$E\left[\{f_{n,X}(x) - E(f_{n,X}(x))\}\{\Phi_{n,X}(x) - E(\Phi_{n,X}(x))\}\right] = \frac{1}{nh_n}\Phi(x)\int_{\mathbb{R}} K^2(u)\,du\,(1+o(1))$$

et

$$Var\left[f_{n,X}(x)\right] = \frac{1}{nh_n}f_X(x)\int_{\mathbb{R}} K^2(u)\,du\,(1+o(1))$$

posons

$$B_n(x) = \begin{pmatrix} f_{n,X}(x) \\ \Phi_{n,X}(x) \end{pmatrix}$$

et

$$A(x) = \left(\frac{-r(x)}{[f_X(x)]^2}, \frac{1}{f_X(x)}\right).$$

La matrice de variance covariance de $B_n(x)$ *est alors donnée par l'expression suivante*

$$\sum := \frac{1}{nh_n} \begin{pmatrix} f_X(x) & \Phi(x) \\ \Phi(x) & \psi(x) \end{pmatrix} \int_{\mathbb{R}} K^2(u)\, du\, (1 + o(1))$$

En remarquant, que

$$\begin{aligned} Var[r_n(x)] &= A \sum A^t \\ &= \frac{1}{nh_n} \left(\frac{\psi(x)}{|f_X(x)|^2} - \frac{(\Phi(x))^2}{|f_X(x)|^3} \right) \int_{\mathbb{R}} K^2(u)\, du\, (1 + o(1)) \end{aligned}$$

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

$$Var[r_n(x)] = \frac{1}{nh_n} \left\{ \frac{\sigma^2(x)}{f_X(x)} \int_{\mathbb{R}} K^2(u)\, du \right\} (1 + o(1))$$

■

2.2.2 Convergence presque complète

En se basant sur la preuve données 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, aux quelles nous rajoutons les hypothèses suivantes.

• f_X, r sont des fonctions continues au voisinage de x, un point fixé de $\mathbb{R}$.

•La densité f_X et la variable Y sont telles que

$$f_X > 0$$

et

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

où M est une constante réelle positive.

•Le paramétre de lissage h_n est tel que

$$\lim_{n\to\infty} h_n(x) = 0 \quad \text{et} \quad \lim \frac{\log n}{nh_n} = 0,$$

Théorème 2.2.1 *Sous les hypothèses ci-dessus et* $(\mathbf{H}.4), (\mathbf{H}.7)$*, on a :*

$$\lim_{n\to+\infty} r_n(x) = r(x). \qquad p.co$$

2.3 Normalité asymptotique de l'éstimateur

Nous supposons que le noyau K et la séquence sont choisis pour satisfaire les conditions :

1. $K(u)$ *et* $|uK(u)|$ sont bornés.
2. $\int uK(u) = 0$.
3. $\int u^2 K(u)\, du < \infty$.
4. $\lim_{n\to+\infty} nh_n^3 = \infty$ *et* $\lim_{n\to+\infty} nh_n^5 = 0$.

on note

$$Var[Y/X = x] = \frac{v(x)}{g(x)} - \frac{w^2(x)}{g^2(x)}$$

où

$$g(x) = \int f(x,y)\, dy \quad w(x) = \int yf(x,y)\, dy \quad \text{et} \quad v(x) = \int y^2 f(x,y)\, dy$$

Théorème 2.3.1 *suppose* $x_1, ..., x_k$ *sont des points distincts et* $g(x_i) > 0$ *pour* $i = 1, 2, ..., k$*.si* $E(Y^3)$ *est finie et si* g', w', v', g' *et* w" *existent et sont bornées*
alors : $\sqrt{nh_n}\,(r_n(x_1) - r(x_1), ..., r_n(x_k) - r(x_k))$ *converge en distribution ver* Z^* *où* Z^* *est multivariée normale avec vecteur moyen 0 et matrice de covariance diagonale* $C = [C_{ij}]$ *où*

$$C_{ii} = V[Y/X = x_i] \int K^2(u)\, du / g(x_i) \qquad (i = 1, 2, ..., k).$$

Preuve. *pour simplifier la preuve ; nous la donnons dans les cas $k = 2$.La méthode reste vraie pour les cas général.*

pour $i = 1, 2, ..., n$ et $s = 1, 2$:

$$
\begin{aligned}
U_{ni}(x_s) &= \sqrt{h_n}\left[K\left(\frac{x_s - X_i}{h_n}\right)/h_n - E\left[K\left(\frac{x_s - X_i}{h_n}\right)/h_n\right]\right] \\
V_{ni}(x_s) &= \sqrt{h_n}\left[Y_i K\left(\frac{x_s - X_i}{h_n}\right)/h_n - E\left[Y_i K\left(\frac{x_s - X_i}{h_n}\right)/h_n\right]\right] \\
U_n(x_s) &= \sum_{i=1}^{n} U_{ni}(x_s), \qquad\qquad , V_n(x_s) = \sum_{i=1}^{n} V_{ni}(x_s) \\
W_{ni} &= (U_{ni}(x_1), V_{ni}(x_1), U_{ni}(x_2), V_{ni}(x_2)) \\
\sqrt{n} Z_n &= (U_n(x_1), V_n(x_1), U_n(x_2), V_n(x_2))^t \\
A &= \int K^2(u)\, du \begin{bmatrix} g(x_1) & w(x_1) & 0 & 0 \\ w(x_1) & v(x_1) & 0 & 0 \\ 0 & 0 & g(x_2) & w(x_2) \\ 0 & 0 & w(x_2) & v(x_2) \end{bmatrix}.
\end{aligned}
$$

■

soit Z une variable aléatoire naturale en demension 4 centré et de matrice de covariance A. Nous aurons besoin des lemmes suivant :

Lemme 2.3.1 *supposons que K staisfait à i) et ii) précidents et $nh_n^3 \longrightarrow +\infty$. soit $E|Y|^3$ fini et $g\prime, w\prime$ et $v\prime$ éxistent et sont bornés. Si pour $x_1 \neq x_2$ et $g(xi) > 0$ pour $i = 1, 2$ et $r = 1, 2$, alors $Z_n \underset{n\to+\infty}{\longrightarrow} Z$.*

Preuve. *D'aprés le théorème de Gramer-Wold , il suffit de montrer de montrer que $\lambda Z_n^t \xrightarrow{L} \lambda Z^t$ pour tout $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ dans $\mathbb{R}^4$. Pour $s = 1, 2$,sous les hypotèses*

du lemme on a ;

$$\begin{aligned}
E\left[U_{ni}^{2}(x_s)\right] &= g(x_s)\int K^{2}(u)\,du + O(h_n)\\
E\left[V_{ni}^{2}(x_s)\right] &= v(x_s)\int K^{2}(u)\,du + O(h_n)\\
E\left[U_{ni}(x_s)\,V_{ni}(x_s)\right] &= W(x_s)\int K^{2}(u)\,du + O(h_n)\\
E\left[U_{ni}(x_s)\,U_{ni}(x_r)\right] &= O(h_n)\\
E\left[V_{ni}(x_s)\,V_{ni}(x_r)\right] &= O(h_n)\\
E\left[U_{ni}(x_s)\,V_{ni}(x_r)\right] &= O(h_n).
\end{aligned}$$

