

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITÉ ABDELHAMID BEN BADIS DE MOSTAGANEM
FACULTÉ DES SCIENCES EXACTES ET DE L'INFORMATIQUE
DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE
FILIÈRE : MATHÉMATIQUES



MÉMOIRE DE FIN D'ÉTUDES

Pour l'Obtention du Diplôme de Master en Mathématiques délivré par

Université de Mostaganem

Spécialité " Modélisation, Contrôle et Optimisation"

présenté par :

Abdelkader Hakim BOUHADI

**Estimation non paramétrique de certains paramètres
fonctionnels**

soutenu le 12/07/2021 devant le jury composé de :

Président :	Abdellaoui RABAB ^c	MCB	UMAB
Examineur :	Oussama BOUANANI	MAB	UMAB
Encadreur :	Mustapha MOHAMMEDI	MCB	UMAB

Année Universitaire : 2020 / 2021

Table des matières

Introduction	3
1 Modèle de fonction de répartition	6
1 Quelques outils de probabilités	6
2 Estimateur empirique	8
2.1 La vitesse de convergence	11
3 Estimation non-paramétrique par la méthode de noyaux	13
3.1 Etude asymptotique du biais et de la variance	15
2 Modèle de densité :	17
1 Estimation par Histogramme	17
2 Estimateur à noyau	18
2.1 Définition et propriétés élémentaires	19
2.2 Etude du biais et de la variance	23
2.3 Quelques remarques :	25
3 Régression non paramétrique	27
1 Estimateur de Nadaraya-Watson	28
2 L'estimation par la méthode des polynômes locaux dans le cas réel . .	34
2.1 Le principe de la méthode	34
2.2 La méthode locale linéaire	37
4 Application sur des données simulées	39
Conclusion	42
Bibliographie	42

Remerciements

En tout premier lieu, je remercie le bon Dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.

Je remercie mon encadreur, *Dr. Mohammedi Mustapha*, pour tous les conseils et instructions qu'ils m'ont donnés, et pour leur présence tout au long de la période de la réalisation de ce mémoire. Merci de m'avoir fait confiance depuis le début.

J'aimerais également adresser mes remerciements les plus profonds à Monsieur le *Dr Bouanani Oussama*, pour son soutien et ses conseils avisés et son écoute qui ont été prépondérants pour la bonne réussite de ce manuscrit.

Je tiens à remercier les membres du jury, *Dr. Bouabdelli Rabab* et *Dr. BOUANANI Oussama*, d'avoir accepté d'évaluer et d'examiner ce travail.

Je remercie particulièrement mes parents, qui m'ont tant soutenu tout au long de ma carrière universitaire et m'ont donné la force et le courage pour réussir. Je n'oublie pas mes frères et sœurs et toute ma famille car ils ont mis à ma disposition tous les moyens, mais surtout pour leur amour.

Mes reconnaissances vont aussi à tous mes chers amis, en particulier *Hammou Mohammed*, *Youcef Taf*, *Oueld Khattab Mohamed*, *Ghalal fatima zohra*, Merci pour tous les merveilleux souvenirs que nous avons eus ensemble.

Enfin, je tiens à remercier tous les professeurs et étudiants du département de mathématiques qui m'ont beaucoup aidé. Merci à tous ceux qui ont participé, de près ou de loin, au couronnement de ce travail.

Résumé

La problématique abordée dans ce manuscrit concerne l'estimation non paramétrique de certains modèles fonctionnels, dans un espace de dimension finie.

Nous supposons pour commencer que l'échantillon que nous étudions est constitué de variables indépendantes et identiquement distribuées. Nous avons fixé comme objectif l'établissement de la convergence en moyenne quadratique de la fonction de régression et la densité par la méthode du noyau.

Dans la deuxième partie de ce manuscrit, nous estimons la densité par la méthode histogramme et la méthode du noyau. Sous certaines des hypothèses, nous étudions une propriété statistique qui concerne la convergence en moyenne quadratique.

Dans la troisième partie de ce manuscrit, nous considérons le même type de indépendance des observations que le cas précédent, en s'intéressant à l'estimation de la fonction de régression par la méthode locale linéaire. Enfin, nous comparons la performance de l'estimateur de la régression à noyau classique et la méthode locale linéaire à l'aide de simulations.

Mots clés : Fonction de régression, estimation non paramétrique, polynômes locaux, Convergence uniforme presque sûre, locale linéaire.

