

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université A.Mira de Béjaïa
Faculté des Sciences Exactes
Département de Mathématiques



Mémoire

En vue de l'Obtention du Diplôme de

Master en Mathématiques

Spécialité : Probabilité Statistique et Application

Thème

Sur la normalité asymptotique d'un estimateur à noyau de la fonction de régression

Présenté par :

MOUSSAOUI Leila

Devant le jury composé de :

M^{me} L. BAICHE	Présidente	M.C.B	U. A. Mira. Béjaïa
M^{me} K. TIMERIDJINE	Examinatrice	Professeur	U. A. Mira. Béjaïa
M^{me} H. TABTI	Rapporteur	M.C.B	U. A. Mira. Béjaïa

Année Universitaire : 2022/2023

※ *Remerciements* ※

Avant tout, je remercie le bon DIEU le tout puissant qui m' à donné l'envie, le courage et la force pour mener à terme ce travail.

Au terme de ce travail, j'exprime toutes mes reconnaissances à mon encadreur M^{me} **H.TABTI**, d'avoir fait l'honneur d'assumé l'encadrement de mon travail et qui m'orientés tout au long de cette période.

J'adresse mes profonds remerciements aux membres de jury qui ont bien voulu évaluer et examiner mon travail :

M^{elle} **L.BAICHE** qui ma fait l'honneur de présider mon jury de soutenance.

M^{me} **K. TIMERIDJINE** qui ma fait l'honneur d'examiner mon travail.

Merci également à ma famille et en particulier mes parents, pour l'encouragement constant et leurs aides moraux sans retenue, et pour leurs soutiens inconditionnels tout au long de ces années d'études. Je remercie aussi tous mes enseignants du département de mathématiques qui nous ont accompagné durant notre cursus universitaire.

Ainsi que toute personne qui à contribué, de près ou de loin à l'aboutissement de ce mémoire.

Sans oublier de remercier tous mes amis et tous les étudiants de la promotion PSA (2023).

※ *Dédicaces* ※

Je dédie ce modeste travail :

À mes très chers parents pour leurs soutiens, leurs patiences, leurs encouragements durant mon
parcours.

À mes très chers frères : NACERDINNE et ABDEREZZAK.

À mes très chères sœurs : KATIBA et KAHINA.

À mes petits chers : YANI et RAYAN.

À mes chers grands parents paternels et maternels.

À tous mes chères ami(e)s et camarades, ainsi qu'à toute personne qui m'a encouragé et aidé de près
ou de loin.

Leila Moussaoui

Table des figures

5.1	Estimateur à noyau de la fonction de régression : h fixé, n varié et K noyau normal	57
5.2	Estimateur à noyau de la fonction de régression : h fixé, n varié et K noyau d'Epanechnikov.	58
5.3	Estimateur à noyau de la fonction de régression : h varié, n fixé et K noyau normal.	59
5.4	Estimateur à noyau de la fonction de régression : h varié, n fixé et K noyau d'Epanechnikov.	60
5.5	L'estimateur à noyau de la densité de Z_n et la densité normale centrée réduite.	61

Table des matières

Table des figures	i
Introduction générale	1
1 Préliminaire	3
1.1 Introduction	3
1.2 Estimation de la fonction de répartition	12
1.3 Estimation de la densité	14
1.3.1 Histogramme	14
1.3.2 L'estimateur noyau	16
2 Estimation de la fonction de régression	22
2.1 Introduction	22
2.2 Construction de l'estimateur de la fonction de régression	23
2.3 Les propriétés de l'estimateur de la fonction de régression	25
2.3.1 Étude asymptotique du biais	25
2.3.2 Étude asymptotique de la variance	26
2.4 Consistances	28
2.4.1 Consistance faible	29
2.4.2 Consistance forte	33
2.5 Convergence presque complète	34

3	Choix du paramètre de lissage	44
3.1	Introduction	44
3.2	Étude de critère d'erreur moyenne de $\hat{r}_n(x)$	44
3.3	Méthodes d'optimisation de h	47
3.3.1	Méthode de validation croisée pour la régression	47
4	Normalité asymptotique et intervalles de confiance	50
4.1	Normalité asymptotique	50
4.1.1	Hypothèses	50
4.2	Intervalles de confiance	52
5	Simulation	55
5.1	Présentation des données	55
5.2	Le paramètre de lissage h fixe, et n varié	56
5.3	Choix du paramètre de lissage	59
5.4	Normalité asymptotique	61
	Conclusion générale	63
A	Annexe	64
B	Annexe	66
C	Annexe	69
	Bibliographie	75

Notations

$r(\cdot)$	Fonction de régression.
$\hat{r}_n(\cdot)$	L'estimateur de la fonction de régression.
K	Noyau.
F	Fonction de répartition de la densité f .
f	Densité de probabilité.
\hat{f}_n	L'estimateur de la densité.
h	Paramètre de lissage ou Fenêtre.
h_{opt}	Fenêtre optimale.
$\mathcal{N}(\mu, \sigma^2)$	Distribution gaussienne de moyenne μ et variance σ^2 .
$\mathcal{N}(0, 1)$	Distribution gaussienne centrée réduite.
$\mathcal{L}(\cdot)$	Loi d'une variable aléatoire.
$\mathbf{1}_A$	Fonction indicatrice.
P	Probabilité.
n	Nombre des échantillons.
\mathbb{E}	Espérance mathématique.
Var	Variance.
MSE	L'erreur moyenne quadratique.
$MISE$	L'erreur moyenne quadratique intégrée.
IC	Intervalle de confiance.
$Q.M$	Convergence en moyenne quadratique.
$\xrightarrow{p.s}$	Convergence Presque sûrement.
\xrightarrow{P}	Convergence en Probabilité.
\xrightarrow{L}	Convergence en Loi.
$a.co$	Convergence Presque Complété.

Introduction générale

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

L'objectif de la normalité asymptotique dans l'estimation non paramétrique est de comprendre le comportement asymptotique des estimateurs non paramétriques lorsque la taille de l'échantillon devient de plus en plus grande. La normalité asymptotique est une propriété importante dans l'inférence statistique, car elle permet d'établir des intervalles de confiance et des tests d'hypothèses valides. Historiquement, le principe des régressions non paramétriques remonte au 19^{ème} siècle selon Cleveland et Loader, toutefois les premiers travaux modernes sur ce sujet datent des années 50. La première application que nous verrons concerne l'estimation de la fonction de densité par des méthodes d'opérateurs à noyau (Kernel) avec les travaux fondateurs de [18] et [4]. Ces premiers travaux ont été étendus à la notion de régression Kernel, imparfaitement traduite en français par le terme de régression avec lissage par opérateur à noyau. Dans ce domaine, on identifie deux articles fondateurs publiés la même année : [9].

l'objet central de ce mémoire est d'exposer la normalité asymptotique d'un estimateur à noyau

de la fonction de régression. Nous considérons le problème d'estimation non paramétrique par la méthode du noyau de la fonction de régression. Nous étudions la consistance de l'estimateur de Nadaraya-Watson à travers les propriétés statistiques et asymptotiques de l'estimateur ainsi que sa normalité asymptotique, Afin d'atteindre ce but le présent mémoire est subdivisé en cinq chapitres.

Dans le premier chapitre nous introduisons quelques rappels sur les différents modes de convergences et sur les inégalités exponentielles de type Bernstein qui permettent de contrôler le comportement limite des déviations des estimateurs par rapport à leur espérances. Nous évoquons, dans ce même chapitre, deux méthodes d'estimation non paramétrique de la densité : la méthode d'estimation par histogramme et la méthode d'estimation par noyau (estimateur de Parzen-Rosenblatt) sur laquelle nous nous concentrons le plus et qui peut être considérée comme une extension de l'estimateur par la méthode de l'histogramme.

Nous présentons, au chapitre 2, l'estimation par la méthode de noyau de la fonction de régression (estimateur de Nadaraya-Watson), qui n'est autre qu'une généralisation de la méthode d'estimation par régressogramme. Afin d'évaluer la qualité de l'estimation, nous traitons les propriétés asymptotiques de ces estimateurs, à savoir la convergence en moyenne quadratique et la convergence presque complète.

Dans le chapitre 3, nous présentons deux méthodes de choix du paramètre de lissage par l'MSE, l'MISE et par la méthode de validation croisé.

Dans le chapitre 4 , nous montrons la normalité asymptotique de la fonction de régression.

On termine le mémoire, avec une simulation par le logiciel **R** qui confortent les résultats théoriques obtenus.

Chapitre 1

Préliminaire

1.1 Introduction

L'estimation non paramétrique connaît un grand essor chez de nombreux auteurs et dans différents domaines. Celle-ci ne fait pas d'hypothèse sur la forme de la loi de probabilité qui régit le phénomène sous étude lorsque nos connaissances sur le modèle ne sont pas précises. Dans ce chapitre, nous rappelons quelques méthodes d'estimation non paramétrique, et nous concentrons sur les méthodes d'estimation de la fonction de répartition et de densité, en donnant les propriétés statistiques des estimateurs de chaque méthode d'estimation.

On s'intéresse à l'estimation non paramétrique à noyaux, et surtout à la convergence de ses estimateurs. Le problème de l'estimation non paramétrique consiste à estimer, à partir des observations, une fonction inconnue, élément d'une certaine classe fonctionnelle assez large, notons cette classe fonctionnelle par \mathcal{F} et la fonction inconnue d'échantillon par $f(\cdot)$ et l'échantillon des observations par (x_1, x_2, \dots, x_n) . On peut poser le problème de l'estimation de f comme suit. Soit $(\Omega; \mathcal{A}; P)$ un espace de probabilité sur lequel est définie une famille de variables aléatoires (X_1, X_2, \dots, X_n) indépendantes suivant la même loi.

La question est : que peut-on dire de f à partir de (x_1, x_2, \dots, x_n) avec (x_1, x_2, \dots, x_n) une réalisation de (X_1, X_2, \dots, X_n) ?

définition 1

Le Biais d'un estimateur \hat{f}_n de la fonction f est la quantité définie par :

$$\text{Biais}(\hat{f}_n) = E(\hat{f}_n) - f.$$

1. On dit qu'un estimateur \hat{f}_n de f est sans biais si :

$$\text{Biais}(\hat{f}_n) = 0.$$

c-à-d :

$$E(\hat{f}_n) = f.$$

2. On dit qu'un estimateur \hat{f}_n de f est asymptotiquement sans biais si :

$$\lim_{n \rightarrow +\infty} E(\hat{f}_n) = f.$$

définition 2

L'erreur moyenne quadratique d'un estimateur \hat{f}_n , notée MSE , est définie comme suit :

$$MSE(\hat{f}_n(x)) = E(f(x) - \hat{f}_n(x))^2.$$

Remarque 1

On peut écrire MSE en fonction du biais et de la variance comme suit :

$$MSE(\hat{f}_n(x)) = \text{Biais}(\hat{f}_n(x))^2 + \text{Var}(\hat{f}_n(x)).$$

définition 3

L'erreur moyenne quadratique intégrée d'un estimateur \hat{f}_n , qui est notée $MISE$, est donné par l'expression suivante :

$$MISE(\hat{f}_n(x)) = \int MSE(\hat{f}_n(x))dx = \int \text{Biais}(\hat{f}_n(x))^2 dx + \int \text{Var}(\hat{f}_n(x))dx.$$

définition 4 [20]

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

$$\forall \varepsilon > 0, \quad \sum_{n \in \mathbb{N}} P[|X_n - X| > \varepsilon] < \infty \quad (1.1)$$

définition 5

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

$$\exists \varepsilon_0 > 0, \quad \sum_{n \in \mathbb{N}} P[|X_n - X| > \varepsilon_0 U_n] < \infty \quad (1.2)$$

Proposition 1 [6]

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

Preuve 1

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

$$\sum_{n \in \mathbb{N}} P[|X_n - X| > \varepsilon] < \infty$$

($P[|X_n - X| > \varepsilon]$ est le terme général d'une série convergente). Le lemme de Borel Cantelli implique que :

$$\forall \varepsilon > 0, \quad P \left[\limsup_{n \rightarrow \infty} |X_n - X| > \varepsilon \right] = 0.$$

De plus, $\lim_{n \rightarrow +\infty} X_n(\omega) \neq X(\omega)$, implique l'existence de $\varepsilon > 0$, tel que

$$\limsup_{n \rightarrow \infty} |X_n(\omega) - X(\omega)| > \varepsilon$$

on alors $P \left[\lim_{n \rightarrow +\infty} X_n = X \right] = 1$, c'est à dire $X_n \rightarrow X$, P.S.

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

Proposition 2

Soient l_x et l_y deux nombres réels déterministes, $(X_n)_n$ et $(Y_n)_n$ deux suites de variable aléatoire et (U_n) une suite de nombres réels telle que :

$$\lim_{n \rightarrow \infty} U_n = 0$$

i) Si $\lim_{n \rightarrow \infty} X_n = l_x$ a.co et $\lim_{n \rightarrow \infty} Y_n = l_y$ a.co, alors :

a) $\lim_{n \rightarrow \infty} (X_n + Y_n) = l_x + l_y$ a.co,

b) $\lim_{n \rightarrow \infty} (X_n \times Y_n) = l_x \times l_y$ a.co,

c) $\lim_{n \rightarrow \infty} \frac{1}{X_n} = \frac{1}{l_x}$ a.co si $l_x \neq 0$.

ii) Si $X_n - l_x = O_{a.co}(U_n)$ et $Y_n - l_y = O_{a.co}(U_n)$, on a :

a) $(X_n + Y_n) - l_x - l_y = O_{a.co}(U_n)$,

b) $(X_n \times Y_n) - l_x \times l_y = O_{a.co}(U_n)$,

c) $\frac{1}{X_n} - \frac{1}{l_x} = O_{a.co}(U_n)$ si $l_x \neq 0$.

iii) Si $X_n = O_{a.co}(U_n)$ et $\lim_{n \rightarrow \infty} Y_n = l_y$ a.co, alors on a :

a) $X_n \times Y_n = O_{a.co}(U_n)$

b) $\frac{X_n}{Y_n} = O_{a.co}(U_n)$, si $l_y \neq 0$.

Preuve 2

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

$$\mathbb{P} [|(X_n + Y_n) - (l_x + l_y)| > \varepsilon] \leq \mathbb{P} \left[|X_n - l_x| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[|Y_n - l_y| > \frac{\varepsilon}{2} \right].$$

ii a) Il suffit d'appliquer la même inégalité à $\varepsilon = \varepsilon_0 U_n$.

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

$$X_n \times Y_n = X_n(Y_n - l_y) + X_n \times l_y$$

nous donne :

$$\begin{aligned} \mathbb{P} [|(X_n \times Y_n)| > \varepsilon] &\leq \mathbb{P} \left[|Y_n - l_y| |X_n| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[|l_y X_n| > \frac{\varepsilon}{2} \right] \\ &\leq \mathbb{P} \left[|Y_n - l_y| > \sqrt{\frac{\varepsilon}{2}} \right] + \mathbb{P} \left[|X_n| > \sqrt{\frac{\varepsilon}{2}} \right] + \mathbb{P} \left[|X_n l_y| > \frac{\varepsilon}{2} \right] \\ &\leq \mathbb{P} \left[|Y_n - l_y| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[|X_n| > \frac{\varepsilon}{2} \right] + \mathbb{P} \left[|X_n l_y| > \frac{\varepsilon}{2} \right]. \end{aligned}$$

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

$$\sum_{n \in \mathbb{N}} \mathbb{P} [|X_n Y_n| > \varepsilon] < \infty$$

ii b) Il suffit d'appliquer le résultat précédent à $\varepsilon = \varepsilon_0 U_n$.

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

tel que :

$$\sum_{n \in \mathbb{N}} \mathbb{P} [|Y_n| \leq \delta] < \infty \quad (1.3)$$

de plus on a :

$$\begin{aligned} \mathbb{P} \left[\left| \frac{1}{Y_n} - \frac{1}{l_y} \right| > \varepsilon \right] &= \mathbb{P} [|Y_n - l_y| > \varepsilon |l_y Y_n|] \\ &\leq \mathbb{P} [|Y_n - l_y| > \varepsilon |l_y Y_n|, |Y_n| > \delta] + \mathbb{P} [|Y_n| \leq \delta] \\ &\leq \mathbb{P} [|Y_n - l_y| > \varepsilon \delta |l_y|] + \mathbb{P} [|Y_n| \leq \delta]. \end{aligned}$$

En utilisant la relation (1.3) et la convergence presque complète de Y_n vers l_y , il vient :

$$\forall \varepsilon > 0, \quad \sum_{n \in \mathbb{N}} P \left[\left| \frac{1}{Y_n} - \frac{1}{l_y} \right| \geq \varepsilon \right] < \infty.$$

ii c) On procède de la même manière pour $\varepsilon = \varepsilon_0 U_n$.