On montre (1) *et* (4) *pour illustrer la méthode.*

Pour (1) *on a.*

$$\begin{aligned}
E\left[U_{ni}^{2}(x_s)\right] &= E\left[\left(\sqrt{h_n}\left[K\left(\frac{x_s-X_i}{h_n}\right)/h_n - E\left[K\left(\frac{x_s-X_i}{h_n}\right)/h_n\right]\right]\right)^{2}\right]\\
&= \frac{1}{h_n}\left[\int K^{2}\left(\frac{x_s-u}{h_n}\right)g(u)\,du - \left(\int K\left(\frac{x_s-u}{h_n}\right)g(u)\,du\right)^{2}\right]\\
&= \frac{1}{h_n}\left[h_n\int K^{2}(t)\,g(x_s-th_n)\,dt - h_n^{2}\left(\int K(t)\,g(x_s-th_n)\,dt\right)^{2}\right]\\
&= h_n\left[\frac{1}{h_n}\int K^{2}(r)\,g(x_s-th_n)\,dt - \left(\int K(r)\,g(x_s-th_n)\,dt\right)^{2}\right]
\end{aligned}$$

comme g′ et $|uK(u)|$ *sont bornées et que* $\int |u|\,K(u)\,du$ *est finie on a* ■

Preuve.

$$\left|\int K(t)\left[g(x_s-th_n)-g(x_s)\right]dt\right| \le \sup_x |g'(x_s)|\,h_n\int |t|\,K(t)\,dt$$

et

$$\left|\int K^{2}(t)\left[g(x_s-th_n)-g(x_s)\right]dt\right| \le \sup_x |g'(x_s)|\,h_n\int |t|\,K^{2}(t)\,dt$$

donc

$$E\left[U_{ni}^{2}\left(x_{s}\right)\right]=g\left(x_{s}\right)\int k^{2}\left(u\right)du+O\left(h_{n}\right)$$

pour (4), *supposons* $x_2 > x_1$ *et soit* $\delta = x_2 - x_1$ *et* $\delta_n = \frac{\delta}{h_n}$.*alors*

$$\begin{aligned}
E\left[U_{ni}\left(x_{1}\right)U_{ni}\left(x_{2}\right)\right] &= \frac{1}{h_{n}}\int K\left(\frac{x_{1}-u}{h_{n}}\right)K\left(\frac{x_{1}-u}{h_{n}}\right)g\left(u\right)du+O\left(h_{n}\right)\\
&= \int K\left(t\right)K\left(\frac{x_{2}-x_{1}}{h_{n}}+t\right)g\left(x_{1}-th_{n}\right)du+O\left(h_{n}\right)\\
&= \int_{|t|\leq\frac{\delta_{n}}{2}} K\left(t\right)K\left(\delta_{n}+t\right)g\left(x_{1}-th_{n}\right)dt\\
&+ \int_{|t|>\frac{\delta_{n}}{2}} K\left(t\right)K\left(\delta_{n}+t\right)g\left(x_{1}-th_{n}\right)dt+O\left(h_{n}\right)\\
&\leq \sup_{|t|\leq\frac{\delta_{n}}{2}}\left|K\left(\delta_{n}+t\right)\right|\int K\left(z\right)g\left(x_{1}-zh_{n}\right)dz\\
&+ \sup_{|t|>\frac{\delta_{n}}{2}}\left|K\left(t\right)\right|\int K\left(\delta_{n}+z\right)g\left(x_{1}-zh_{n}\right)dz+O\left(h_{n}\right)\\
&\leq \sup_{|t|>\frac{\delta_{n}}{2}}\left|K\left(t\right)\right|.O\left(1\right)+\sup_{|t|>\frac{\delta_{n}}{2}}\left|K\left(r\right)\right|\int K\left(z\right)g\left(x_{1}-zh_{n}\right)dz+O\left(h_{n}\right)\\
&\leq 2\sup\left|K\left(t\right)\right|O\left(1\right)+O\left(h_{n}\right)\\
&\leq 4\delta_{n}^{-1}\times\sup\left|tK\left(t\right)\right|.O\left(1\right)+O\left(h_{n}\right)=O\left(h_{n}\right)
\end{aligned}$$

maintenant soit $\sigma_n^2 = Var\left(\lambda Z_n^t\right)$ *d'aprés* (1) − (6)

$$\sigma_{n}^{2}=\int K^{2}\left(u\right)du\sum_{s=1}^{2}\left(\lambda_{s}^{2}g\left(x_{s}\right)+\mu_{s}^{2}v\left(xs\right)+2\lambda_{s}\mu_{s}w\left(x_{s}\right)\right)+O\left(h_{n}\right)$$

supposons $\rho_{ni}^3 = E\left|\frac{\lambda}{\sqrt{n}}W_{ni}\right|^3$ *et* $\rho_n^3 = \sum_{i=1}^{n}\rho_{ni}^3$

$$\rho_n^3 = \frac{1}{\sqrt{n}} E\left|\lambda W_{n1}\right|^3 \leq \frac{1}{\sqrt{n}} \left|\lambda\right|^3 E\left|W_{n1}\right|^3$$
$$\leq \frac{8}{\sqrt{n}} \left|\lambda\right|^3 \max\left\{E\left|U_{n1}\left(x_s\right)\right|^3, E\left|V_{n1}\left(x_s\right)\right|^3\right\}$$

Puisque g', w', v' *et* k *sont bornés et que* $E\left|Y\right|^3$ *est fini, il s'ensuit en utilisant les mêmes arguments que :*

$$E\left|U_{ni}\left(x_s\right)\right|^3 = O\left(h_n^{-\frac{1}{2}}\right)$$

et

$$E\left|V_{n1}\left(x_s\right)\right|^3 = O\left(h_n^{-\frac{3}{2}}\right), \quad s = 1, 2$$

ainsi

$$\rho_n^3 = O\left(n^{-\frac{1}{2}} h_n^{-\frac{3}{2}}\right) = O\left(\frac{1}{\sqrt{nh_n^3}}\right).$$

Puisque

$$g\left(x\right) v\left(x\right) - w^2\left(x\right) = g^2\left(x\right) V\left[Y/X = x\right]$$

on déduit que A *est définie positive que* $g\left(x_1\right) > g\left(x_2\right) > 0$.*Ainsi pour* $\lambda \neq 0$, $\lim \sigma_n^2 = \lambda A \lambda^t > 0$ *.il s'ensuit que* $\lim_{n\to+\infty} \frac{\rho_n}{\sigma_n} = 0$.*Le théorème de Berry-Essen permet de conclure la preuve du lemme.*

Posons maintenant.

$$Z_n^* = \frac{1}{\sqrt{nh_n}} \left\{\sum_{i=1}^{n} \left(K\left(\frac{x_1 - X_i}{h_n}\right) - g\left(x_1\right)\right); \sum_{i=1}^{n} \left(Y_i K\left(\frac{x_1 - X_i}{h_n}\right) - w\left(x_1\right)\right);\right.$$
$$\left.\sum_{i=1}^{n} \left(K\left(\frac{x_2 - X_i}{h_n}\right) - g\left(x_2\right)\right); \sum_{i=1}^{n} \left(Y_i K\left(\frac{x_2 - X_i}{h_n}\right) - w\left(x_2\right)\right)\right\}^t$$

Lemme 2.3.2 *Supposons que* $\int uK\left(u\right) du = 0$, $\int u^2 K\left(u\right) du < +\infty$ *et* $nh_n^5 \to 0$. *Si g"*

et w" éxistent et sont bornées alors sous les conditions du lumme Précédent : $Z_n^* \xrightarrow{L} Z$.