Introduction

La théorie de l'estimation est une branche des mathématiques qui est très importants dans la recherche en statistique. elle permet de faire des prévisions et réduire la complexité du phénomène étudié. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique.

La statistique paramétrique est le cadre classique de la statistique. Le modèle statistique est décrit par un nombre fini de paramètres, c'est à dire la structure du modèle est spécifiée a priori. Par contre **la statistique non paramétrique** modéliser les caractéristiques inconnues d'une population à partir d'un échantillon de celle-ci. c'est à dire cette méthode consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue, élément d' une certaine classe fonctionnelle.

On dit également de ces méthodes qu'elles permettent de " laisser parler les données" (traduction de " letting the data speak for themselves "). Bien qu'en pratique, ces méthodes requièrent des jeux de données de taille importante, elles présentent l'avantage de ne nécessiter que d'hypothèses de régularité (continuité, dérivabilité, . . .) sur la fonction de lien entre les variables.

Dans ce mémoire nous nous focaliserons sur l'estimation non paramétrique des fonctions de densité et de régression. Ces modèles ont connu un essor important dans l'analyse des données avec un large gamme d'applications en filtrage et la prévision dans les communications et le contrôle des systèmes.

Historiquement, La première application que nous verrons relève de l'estimation de fonctions de densité par des méthodes d'opérateur 'à noyau (kernel) avec les travaux fondateurs de Rosenblatt [15] et de Parzen [17]. Ces premiers travaux ont été étendus à la notion de régression kernel, imparfaitement traduit en français par le terme de régression avec lissage par opérateur 'à noyau. Dans ce domaine, on identifie deux papiers fondateurs publiés la même année : Nadaraya [14] et Watson.

Organisation du manuscrit :

1. **Chapitre 1** : Le premier chapitre est consacré aux outils mathématiques, les inégalités et techniques utilisées et existantes pour nous aider à démontrer les théorèmes. Cette partie a été suivie par une présentation sur les types d'estimation non paramétrique de la fonction de répartition. Comme résultats asymptotiques nous établissons la convergence presque complète et erreur quadratique moyenne(MSE).
2. **Chapitre 2** : Dans ce chapitre, nous concentrons sur les méthodes d'estimation de la densité : La méthode d'estimation par histogramme et la méthode d'estimation par noyau (estimateur de densité Parzen-Rosenblatt) qui peut être vue comme une extension de la méthode d'estimation par histogramme. De plus, nous présentons également les propriétés fondamentales de Estimation à noyau de la densité.
3. **Chapitre 3** : Dans cette partie nous présentons l'estimation non paramétrique de la régression par la méthode deNadaraya Watson et celle de polynômes locaux qui est une généralisation de la méthode de Nadaraya Watson. Ensuite, nous avons établi sous certaines conditions, une propriété statistique qui concerne erreur quadratique moyen de l'estimateur de Nadaraya-Watson.

Nous terminons notre manuscrit par un quatrième chapitre (Simulation) : Dans ce chapitre nous avons illustré notre méthode par une application sur des données simulées, l'objectif est l'étude comparative entre la méthode locale constante et la méthode locale linéaire.

Chapitre 1

Modèle de fonction de répartition

Dans ce chapitre, nous avons commencé par une partie introductif, en s'intéressant d'abord aux notions élémentaires sur la statistique non paramétrique. Ensuite nous avons exposé un bref descriptif sur l'estimateur empirique de la fonction de répartition, pour lequel nous avons énoncée le théorème de la vitesse de convergence. Dans la dernière sections , On suppose qu'on dispose des observations indépendantes et nous présentons l'estimation non paramétrique de la fonction de répartition par la méthode de noyaux et les propriétés de l' estimateur.

1 Quelques outils de probabilités

Dans cette partie, nous allons rappeler quelques définitions et théorèmes utilisés dans ce manuscrit.

Notions de convergence

Définition 1.1. (*Définition A.1.[12]*).

Soit $(X_n)_{n \in \mathbb{N}}$, une suite de variables aléatoires réelles définie sur un espace de probabilité $(\Omega; \mathcal{A}; P)$. On dit que la suite $(X_n)_{n \in \mathbb{N}}$ converge presque complètement (*p, co.*) vers X une variable aléatoire réelle définie sur le même espace, si et seulement si

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

et nous écrivons $X_n \xrightarrow[n \rightarrow +\infty]{p,co.} X$.