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

tel que

$$\sum_{n \in \mathbb{N}} P[|Y_n| > \delta] < \infty$$

La décomposition suivante

$$\begin{aligned} P[|Y_n X_n| > \varepsilon U_n] &= P[|Y_n X_n| > \varepsilon U_n, |Y_n| \leq \delta] + P(|X_n Y_n| > \varepsilon U_n, |Y_n| > \delta) \\ &\leq P(|X_n| > \varepsilon \delta^{-1} U_n) + P(|Y_n| > \delta) \end{aligned}$$

associée à l'inégalité précédente et à l'hypothèse $X_n = O_{a.co}(U_n)$, conduit à

$$X_n Y_n = O_{a.co}(U_n)$$

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

définition 6

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

1. faiblement consistant si

$$\forall x \in \mathbb{R} \quad \hat{\theta}_n(x) \xrightarrow{P} \theta(x), \quad \text{quand } n \rightarrow \infty.$$

2. faiblement et uniformément consistant si

$$\sup_{x \in \mathbb{R}} |\hat{\theta}_n(x) - \theta(x)| \xrightarrow{P} 0, \quad \text{quand } n \rightarrow \infty.$$

3. fortement consistant si

$$\forall x \in \mathbb{R}, \quad \hat{\theta}_n(x) \xrightarrow{P.S} \theta(x), \quad \text{quand } n \rightarrow \infty.$$

4. *fortement et uniformément consistant si*

$$\sup_{x \in \mathbb{R}} |\hat{\theta}_n(x) - \theta(x)| \xrightarrow{P.S.} 0, \quad \text{quand } n \rightarrow \infty.$$

5. *asymptotiquement sans biais si*

$$\forall x \in \mathbb{R}, \quad \lim_{n \rightarrow \infty} \mathbf{E}(\hat{\theta}_n(x)) = \theta(x).$$

6. *asymptotiquement et uniformément sans biais si*

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} \mathbf{E}(\hat{\theta}_n(x) - \theta(x)) = 0$$

7. *convergent en moyenne quadratique si*

$$\lim_{n \rightarrow \infty} \mathbf{E}(\hat{\theta}_n(x) - \theta(x))^2 = 0$$

Nous allons donner deux versions des inégalités exponentielles de type Bernstein qui nous seront utiles pour l'établissement des résultats que nous avons choisi de reprendre. Nous supposons que X_1, X_2, \dots, X_n est une suite de variables aléatoires réelles, indépendantes et centrées.

Corollaire 1 [13]

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

$$\mathbf{E}|X_1^m| \leq C_m a^{2(m-1)}$$

alors on a

$$\forall \varepsilon > 0, \quad P \left[\left| \sum_{i=1}^n X_i \right| > \varepsilon n \right] \leq 2 \exp \left\{ \frac{-\varepsilon^2 n}{2a^2(1 + \varepsilon)} \right\}.$$

b) *Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de n ($X_i = X_{i,n}$).*

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

$$\mathbf{E}|X_1^m| \leq C_m a_n^{2(m-1)}$$

et si $U_n = n^{-1}a_n^2 \log n$, vérifie $\lim_{n \rightarrow \infty} U_n = 0$, alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O_{a.co}(\sqrt{U_n}).$$

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

Corollaire 2

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

$$|X_1| \leq M$$

alors on a :

$$\forall \varepsilon \geq 0, \quad P \left[\left| \sum_{i=1}^n X_i \right| > \varepsilon n \right] \leq 2 \exp \left\{ \frac{-\varepsilon^2 n}{2\sigma^2 \left(1 + \frac{M\varepsilon}{\sigma^2}\right)} \right\}$$

où

$$\sigma^2 = EX_1^2$$

b) Supposons que les $(X_i)_{1 \leq i \leq n}$ dépendent de n et que $\sigma_n^2 = EX_i^2$, s'il existe $M = M_n < \infty$ telle que :

$$|X_1| \leq M$$

et

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

et si $U_n = n^{-1}\sigma_n^2 \log n$, vérifie $\lim_{n \rightarrow \infty} U_n = 0$, alors on a

$$\frac{1}{n} \sum_{i=1}^n X_i = O_{a.co}(\sqrt{U_n}).$$

Les démonstrations de ces corollaires sont basées sur la proposition suivante dont la preuve peut être trouvée dans [13].

Proposition 1 [23]

Si

$$\forall m \geq 2, |\mathbb{E}(X_i^m)| \leq \left(\frac{m!}{2}\right) (a_i)^2 b^{m-2}$$

alors

$$\forall \varepsilon \geq 0, \quad \mathbb{P} \left[\sum_{i=1}^n |X_i| > \varepsilon A_n \right] \leq 2 \exp \left\{ \frac{-\varepsilon^2}{2 \left(1 + \frac{bc}{A_n}\right)} \right\}$$

où $(a_i)_{1 \leq i \leq n}$ sont des réels positifs, $b \in \mathbb{R}^+$ et $A_n^2 = a_1^2 + a_2^2 + a_3^2 + \dots + a_n^2$.

La démonstration du corollaire (1)

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

b) En posant $\varepsilon = \varepsilon_0 \sqrt{U_n}$ dans a) et comme U_n tend vers zéro, pour une certaine constante C' on a :

$$\begin{aligned} \mathbb{P} \left[\frac{1}{n} \left| \sum_{i=1}^n X_i \right| > \varepsilon_0 U_n \right] &\leq 2 \exp \left\{ \frac{-\varepsilon_0^2 \log n}{2 \left(1 + \varepsilon_0 \sqrt{U_n}\right)} \right\} \\ &\leq 2n^{-C'} \varepsilon_0^2 \end{aligned}$$

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

$$\frac{1}{n} \sum_{i=1}^n X_i = O_{a.co}(\sqrt{U_n}).$$

La démonstration du corollaire (2)

a) En appliquant la proposition (1) da $a_i^2 = \sigma^2$, $A_n^2 = n\sigma^2$ et $b = M$ on aboutit à a).

b) Comme $\frac{MU_n}{\sigma^2}$ tend vers zero, il suffit de reprendre le résultat a) pour $\varepsilon = \varepsilon_0 \sqrt{D_n}$, on arrive donc a

l'existence d'une constante C' telle que :

$$\begin{aligned} \mathbb{P} \left[\frac{1}{n} \left| \sum_{i=1}^n X_i \right| > \varepsilon_0 U_n \right] &\leq 2 \exp \left\{ \frac{-\varepsilon_0^2 \log n}{2 \left(1 + \varepsilon_0 \sqrt{\frac{MU_n}{\sigma^2}}\right)} \right\} \\ &\leq 2n^{-C'} \varepsilon_0^2 \end{aligned}$$

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

1.2 Estimation de la fonction de répartition

Soit $X_{1,n} < X_{2,n} < \dots < X_{n,n}$ la statistique d'ordre associée à $X_1; X_2; \dots; X_n$, une suite de variables aléatoires indépendantes identiquement distribuées de même loi que X_1 de fonction de répartition $F(x) = P(X_1 \leq x)$, et de densité de probabilité (par rapport à la mesure de Lebesgue) f .

F et f sont complètement inconnues (cadre non paramétrique). Avant d'étudier la construction et les propriétés de l'estimateur à noyau de f (une fonction mesurable par rapport à la tribu engendrée par (X_1, X_2, \dots, X_n)), nous donnons brièvement un aperçu sur l'estimateur de la fonction de répartition.

Un estimateur sans biais classique de la fonction de répartition $F(x)$ est la fonction de répartition empirique \hat{F}_n définie par :

$$\begin{aligned} \hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_{i,n} \leq x\}} \\ &= \begin{cases} 0 & \text{si } x < X_{1,n} \\ \frac{k}{n} & \text{si } X_{k,n} \leq x < X_{k+1,n} \quad k = 1, \dots, n-1. \\ 1 & \text{si } x \geq X_{n,n}. \end{cases} \end{aligned}$$

Voici quelques propriétés de la \hat{F}_n :

Propriété 1

L'estimateur \hat{F}_n possède les propriétés suivantes :

1. $\forall x \in \mathbb{R}, n\hat{F}_n(x)$ est de loi binomiale $B(n, F(x))$.
2. $\forall x \in \mathbb{R}, \hat{F}_n(x)$ est un estimateur sans biais et convergent quadratique.
3. $\forall x \in \mathbb{R}, \hat{F}_n(x)$ converge presque sûrement vers $F(x)$.

$$\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{\text{P.s.}} F(x).$$

4. $\forall x \in \mathbb{R}, \hat{F}_n(x)$ converge en probabilité vers $F(x)$.

$$\forall x \in \mathbb{R}, \hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{\text{P}} F(x).$$

Preuve 3

1) On a :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} \implies n\hat{F}_n(x) = \sum_{i=1}^n 1_{\{X_i \leq x\}}$$

c'est la somme de n variables aléatoires indépendantes de même loi de Bernoulli de paramètre

$P(X_i \leq x) = F(x)$, alors

$$n\hat{F}_n(x) \rightsquigarrow B(n, F(x))$$

2) Calculons le biais et la variance :

$$\begin{aligned} E(\hat{F}_n(x)) &= E\left(\frac{1}{n}(n\hat{F}_n(x))\right) \\ &= \frac{1}{n}E(n\hat{F}_n(x)) \\ &= \frac{1}{n}nF(x) \\ &= F(x) \end{aligned}$$

Donc $\hat{F}_n(x)$ est un estimateur sans Biais.

$$\begin{aligned} \text{Var}(\hat{F}_n(x)) &= \text{Var}\left(\frac{1}{n}(n\hat{F}_n(x))\right) \\ &= \left(\frac{1}{n}\right)^2 \text{Var}(n\hat{F}_n(x)) \\ &= \frac{1}{n^2}(nF(x)(1-F(x))) \\ &= \frac{F(x)(1-F(x))}{n}, \end{aligned}$$

en remarque que $\lim_{n \rightarrow \infty} \text{Var}(\hat{F}_n(x)) = 0$ et puisque $\hat{F}_n(x)$ est non Biaisé ,

alors

$$\hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{MQ} F(x), \forall x.$$

3) $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}}$ En appliquant la loi forte de grand nombres sur la suite indépendante de variables aléatoires de Bernoulli $1_{\{X_i \leq x\}}$, on obtient :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x\}} \xrightarrow[n \rightarrow +\infty]{P.S} E(1_{\{X_i \leq x\}}) = P(X_i \leq x) = F(x).$$

D'où, la convergence presque sûrement de $\hat{F}_n(x)$ vers $F(x)$.

4) D'après l'inégalité de Bienaymé-Tchebychev.

$$\begin{aligned} \forall \varepsilon > 0, P(|\hat{F}_n(x) - F(x)| \geq \varepsilon) &\leq \frac{\text{Var}(\hat{F}_n(x))}{\varepsilon^2} \\ &\leq \frac{F(x)(1-F(x))}{n\varepsilon^2} \xrightarrow{n \rightarrow +\infty} 0. \end{aligned}$$

Alors on a : $\hat{F}_n(x) \xrightarrow[n \rightarrow +\infty]{P} F(x)$

1.3 Estimation de la densité

Dans cette section, on suppose que la loi de l'échantillon (X_1, \dots, X_n) est continue et on cherche à estimer sa densité f . Par définition f est la dérivée de la fonction de répartition F . Mais la fonction de répartition empirique \hat{F}_n n'est pas dérivable, car c'est une fonction en escalier, on ne peut pas donc utiliser directement les résultats sur la fonction de répartition empirique pour estimer la densité. On peut se demander quelle est l'utilité d'estimer la densité alors que l'on a déjà un très bon estimateur de la fonction de répartition, La principale raison est que la forme d'une densité est beaucoup plus facile à interpréter que celle d'une fonction de répartition.

1.3.1 Histogramme

L'histogramme est l'estimateur de la densité non paramétrique le plus ancien et le plus simple, il est introduit par John Graunt au XVIIe siècle, il est considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini. Cette méthode consiste à estimer f en un point x par la proportion de variables aléatoires (X_1, \dots, X_n) qui se trouve dans un intervalle de longueur h_n , et qui contient x . Un histogramme est défini à partir d'une suite double de valeurs croissantes $\{., \dots, a_{i-1}, \dots, a_{-1}, a_0, a_1, \dots, a_i\}$, consistant un découpage, en intervalles $]a_k, a_{k+1}]$ de la droite réelle. soit n_k la fréquence des observations situées dans l'intervalle $[a_k; a_{k+1}]$ pour un échantillon de taille

n , alors l'histogramme est la fonction constante par morceaux \hat{f}_n définie pour tout $k \in Z$ par :

$$\hat{f}_n(x) = \frac{\frac{n_k}{n}}{(a_{k+1} - a_k)}, \quad \text{pour } x \in]a_k, a_{k+1}] \quad (1.4)$$

qui nous donne la représentation graphique en rectangles. Pour estimer f par un histogramme, on fixe d'abord une borne inférieure de l'échantillon $a_0 < x_1$ et une borne supérieure $a_k > x_n$, puis on partitionne l'intervalle $]a_0, a_k]$ contenant toutes les observations en k classes $]a_{k-1}, a_k]$, dont la largeur de la classe j est $h_j = a_j - a_{j-1}$.

définition 7

L'estimateur de f sur $]a_k, a_{k+1}]$, de type histogramme, noté \hat{f}_{h_n} , est définie par :

$$\begin{aligned} \hat{f}_{h_n}(x) &= \frac{n_k}{nh_n}, \quad x \in]a_k, a_{k+1}] \\ &= \frac{\sum_{k=1}^n \mathbf{I}_{]a_k, a_{k+1}]}(x)}{nh_n} \\ &= \frac{N_k}{nh_n}. \end{aligned}$$

Où N_k est nombre aléatoire de valeurs tombant dans $]a_k, a_{k+1}]$, telle que $N_k \rightsquigarrow B(n, p_k)$, où p_k est la probabilité estimation par $\frac{n_k}{n}$.

Propriété 2

Les propriétés asymptotiques de cet estimateur ont été détaillées dans [3] et [15] Pour tout $x \in [a_k, a_{k+1}]$, l'étude asymptotique de l'erreur quadratique moyenne de \hat{f}_{h_n} et l'erreur quadratique moyenne intégrée de \hat{f}_{h_n} ont été établies par [14]. Nous avons pour tout $x \in [a_k, a_{k+1}]$, pour tout $k \in (1, \dots, p)$ et si la densité f vérifie certaines hypothèses de régularité alors les deux critères d'erreurs cités précédemment sont donnés respectivement par les expressions suivantes :

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

et

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

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

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

Geoffrey (1974)[12] à étudié la convergence uniforme et presque complète de cet estimateur. Nous constatons que l'histogramme à de bonnes propriétés statistiques, par contre il n'est robuste ni pour le choix du paramètre de lissage h , ni pour celui de a_0 . Le deuxième désavantage est sa discontinuité qui ne peut pas s'adapter au cas où f , la densité à estimer, vérifie certaines hypothèses de régularité. Afin de résoudre ce problème, l'estimateur de Parzen Rosenblatt à été introduit, il généralise intuitivement la méthode d'estimation par histogramme, et il est très utilisé en estimation non paramétrique.

1.3.2 L'estimateur noyau

Du fait que

$$f(x) = \frac{F(x+h) - F(x-h)}{2h} \quad \text{pour } h \text{ petit.}$$

Rosenblatt (1956) á donné un estimateur de f , en remplaçant F par son estimateur F_n [18].

D'où

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

Où \hat{f}_n est la fonction de répartition empirique.

définition 8

L'estimateur de Resonblatt, de la fonction de densité f est définie par :

$$\begin{aligned}\hat{f}_n(x) &= \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \\ &= \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}_{\{-1 \leq \frac{x-X_i}{h} < 1\}}.\end{aligned}$$

Remarques 1

Notons que cet estimateur peut encore s'écrire comme

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

en ayant posé

$$K(u) = \frac{1}{2} \mathbf{1}_{\{-1 \leq u < 1\}}.$$

la méthode du noyau consiste à généraliser cette approche à d'autres fonction k .

définition 9

Un estimateur à noyau de la densité f est une fonction \hat{f}_{h_n} définie par :

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right). \quad (1.5)$$

Où $\{h\}_{n \geq 1}$ est une suite de réels positifs appelées paramètres de lissage ou largeurs de la fenêtre qui tend vers 0 quand n tend vers l'infini et K est une densité de probabilité appelée noyau.