Preuve. *Soit*

$$B_n = \left\{ \left(g(x_1) - E\left[\frac{1}{h_n} K\left(\frac{x_1 - X_1}{h_n} \right) \right] \right); \left(w(x_1) - E\left[Y_1 \frac{1}{h_n} K\left(\frac{x_1 - X_1}{h_n} \right) \right] \right); \right.$$
$$\left. \left(g(x_2) - E\left[\frac{1}{h_n} K\left(\frac{x_2 - X_1}{h_n} \right) \right] \right); \left(w(x_2) - E\left[Y_1 \frac{1}{h_n} K\left(\frac{x_2 - X_1}{h_n} \right) \right] \right) \right\}^t$$

D'aprés les hypothèses on a

$$\begin{aligned} \left| E\left(\frac{1}{h_n} K\left(\frac{x_i - X_1}{h_n} \right) - g(x_1) \right) \right| &= \left| \int \frac{1}{h_n} K\left(\frac{x_i - u}{h_n} \right) g(u)\, du - g(x_1) \right| \\ &= \left| \int K(t) \left\{ g(x_i - t h_n) - g(x_i)\, dt \right\} \right| \\ &\leq \sup |g"(x)| \frac{h_n^2}{2} \int u^2 K(u)\, du = O\left(h_n^2\right) \qquad i = 1, 2. \end{aligned}$$

De maniére sumilaire

$$\left| E\left(Y_1 \frac{1}{h_n} K\left(\frac{x_i - X_i}{h_n} \right) - w(x_i) \right) \right| = O\left(h_n^2\right)$$

Ains $B_n = O(h_n^2)$. Par suite

$$Z_n - Z_n^* = \sqrt{n h_n} B_n = O\left[\left(n h_n^5 \right)^{\frac{1}{2}} \right] = O(1)$$

puisque $n h_n^5 \to 0$.

Maintenant, on peut donner la preuve du théorème.

Preuve du Théorème*.Soit*

$$H : \mathbb{R}^4 \to \mathbb{R}^2$$

$$(y_1, y_2, y_3, y_4) \to \left(H_1(y_1, y_2, y_3, y_4), H_2(y_1, y_2, y_3, y_4) \right)^t.$$

où

$$H_1(y_1, y_2, y_3, y_4) = \frac{y_2}{y_1} \quad et \quad H_2(y_1, y_2, y_3, y_4) = \frac{y_4}{y_3}$$
$$\theta = H_1(g(x_1), g(x_2), g(x_3), g(x_4)).$$

soit

$$Z_n^* = \sqrt{nh_n}(T_n - \theta)^t$$

où

$$T_n = (T_{n1}, T_{n2}, T_{n3}, T_{n4})$$

avec

$$T_{n1} = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_1 - X_i}{h_n}\right), \quad T_{n2} = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x_1 - X_i}{h_n}\right),$$
$$T_{n3} = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x_2 - X_i}{h_n}\right), \quad T_{n4} = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{x_2 - X_i}{h_n}\right),$$

alors le théorème Mann-Wold avec $\sqrt{nh_n}$ *au lien de* $\sqrt{n}$ *et le lemme précident permet de conclure que*

$$\sqrt{nh_n}(H(T_n) - H(\theta)) \xrightarrow{L} Z^*$$

où $Z^* = N(0, DAD^t)$ *où* D *est la matrice des dérivées de* H *calculées en* θ .

On vérifié que $DAD^t = C$ *et que*

$$(H(T_n) - H(\theta)) = (r_n(x_1) - r(x_1); r_n(x_2) - r(x_2))^t.$$

■

■

2.4 Convergence en moyenne quadratique

Le théorème suivant donne l'expression du biais asymptotique de l'éstimateur à noyau :

Théorème 2.4.1 *On suppose que f et Φ sont deux fois dérivables en x, $f(x) \neq 0$ et Y bornée. K est supposée paire, positive et à support borné. Si $h \to 0$ et $nh \to \infty$ alors*

$$Biais(r_n(x)) = \mathbb{E}[r_n(x) - r(x)] = \frac{h^2}{2}\int u^2 K(u)\, du \frac{\Phi''(x) - r(x) f''(x)}{f(x)} + o\left(h^2\right) + O\left(\frac{1}{nh}\right).$$

Le théorème suivant donne l'expression de la variance asymptotique de l'éstimateur à noyau :

Théorème 2.4.2 *On suppose que f et r sont continues en x, $f(x) \neq 0$ et Y bornée. K est supposée paire, positive et à support borné. On suppose que $\mathbb{E}[Y^2/X = x]$ est continue en x. Si $h \to 0$ et $nh \to \infty$ alors*

$$Var(r_n(x)) = \frac{1}{nh}\frac{\Phi(x)}{f(x)}\int K^2(u)\, du + o\left(\frac{1}{nh}\right),$$

où $\Phi(x) = \mathbb{E}\left[(Y - r(x))^2 / X = x\right]$ est la variance conditionnelle au point x.

Le théorème suivant donne l'expression de l'erreur quadratique moyenne (Mean Square Error) :

Théorème 2.4.3 *Sous les hypothèses jointes des Théorèmes 2.4.1 et 2.4.2, et si $h \to 0$ et $nh \to \infty$ alors*

$$\begin{aligned} EQM(r_n(x)) &= Biais^2(r_n(x)) + Var(r_n(x)) \\ &= \frac{h^4}{4}\left(\int u^2 K(u)\, du \frac{\Phi''(x) - r(x) f''(x)}{f(x)}\right)^2 + \frac{1}{nh}\frac{\Phi(x)}{f(x)}\int K^2(u)\, du \\ &+ o\left(h^4\right) + o\left(\frac{1}{nh}\right) \end{aligned}$$

Nous allons maintenant nous intéresser à des résultats asymptotiques en termes de convergence quadratique.