Définition 1.2. (Définition A.3.[12]).

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles, $(u_n)_{n \in \mathbb{N}}$ une suite de nombres réelles. On dit que $X_n = O(u_n)$ en $p, co.$, si est seulement si

$$\forall \epsilon > 0, \sum_{n=0}^{\infty} P(|X_n| > \epsilon u_n) < \infty.$$

Définition 1.3. (Proposition A.5.[12]).

Supposons que $\lim_{n \rightarrow +\infty} u_n = 0$, $X_n = O_{p,co.}(u_n)$ et $\lim_{n \rightarrow +\infty} Y_n = l$ ($p, co.$) où l est un nombre réel. Nous avons :

1. $X_n Y_n = O_{p,co.}(u_n)$.
2. $\frac{X_n}{Y_n} = O_{p,co.}(u_n)$, à condition que $l \neq 0$.

Ordre de grandeur

Pour comparer les vitesses de convergence asymptotique de différents estimateurs, on utilise la famille de notation de Landau. Notons les séquences stochastiques que l'on exprimera avec les notions de "petit o_p " et de "grand O_p ".

Définition 1.4. Soit X_n une séquence de variables aléatoires réelles indexée par un entier positif $n = 1, 2, \dots$ et C une constante positive. On dit que X_n est bornée en probabilité, si pour tout $\epsilon > 0$, il existe une constante positive C est un entier positif N tel que

$$P[|X_n| > C] \leq \epsilon,$$

pour tout $n > N$, où ϵ est un petit nombre positif arbitraire. On écrit

1. $X_n = O_p(1)$ pour indiquer que X_n est bornée en probabilité ,
2. $X_n = o_p(1)$ si $X_n \xrightarrow[n \rightarrow +\infty]{P} 0$ où $P \rightarrow$ dénote la convergence en probabilité (i.e. $\mathbb{P}[|X_n| > \epsilon] \rightarrow 0$ quand $n \rightarrow \infty$),
3. $X_n = O_p(Y_n)$ si $X_n/Y_n = O_p(1)$, et $X_n = o_p(Y_n)$ si $X_n/Y_n = o_p(1)$.

Quelques inégalités utiles

Théorème 1.1. (Markov)[13].

Soient X une variable aléatoire réelle. Alors, pour tout $a > 0$,

$$P(|X| > a) < \frac{E(|X|)}{a}.$$

Théorème 1.2. (Bienaymé-Tchebychev)[4].

Soient X une variable aléatoire réelle. Alors, pour tout $a > 0$:

$$P(|X - E(X)| > a) < \frac{\text{var}(X)}{a^2}.$$

Théorème 1.3. (Inégalité de Bernshtein Fréchet)[1] Soit X_1, X_2, \dots, X_n une suite de variables aléatoires i.i.d, tel que, $\alpha_i \leq X_i \leq \beta_i$ et $\alpha_i, \beta_i \in \mathbb{R}$. Alors, $\forall t > 0$ on a :

$$P \left[\left| \sum_{i=1}^n (X_i - E(X_i)) \right| \geq t \right] \leq 2 \exp \left(\frac{-2t^2}{\sum_{i=1}^n (\beta_i - \alpha_i)^2} \right)$$

2 Estimateur empirique

Soit X une variable aléatoire (v.a.) de loi F ; avec $F(x) = P(X \leq x)$ la fonction de répartition de X : Soit X_1, X_2, \dots, X_n un échantillon de X et soient $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ l'échantillon ordonné. Supposons que F soit complètement inconnue (non paramétrique), un estimateur pour F est la fonction de répartition empirique, notée F_n qui est définie par :

$$\begin{aligned} F_n(x) &= \frac{1}{n} \# \{i =: X_i \leq x\}, \\ &= \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(X_i), \\ &= \begin{cases} 0 & \text{si } x < X_{(1)}, \\ k/n & \text{si } X_{(k)} < x < X_{(k+1)}, \\ 1 & \text{si } x > X_{(n)}. \end{cases} \end{aligned} \tag{1.1}$$

Exemple fondamentale 2.1. *Le programme R suivant permet de simuler la distribution empirique pour un échantillon de taille 100 ; issu d'une v.a. de loi exponentielle de paramètre λ , dont la distribution théorique est :*

$$F(x) = \int_{-\infty}^x \lambda \exp(-\lambda t) dt, \text{ avec : } \lambda = 1$$