Exemple des noyaux

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

-Noyau rectangulaire :

$$K_1(x) = \begin{cases} \frac{1}{2}, & \text{si } |x| \leq 1 \\ 0, & \text{si } |x| > 1 \end{cases}$$

-Noyau triangulaire :

$$K_2(x) = \begin{cases} 1 - |x|, & \text{si } |x| \leq 1; \\ 0, & \text{si } |x| > 1. \end{cases}$$

-Noyau d'Epanechnikov ou parabolique :

$$K_3(x) = \begin{cases} \frac{3}{4}(1-x^2), & \text{si } x \in [-1, 1] \\ 0, & \text{sinon.} \end{cases}$$

-Noyau quadratique :

$$K_4(x) = \begin{cases} \frac{15}{16}(1-x^2)^2, & \text{si } x \in [-1, 1] \\ 0, & \text{sinon} \end{cases}$$

-Noyau cubique :

$$K_5(x) = \begin{cases} \frac{35}{32}(1-x^2)^3, & \text{si } x \in [-1, 1]; \\ 0, & \text{sinon.} \end{cases}$$

-Noyau gaussien :

$$K_6(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad x \in \mathbb{R}.$$

-Noyau sinus :

$$K_7(x) = \begin{cases} \frac{1}{2\pi} \left(\frac{\sin(\frac{\pi x}{2})}{\frac{x}{2}}\right)^2, & \text{si } x \neq 0; \\ \frac{1}{2\pi}, & \text{si } x = 0. \end{cases}$$

-Noyau cosinus :

$$K_8(x) = \begin{cases} \frac{\pi}{4} \cos\left(\frac{\pi x}{2}\right), & \text{si } -1 \leq x \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

-Noyau de Silverman :

$$K_9(x) = \frac{1}{2} \exp(-|x|/\sqrt{2}) \sin\left(|x|/\sqrt{2} + \frac{\pi}{4}\right), \quad x \in \mathbb{R}$$

Propriété 3

1. $\text{Biais}(\hat{f}_n(x)) = \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + O(h^2).$
2. $\text{Var}(\hat{f}_n(x)) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O\left(\frac{1}{nh}\right).$
3. $\text{MSE}(\hat{f}_n(x)) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O\left(\frac{1}{nh}\right) + \frac{h^4}{4} f''(x)^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du\right]^2 + O(h^4).$

Propriété 4 Il y a un certain nombre des conditions qui sont considérées comme usuelles pour les noyaux et qui permettent d'analyser le risque de l'estimateur à noyau K : [10]

$$H_1) K \text{ est bornée ; i.e, } \sup_{u \in \mathbb{R}} |K(u)| \leq M < \infty;$$

$$H_2) \lim_{|u| \rightarrow \infty} |u|K(u) = 0;$$

$$H_3) \int |K(u)| du < \infty;$$

$$H_4) \int_{\mathbb{R}} K(u) du = 1;$$

$$H_5) \forall u \in \mathbb{R}, K(u) = K(-u);$$

$$H_6) \int_{\mathbb{R}} u^2 |K(u)| du \leq \infty;$$

$$H_7) \int_{\mathbb{R}} uK(u) du = 0.$$

Preuve 4

Sous les conditions On a :

1) L'espérance et le Biais

$$\begin{aligned} \mathbb{E}(\hat{f}_n(x)) &= \frac{1}{nh} \mathbb{E} \left[\sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) \right] \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{y-x}{h} \right) f(y) dy \\ &= \int_{-\infty}^{+\infty} K(u) f(x + uh_n) du, \end{aligned} \quad (1.6)$$

En posant $u = \frac{y-x}{h}$. En utilisant le développement de Taylor de f au voisinage de x on obtient :

$$\begin{aligned} f(x + uh) &= f(x) + hf'(x) + \frac{(uh)^2}{2} f''(x) + O(h^2) \\ \mathbb{E}(\hat{f}_n(x)) &= \int_{-\infty}^{+\infty} K(u) [f(x) + uhf'(x) + \frac{1}{2}(uh)^2 f''(x) + O(h^2)] du \\ &= f(x) \int_{-\infty}^{+\infty} K(u) du + hf'(x) \int_{-\infty}^{+\infty} uK(u) du + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du \\ &= f(x) + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + O(h^2), \end{aligned} \quad (1.7)$$

Alors :

$$\mathbb{Biais}(\hat{f}_n(x)) = \mathbb{E}(\hat{f}_h(x)) - f(x) = \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + O(h^2). \quad (1.8)$$

D'où

$$\lim \mathbb{Biais}(\hat{f}_{h_n}(x)) = 0, \text{ lorsque } n \rightarrow \infty \text{ et } h \rightarrow 0.$$

2) La variance

$$\begin{aligned} \text{Var}(\hat{f}_n(x)) &= \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)\right) \\ &= \frac{1}{n} \left\{ \frac{1}{h^2} \mathbb{E}\left[K\left(\frac{X_i - x}{h}\right)^2\right] - \frac{1}{h} \mathbb{E}^2\left[K\left(\frac{X_i - x}{h}\right)\right] \right\} \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} (K(u))^2 f(x+uh) du - \frac{1}{n} \left[\int_{-\infty}^{+\infty} K(u) f(x+uh) \right]^2, \end{aligned} \quad (1.9)$$

Le terme $\frac{1}{n} \left[\int_{-\infty}^{+\infty} K(u) f(x+uh) \right]^2 \rightarrow 0$ lorsque $n \rightarrow +\infty$.

Donc :

$$\text{Var}(\hat{f}_n(x)) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O\left(\frac{1}{nh}\right). \quad (1.10)$$

D'où :

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{f}_n(x)) = 0, \text{ quand } nh \rightarrow \infty.$$

3) L'erreur moyenne quadratique

$$\begin{aligned} \text{MSE}(\hat{f}_n(x)) &= \mathbb{E}(\hat{f}_n(x) - f(x))^2 \\ &= \mathbb{E}(\hat{f}_n(x) - \hat{f}_n(x) + \hat{f}_n(x) - f(x))^2 \\ &= \text{Var}(\hat{f}_n(x)) + \text{Biais}(\hat{f}_n(x))^2 \\ &= \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + O\left(\frac{1}{nh}\right) + \frac{h^4}{4} f''(x)^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + O(h^4). \end{aligned} \quad (1.11)$$

Remarque 2

$$\begin{aligned} \text{MISE}(\hat{f}_n(x)) &= \frac{1}{nh} \int_{-\infty}^{+\infty} \left(f(x) \left[\int_{-\infty}^{+\infty} K^2(u) du \right] + O\left(\frac{1}{nh}\right) \right) dx \\ &\quad + \frac{h^4}{4} \int_{-\infty}^{+\infty} \left(f''(x)^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + O(h^4) \right) dx. \end{aligned} \quad (1.12)$$

$h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$ alors $\text{MISE}(\hat{f}_n(x)) \rightarrow 0$.

On dit que $\hat{f}_n(x)$ converge en probabilité vers $f(x)$.

Chapitre 2

Estimation de la fonction de régression

2.1 Introduction

Le modèle de la régression est l'un des modèles les plus fréquemment rencontrés en statistique paramétrique et non paramétrique. Soit un échantillon aléatoire composé des couples (x_i, y_i) , $i = 1, \dots, n$, où les (x_i) représentent les valeurs observées de la variable explicative X qui est une variable aléatoire réelle de loi donnée par la fonction de répartition F supposée absolument continue dérivable, de dérivée f (qui est donc la densité de X par rapport à la mesure de Lebesgue sur \mathbb{R}), et les y_i représentent celles de la variable dépendante Y (dite aussi la variable d'intérêt).

Alors, le modèle de régression non paramétrique est donné par :

$$y_i = r(x_i) + \varepsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

Où $r(\cdot)$ est une fonction inconnue appelé fonction de régression et les ε_i sont les erreurs aléatoires réelles, appelées "bruit" sous les hypothèses suivantes :

1. $\mathbb{E}(\varepsilon_i) = 0, \forall i$
2. X indépendante de ε_i
3. L'homogénéité des variance, i.e, $\text{Var}(\varepsilon_i) = \sigma^2, \forall i$,
4. La non auto-corrélation des erreurs, i.e, $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$, pour $i \neq j$.

Dans notre travail, nous nous intéressons au modèle où les données $\{X_i, 1 \leq i \leq n\}$ sont strictement

aléatoires et non dégénérées et nous supposons aucune hypothèse paramétrique sur la loi du couple (X, Y) ; d'une manière plus explicite, (X, Y) est à valeurs dans \mathbb{R}^2 , il admet une densité jointe $f(x, y)$ sur \mathbb{R}^2 et une densité marginale $f(x) > 0$ (par rapport à la mesure de Lebesgue sur \mathbb{R}).

La variable Y est supposée intégrable, c'est à dire $E(|Y|) < \infty$, on peut alors définir la fonction de régression $r(x)$ par :

$$r(x) = E(Y/X = x) = \frac{\int_{\mathbb{R}} y f(x, y) dy}{\int_{\mathbb{R}} f(x, y) dy} = \frac{\int_{\mathbb{R}} y f(x, y) dy}{f(x)} \quad (2.2)$$

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

2.2 Construction de l'estimateur de la fonction de régression

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

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i 1_{[t_k, t_{k+1}]}(X_i)}{\sum_{i=1}^n 1_{[t_k, t_{k+1}]}(X_i)} \text{ pour } x \in [t_k, t_{k+1}[\quad (2.3)$$

où $[t_k, t_{k+1}[, k \in N$ est une partition de support de x . Bosq (1969) fut le premier à donner une étude des propriétés statistiques de cet estimateur, il a montré la convergence uniforme presque sûre du regressogramme sur l'intervalle $[a, b]$, quand

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

Sabry (1978) à montré la convergence uniforme presque sûre sur $\left[0, \sqrt{\frac{\log n}{\log \log n}}\right]$ et Lecoultre (1982) à procédé à l'extension de tous ces résultants à R . (pour une étude plus approfondie de cet estimateur on

se réfère au livre de Bosq et Lecoutre (1987)). Pour éviter le problème de positionnement des bornes dans intervalles de la partition, un autre estimateur à été construit comme suit :

$$\forall x, \hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i 1_{[x-h, x+h]}(X_i)}{\sum_{i=1}^n 1_{[x-h, x+h]}(X_i)}, \text{ où } h \in R_*^+.$$

L'inconvénient de ce nouveau estimateur est sa discontinuité. Sa généralisation à été introduite et étudiée par Nadaraya (1964) et Watson (1964). Une façon de construire l'estimateur de Nadaraya-Watson et l'utiliser les estimateurs de $f(x, y)$ et de $f(x)$ dans (1.3) comme suit :

$$r(x) = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f(x)}$$

où $f_{X,Y}(\cdot, \cdot)$ est la densité jointe sur \mathbb{R}^2 et nous désignons par $f(x)$ la densité marginale de X (par rapport à la mesure de Lebesgue sur \mathbb{R}), Un estimateur à noyau de la densité joint $f_{(X,Y)}$, noté $\hat{f}_{(X,Y)}$ peut être obtenu en suivant les mêmes étapes utilisés pour obtenir $f(x)$ estimateur de $f(x)$, et on obtient :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right), \tag{2.4}$$

d'où l'estimateur de Nadaraya-Watson (1964) de la fonction r est :

$$\hat{r}_n(x) = \frac{\frac{1}{nh} \sum_{i=0}^n Y_i K\left(\frac{X-x}{h}\right)}{\frac{1}{nh} \sum_{i=0}^n k\left(\frac{X-x}{h}\right)}. \tag{2.5}$$

définition 10

L'estimateur à noyau, de Nadaraya-Watson (1964) notée $\hat{r}_n(x)$ de la fonction de régression r est donnée par :

$$\hat{r}_n(x) = \begin{cases} \frac{\sum_{i=0}^n Y_i K\left(\frac{x-X_i}{h}\right)}{\sum_{i=0}^n K\left(\frac{x-X_i}{h}\right)} = \frac{\phi_n(x)}{f_n(x)}, & \text{si } 1\left(\sum_{i=0}^n K\left(\frac{x-X_i}{h}\right)\right) \neq 0 \\ \frac{1}{n} \sum_{i=0}^n Y_i, & \text{sinon} \end{cases} \tag{2.6}$$

où

$$\phi_n(x) = \frac{1}{nh} \sum_{i=0}^n Y_i K\left(\frac{x-X_i}{h}\right) \tag{2.7}$$

Remarque 3

L'estimateur \hat{r}_n dépend de deux paramètres le paramètre de lissage h et noyau K

définition 11

Un estimateur $\hat{r}_n(x)$ de $r(x)$ est dit estimateur linéaire de la régression non-paramétrique si :

$$\hat{r}_n(x) = \sum_{i=0}^n Y_i w_i, \quad (2.8)$$

Où la fonction de poids w_i ne dépend pas des observations Y_i .

définition 12

L'estimateur à noyau introduit par Watson Nadaraya, de la fonction de régression évaluée au point x_0 , noté $\hat{r}_n(x)$, est défini par :

$$\hat{r}_n(x) = \sum_i^n Y_i w_i(x), \quad (2.9)$$

avec

$$w_i(x_0) = \frac{K\left(\frac{x_0 - X_i}{h}\right)}{\sum_{i=0}^n K\left(\frac{x_0 - X_i}{h}\right)}. \quad (2.10)$$

Où $K(\cdot)$ désigne une fonction noyau et h un paramètre de lissage ($h > 0$).

2.3 Les propriétés de l'estimateur de la fonction de régression**2.3.1 Étude asymptotique du biais**

Le traitement du biais est purement analytique et repose essentiellement sur le développement de Taylor. Il nous faut supposer certaines conditions de régularités sur les fonctions $g(\cdot)$ et $f(\cdot)$ qui détermineront l'ordre du biais asymptotique en fonction du paramètre de lissage h . [21]

Propriété 5

Supposons que $g(\cdot)$ et $f(\cdot)$ sont de classe $C^2(\mathbb{R})$ et que le noyau K est d'ordre 2, i.e, tel que : les conditions (2,4,6) les propriétés de noyau K sont vérifiées. Nous avons alors, lorsque $h \rightarrow 0$ et $nh \rightarrow \infty$

$$\begin{aligned} \text{Biais}(\hat{r}_n(x)) &= \mathbb{E}[\hat{r}_n(x) - r(x)] \\ &= \frac{h^2}{2} \left(r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right) \int u^2 K(u) du + o(h^2) \end{aligned} \tag{2.11}$$

Démonstration

$$\begin{aligned} \text{Biais}(\hat{r}_n(x)) &= \mathbb{E}(\hat{r}(x)) - r(x) \\ &= \left[\mathbb{E} \left(K \left(\frac{x-X}{h} \right) \right) \right]^{-1} \left[\int \frac{1}{h} K \left(\frac{x-t}{h} \right) g(t) dt - g(x) + g(x) - r(x) \int \frac{1}{h} K \left(\frac{x-t}{h} \right) f(t) dt \right] \\ &\simeq \frac{1}{h^2} [f(x)]^{-1} [g''(x) - r(x)f''(x)] \int_{\mathbb{R}} u^2 K(u) du \\ &= \frac{1}{h^2} \left[r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right] \int_{\mathbb{R}} u^2 K(u) du \end{aligned}$$

2.3.2 Étude asymptotique de la variance

Les hypothèses (1),(3) de noyau sont supposés vérifier le fait que $K(\cdot)$ soit de carré intégrable.