Théorème 2.4.4 (convergence en MQ sous condition de continuité) *Supposons* $f(x) > 0$, $|Y| < M < \infty$, $h \to 0$, $nh \to \infty$ *et* K *est borné, intégrable, positif, symétrique et à support compact. On a*

$$\mathbb{E}\left[r_n(x) - r(x)\right]^2 \to 0.$$

Nous allons donner des versions uniformes sur un compact de théorème précedent. Pour cela, nous allons nous intéresser aux erreurs quadratiques moyennes intégrées, souvent notées EQMI (ou bien MISE) et définies par :

$$EQMI\left(r_n(x)\right) = \mathbb{E}\left(\int \left[r_n(x) - r(x)\right]^2 w(x)\, dx\right).$$

La fonction w est une fonction de poids vérifiant

$$w \text{ est positive, bornée et à support compact } C.$$

Théorème 2.4.5 (EQMI sous condition de continuité) *Supposons que*

$$\exists \beta > 0 \quad \textit{telle que} \quad \inf_{x \in C} f(x) > \beta$$

et $h \to 0$, $nh \to \infty$ *et* K *est borné, intégrable, positif, symétrique et à support compact. On a*

$$EQMI\left(r_n(x)\right) \to 0.$$

2.5 Choix du noyau et de la largeur de fenêtre

A la fin de ce chapitre nous parlons du choix de noyau et de la de fenêtre de lissage. On suppose que K est une fonction à support compact la plupart du temps. Ceci dit, cela n'est pas une hypothèse nécessaire : K peut être une fonction à support non compact. On peut par exemple prendre la densité d'une loi normale centrée en x.

Pour $s = 1$, on rencontre les noyau suivantes :

$$k(u) = \frac{3}{4}\left(1 - u^2\right)\mathbf{1}_{\{|u|\leq 1\}}$$

et

$$k(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-u^2/2\right) \ , \qquad u \in \mathbb{R}$$

Le première c'est le noyau d'epanechenkov, il n'est pas dérivable aux points $\{1, -1\}$ contrairement au seconde qui est le noyau gaussien.

Par construction, les estimateurs à noyau r_n dépendent de deux paramètres : le noyau K et la largeur de fenêtre h. Pratiquement, on a besoin de décider quels choix effectuer pources deux paramètres.

Il est évident que le paramètre h contrôle la régularité de la fonction estimée, joue un rôle essentiel. Lorsque h est trop grand, le biais est grand et la variance est petite ; lorsque h est trop petit, le biais est petit et la variance est grande. La valeur optimale de h dépend de ces deux quantités biais $B(.)$ et variance $V(.)$ et donné par

$$h_{opt} = \left(\frac{\int V(x)\, w(x)\, dx}{4n\int B^2(x)\, w(x)\, dx}\right)^{\frac{1}{5}} \backsimeq Cn^{-1/5}$$

où

$$V(x) = \frac{\Phi(x)}{f(x)}\int K^2(u)\, du$$

et

$$B(x) = \frac{\int u^2 K(u)\, du}{2}\frac{\Phi"(x) - m(x) f"(x)}{f(x)}$$

Chapitre 3

Application sous R

Dans ce dernier chapitre, nous utilisons le logiciel **R**, pour calculer et représenter graphiquement la fonction de regression et son estimateur en vue de les comparer dans des situations simulées. Il s'agit de l'estimateur proposé par Nadaraya-Watson (1964) et présenté au chapitre 2. Nous donnons des exemples sur cet estimateur qui expriment l'importance de paramétre de lissage h, du noyau K.

Ensuite, nous présentons les résultats obtenus pour les différents jeux de données ainsi que pour les différentes noyaux K (noyau Gaussien : à support non compact et noyau Epanichnekov : à support compact), différents valeurs de h strictement positif (h fixé ou h varié), régression linéaire et non linéaire.

Rappelons qu'on suppose que l'on a observé un échantillon $\{(X_i; Y_i) \ \ ; i = 1, ..., n\}$ et on veut expliquer la variable aléatoire Y_i par X_i. De plus, on suppose que le modèle est donné par l'expression :

$$Y_i = r(X_i) + \varepsilon_i$$

où ε_i est l'aléatoire centré et indépendente de X_i. Aussi la fonction de regression

$$r(x) = E[Y/X = x] = \frac{\int y f(x; y) dy}{f_X(x)} \tag{3.1}$$

où $f_X(x)$ est la densité de la variable X.

Nous avons vu que $r(x)$ est estimé par la quantité :

$$r_n(x) = \frac{\sum_{i=1}^{n} Y_i K_{h_n}(X_i - x)}{\sum_{i=1}^{n} K_{h_n}(X_i - x)} = \frac{\Phi_{n,X}(x)}{f_{n,X}(x)} \tag{3.2}$$

Il dépend de la taille de l'échantillon n ; et aussi du noyau K et de la fenetre h_n qu'il faut choisir pour calculer $r_n(x)$: avec $\Phi_{n,X}(x)$ est l'estimateur naturel de $\Phi_X(x)$:

$$\Phi_{n,X}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} Y_i K_{h_n}(X_i - x)$$

et $f_{n,X}(x)$ l'estimateur à noyau de la densité

$$f_{n,X}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_{h_n}(X_i - x)$$

Dans la suite de ce chapitre, nous supposons que notre modèle à la forme

$$y = r(x) + \varepsilon, \text{ où } \qquad \varepsilon \to \mathcal{N}(0, \sigma^2) \tag{3.3}$$

et nous étudions les deux cas :

• Régression linéaire : $r(x) = 3 + 0.8x + \varepsilon$.

• Régression non linéaire : $r(x) = \sin(x) + \varepsilon$.

on supposons que : X est de loi normale centré de variance $\sigma^2 = 0.2$ et ε un terme d'erreur de loi $N(0; 1)$.

Nous allons donc étudier les cas suivants dans chaque modèle :

-Paramètre de lissage ou fenêtre h fixe, noyau normal (noyau à support non compact) et n varié.

-Paramètre de lissage ou fenêtre h fixe, noyau d'Epanechnikov (noyau à support non compact) et n varié.

-n fixe et fenêtre, h varié (noyau normal).

-n fixe et fenêtre, h varié (noyau d'Epanechnikov).

3.1 Régression linéaire

On veut estimer le modèle linéaire

$$y = 3 + 0.8x + \varepsilon.$$

Dans les résultats graphiques de cette section, on a :

- La droite noire exprime la fonction de régression $r(x)$.