Nous posons, par convenance que :

$$\sigma^2(x) = \text{Var}[Y \setminus_{X=x}] \tag{2.12}$$

Lorsque cette expression est bien définie.

$$\text{var}(\hat{r}_n(x)) = \frac{1}{nh} \left\{ \frac{\sigma^2(x)}{f(x)} \int_{\mathbb{R}} K^2(u) du \right\} (o(1) + 1) \tag{2.13}$$

où

$$f(x) > 0.$$

Démonstration

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

Lemme 1 { Bochner }[10]

soit $K : (\mathbb{R}^m, \mathcal{B}^m) \rightarrow (\mathbb{R}, \mathcal{B})$ une fonction mesurable, où \mathcal{B}^p est la tribu borélienne de \mathbb{R}^p , vérifiant :

$\exists M$ (constante) telle-que,

$$\forall z \in \mathbb{R}^m, |K(z)| \leq M$$

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

Et

$$\|z\|^m |K(z)| \rightarrow 0 : \text{ quand } \|z\| \rightarrow \infty$$

Par ailleurs, soit $g : (\mathbb{R}^m, \beta) \rightarrow (\mathbb{R}, \beta)$ une fonction tq

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

Si g est continue, et si $0 < h \rightarrow 0$, quand $n \rightarrow \infty$ alors :

$$\lim_{n \rightarrow \infty} \frac{1}{h^m} \int \mathbb{R}^m K\left(\frac{z}{h}\right) g(x-z) dz = g(x) \int \mathbb{R}^m K(z) dz$$

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

On a :

$$\begin{aligned} \text{Var}[\hat{f}_n(x)] &= \frac{1}{nh^2} \left\{ \mathbb{E} \left[K^2\left(\frac{x-X}{h}\right) \right] - \mathbb{E} \left[K\left(\frac{x-X}{h}\right) \right]^2 \right\} \\ &= \frac{1}{nh} \left\{ \int_{\mathbb{R}} K^2(u) f(x-hu) du - h \left\{ \int_{\mathbb{R}} K(u) f(x-hu) du \right\}^2 \right\} \\ &= \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2 du (1 + o(1)) \end{aligned}$$

Lorsque $h \rightarrow 0$. Soit la fonction $s(x) = \int_{\mathbb{R}} y^2 f(x,y) dy$. nous avons :

$$\begin{aligned} \text{Var}[\hat{g}_n(x)] &= \frac{1}{nh^2} \left\{ \mathbb{E} \left[Y^2 K^2\left(\frac{x-X}{h}\right) \right] - \mathbb{E} \left[Y K\left(\frac{x-X}{h}\right) \right]^2 \right\} \\ &= \frac{1}{nh} \left\{ \int_{\mathbb{R}} K^2(u) s(x-hu) du - h \left\{ \int_{\mathbb{R}} K(u) r(x-hu) du \right\}^2 \right\} \\ &= \frac{1}{nh} s(x) \int_{\mathbb{R}} K^2 du (1 + o(1)) \end{aligned}$$

De même,

$$\mathbb{E} [\{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\} \{\hat{g}_n(x) - \mathbb{E}[\hat{g}_n(x)]\}] = \frac{1}{nh} g(x) \int_{\mathbb{R}} K(u)^2(u) du (1 + o(1))$$

Soit le vecteur

$$A(x) = \begin{pmatrix} \hat{f}_n(x) \\ \hat{g}_n(x) \end{pmatrix}$$

et $\Sigma[A(x)]$ sa matrice de variance covariance. Il s'ensuit :

$$\Sigma[A_n(x)] = \frac{1}{nh} \begin{pmatrix} f(x) & g(x) \\ g(x) & s(x) \end{pmatrix} \int_{\mathbb{R}} K^2(u) du (1 + o(1))$$

En remarquant que,

$$\begin{pmatrix} -\frac{g(x)}{f(x)^2} & \frac{1}{f(x)} \end{pmatrix} \begin{pmatrix} f(x) & g(x) \\ g(x) & s(x) \end{pmatrix} \begin{pmatrix} -\frac{g(x)}{f(x)^2} \\ \frac{1}{f(x)} \end{pmatrix} = \frac{s(x)}{f(x)^2} - \frac{g(x)^2}{f(x)^2} \quad (2.14)$$

On obtient alors,

$$\begin{aligned} \text{Var}(\hat{r}_n(x)) &= \frac{1}{nh} \left[\frac{s(x)}{f(x) - \frac{r(x)^2}{f(x)^2}} \right] \int_{\mathbb{R}} K^2(u) du (1 + o(1)) \\ &= \frac{1}{nh} \left[\frac{\sigma^2(x)}{\{f(x)\}} \int_{\mathbb{R}} K^2(u) du \right] (1 + o(1)) \end{aligned}$$

Remarque 4

o(1) dans la démonstration précédente vérifiée l'égalité suivante :

$$o(1) = o(h) + o((nh)^{-1})$$

2.4 Consistances

La consistance est la propriété la plus importante d'un estimateur. On distingue une consistance forte et aussi une consistance faible. L'estimateur à noyau de la fonction de régression dépend de deux paramètres : la fenêtre h et le noyau K . Le noyau K établit l'aspect du voisinage de x et h contrôle la taille de ce voisinage, donc h est le paramètre prédominant pour avoir de bonnes propriétés asymptotiques, néanmoins le noyau K ne doit pas être négligé en se basant alors sur l'étude du biais, et de la variance.

2.4.1 Consistance faible

Théorème 1

Supposons que le noyau K vérifie les propriétés (1,6), que $E(Y^2) < 1$ et que $f(x)$ est strictement positive. Si $h \rightarrow 0, nh \rightarrow \infty$ (quand $n \rightarrow \infty$), alors $\hat{r}_n(x)$ est un estimateur consistant de $r(x)$ [10].

Preuve 5

Nous déduisons du théorème de Bochner que, lorsque $h \rightarrow 0$

$$\begin{aligned} \mathbb{E}[\hat{f}_n(x)] &= \mathbb{E}\left[\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right)\right] \\ &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{x-X}{h}\right)\right] \\ &= \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{x-t}{h}\right) f(t) dt \longrightarrow f(x) \int_{\mathbb{R}} K(t) dt \\ &= f(x). \end{aligned}$$

Nous constatons que le biais de l'estimateur converge vers zéro quand la fenêtre tend vers zéro, de plus, on constate qu'il ne dépend pas du nombre de variables, il dépend surtout du noyau K . Donc \hat{f}_n est un estimateur asymptotiquement sans biais. D'autre part, comme les X_i sont indépendantes et identiquement distribuées, il vient que :

$$\begin{aligned} \text{Var}(\hat{f}_n(x)) &= \frac{1}{n} \text{Var}\left[\frac{1}{h} K\left(\frac{x-X}{h}\right)\right] \\ &\leq \frac{1}{n} \mathbb{E}\left[\frac{1}{h} K\left(\frac{x-X}{h}\right)\right]^2 \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} \frac{1}{h} K^2\left(\frac{x-t}{h}\right) f_X(t) dt \end{aligned}$$

D'après le théorème de Bochner :

$$\int_{-\infty}^{+\infty} \frac{1}{h} K^2\left(\frac{x-t}{h}\right) f(t) dt \longrightarrow f(t) \int_{\mathbb{R}} K^2(t) dt < \infty, \text{ quand } n \rightarrow \infty$$

donc

$$\text{Var}(\hat{f}_n(x)) \rightarrow 0 \text{ quand } nh \rightarrow \infty$$

Donc la variance est convergente puisque, $\hat{f}_n(x)$ est un estimateur consistant de $f(x)$. Puisque $\hat{f}_n(x)$ est un estimateur consistant de $f(x)$, il suffit donc de montrer que $\hat{g}_n(x)$ est un estimateur consistant

de

$$g(x) = \int_{-\infty}^{+\infty} y f_{X,Y}(x,y) dy$$

Nous avons

$$\begin{aligned} \mathbb{E}[\hat{g}_n(x)] &= \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{x-x_i}{h} \right) \right] \\ &= \frac{1}{h} \mathbb{E} \left[Y K \left(\frac{x-X}{h} \right) \right] \\ &= \frac{1}{h} \int_{\mathbb{R}} \mathbb{E}(Y/X = x) K \left(\frac{x-t}{h} \right) f(t) dt \\ &= \frac{1}{h} \int_{\mathbb{R}} r(t) K \left(\frac{x-t}{h} \right) f(t) dt \\ &= \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{x-t}{h} \right) r(t) f(t) dt \\ &= \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{x-t}{h} \right) g(t) dt \longrightarrow g(x) \end{aligned}$$

par le théorème de Bochner De plus, si $s(x) = \int y^2 f_{X,Y}(x,y) dy$

$$\text{Var}(\hat{g}_n(x)) = \mathbb{E}[\hat{g}_n(x)]^2 - \mathbb{E}^2[\hat{g}_n(x)] \simeq \frac{1}{h} s(x) \int_{\mathbb{R}} K^2(t) dt \longrightarrow 0 \text{ quand } n \longrightarrow \infty$$

Donc

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{g}_n(x)) = 0$$

d'où $\hat{r}_n(x) = \frac{\hat{g}_n(x)}{\hat{f}_n(x)} \longrightarrow \frac{g(x)}{f(x)} = r(x)$ en probabilité

Absence de biais asymptotique

Théorème 2

Sous les conditions (1,6) et si $f(x)$ est strictement positive, il vient que :

a- Lorsque Y est bornée, $p.s, h \rightarrow 0$ et $nh \rightarrow \infty$ (quand $n \rightarrow \infty$)

alors

$$\mathbb{E}(\hat{r}_n(x)) = \mathbb{E}[\hat{g}_n(x)] / \mathbb{E}[\hat{f}_n(x)] + o((nh)^{-1}) \tag{2.15}$$

b- Lorsque $\mathbb{E}[Y^2] < \infty, h \rightarrow \infty$ et $nh^2 \rightarrow \infty$ (quand $n \rightarrow \infty$) alors :

$$\mathbb{E}(\hat{r}_n(x)) = \mathbb{E}[\hat{g}_n(x)] / \mathbb{E}[\hat{f}_n(x)] + o\left(\left(n^{\frac{1}{2}}h\right)^{-1}\right) \tag{2.16}$$

(a), (b) et le théorème de Bochner impliquent que $\hat{f}_n(x)$ est un estimateur asymptotiquement sans biais de $r(x)$

Preuve 6

En utilisant l'identité suivante

$$\frac{1}{\hat{f}_n(x)} = \frac{1}{\mathbb{E}[\hat{f}_n(x)]} - \frac{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]}{\{\mathbb{E}[\hat{f}_n(x)]\}^2} + \frac{\{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\}^2}{\hat{f}_n(x)\{\mathbb{E}[\hat{f}_n(x)]\}^2}$$

On multiplie par $\hat{g}_n(x)$ des deux côtés, puis on passe à l'espérance

$$\begin{aligned} \mathbb{E}[\hat{f}(x)] &= \frac{\mathbb{E}[\hat{g}_n(x)]}{\mathbb{E}[\hat{f}_n(x)]} - \mathbb{E}\left[\hat{g}_n(x) \left(\frac{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]}{\{\mathbb{E}[\hat{f}_n(x)]\}^2}\right)\right] \\ &\quad + \mathbb{E}\left[\hat{g}_n(x) \frac{\{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\}^2}{\hat{f}_n(x)\{\mathbb{E}[\hat{f}_n(x)]\}^2}\right] \\ &= \frac{\mathbb{E}[\hat{g}_n(x)]}{\mathbb{E}[\hat{f}_n(x)]} - \frac{a_n(x) + b_n(x)}{\{\mathbb{E}[\hat{f}_n(x)]\}^2} \end{aligned}$$

tel que

$$a_n(x) = [\mathbb{E}(\hat{g}_n(x)) - \mathbb{E}(\hat{g}_n(x))] - [\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x))];$$

Et

$$b_n(x) = \mathbb{E}[(\hat{f}_n(x))^{-1} \hat{g}_n(x) (\hat{f}_n(x) - \mathbb{E}(\hat{f}_n(x)))^2]$$

Soit $s(x) = \int_{\mathbb{R}} y^2 f_{X,Y}(x,y) dy$. Nous calculons la variance asymptotique de $\hat{g}_n(x)$ puis $\hat{f}_n(x)$ via le théorème de Bochner

$$\begin{aligned} \text{Var}[\hat{g}_n(x)] &= \frac{1}{nh} \int_{\mathbb{R}} K^2(u) s(x-uh) du - \frac{1}{n} \left\{ \int_{\mathbb{R}} K(u) g(x-uh) du \right\}^2 \\ &\approx \frac{1}{nh} s(x) \int_{\mathbb{R}} K^2 du \\ \text{Var}[\hat{f}_n(x)] &= \frac{1}{nh} \int_{\mathbb{R}} K^2(u) f(x-uh) du - \frac{1}{n} \left\{ \int_{\mathbb{R}} K(u) f(x-uh) du \right\}^2 \\ &\approx \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2(u) du \end{aligned}$$

En utilisant l'inégalité de Cauchy-Schwartz combinée aux formules cidessus, on obtient

$$a_n(x) = o((nh)^{-1}) \tag{2.17}$$

Lorsque la variable Y est bornée, i.e $|y| \leq M$ pour une certaine constante M fixée, nous remarquons que l'estimateur de $[N - W]$ est lui aussi naturellement borné,

$$\frac{\hat{g}_n(x)}{\hat{f}_n(x)} = \frac{\sum_1^n y_i K\left(\frac{x-x_i}{h}\right)}{\sum_1^n K\left(\frac{x-x_i}{h}\right)} \leq \frac{\sum_1^n MK\left(\frac{x-x_i}{h}\right)}{\sum_1^n K\left(\frac{x-x_i}{h}\right)} = M \quad (2.18)$$

Cette dernière inégalité (2.18) permet de borner $b_n(x)$

$$b_n(x) \leq M \times \mathbb{E} [\{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\}^2] \approx \frac{M}{Mh} f(x) \int_{\mathbb{R}} K^2(u) du = o((nh)^{-1})$$

Les relation (2.17) et (2.19) entraînent (2.15). Pour démontrer le cas b); il suffit de remarquer que la relation (2.17) est toujours valable mais la relation (2.18) devient

$$b_n(x) \leq M \times \mathbb{E}[\hat{f}_n(x) - E[\hat{f}_n(x)]]^2 \approx \frac{M}{Mh} f(x) \int_{\mathbb{R}} K^2(u) du = o((nh)^{-1}) \quad (2.19)$$

$$\begin{aligned} |b_n(x)| &\leq \mathbb{E} \left[\max_{1 \leq i \leq n} |Y_i| \{\hat{f}_n(x) - \mathbb{E}[\hat{f}_n(x)]\}^2 \right] \\ &\leq \left\{ \sum_1^n Y_i^2 \right\}^{1/2} \times \{\mathbb{E} [\{\hat{f}_n(x) - E[\hat{f}_n(x)]\}^4]\}^{1/2} \\ &= \sqrt{n} \{\mathbb{E} [Y_i^2]\}^{1/2} \times o((nh)^{-1}) \\ &= o\left(\left(n^{1/2}h\right)^{-1}\right) \end{aligned} \quad (2.20)$$

Les relations (2.20) et (2.17) impliquent (2.16), d'ou le résultat énoncé

Remarque 5

1. Dans l'estimation de $r(x)$ par le noyau rectangulaire , le même poids est accordé à toutes les observations comprise entre $x - h$ et $x + h$. Dans les autres noyaux , le poids d'une observation est d'autant plus fort qu'elle est proche de x .
2. \hat{r}_n a les mêmes propriétés de continuité et de différentiabilité que K . Par exemple, si K est le noyau gaussien \hat{f}_n admet des dérivées de tous ordres
3. Pour choisir quel noyau prendre et surtout choisir le paramètre de lissage h , il faut étudier la qualité de l'estimation de r par \hat{r}_n .

2.4.2 Consistance forte

Parzen a établi la normalité asymptotique, ainsi que la convergence uniforme en probabilité.

Son travail est un outil important et a été largement développé par plusieurs chercheurs (Devroye et Györfi (1985)[16], Silverman (1986)[1]). En (1976) Nadaraya [5] a énoncé le théorème de la consistance forte de l'estimateur.

Théorème 3

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

$$v_n = \sup_{x \in \mathbb{R}} |\hat{f}_n(x, h) - f(x)| \xrightarrow{P} 0, \quad \text{quand } n \rightarrow \infty.$$

si et seulement si la densité f est uniformément continue. La démonstration de ce théorème est basée sur la définition et les lemmes suivants.

définition 13

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

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

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

Lemme 2

$\forall n > 0, \forall \lambda > 0, \exists C > 0$ tel que :

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

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

Lemme 3

Pour toute fonction g , on a :

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{f}_n(x, h) - g(x)| \xrightarrow{P} 0$$

si seulement si

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |E \hat{f}_n(x, h) - g(x)| = 0$$

Lemme 4

Si

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |\hat{f}_n(x, h) - g(x)| = 0 \quad P.S$$

L'étude de la convergence presque complète des estimateurs à noyau introduite par Ferraty et Vieu (2000) [7], s'inspire énormément du cas réel exposé ci dessous. Cette convergence implique la convergence en probabilité et la convergence presque sûre mais n'implique pas la convergence en moyenne quadratique.

2.5 Convergence presque complète

En se basant sur la preuve donnée dans Ferraty et Vieu (2003) [8], nous traitons dans ce paragraphe la convergence presque complète de l'estimateur à noyau de la fonction de régression. Nous gardons quelques conditions précédentes, et nous introduisons des hypothèses de base permettant de donner un théorème général sur la convergence presque complète.

1. $f(x)$ et r sont des fonctions continues au voisinage de la V.a.X, un point fixé de \mathbb{R} .
2. 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.

3. Le paramètre de lissage h est tel que

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

Théorème 4 [11]

Sous les hypothèses (1),(2),(3) de plus (H_4) et le fait que K est borné, intégrable et à support compact.

$$\lim_{n \rightarrow \infty} \hat{r}_n(x) = r(x) \cdot p \cdot co$$

Démonstration

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

$$\begin{aligned} \hat{r}_n(x, h) - r(x) &= \frac{1}{\hat{f}_n(x, h)} [(\phi_n(x, h) - E\phi_n(x, h)) + (E\phi_n(x, h) - \phi(x))] \\ &+ \left[(f(x) - E\hat{f}_n(x, h)) + (E\hat{f}_n(x, h) - \hat{f}_n(x, h)) \right] \frac{r(x)}{\hat{f}_n(x, h)} \end{aligned}$$

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

Lemme 5

Sous les hypothèses (H_4) et (H_5) on a :

$$\lim_{n \rightarrow \infty} E\phi_n(x, h) = \phi(x)$$

Preuve 7

Nous avons :

$$\begin{aligned} E\phi_n(x, h) &= E \left[\frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{x - X_i}{h} \right) \right] \\ &= \frac{1}{h} E \left[YK \left(\frac{x - X}{h} \right) \right] \end{aligned}$$

Le conditionnement par rapport à $X = x$ nous donne :

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

où

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

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

$$Eg(x) = \int \phi(x-uh) K(u) du$$

Comme K est à support compact, la continuité uniforme de ϕ et la condition (H_4) , nous donnent :

$$\lim_{n \rightarrow \infty} E\phi_n(x, h) = \phi(x)$$

Lemme 6

Sous les hypothèses (H_5) et (H_4) on a :

$$\lim_{n \rightarrow \infty} (E\phi_n(x, h) - \phi(x, h)) = 0 \quad a.co$$

Preuve 8

On a :

$$\phi_n(x, h) - E\phi_n(x, h) = \frac{1}{n} \sum_{i=1}^n Z_i$$

où

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

De plus, la condition (H_5) nous donnent :

$$|Z_i| \leq \frac{C}{h}$$

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

$$EZ_i^2 = \text{var}\left(\frac{1}{h} Y_i K\left(\frac{x-X_i}{h}\right)\right) \leq E_i^2$$

où

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

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

$$\begin{aligned} E_i^2 &= h^{-2} E\left[\Phi(X) K^2\left(\frac{x-X_i}{h}\right)\right] \\ &= h^{-2} \int \Phi(t) K^2\left(\frac{x-t}{h}\right) f(t) dt \end{aligned}$$

où

$$\Phi(x) = E[Y^2/X = x]$$

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

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

la continuité de f sur le support compact K , la condition (H_4) impliquent :

$$E\Gamma_i^2 \leq \frac{C}{h}$$

Les conditions du corollaire (1) étant satisfaites, nous déduisons :

$$\frac{1}{n} \sum_i Z_i = O_{a.co} \left(\sqrt{\frac{\log n}{nh}} \right).$$

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

$$\lim_{n \rightarrow \infty} E[\hat{f}_n(x, h)] = f(x),$$

$$\lim_{n \rightarrow \infty} E[\hat{f}_n(x, h)] - f(x) = 0 \quad a.co.$$

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

Lemme 7

Sous les conditions (H_4) et (H_5) on a

$$\exists \delta > 0, \quad \sum_{n=1} P[\hat{f}_n(x, h) \leq \delta] < \infty. \tag{2.21}$$

Théorème 5

entraîne la convergence presque complète de $\hat{f}_n(x, h)$ vers $f(x)$, c'est à dire

$$\forall \varepsilon, \quad \sum_{n=1} P[|\hat{f}_n(x, h) - f(x)| > \varepsilon] < \infty$$

De plus on a

$$|\hat{f}_n(x, h)| \leq \frac{f(x)}{2} \implies |\hat{f}_n(x, h) - f(x)| > \frac{f(x)}{2}.$$

$$\mathbb{P} \left[\left| \hat{f}_n(x, h) - \frac{f(x)}{2} \right| \leq \frac{f(x)}{2} \right] \leq \mathbb{P} \left[|\hat{f}_n(x, h) - f(x)| > \frac{f(x)}{2} \right].$$

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

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

$$\mathbf{r \text{ et } f \text{ sont } k \text{ fois continument dérivables autour de } x.} \tag{2.22}$$

Théorème 6

Considérons le modèle (2.22) avec $k > 0$ et supposons que les hypothèse (2.12), (2.11), (H_5) soient réalisées, alors on a

$$|\hat{r}_n(x, h) - r(x)| = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right) \quad a.co.$$

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

Lemme 8

Sous les hypothèses (2.22) et (H_5)

$$\mathbb{E}\phi_n(x, h) - \phi(x) = O(h^k) \tag{2.23}$$

Preuve 9

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

$$\mathbb{E}\phi_n(x, h) = \int \phi(x - zh) \mathbf{K}(z) dz.$$

Le modèle (2.22), nous permet de développer ϕ au voisinage de x , ceci nous permet d'écrire :

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

où θ_z est entre x et $x - zh$. La condition (6) sur K , implique :

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

La convergence uniforme de $\phi^k(\theta_z)$ vers $\phi^k(x)$ (assurée par le modèle (2.22)) et la condition (H_5) , nous donnent

$$E\phi_n(x, h) - \phi(x) = O(h^k)$$

Les lemmes (2), (3) et (4) nous assurent que :

$$E\hat{f}_n(x, h) - \hat{f}_n(x, h) = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad a.co$$

$$Eg_n(x, h) - g_n(x, h) = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad a.co$$

et

$$\exists \delta > 0 \quad \sum_{n=1} P[\hat{f}_n(x, h) \leq \delta] < \infty.$$

En combinant tous les résultats cités précédemment, on arrive au résultat cherché. Nous essayons maintenant d'établir une vitesse convergence presque complète uniforme. Il suffit d'une part de considérer un compact S de \mathbb{R} tel que la condition (2.22) soit remplacée par le modèle suivant

$$r \text{ et } f \text{ sont } k \text{ continument dérivables sur } S. \tag{2.24}$$

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

$$\inf_{x \in S} f(x) > \theta \tag{2.25}$$

Théorème 7

Soient les modèles (2.24), (2.25) avec $k > 0$ et les conditions (H_5) et (2.12), on a

$$\sup_{x \in S} |\hat{r}_n(x, h) - r(x)| = O(h^k) + O\left(\sqrt{\frac{\log n}{nh}}\right), \quad a.co$$

Preuve 10

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

$$\begin{aligned} \sup_{x \in S} |\hat{f}_n(x, h) - r(x)| &\leq \frac{\sup_{x \in S} |\phi_n(x, h) - \phi(x)|}{\inf_{x \in S} |\hat{f}_n(x, h)|} + \sup_{x \in S} |f(x) - \hat{f}_n(x, h)| \frac{\sup_{x \in S} |r(x)|}{\inf_{x \in S} |\hat{f}_n(x, h)|} \\ &\leq \frac{\sup_{x \in S} |\phi_n(x, h) - E(\phi_n(x, h))|}{\inf_{x \in S} |\hat{f}_n(x, h)|} + \frac{\sup_{x \in S} |E(\phi_n(x, h)) - \phi(x)|}{\inf_{x \in S} |\hat{f}_n(x, h)|} \\ &\quad + \left\{ \sup_{x \in S} |f(x) - E(\hat{f}_n(x, h))| + \sup_{x \in S} |E(\hat{f}_n(x, h)) - \hat{f}_n(x, h)| \right\} \frac{\sup_{x \in S} |r(x)|}{\inf_{x \in S} |\hat{f}_n(x, h)|} \end{aligned}$$

Les approximations en $O(h^k)$ traitées précédemment peuvent se généraliser via (H_5) et (2.22) comme suit

$$\sup_{x \in S} |Eg_n(x, h) - g(x)| = O(h^k)$$

et

$$\sup_{x \in S} |E\hat{f}_n(x, h) - f(x)| = O(h^k)$$

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

Lemme 9

Sous les hypothèses (1.8), (2.24), (1.19) et (2.20) on a

$$\sup_{x \in S} |E\phi_n(x, h) - \phi_n(x, h)| = O\left(\sqrt{\frac{\log n}{nh}}\right), \quad a.co$$

Preuve 11

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

$$S \subset \bigcup_{k=1}^{z_n} S_k$$

où

$$S_k =]t_k - l_n, t_k + l_n[\quad \text{et} \quad l_n = n^{-\beta}$$

Posons

$$t_x = \arg \min_{t \in \{t_1, t_2, \dots, t_n\}} |x - t|$$

avec

$$l_n = n^{-2\xi}, l_n = Cz_n^{-1}$$

On a

$$\sup_{y \in S} |E\phi_n(x, h) - \phi_n(x, h)| \leq A_1 + A_2 + A_3$$

où

$$A_1 = \sup_{y \in S} |\phi_n(x, h) - \phi_n(t_x, h)|,$$

$$A_2 = \sup_{y \in S} |\phi_n(t_x, h) - E\phi_n(t_x, h)|,$$

$$A_3 = \sup_{y \in S} |E(\phi_n(t_x, h)) - E\phi_n(x, h)|.$$

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

$$\begin{aligned} |\phi_n(t_x, h) - \phi_n(x, h)| &= \frac{1}{nh} \sum_{i=1}^n |Y_i| \left| \left[K\left(\frac{t_x - X_i}{h}\right) - K\left(\frac{x - X_i}{h}\right) \right] \right| \\ &\leq \frac{C}{h} \frac{|t_x - x|}{h^2} \\ &= \frac{Cl_n}{h^2} \end{aligned}$$

d'où :

$$A_1 = o\left(\frac{\log n}{nh}\right).$$

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

$$A_3 = o\left(\frac{\log n}{nh}\right).$$

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

$$\begin{aligned} P \left[\sup_{y \in S} |\phi_n(t_x, h) - E\phi_n(t_x, h)| > \varepsilon \right] &= P \left[\max_{\{j=1, \dots, z_n\}} |\phi_n(t_j, h) - E\phi_n(t_j, h)| > \varepsilon \right] \\ &\leq z_n P \left[|\phi_n(t_j, h) - E\phi_n(t_j, h)| > \varepsilon \right] \\ &\leq z_n P \left[\frac{1}{nh} \sum_{i=1}^n |U_i - EU_i| > \varepsilon \right]. \end{aligned}$$

où

$$U_i = Y_i K\left(\frac{X_i - t_j}{h}\right)$$

Il suffit de trouver des majorants pour U_i et EU_i^2 , pour pouvoir appliquer le corollaire (2). D'après la démonstration du lemme (5), on a .

$$|U_i| \leq \frac{C}{h} \quad \text{et} \quad EU_i^2 \leq \frac{C}{h}$$

Maintenant nous sommes en mesure d'appliquer le corollaire (2) :

$$\forall \varepsilon > 0, \quad \mathbf{P} \left[\sup_{y \in S} |\phi_n(t_x, h) - \mathbf{E}\phi_n(t_x, h)| > \varepsilon \right] \leq n^{2\xi} \exp(-Cn\varepsilon^2h)$$

Comme

$$\lim_{n \rightarrow \infty} \sqrt{\frac{\log(n)}{nh}} = 0$$

en posant

$$\varepsilon = \varepsilon_0 \sqrt{\frac{\log(n)}{nh}}$$

on obtient alors, pour un choix adéquat de ε_0

$$\forall \varepsilon > 0, \quad \sum_n \mathbf{P} \left[\sup_{x \in S} |\phi_n(t_x, h) - \mathbf{E}\phi_n(t_x, h)| > \varepsilon_0 \sqrt{\frac{\log(n)}{nh}} \right] < \infty.$$

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

Lemme 10

Sous les conditions (H_5) , (2.24) et (2.25) on a :

$$\exists \delta > 0, \quad \sum_{n=1}^{\infty} \mathbf{P} \left[\inf_{x \in S} |\hat{f}_n(x, h)| \leq \delta \right] < \infty.$$

Preuve 12

On a :

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{\infty} \mathbf{P} \left[\sup_{x \in S} |\hat{f}_n(x, h) - f(x)| > \varepsilon \right] < \infty.$$

Si

$$\inf_{x \in S} |\hat{f}_n(x, h)| \leq \frac{\theta}{2}$$

alors

$$\sup_{x \in S} |\hat{f}_n(x, h) - f(x)| > \frac{\theta}{2}$$

où θ vérifie (2.25). Pour $\varepsilon = \frac{\theta}{2}$, la convergence uniforme complète $\hat{f}_n(x, h)$ vers $f(x)$ nous donne

$$\sum_{n=1}^{\infty} \mathbf{P} \left[\inf_{x \in S} |\hat{f}_n(x, h)| \leq \frac{\theta}{2} \right] < +\infty.$$

Chapitre 3

Choix du paramètre de lissage

3.1 Introduction

Par construction l'estimateur à noyau dépend de deux paramètres : Le noyau K et le paramètre de lissage h . Dans la pratique, il faut décider du choix à faire pour ces deux paramètres. Comme d'habitude le noyau n'a pas une grande influence sur l'estimateur, par contre le choix de h est essentiel. Il est utile de rappeler les hypothèses, précédemment introduites, et qui sont nécessaires pour la suite de notre travail.

- $f(\cdot)$ et $r(\cdot)$ sont de classes $C^2(\mathbb{R})$,
- $\lim_{n \rightarrow \infty} h = 0$ quand $n \rightarrow \infty$,
- $|Y|$ est bornée.
- K satisfait les propriétés suivantes : (H_1) , (H_4) et (H_6) , (H_7) .

3.2 Étude de critère d'erreur moyenne de $\hat{r}_n(x)$

L'erreur quadratique moyenne MSE (mean square error) est une mesure permettant d'évaluer la similarité de $\hat{r}_n(x)$ par rapport à la fonction de régression inconnue $r(x)$, au point x donné de $\mathbb{R}[11]$.

Notre but est de minimiser

$$MSE(\hat{r}_n(x)) = \mathbb{E}(\hat{r}_n(x) - r(x))^2$$

Le développement de cette expression faite précédemment, nous donne

$$\text{MSE}(\hat{r}_n(x)) = \mathbb{B}iais(\hat{r}_n(x))^2 + \text{Var}(\hat{r}_n(x))$$

Nous constatons d'une part que les expressions du biais de $\hat{r}_n(x)$ et de la variance de $\hat{r}_n(x)$ données dans le chapitre (2), permettent de conclure qu'une grande valeur de h donne une augmentation du biais et une diminution de la variance (estimation fortement biaisée) et qu'un faible paramètre h , donne une diminution du biais et une augmentation de la variance (phénomène de sous lissage).

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

$$\text{MSE}(\hat{r}_n(x, h)) = \frac{h^4}{4} \left\{ \left[r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right] [u^2 \mathbf{K}(u)] + o(1) \right\}^2 + \frac{1}{nh} \left\{ \frac{\sigma^2(x)}{f(x)} [\mathbf{K}^2] \right\} (1 + o(1)),$$

où $[u^p \mathbf{K}^q(u)] = \int t^p \mathbf{K}^q(t) dt$.

Pour trouver donc un compromis entre le biais et la variance nous minimisons par rapport à h l'expression de l'erreur quadratique moyenne asymptotique AMSE (asymptotique mean squared error) donnée par

$$\text{AMSE}[\hat{r}_n(x, h)] = \frac{h^4}{4} \left\{ r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right\}^2 [u^2 \mathbf{K}]^2 + \frac{1}{nh} \times \frac{\sigma^2(x)}{f(x)} [\mathbf{K}^2(u)].$$

Comme AMSE est une fonction convexe. La fenêtre $h_{opt}^{\text{MSE}}(\hat{r}_n(x, h)) = \arg \min_h [\text{AMSE}(\hat{r}_n(x, h))]$ est solution de l'équation suivante

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

lorsque

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

d'où

$$h_{opt}^{\text{MSE}}(\hat{r}_n(x, h)) = n^{-\frac{1}{5}} \left\{ \frac{\frac{\sigma^2(x)}{f(x)} [\mathbf{K}^2(x)]}{\left\{ \left[r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right] [t^2 \mathbf{K}] \right\}^2} \right\}^{\frac{1}{5}}.$$

Démonstration.