- La droite en rouge exprime la fonction de régression empirique $r_n(x)$

3.1.1 Paramètre de lissage h fixé, n varié

En choisissant le paramètre de lissage $h_n = n^{-\frac{1}{5}}$(fixé) et n varié ($n = 50, 100, 500$)

K à support non compact

Dans ce premier cas, on pose un noyau gaussien $K(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right)$ et on va utiliser le code ci-dessous pour estimer ce modèle, et le resultat graphique obtenu représenté dans la figure [FIG3.1]

Code R :

```
rn(list=ls(all=TRUE)) # Nouveau programme
n=50 # taille de l'\'{e}chantillon (X,Y)
X=rnorm(n,0,2)
E=rnorm(n)
Y=3+.8*X+E # Mod\‘{e}le lin\'{e}aire
# Noyau Normale K(t) c'est une densit\'{e}
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
```

```
# param\'{e}tre de lissage h

h=n^-.2

# 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=numeric(n)

fn=numeric(s)

# Densit\'{e} fn(.)

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\'{e}gression Rn(.)

op=par(mfrow=c(1,3))

plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=2, lwd= 2)

abline(3,.8,lwd= 2)


###Pour n =100###

n=100

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

```
E=rnorm(n)
Y=3+.8*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)
abline(3,.8,lwd= 2)\bigskip
####Pour n =500###
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=3+.8*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(3,.8,lwd= 2)
par(op)
```

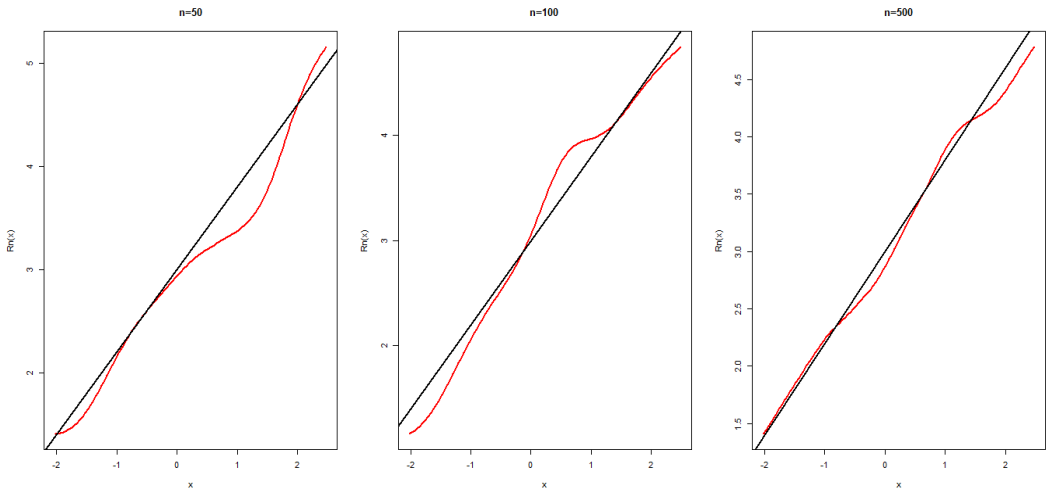

FIG. 3.1 – Régression linéaire : h fixé, n varié et K noyau normal

L'axe des absices représente les valeurs des x et l'axe des coordonnées les valeurs des r_n(et r). Par la comparaison graphique, on remarque que le graphe rouge de r_n est approche beaucoup à la droite noire de r dans le troisième graphe, donc ce graphe exprime la convergence de l'estimateur r_n vers r.

K à support compact

Dans ce second cas, on choisit le noyau d'Epanechnikov :$K(t) = \frac{3}{4}(1-t^2)\,\mathbf{1}_{\{|t|\leq 1\}}$. Ensuite, on modifie seulement cette partie dans le programme **R** précédent :

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

On obtient la figure [FIG 3.2] suivante :

même conclusion de la convergence de l'estimateur (voir la [FIG-3 :1], i.e ; convergence de l'estimateur pour n assez grand).

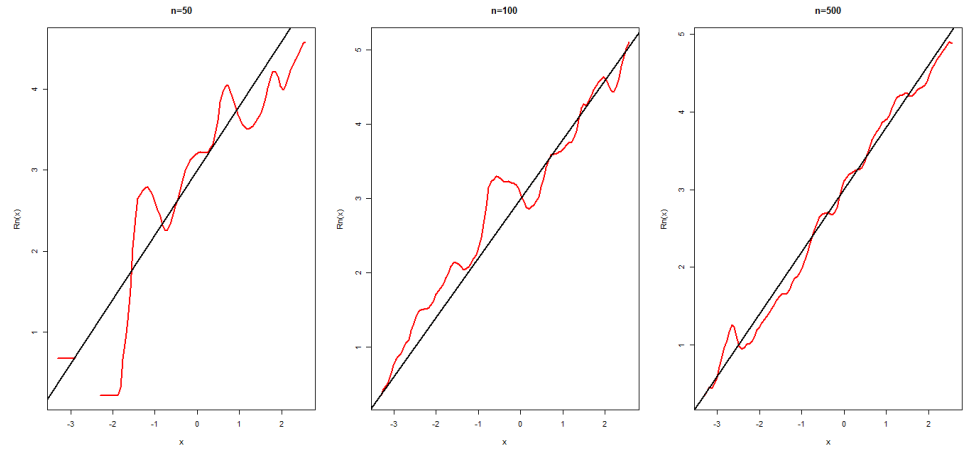

FIG. 3.2 – Régression linéaire : h fixe, n varié et K noyau d'Epanechnikov

3.1.2 Choix graphique du paramètre de lissage

Dans cette section, nous prenons le paramètre de lissage dans l'intervalle $]0;1[$ et avec des tests graphique en va diterminer 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'\'{e}chantillon
X=rnorm(n,0,2)
E=rnorm(n)
Y=3+.8*X+E
# Noyau Normale K(t) c'est une densit\'{e}
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
# param\'{e}tre 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 fenettre 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(3,.8,lwd= 2)
}
par(op)
```

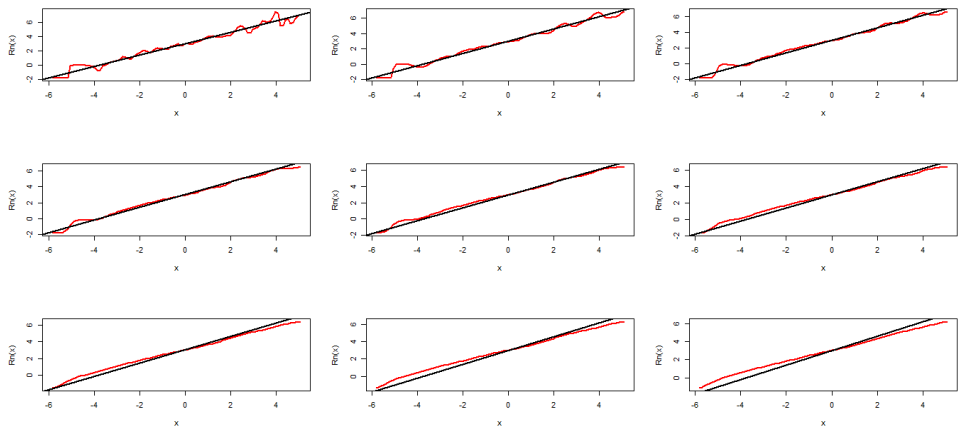

FIG. 3.3 – Régression linéaire avec h varié, n fixé et K noyau gaussion.

Identique aux choix précédents, mais on change le noyau : $K(t) = \frac{3}{4}(1-t^2)\,\mathbf{1}_{\{|t|\leq 1\}}$ (noyau d'Epanechnikov). On obtenu la figure [FIG3.4] qui explique l'estimation obtenue avec les valeurs de h varié de 0.1 à 0.9.