On pose :

$$B^2 = \left[r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right]^2 [u^2 K(u)]^2, \quad \text{et} \quad V = \frac{\sigma^2(x)}{f(x)} [K^2(u)].$$

Donc :

$$AMSE = \frac{1}{4} B^2 h^4 + \frac{V}{nh}.$$

On calcule la dérivée de l'AMSE :

$$\frac{\partial}{\partial h} \left[\frac{1}{4} B^2 h^4 + \frac{V}{nh} \right] = B^2 h^3 - \frac{V}{nh^2}.$$

Il est clair que la deuxième dérivée de l'AMSE est positive, alors on a une certitude qu'elle admet un minimum :

$$\begin{aligned} B^2 h^3 - \frac{V}{nh^2} &= B^2 h^5 - \frac{V}{n} \\ &= 0, \end{aligned}$$

on trouve :

$$\begin{aligned} h_{opt} &= \left(\frac{V}{nB^2} \right)^{\frac{1}{5}} \\ &= \left(\frac{\frac{\sigma^2(x)}{f(x)} [K^2(u)]}{n \left\{ r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right\}^2 [u^2 K(u)]^2} \right)^{\frac{1}{5}} \end{aligned}$$

On s'intéresse maintenant à l'approche globale pour la sélection du paramètre hn , pour cela on introduit le critère d'erreur quadratique intégrée moyenne ou **MISE** (mean integrated squared error) de $\hat{r}_n(x, h)$.

$$MISE[\hat{r}_n(x, h)] = E \left[\int_{\mathbb{R}} (\hat{r}_n(x, h) - r(x))^2 dx \right]$$

En appliquant le théorème de Fubini, On a

$$MISE[\hat{r}_n(x, h)] = \left[\int_{\mathbb{R}} E (\hat{r}_n(x, h) - r(x))^2 dx \right],$$

On a

$$AMISE[\hat{r}_n(x, h)] = \frac{h^4}{4} \int \left\{ r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right\}^2 dx [u^2 K] + \frac{1}{nh} \int \frac{\sigma(x)^2}{f(x)} dx [K^2(u)].$$

La fenêtre $h_{opt}^{AMISE}(\hat{r}_n(x, h))$ minimisant l'AMISE du critère global est :

$$h_{opt}^{AMISE}(\hat{r}_n(x, h)) = n^{-\frac{1}{5}} \left\{ \frac{\int \frac{\sigma_f^2(x)}{f(x)} [K^2] dx}{\int \left\{ r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right\}^2 dx [t^2 K]} \right\}^{\frac{1}{5}}.$$

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

$$h_{opt}^{MSE}(\hat{f}_n(x, h)) = n^{-\frac{1}{5}} \left\{ \frac{f(x) [K^2]}{(f''(x))^2 [t^2 K]^2} \right\}^{\frac{1}{5}},$$

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

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

3.3 Méthodes d'optimisation de h

Les méthodes de la minimisation de l'estimation par l'Erreur Moyenne Quadratique (MSE) ou par l'Erreur Intégrée Moyenne Quadratique (MISE) restent incomplètes ou autrement dit, ces méthodes ont toujours des inconvénients, soit au sens de la qualité de l'estimateur par rapport à une norme d'erreur bien déterminée, soit par l'allure graphique de la courbe (lisse ou non). Parmi ces méthodes, Pour parer à cette difficulté, ils existent trois méthodes célèbres. La famille des méthodes de validation croisées (cross validation) et la famille des méthodes Plug in et aussi la famille Bootstrap. Nous nous limitons à la méthode de la validation croisée pour la régression.

3.3.1 Méthode de validation croisée pour la régression

La méthode de la validation croisée, concernant l'estimation non paramétrique de la régression, a été étudiée par Hall (1984)[19], Hördle et Marron (1985)[26], Hördle et Kelly (1987)[25]. L'idée de la validation croisée consiste à minimiser par rapport à h la distance $d_1(\hat{r}_n, r)$ définie par

$$d_1(\hat{r}_n, r) = \int_{\mathbb{R}} (\hat{r}_n(x, h) - r(x))^2 \omega(x) f(x) dx,$$

où $\omega(\cdot)$ est une fonction de poids arbitraire (voir Härdle et Kelly (1987)[25]). On a

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

Comme le dernier terme ne dépend pas de h , la minimisation $d_1(\hat{r}_n, r)$ revient à optimiser :

$$\int_{\mathbb{R}} \hat{r}_n(x, h)^2 \omega(x) f(x) dx - 2 \int_{\mathbb{R}} \hat{r}_n(x, h) r(x) \omega(x) f(x) dx. \quad (3.1)$$

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

$$\int_{\mathbb{R}} \hat{r}_n(x, h) r(x) \omega(x) f(x) dx = E(\hat{r}_n(x, h) Y \omega(X))$$

d'où un estimateur du deuxième terme de l'équation (3.1) est :

$$\frac{1}{n} \sum_{i=1}^n r_i(X_i) Y_i \omega(X_i)$$

avec :

$$r_i(X_i) = \frac{\sum_{j \neq i}^n Y_j \mathbf{K}\left(\frac{x-X_j}{h}\right)}{\sum_{j \neq i}^n \mathbf{K}\left(\frac{x-X_j}{h}\right)}$$

D'une manière similaire le premier terme de l'équation (3.1) est estimé par :

$$\frac{1}{n} \sum_{i=1}^n \{r_i^2(X_i) \omega(X_i)\}$$

Maintenant, nous sommes en mesure de minimiser la quantité suivante

$$\frac{1}{n} \sum_{i=1}^n r_i^2(X_i) \omega(X_i) - \frac{2}{n} \sum_{i=1}^n r_i(X_i) Y_i \omega(X_i),$$

sans contraintes. En plus nous remarquons que :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n r_i^2(X_i) \omega(X_i) - \frac{2}{n} \sum_{i=1}^n r_i(X_i) Y_i \omega(X_i) &= \frac{1}{n} \sum_{i=1}^n \{r_i(X_i) - Y_i\}^2 \omega(X_i) \\ &\quad - \frac{1}{n} \sum_{i=1}^n Y_i^2 \omega(X_i) \end{aligned}$$

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

$$\frac{1}{n} \sum_{i=1}^n \{Y_i - r_i(X_i)\}^2 \omega(X_i).$$

Remarque Les méthodes les plus connues et les plus utilisées, pour choisir le paramètre de lissage en plus de la méthode de validation croisée, on a :

1. Les méthodes plug-in (ré-injection). Le principe de ces méthodes repose sur l'estimation d'une quantité qui dépend de la dérivée seconde de la densité de probabilité inconnue r qu'on désire estimer. La méthode donne de meilleurs résultats théorique set pratiques[20].
2. La méthode de bootstrap : L'application de la méthode bootstrap permettent d'estimer une région de confiance pour les courbes des fonctions r . cette méthode est plus couteux en calcul et s'applique aussi bien dans le cas discret que dans le cas continu [17].

On donnera les deux méthodes numériques pour pouvoir les comparer avec l'estimateur à noyau.

Chapitre 4

Normalité asymptotique et intervalles de confiance

4.1 Normalité asymptotique

La première démonstration de la normalité asymptotique de l'estimateur à noyau a été fournie par Schuster et al. (1972) qui a étendu le résultat de Nadaraya (1964), dans lequel celui-ci montre la normalité asymptotique sous la condition que Y est borné et que $nh^2 \rightarrow \infty$. D'autres démonstrations sont disponibles dans la littérature statistique (Härdle, 1990; Nadaraya, 1989). Dans le cas où la variable réponse est censurée à droite, la normalité asymptotique de l'estimateur de type Nadaraya-Watson, introduit dans Kohler et al. (2002), a été prouvée par Guessoum et Ould Saïd (2008) pour des données indépendantes et identiquement distribuées par Guessoum et Ould Saïd (2012) pour des données α -mélangeantes. Cette étude leur a permis de construire des bandes de confiance asymptotiques uniformes basées sur la consistance uniforme des estimateurs sur des intervalles fermés de \mathbb{R} . Commençons par les hypothèses nécessaires à ce travail et qui sont standard dans le contexte de notre étude.

4.1.1 Hypothèses

- \mathcal{H}_1 : La fenêtre h vérifie : $h > 0$, $\lim_{n \rightarrow \infty} h = 0$, $\lim_{n \rightarrow \infty} \frac{nh}{\log n} = +\infty$, $\log \log n = o\left(\frac{1}{h^\mu}\right)$,
où $0 < \mu < 1$.
- \mathcal{H}_2 : K est borné et à support compacte, $\int K(t)dt = 1$, il existe ℓ : $\int t^j K(t)dt = 0$,

$$\forall j = 1, \dots, \ell - 1 \text{ et } 0 < |f t^\ell K(t)| < \infty.$$

— \mathcal{H}_3 : $f(x) > 0$, g et f sont ℓ fois différentiables dans un voisinage de x , où $g(x) = \int y f_{X,Y}(x,y) dy$ et $f_{X,Y}(x,y)$ est la densité du couple (X,Y) dont nous supposons l'existence.

— \mathcal{H}_4 : $\lim_{n \rightarrow 0} nh^{2\ell+1} = 0$, où ℓ est donnée en \mathcal{H}_2 et $\ell > 2$.

— \mathcal{H}_5 : $\int_{\{F_L(y)S_R(y) \neq 0\}} \frac{y^2 f_{X,Y}(x,y)}{F_L(y)S_R(y)} dy = q(x)$ deux fois différentiable.

— \mathcal{H}_6 : Les fonctions de répartition des variables Y, R et L sont continues, $\sup(I_L, I_Y) < I_R$ et $T_R < T_Y$ Rappelons les hypothèses H_2 et H_3 déjà utilisées aux chapitres 1 et 2,

H_2 : (R, L) et (X, Y) sont indépendantes.

$H_{1,3}$: $\exists T < T_R$ et $I > I_L$ telle que $\forall n \in \mathbb{N}, \forall i (1 \leq i \leq n), A_i = 0 \implies I \leq Z_i \leq T$ p.s.

Théorème 8

Un résultat important de Härdle (1990) [24], montre que si $f(x)$ et $r(x)$ sont deux fois continument dérivables sur \mathbb{R} , $f(x) > 0$ et si la condition (F, 3) est vérifiée, alors pour $h = n^{-\alpha}$, $\alpha \in [1/5, 1]$, on a la normalité asymptotique de l'estimateur de Nadaraya-Watson c'est à dire :

$$\sqrt{nh} [\hat{r}_n(x; h(x)) - r(x)] \xrightarrow{\mathcal{L}} \mathcal{N} (D(x), v^2(x))$$

avec :

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

et

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

Comme le terme du biais de $\hat{r}_n(x, h)$ est négligeable par rapport au terme de la variance, l'intervalle de confiance au seuil de α de la fonction régression, est donné par

$$\left[\hat{r}_n(x; h) - z_{1-\alpha/2} \sqrt{\frac{\sigma^2(x) [K^2]}{nhf(x)}}, \hat{r}_n(x; h) + z_{1-\alpha/2} \sqrt{\frac{\sigma^2(x) [K^2]}{nhf(x)}} \right],$$

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

$$P \left(r(x_0) \in \left[\hat{r}_n(x_0;h) - 1.96 \sqrt{\frac{\sigma^2(x_0) [\mathbf{K}^2]}{nhf(x_0)}}, \hat{r}_n(x_0;h) + 1.96 \sqrt{\frac{\sigma^2(x_0) [\mathbf{K}^2]}{nhf(x_0)}} \right] \right) \approx 95\%$$

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

4.2 Intervalles de confiance

Pour construire un intervalle de confiance, i.e, IC sur $r(x_i)$, on admet le résultat suivant :

Théorème 9 [2]

L'estimateur à noyau de Nadaraya-Watson vérifie :

$$\sqrt{hn} \left\{ \hat{r}_n(x) - r(x) - \frac{1}{2} a_k h^2 \left[r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right] \right\} \rightarrow \mathcal{N} \left(0, \frac{b_k \sigma^2}{f(x)} \right)$$

où $f(\cdot)$ désigne la densité de x et

$$a_k = \int u^2 K(u) du \quad b_k = \int k^2(u) du$$

Voir Wand et Jones (1995) pour les valeurs de a_K et b_K pour de nombreux noyaux. On admettra en particulier que :

Noyau	b_k
Uniforme	1/2
Triangulaire	2/3
Gaussien	1/2π
Biweight	5/7
TRIweight	350/429
Espanechinkov	3/5

Des simplifications peuvent être apportées si la valeur de h décroît avec n plus rapidement que $h = n^{-1/5}$. Dans ce cas, le terme de biais disparaît et donc on obtient le résultat suivant dans le cas d'un noyau uniforme $b_k = 1/2$

$$\sqrt{hn} [\hat{r}_n(x) - r(x)] \longrightarrow N\left(0, \frac{\sigma^2}{2f(x)}\right)$$

peut donc en déduire la manière de construire des *IC* sur les valeurs de $r(x_i)$.

Théorème 10 [15] *Sous l'hypothèse que h vérifie $h^{-1/5} \rightarrow 0$, l'écart type de l'estimateur à noyau de Nadaraya-Watson $\hat{r}(x)$ vérifie :*

$$\sqrt{hn} [\hat{r}_n(x) - r(x)] \longrightarrow \mathcal{N}\left(0, \frac{b_k \sigma^2}{f(x)}\right)$$

$f(\cdot)$ est la densité de x . dans un intervalle de confiance sur $\hat{r}(x)$ au seuil de $\alpha\%$ donc on définit par :

$$IC_\alpha = [\hat{r}_n(x) - Z_{1-\alpha/2} S_{\hat{r}(x)}, \hat{r}_n(x) + Z_{1+\alpha/2} S_{\hat{r}(x)}]$$

où $Z_{1-\alpha/2}$ désigne le fractile de la loi $\mathcal{N}(0; 1)$ où :

$$S_{\hat{r}(x)} = \sqrt{\frac{b_k \sigma^2}{\hat{f}(x)}}$$

avec

$$\hat{f}_n(x) = K\left(\frac{x_i - x}{h}\right)$$

La procédure pour obtenir les *IC* est donc la suivante :

1. On choisit h vérifie $h^{1/5} \rightarrow 0$ et une fonction noyau $K(u)$; d'où l'on déduit b_k .
2. On construit l'estimateur à noyau de Nadaraya-Watson $\hat{r}_n(x)$. On recommence pour toutes les valeurs x_1, \dots, x_n .
3. On calcule l'estimateur de la variance des résidus

$$\sigma^2 = 1/n [y_i - \hat{r}_n(x_i)]^2$$

4. On estime la valeur de $f(x)$, densité de X au point x par :

5. On calcule IC sur $\hat{f}_n(x)$ au seuil défini par :

$$IC_\alpha = \left[\hat{r}_n(x) - Z_{1-\alpha/2} \sqrt{\frac{b_k \sigma^2}{\hat{f}_n(x)}}, \hat{r}_n(x) + Z_{1-\alpha/2} \sqrt{\frac{b_k \sigma^2}{\hat{f}_n(x)}} \right]$$

Il existe des procédures numériques de choix d'un h optimal. La première méthode consiste à choisir h de sorte à minimiser la MISE (Mean Integrated Squared Error). C'est la définition même du paramètre de lissage optimale.

Chapitre 5

Simulation

Nous terminons notre mémoire par ce cinquième chapitre, dont nous essayons par des simulations en utilisant le logiciel R, donner des différentes notions rencontrées au deuxième chapitre sur la fonction de régression, et la normalité asymptotique. Nous donnons l'importance de paramètre de lissage h , du noyau K dans l'estimation non paramétrique de la fonction de régression.

5.1 Présentation des données

Dans ce chapitre, nous avons présenté les résultats obtenus pour les différents jeux de données ainsi que pour différentes valeurs de h strictement positif (h fixé ou h varié), différents noyaux K (noyau Gaussien et noyau d'Epanechnikov), Dans ce qui suit, on va estimer une fonction de régression (non-linéaire), en utilisant l'estimateur de Nadarya Watson .

On suppose que l'on a observé un échantillon $X_1; \dots; X_n$,

$\hat{f}(\cdot)$ l'estimateur à noyau de la fonction de régression donnée par la forme :

$$\hat{f}(x) = \frac{\frac{1}{nh} \sum_{i=0}^n Y_i K\left(\frac{X-x}{h}\right)}{\frac{1}{nh} \sum_{i=0}^n K\left(\frac{X-x}{h}\right)}.$$

Nous allons donc étudier les cas suivants :

1. Paramètre de lissage ou fenêtre h fixé, noyau normal (noyau à support non compact) et n varié.
2. Paramètre de lissage ou fenêtre h fixé, noyau d'Epanechnikov (noyau à support compact) et n varié.
3. n fixé et fenêtre h varié (noyau Normal).

4. n fixé et fenêtre h varié (noyau d'Epanechnikov).

Dans les résultats graphiques de cette section :

- Le graphe noir exprime la fonction de régression $r(x)$
- Le graphe en bleu exprime l'estimateur à noyau de la régression $\hat{r}(x)$

Nous avons présenté les résultats obtenus pour les différents données.

5.2 Le paramètre de lissage h fixe, et n varié

On définit le modèle non linéaire suivant :

$$Y = \sin(\exp(x)) + 0.2\varepsilon$$

Dans ce premier cas, le paramètre de lissage ou la fenêtre h est fixé $h = n^{-\frac{1}{5}}$ et nous prenons différentes valeurs de la taille de l'échantillon ($n = 40; n = 200; n = 400$);

et K est un noyau normal $K(t) = \exp\{-t^2/2\}/\sqrt{2\pi}$; En utilisant le code **R**(Voir Annexe C.1)

Puis, on refait le programme pour $n = 200$ et $n = 400$, on obtient le graphe suivant :

La figure(5.1), représente la régression théorique en noir et l'estimateur à noyau de la régression en bleu.

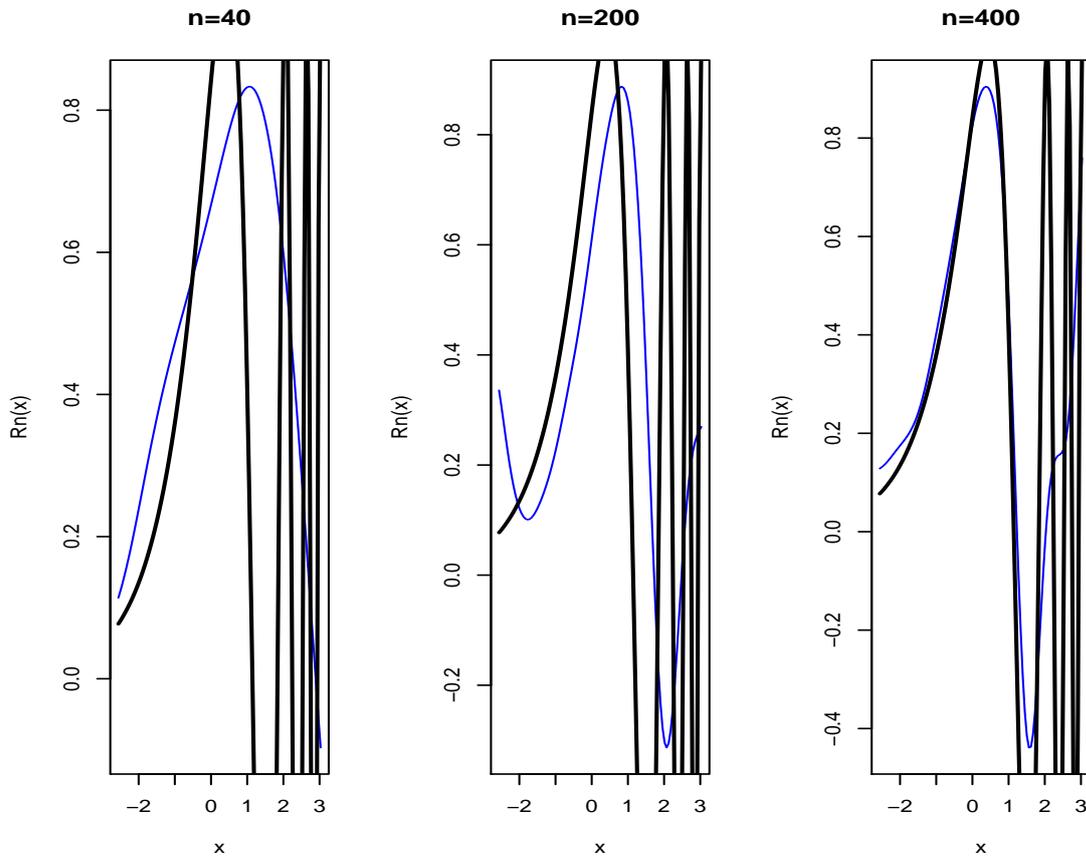


FIGURE 5.1 – Estimateur à noyau de la fonction de régression : h fixé, n varié et K noyau normal

Remarque

Nous remarquons sur le graphe de la Figure(5.1), La courbe de l'estimateur $\hat{r}(x)$ s'approche vers la courbe de la fonction $r(x)$, alors le noyau gaussien peut être le noyau optimale dans ce cas. Et que quand n est grand l'estimateur $\hat{r}(x)$ est plus proche de la fonction $r(x)$ (estimateur lisse), ce qui implique que la convergence de l'estimateur.

On va refaire le même travail précédent, remplaçant seulement le noyau normal par le noyau d'Epanechnikov : $K(x) = \frac{3}{4}(1-t^2)I_{(|t|<1)}$ ensuite, on modifié seulement cette partie dans le programme **R** précédent :

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

On obtient la figure(5.2) suivante :

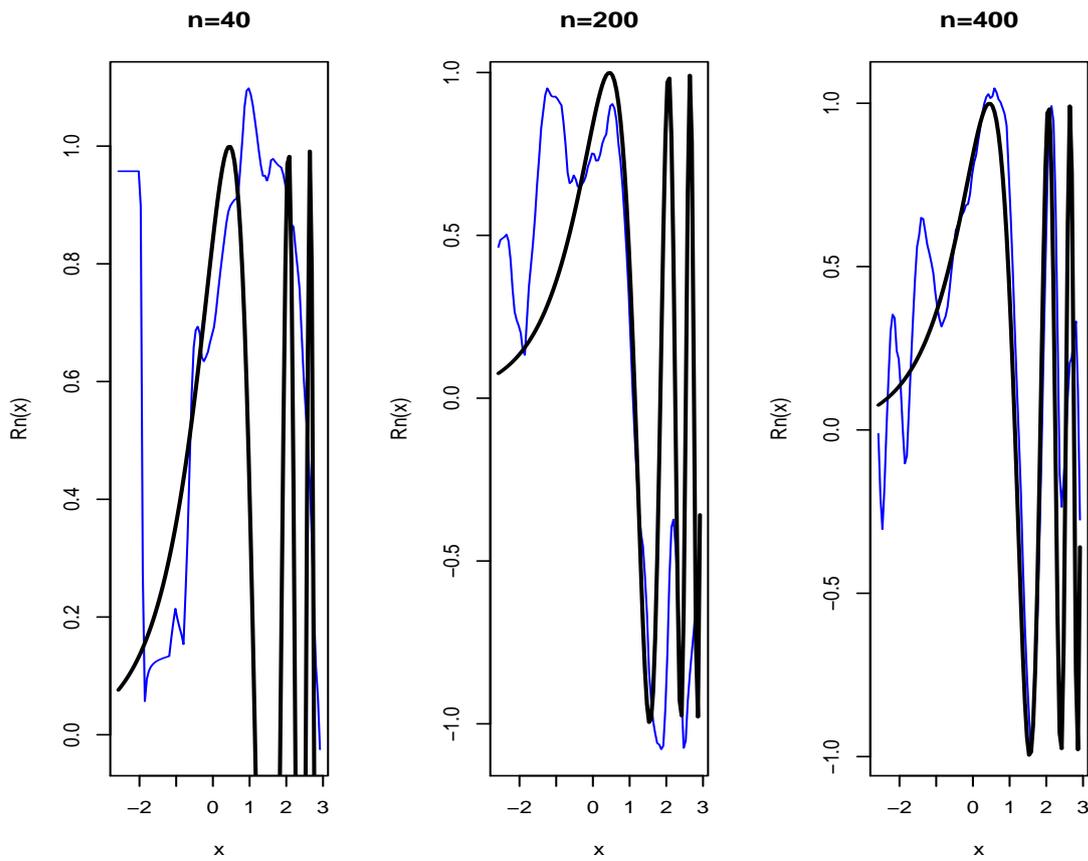


FIGURE 5.2 – Estimateur à noyau de la fonction de régression : h fixé, n varié et K noyau d’Epanechnikov.

Remarque

Nous remarquons ici qu’il n’y a pas beaucoup de variabilité dans le cas précédent, notamment avec l’estimateur lisse lorsque n est grand ($n = 400$). Par conséquent, le noyau d’Epanechnikov n’est pas le meilleur choix pour estimer $r(x)$.

5.3 Choix du paramètre de lissage

Considérons un échantillon de taille $n = 400$ et le noyau K est gaussien. Nous choisissons h variée dans l'intervalle $[0.1; 0.9]$. La comparaison graphique entre les graphes de la régression théorique et son estimateur, permet trouver une valeur h optimal (au sens graphique) ; En utilisant le code **R** (Voir Annexe C.2)

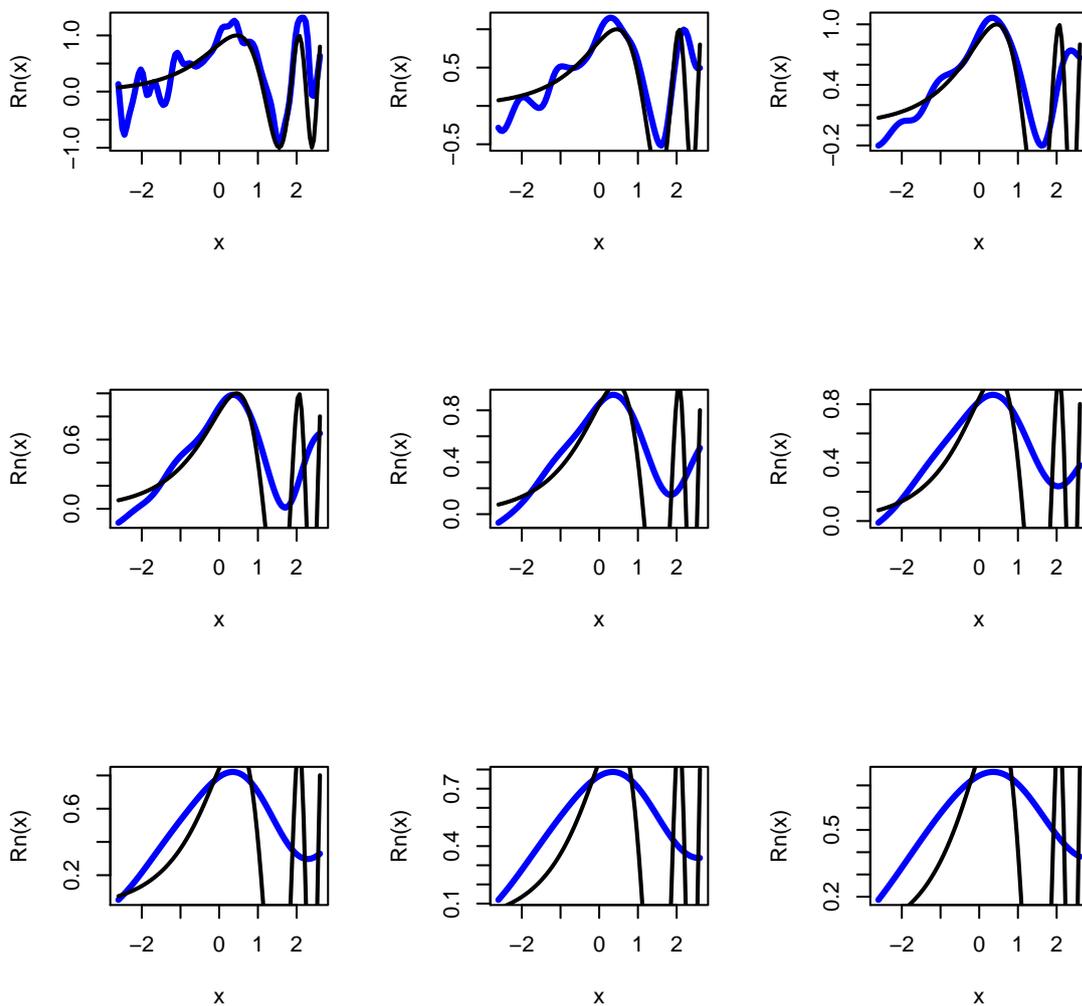


FIGURE 5.3 – Estimateur à noyau de la fonction de régression : h varié, n fixé et K noyau normal.

Nous remarquons sur les graphes de la Figure(5.3), qu'il est claire que la valeur du h optimale est représenté dans le 4ème graphe (ligne 2, colonne 1), pour $h = 0.4$.

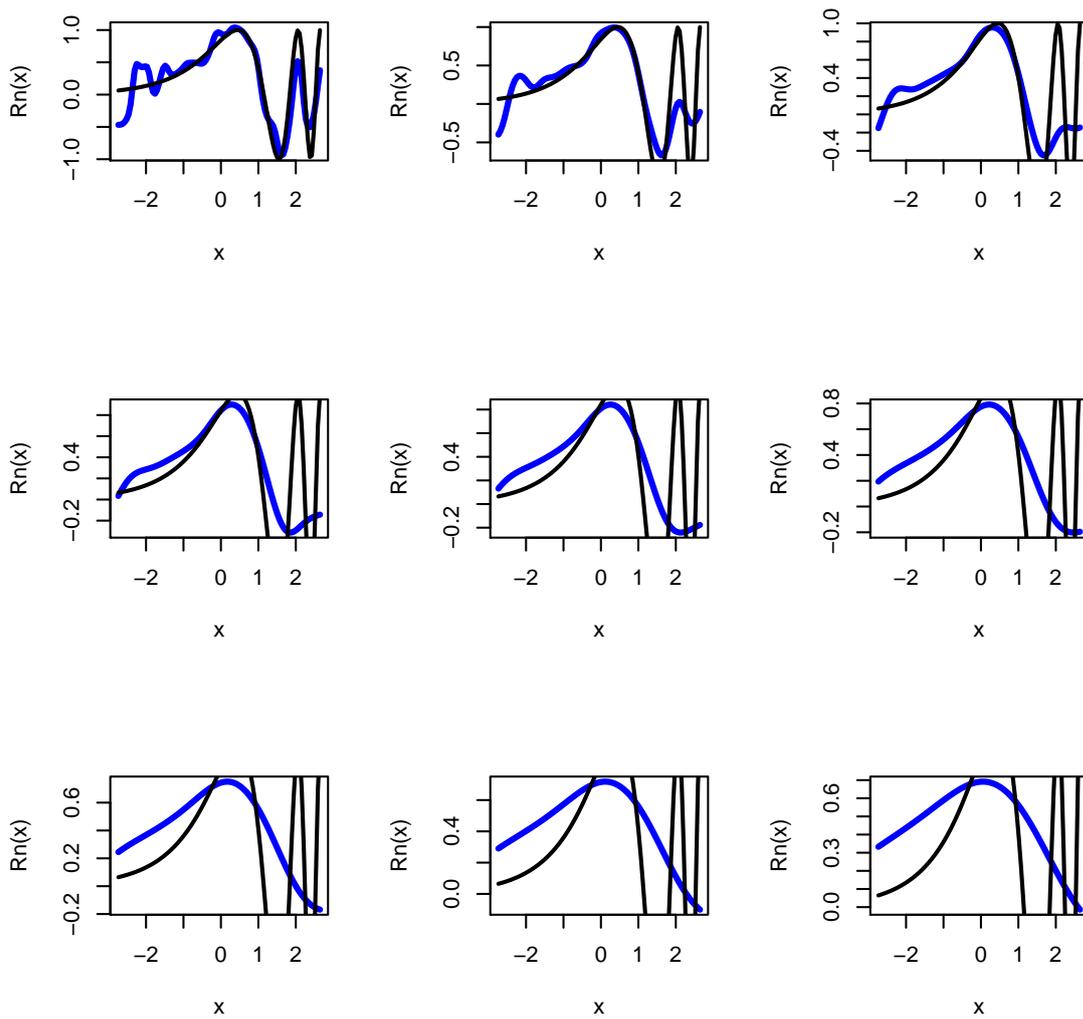


FIGURE 5.4 – Estimateur à noyau de la fonction de régression : h varié, n fixé et K noyau d’Epanechnikov.

Remarque

A noté, que pour les mêmes données, pour le noyau d’Epanechnikov L’estimation obtenue avec les valeurs de h varié de 0.1 à 0.9 sont données dans la Figure(5.4). Il est claire que la valeur du h optimale est de $h = 0.2$ (ligne 1, colonne 2).

En conclusion, le type du noyau n’est pas très influent sur la qualité de l’estimation contrairement à la valeur de h , c’est le choix de la fenêtre h , qui est très important, par rapport au choix du noyau.

5.4 Normalité asymptotique

Nous considérons :

$$Z_n(x) = \sqrt{nh}(\hat{r}(x) - r(x)).$$

Nous comparons la distribution d'échantillonnage de $Z_n(x)$ à une distribution normale centrée réduite,

En utilisant le code **R** (Voir Annexe C.3) :

la Figure(5.5), représente la comparons de l'estimateur à noyau de la densité de $Z_n(x)$ à la densité normale centrée réduite.