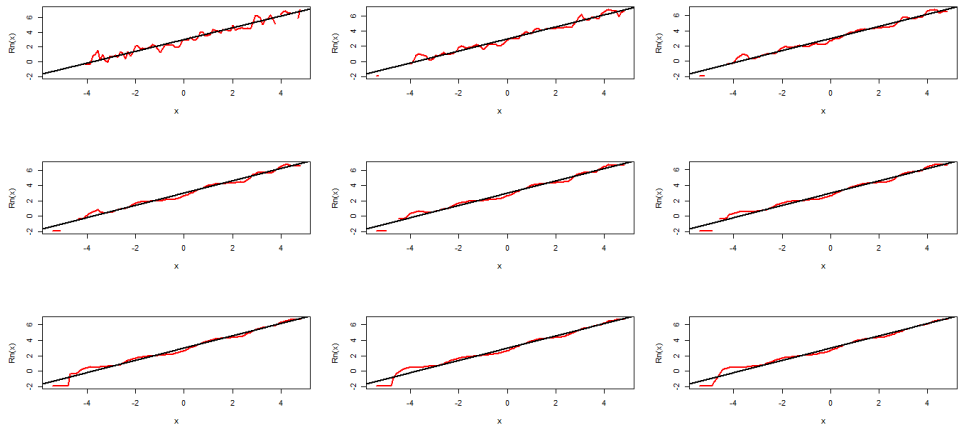

FIG. 3.4 – Régression linéaire avec h varié, n fixé et K d'Epanechnikov

Il est claire que la valeur du h optimale est de $h = 0.9$ (ligne 3, colonne 3).

3.2 Régression non linéaire

Dans cette section, nous allons répeter les mêmes étapes que dans la régression linéaire mais avec un modèle non linéaire :

$$y = \sin x + \varepsilon$$

où ε un terme d'erreur de loi $\mathcal{N}(0;1)$.

Toujours, la ligne noire exprime la fonction de régression théorique $r(x)$ [Eq.(3.1)]et la ligne rouge exprime la fonction de régression empirique $r_n(x)$ donnée par l'équation [Eq.(3.2)].

3.2.1 Paramètre de lissage h fixé, n varié

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 :

```
rn(list=ls(all=TRUE)) # Nouveau programme
n=50 # taille de l'\'{e}chantillon (X,Y)
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\‘{e}le Sinus Non Lin\'{e}aire
# Noyau Normale K(t) c'est une densit\'{e}
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
# param\‘{e}tre de lissage h
h=n^-.2
# 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=numeric(n)
fn=numeric(s)
# Densit\'{e} fn(.)
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\'{e}gression Rn(.)
# Graphes
x11() # nouvelle fenettre 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,sin(x),lwd= 2)
#####Pour n =100 ####
n=100
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\`{e}le Sinus Non Lin\'{e}aire
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,sin(x),lwd= 2)
#####Pour n =500
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\'{e}le Sinus Non Lin\'{e}aire
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=4, lwd= 2)
lines(x,sin(x),lwd= 2)
par(op)
```

```
rn(list=ls(all=TRUE)) # Nouveau programme
n=50 # taille de l'\'{e}chantillon (X,Y)
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\‘{e}le Sinus Non Lin\'{e}aire
# Noyau Normale K(t) c'est une densit\'{e}
K=function(t){(1/sqrt(2*pi))*exp(-0.5*t^2)}
# param\‘{e}tre de lissage h
h=n^-.2
# 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=numeric(n)
fn=numeric(s)
# Densit\'{e} fn(.)
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\'{e}gression Rn(.)
```

```
# Graphes
x11() # nouvelle fenettre graphique
op=par(mfrow=c(1,3))
plot(x,Rn,xlab="x", ylab="Rn(x)", main="n=50",type='l',col=2, lwd= 2)
lines(x,sin(x),lwd= 2)\bigskip
###Pour n =100###
n=100
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\'{e}le Sinus Non Lin\'{e}aire
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,sin(x),lwd= 2)
#####Pour n =500
n=500
X=rnorm(n,0,2)
E=rnorm(n)
Y=sin(X)+E # Mod\'{e}le Sinus Non Lin\'{e}aire
```

```
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,sin(x),lwd= 2)
par(op)
```

On obtenu la figure [FIG-3.5], On remarque la même conclusion pour le cas non linéaire que le cas linéaire (i.e ; convergence de l'estimateur pour n assez grand).

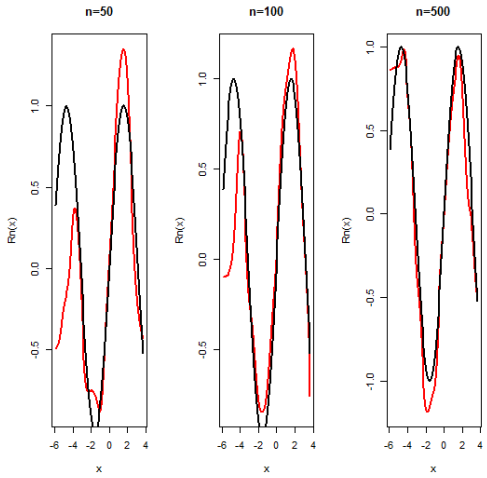

FIG. 3.5 – Régression non linéaire : h fixé, n varié et K noyau normal

Dans ce second cas, on choisit le noyau d'Epanechnikov : $K(t) = \frac{3}{4}(1-t^2)\,\mathbf{1}_{\{|t|\leq 1\}}$. En

suite, on modifie seulement cette partie dans le programme **R** précédent :

Noyau Epanechnikov $K(t)$

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

On obtient la figure [FIG 3.6] ; et on arrive au même conclusion de la convergence de l'estimateur (voir la [FIG3.5], i,e ; convergence de l'estimateur pour n assez grand).

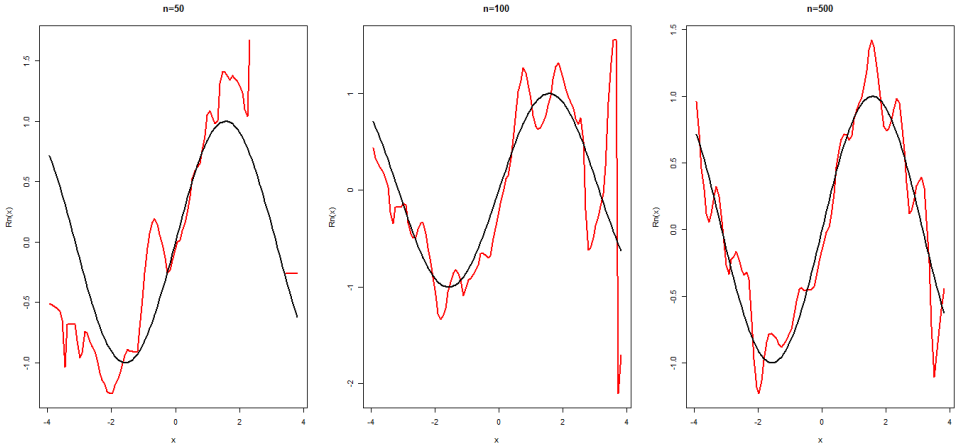

FIG. 3.6 – Régression non linéaire : h fixé, n varié et K noyau d'Epanechnikov

3.2.2 Choix graphique du paramètre de lissage

Dans cette partie, on va prendre le paramètre de lissage dans l'intervalle $[0;1]$ de même façon pour la régression linéaire, et avec des tests graphiques en va diterminer 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 [FIG3.7]. Il est clair que la valeur du h optimale est de $h = 0.5$ (ligne 2, colonne 2).

Code R :

```
n=250#taille de l'\'{e}chantillon
X=rnorm(n,0,2)
```

```
Y=sin(X)+E # Mod\‘{e}le Sinus Non Lin\’{e}aire

# Noyau Normal K(t) c’est une densit\’{e}

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

# param\‘{e}tre 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 fenettre 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,sin(x),lwd= 2)
}
par(op)
```

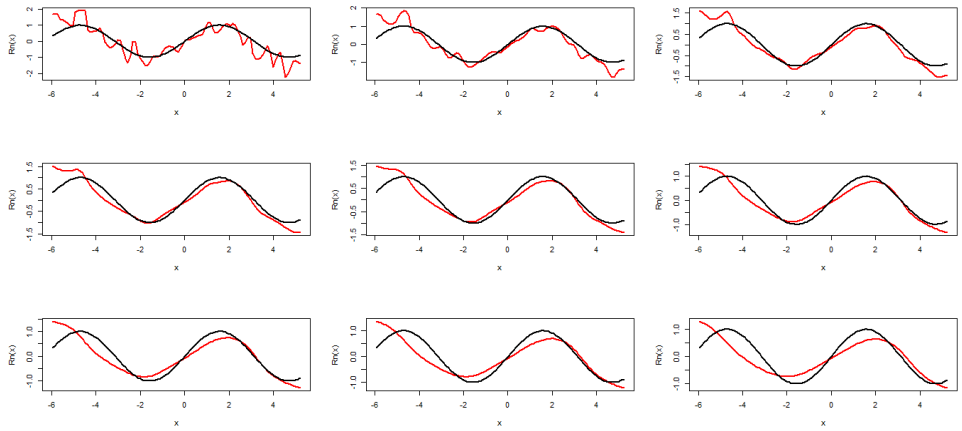

FIG. 3.7 – Régression non linéaire avec h varié, n fixé et K gaussien.

Si nous gardons le même modèle non linéaire $y = sinx + \varepsilon$; mais avec le noyau d'Epanechikov. On note, que la valeur du h optimale est de $h = 0.9$ (ligne 3 ; colonne 3 ; voir la FIG-3.8).

Finallement, ce chapitre montre l'importance de paramètre de lissage h et du noyau K dans l'estimation non paramétrique de la régression linéaire et non linéaire. Mais à noté que le choix de h est plus crucial que le choix de noyau.

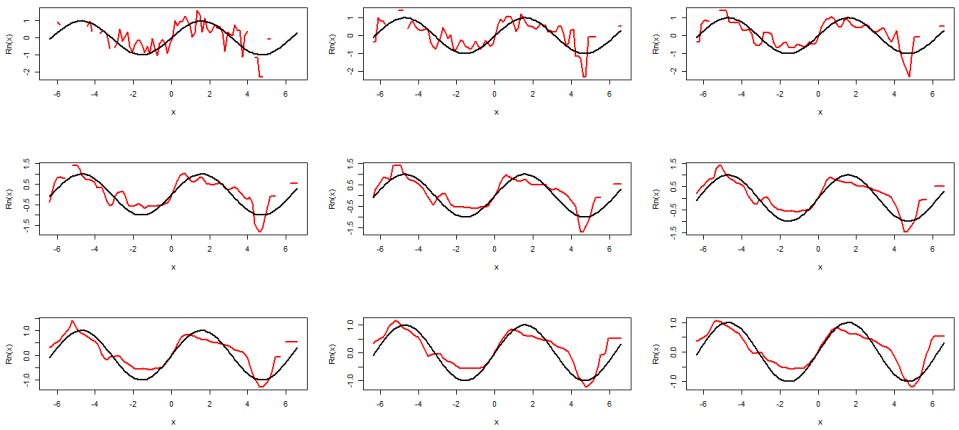

FIG. 3.8 – Régression non linéaire avec h varié, n fixé et K d'Epanechnikov.

Conclusion

Dans ce mémoire, on a présenté la méthode d'estimation à noyau, qui permettant d'effectuer de la régression non paramétrique. Ce travail a montré que la méthode d'estimation de régression non paramétrique est simple et peut être très utile dans plusieurs situations. Par exemple, dans l'analyse des données, lorsque l'on désire comprendre et observer les relations qui existent entre les variables.

Dans la régression non paramétrique, la méthode du noyau joue un grand rôle. Pour que son soit plus utilisée par les praticiens, il est nécessaire que les programmes informatiques permettant d'appliquer ces méthodes soient facilement accessibles et assez simples d'utilisation. Cela favorise aussi les échanges entre statisticiens et utilisateur. L'estimateur à noyau de la régression non paramétrique d'pend de deux paramètres le noyau K et le paramètre de lissage h.

Dans la pratique, on utilisé le logiciel R pour présenté des exemples sur cet estimateur, et à travers les résultats obtenus, nous concluons que : le noyau K est peu influence sur l'estimateur, par contre le paramètre h est un grand influence, et dont le choix est crucial

.

Bibliographie

[1] Bochner, S. (1946) Vector fields and Ricci curvature. *Bulletin of the American Mathematical Society*, 52(9), 776-797.

[2] Bosq D. (1998). Nonparametric Statistics for Stochastic Processes. Springer, New York, Berlin, Heidelberg.

[3] Carbonez A , Györfi, L., van derMeulin, E.C. (1995). Partition-estimate of a regression function under random censoring. Statistics and Decisions, 13 : 21–37..

[4] Collomb,G.(1981). Estimation non paramétrique de la régression : Revue bibliographique,ISI49 :75-93.

[5] Devroye, L., Györfi, L., (1985). Nonparametric density estimation. The L1 view. Wiley, New York.

[6] Carbonez A , Györfi, L., van derMeulin, E.C. (1995). Partition-estimate of a regression function under random censoring. Statistics and Decisions, 13 : 21–37.

[7] Epanechnikov, V.A. (1969) Nonparametric estimation of a multivariate probability density. *Theory Probab. Appl.* 14, 153-158.

[8] Kohler M., Mathé K., Pintér M. (2002). Prediction from randomly right censored data. Journal of Multivariate Analysis, 80 : 73–100..

[9] Ould-Saïd E., Lemdani M. (2006). Asymptotic properties of a nonparametric regression function estimator with randomly truncated data. J. of the Institute of Statistical Mathematics, 58 : 357–378.