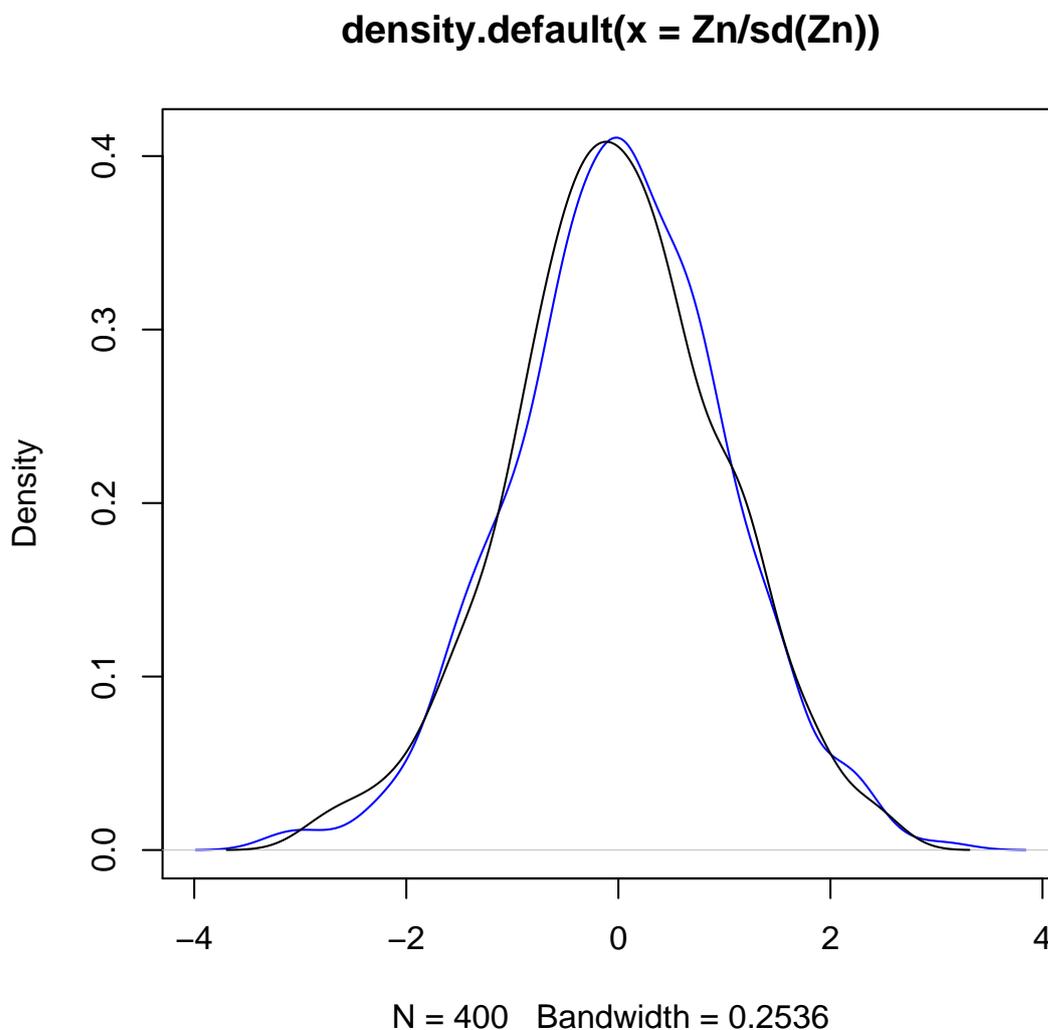


FIGURE 5.5 – L'estimateur à noyau de la densité de Z_n et la densité normale centrée réduite.

La figure (5.5), montre le graphe en noir de la densité normale centrée réduite superposée à l'estimateur à noyau de la densité \hat{Z}_n en bleu. Pour un échantillon de taille $n = 400$, les résultats suggèrent une bonne approximation de la loi normale pour l'estimateur de la fonction de régression.

Conclusion générale

Dans ce mémoire, nous avons examiné la méthode d'estimation à noyau, qui joue un rôle crucial dans la régression non paramétrique. Notre travail a démontré la simplicité et l'utilité de la méthode d'estimation de régression non paramétrique dans diverses situations.

L'estimateur à noyau de la régression non paramétrique dépend de deux paramètres : le noyau K et le paramètre de lissage h . Nos résultats mettent en évidence que le choix du noyau K a peu d'influence sur l'estimateur, tandis que le paramètre de lissage h joue un rôle essentiel. Ainsi, le choix de ce paramètre est crucial pour obtenir des estimations précises. Nous avons examiné plusieurs méthodes couramment utilisées pour sélectionner la valeur optimale de h , telles que la minimisation des erreurs quadratiques moyennes (MSE), des erreurs quadratiques intégrées ($MISE$) et la méthode de validation croisée.

Enfin, nous avons constaté que la normalité asymptotique dans l'estimation non paramétrique permet d'obtenir des approximations de la distribution de l'estimateur à une distribution gaussienne, lorsque la taille de l'échantillon augmente. Cela ouvre la voie à des calculs statistiques plus précis et à l'évaluation de la précision de l'estimation.

Annexe A

Annexe

Présentation de Logiciel R

Le logiciel **R** (un logiciel de statistique), a été initialement créé en 1977 par Robert Gentleman et Ross Ihaka du département de statistique de l'Université d'Auckland en Nouvelle Zélande. Il est la fois un langage informatique et un environnement de travail. Les commandes sont exécutées grâce des instructions codées dans un langage relativement simple, les résultats sont affichés sous forme de texte et les graphiques sont visualisés directement dans une fenêtre qui leur est propre.

C'est un clone du logiciel S-plus qui est fondé sur le langage de programmation orienté objet S, développé par AT et T Bell laboratoires en 1988. Ce logiciel sert manipuler des données, tracer des graphiques et faire des analyses statistiques sur ces données.

C'est aussi un outil très puissant et très complet, particulièrement bien adapté pour la mise en œuvre informatique de méthodes statistiques. Il est plus difficile d'accès que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de clics de souris dans des menus. L'avantage en est toutefois double :

L'approche est pédagogique puisqu'il faut maîtriser les méthodes statistiques pour parvenir à les mettre en œuvre ;

L'outil est très efficace lorsque l'on domine le langage R l'on devient alors capable de créer ses propres

outils, ce qui permet ainsi d'opérer des analyses très sophistiquées sur les données.

Annexe B

Annexe

LES CONVERGENCES

B.1 Convergence en loi

définition 14

Soit $(X_n)_n$ une suite de variables aléatoires réelles définies sur un même espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$. On dit que la suite $(X_n)_n$ 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 :

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

B.2 Convergence en probabilité

définition 15

On dit que X_n converge vers X en probabilité

$$\text{si, } \forall \varepsilon > 0 \lim_{n \rightarrow \infty} (\mathbb{P} |X_n - X| \geq \varepsilon) = 0$$

On note aussi : $X_n \xrightarrow{P} X$

B.3 Convergence presque sûre

définition 16

On dit que X_n converge presque sûrement vers X si : $\mathbb{P}(\lim_{n \rightarrow \infty} (X_n = X)) = 1$.

Ou de manière équivalente, s'il existe un sous-ensemble \mathbb{P} -négligeable N dans Ω tel que : $\forall \omega \in \Omega \setminus N$, on a $(X_n)(\omega) \xrightarrow{P.s} X(\omega)$ On parle aussi de convergence presque partout ou avec probabilité 1 ou forte, et on écrit : $X_n \xrightarrow{P.s} X$

B.4 Convergence presque complète

définition 17

La convergence presque complète implique la convergence presque sûre et se prête bien aux calculs faisant intervenir des sommes de variables aléatoires. Malgré cela, elle ne commence à devenir populaire dans la communauté statistique. Elle est utilisée surtout en statistique non-paramétrique. On dit que la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \rightarrow \infty$ si :

$$\forall \varepsilon > 0, \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \varepsilon] < \infty$$

et on dit que la vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ vers X est d'ordre (u_n) si :

$$\forall \varepsilon_0 > 0, \sum_{n \in \mathbb{N}} \mathbb{P}[|X_n - X| > \varepsilon_0 u_n] < \infty$$

Cette définition du taux a été 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.

Remarque 6

La convergence presque complète est plus forte que la convergence presque sûre et la convergence en probabilité.

Annexe C

Annexe

Les code de simulation sur R

C.1 Paramètre de lissage ou fenêtre h fixé, noyau normal et n varié

$n = 40$

$X = rnorm(n)$

$E = rnorm(n)$

$Y = \sin(\exp(X)) + (0.2) * E$

Noyau Normale $K(t)$

$dev.new()$

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

$h = n^{-.5}$

Initiation

taille de l'intervalle $[a, b]$

$s = 100$

borne – inf

$a = \min(X)$

borne – sup

$b = \max(X)$

$\text{Intervalle}[a,b]$

$x = \text{seq}(a,b, \text{length} = s)$

$V = \text{numeric}(n)$

$f_n = \text{numeric}(s)$

for(j in 1 :s)

for(i in 1 :n)

$V[i] = K((x[j] - X[i])/h)$

$f_n[j] = \text{sum}(V)/(n * h)$

Fonction Gn(.)

$W = \text{numeric}(n)$

$G_n = \text{numeric}(s)$

for(j in 1 :s)

for(i in 1 :n)

$W[i] = K((x[j] - X[i])/h) * Y[i]$

$G_n[j] = \text{sum}(W)/(n * h)$

$Rn = G_n / f_n$

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

plot(x,Rn,xlab="x", ylab="Rn(x)",type="l", main="n=40",col=4)

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

C.2 n fixé, fenêtre h varié et noyau Normal

```

n=400
X=rnorm(n)
E=rnorm(n)
Y=sin(exp(X))+(0.2)*E
Noyau Normale K(t)
K = function(t)(1/sqrt(2 * pi)) * exp(-0.5 * t^2)
h=seq(.1,.9,length=9)
s=100
a=min(X)
b=max(X)
x=seq(a,b,length=s)
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))
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])
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]  
op=par(mfrow=c(3,3))  
for(k in 1 :9)  
plot(x,Rn[,k],xlab="x",ylab="Rn(x)",col="4",lwd=3,type="l",main="")  
lines(x,sin(exp(x)),lwd= 2)  
}  
par(op)
```

C.3 Normalité asymptotique

```
norma = function(n,h){
```

```
  X = rnorm(n,0,1)
```

```
  E = rnorm(n,0,1)
```

```
  Y = sin(exp(X)) + (0.2) * E
```

```
  dev.new()
```

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

```
  s = 100
```

```
  a = min(X)
```

```
  b = max(X)
```

```
  x = seq(a,b,length = s)
```

```
  V = numeric(n)
```

```
  fn = numeric(s)
```

```
  for(jin1 : s){
```

```
    for(iin1 : n){
```

```
      V[i]=K((x[j]-X[i])/h)
```

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

```
    W=numeric(n)
```

```
    Gn=numeric(s)
```

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

```
      for(i in 1 :n)
```

```
        W[i]=K((x[j]-X[i])/h)*Y[i]
```

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

```
Rn =Gn/fn
```

```
Zn=sqrt(h*n)*(Rn-Y)
```

```
b=rnorm(n)
```

```
density(Zn)
```

```
Z=density(Zn)/sd(Zn)
```

```
plot(density(Z),col=4)
```

```
lines(density(b))
```

```
}
```

```
norma(400, 4.5)
```

Bibliographie

- [1] B. W.Silverman, Density Estimation for Statistics and Data Analysis. Chapman Hall, London, (1986).
- [2] C.HURLIN, Régressions non paramétriques univariées. Master Econométrie et Statistique Appliquée (ESA), Université d'Orléans, (2007).
- [3] D.Bosq, and J.,P.Lecoutre, Théorie de l'Estimation Fonctionnelle, Economica, Paris, (1987).
- [4] E.Parzen, On estimation of a probability density function and mode. Ann. Math. Statist., 33, 1065-1076,(1962).
- [5] E. A.Nadaraya, On the nonparametric estimator of Bayesian risk in the classificationproblem. Proc. AN. Georg SSR, 82(2), 277-280 (in Russian),(1976).
- [6] F.Ferraty, et P.Vieu, Nonparametric functional data analysis, Theory and practice. Springer, (2006).
- [7] F.Ferraty, et P.Vieu, Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. C. R . Acad. Sci., Paris. 330, No. 2, 139-142,(2000).
- [8] F.Ferraty, et P.Vieu, Modèles Non paramétrique de Régression,(2002).
- [9] G. S.Watson, Smooth Regression analysis. Sankhyà Ser. A, 26, 359-372, (1964).
- [10] H.Beghriche, L'estimation de la Fonction de Régression, Mémoire de magistère en mathématique université de Mentouri-Constantine.

- [11] I.Ben Khalifa, Estimation non-paramétrique par noyaux associés et données de panel en marketing, université du 7 Novembre à Carthage. Tunisie, (2007).
- [12] J.Geoffrey, Sur l'estimation d'une densité dans un espace métrique. C.R. Acad.Paris Sér. A, 278 :1449-1452, (1974).
- [13] J.USPENSKY, Introduction to mathematical probability. McGraw-Hill, New York, (1937).
- [14] J.Lecoutre, Contribution à l'estimation non paramétrique de la régression. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France, (1982).
- [15] J. S.Simonoff, Smoothing Methods in Statistics. Springer Series in Statistics. Springer-Verlag, New York,(1996).
- [16] L.Devroye, and L.Györfi, , Nonparametric Density Estimation. New York. (1985).
- [17] M.Lejeune, Statistique. La théorie et ses applications : La théorie et ses applications. Springer Science & Business Media.(2011).
- [18] M.Rosenblatt, Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27, 832-837,(1956).
- [19] P.Hall, Asymptotic properties of integrated square error and cross validation for kernel estimation of a regression function. Z. Wahrsch. Verw. Gebiete, 67, 175-196,(1984).
- [20] P. L. HSU et H. ROBBINS, Complete convergence and the law of large numbers. Proceedings of the national academy of sciences, 33(2), (1947).
- [21] S.Amroun, Sur L'Estimation De La Courbe De Régression De La Moyenne, Mémoire De Magister en Mathématiques Appliquées université de Bêjaïa,(2011).
- [22] S. BERNSTEIN : Probability Theory, 4th ed. (in russian). M. L. Gostechizdat, (1946).
- [23] V. YURINSKII : Exponential inequalities for sums of random vectors. Journal of Multivariate Analysis, 6 :475-499, (1976).

- [24] W. Härdle, Applied Nonparametric Regression. universitè de Berlin, (1990).
- [25] W.Härdle, et G.Kelly, Nonparametric kernel regression estimation -optimal choice of bandwidth. Statistics, 18.1, 21-35, (1987).
- [26] W.Härdle,J. S. Marron,Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist., 13.4, 1465-1481, (1985).

Résumé

L'objectif principal de ce travail est de présenter la normalité asymptotique d'un estimateur à noyau de la fonction de régression. Nous commençons par fournir un rappel des notions et des définitions de base en statistique non paramétrique .

Nous décrivons ensuite la construction de l'estimateur à noyau et ses propriétés fondamentales, notamment les théorèmes de convergence presque complète , la vitesse de convergence, ainsi que le choix du noyau K et du paramètre de lissage h , qui jouent un rôle crucial dans la qualité de l'estimation.

Ensuite, nous effectuons des simulations en utilisant le logiciel R, ce qui nous permet d'observer l'influence de la taille de l'échantillon et des valeurs choisies pour les paramètres de l'estimateur.

Nous examinons également les résultats asymptotiques d'un estimateur à noyau de la fonction de régression similaire à l'estimateur de Nadaraya-Watson. Cela comprend la vitesse de convergence ponctuelle et uniforme, ainsi que la normalité asymptotique. Le mode de convergence utilisé est celui de la convergence presque complète.

Mots clés

Fonction de régression, Estimation non-paramétrique, Estimateur à noyau, normalité asymptotique, convergence presque complète.

Abstract

The main objective of this work is to present the asymptotic normality of a kernel estimator of the regression function. We begin by providing a reminder of the concepts and basic definitions in nonparametric statistics.

We then describe the construction of the kernel estimator and its fundamental properties, including the theorems of almost complete convergence, convergence rate, as well as the choice of the kernel function K and the smoothing parameter h , which play a crucial role in the quality of the estimation.

Next, we perform simulations using the software R, which allows us to observe the influence of sample size and the values chosen for the estimator's parameters.

We also examine the asymptotic results of a kernel estimator of the regression function similar to the Nadaraya-Watson estimator. This includes the pointwise and uniform convergence rates, as well as the asymptotic normality. The mode of convergence used is almost complete convergence.

Keywords

Regression function, non-parametric estimation, kernel estimator, asymptotic normality, almost complete convergence.