[10] Nadaraya, E.A. (1964). On estimating regression. Theory Probab. Appl. 9 : 141–142..

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

[12] Parzen, E. (1962). On estimation of a probability density function and mode. The annals of mathematical statistics, 33(3),1065-1076.

[13] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 832-837.

[14] Rao, P. (1983). Nonparametric Functional Estimation. Academic Press, Inc, London.

[15] Schuster, E.F. (1972), Joint asymptotic distribution of the estimated regression function at a finite number of distinct points. The Annals of Mathematical Statistics,43(1) :84-88.

[16] Silverman B.W. (1986). Density Estimation. London : Chapman and Hall.

[17] Watson, G.S., (1964). Smooth regression analysis. Sankhya Ser. A 26 : 359–372

Annexe A : Rappels

A 1. Convergence en probabilité : On dit que la suite de v.a. (X_n) converge en probabilité vers une v.a. X si, pour tout $\varepsilon > 0$:

$$\mathbb{P}\left(|X_n - X| < \varepsilon\right) \longrightarrow 1 \quad \text{quand } n \longrightarrow \infty, \quad \text{on écrit alors,} \quad X_n \xrightarrow{\mathbb{P}} X.$$

A 2. Convergence en loi : On dit que la suite de v.a. (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 : $X_n \xrightarrow{\mathcal{L}} X$, quand $n \longrightarrow \infty$.

A 3. Convergence en moyenne quadratique : On dit que la suite de va (X_n) converge en moyenne quadratique vers une $v.a.$ X si :

$$\mathbb{E}\left|X_n - X\right|^2 \longrightarrow 0, \quad \text{quand } n \longrightarrow \infty, \quad \text{on écrit alors, } X_n \xrightarrow{mq} X.$$

A 4.Théorème de la limite centrale : Si $X_1, X_2, ..., X_n$ est une suit de $v.a.$ $i.i.d.$ d'espérance $\mu < \infty$ et de variance $\sigma^2 < \infty$, alors

$$\sqrt{n}\left(\bar{X} - \mu\right)/\sigma \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \text{quand} \quad n \longrightarrow \infty, \qquad \text{où } \bar{X} = \tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i.$$

A 5.Théorème (Lois des grands nombres) : Si $(X_1, X_2, ..., X_n)$ est un échantillon provenant d'une v.a. X telle que $\mathbb{E}|X| < \infty$, alors :

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}(X) \quad \text{quand } n \longrightarrow \infty, \qquad \text{(loi faible)}$$

$$\bar{X}_n \xrightarrow{\mathbb{P}s.} \mathbb{E}(X) \quad \text{quand } n \longrightarrow \infty, \qquad \text{(loi forte).}$$

A 6. Définition (Biais d'un estimateur) : Un estimateur $\hat{\theta}_n$ de θ est dite sans biais si pour tout $\theta \in \Theta$ et tout entier positif n : $\mathbb{E}\left(\hat{\theta}_n\right) = \theta$.
De même, $\hat{\theta}_n$ est dite asymptotiquement sans biais si pour tout $\theta \in \Theta$:

$$\mathbb{E}\left(\hat{\theta}_n\right) \longrightarrow \theta \text{ quand } n \longrightarrow \infty.$$

La quantité : $\mathbb{E}\left(\hat{\theta}_n\right) - \theta$, est appelée le biais de l'estimateur $\hat{\theta}_n$.

A 7. Définition les notations $O(h_n)$, $o(h_n)$**,** $O_p(H_n)$ et $o_p(H_n)$ se lit de la façon qui suit :

1) Soit x_n et y_n deux suites de nombres réels. Alors, lorsque $n \to \infty$,

$$i) \quad x_n = O(y_n) \Leftrightarrow \limsup_{n\to\infty} |x_n/y_n| < \infty,$$
$$ii) \quad x_n = o(y_n) \Leftrightarrow \lim_{n\to\infty} |x_n/y_n| = 0.$$

2) Soit X_n et Y_n deux suites aléatoires. Alors, lorsque $n \to \infty$,

$$i) \quad X_n = O_p(Y_n) \Leftrightarrow \forall \beta > 0,\ \exists \delta, N\ : P(|X_n/Y_n| > \beta) < \delta,\ \forall n > N,$$
$$ii) \quad X_n = o_p(Y_n) \Leftrightarrow \forall \beta > 0,\ \lim_{n\to\infty} P(|X_n/Y_n| > \beta) = 0.$$

Annexe B : Logiciel *R*

Le chapitre 3 de ce mémoire comprend des simulations effectuées en utilisant le Logiciel R. Les codes R utilisés sont donnés avec les sorties graphiques correspondantes. L'étude de simulation est basé sur l'observation des résultats d'une estimation de la régression avec la méthode du noyau. L'influence de plusieurs paramètres tels que le nombre de données générées (taille de l'échantillon noté n), la valeur choisie pour la fenêtre h et le noyau K est bient détaller.

 est un système qui est communément appelé logiciel : R Development Core Team (2010). : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http ://www.R-project.org/.

 permet de réaliser des analyses des statistiques. Plus particulièrement, il comporte des moyennes qui rendent possibles la manipulation des données, les calculs et les représentations graphique, R à aussi la possibilité d'exécuter des programmes stokes dans des fichiers textes. En effet R possède :

1. déférentes opérateurs pour calcul sur tableaux, en particulier les matrices,
2. un grand nombre d'outils pour l'analyse des données et les méthodes statistique,
3. des moyennes graphiques pour visualiser les analyses,
4. un langage de programmation simple et performât comportant,
5. Conditions, boucles, moyennes d'entrées sorties...

Annexe C : Abréviations et Notations

Les différentes abrévariations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$(X_1, ..., X_n)$	échantillon de taille n de v.a's.
$h := h_n$	Paramètre de lissage ou Fenêtre
$K(.)$	Noyau
$\mathcal{A}$	Tribu
$\mathbb{E}$	Espace des observations
$\mathcal{B}$	Tribu borélienne
$\mathbb{P}$	une famille de probabilité
Θ	Ensemble des paramètres
L^1	Espace des fonctions intégrables
iid	Indépendantes et identiquement distribu
E	Espérance de probabilité
$Biais$	Biais d'un estimateur
EQM	Erreur Quadratique Moyenne
$\mathbb{R}$	ensemble des nombres réels

Var	Variance d'une stimateur
f_X	Densité de X
F	Fonction de répartition
F^{-1}	Fonction des quantiles
$f_{n,X}$	Estimateur de f
Var	Variance d'une stimateur
F	Fonction de répartition
F^{-1}	Fonction des quantiles
$v.a$	Variable aléatoire
r	Fonction de regression
r_n	Estimateur de r
$\mathbf{1}(.)$	Fonction indécatrice
$\|.\|$	Norme euclidienne
$\xrightarrow{p}$	Convergence en probabilité
$\xrightarrow{\mathcal{L}}$	Convergence en loi.
$\xrightarrow{p.s}$	Convergence presque sûre.