Code R

```
N=100 taille de lechantillon
X=rexp( N, rate=1);Y=numeric(N);V=numeric(N)
FN=function(x)
for(i in 1 :N)
if(X[i]<=x)
Y[i]=1

else
Y[i]=0

mean(Y)

a=min(X);b=max(X);x=seq(a,b,length=N)
for(i in 1 :N)
V[i]=FN(x[i])
plot(x,V,xlab="x",ylab="Fn(X)",      main="Distribution      Non      parame-
trique",type="l",col=4, lwd= 2)
lines(x,pexp(x), lty=2,lwd=2)
```

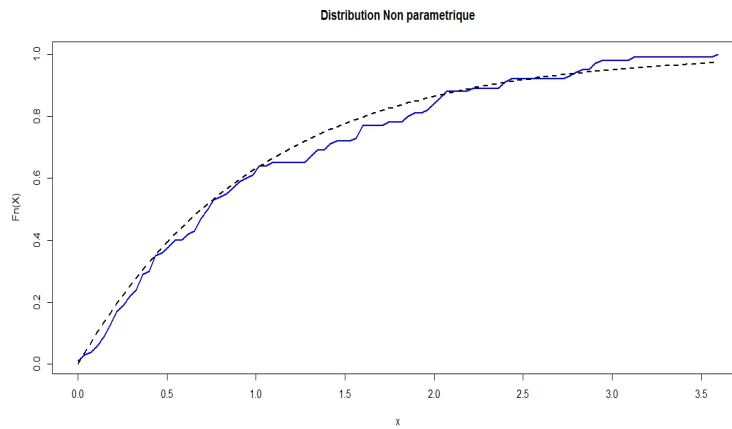


FIGURE 1.1 – Simulation d'une distribution empirique

Les Propriétés élémentaires de la fonction de répartition empirique

1. $F_n(x)$ est un estimateur sans biais de $F(x)$:

$$\begin{aligned}
 E(F_n(x)) &= \frac{1}{n} \sum_{i=1}^n E(I_{]-\infty, x]}(X_i), \\
 &= P(X < x), \\
 &= F(x)
 \end{aligned}$$

2. Variance de l'estimateur $F_n(x)$.

Il est facile de montrer que, pour tout x , la variance de l'estimateur $F_n(x)$ est donnée par :

$$\text{Var}[F_n(x)] = \frac{F(x)(1 - F(x))}{n} \longrightarrow 0, \quad n \rightarrow \infty.$$

3. L'erreur quadratique moyenne de l'estimateur $F_n(x)$

$$\begin{aligned}\mathbb{E}[F_n(x) - F(x)]^2 &= \mathbb{E}[F_n(x) - \mathbb{E}[F_n(x)] + \mathbb{E}[F_n(x)] - F(x)]^2, \\ &= \text{Var}[F_n(x)] + [\text{Biais}\{F_n(x)\}]^2, \\ &= \frac{1}{n}F(x)(1 - F(x)).\end{aligned}\tag{1.2}$$

Donc, quand $n \rightarrow \infty$, on a : $E[F_n(x) - F(x)]^2 \rightarrow 0$. Alors pour tout point x . L'estimateur $F_n(x)$ est un estimateur consistant de $F(x)$.

4. Le théorème central-limite donne : $\sqrt{n}(F_n(x) - F(x)) \xrightarrow{\text{loi}} N(0, 1)$.

5. La loi des grands nombres donne : $\forall x \in \mathbb{R} \quad F_n(x) \xrightarrow{P} F(x)$.

6. Théorème de Glivenko-Cantelli fournit la convergence presque sûre uniforme de la suite des fonctions de répartition empiriques : $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0$, ps.

Ensuite, nous présenterons, en détails un résultat concernant la vitesse de convergence de F_n .

2.1 La vitesse de convergence

Théorème 2.1. *soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F et F_n la fonction empirique. Alors, pour tout x on a :*

$$F_n(x) - F(x) = o\left(\sqrt{\frac{\log n}{n}}\right) \quad (P.C.O)$$

Preuve 2.1. *D'après la définition précédente de vitesse de convergence, il suffit de motrer que : $\exists \varepsilon > 0$, tel que*

$$\sum_{n=1}^{\infty} P\left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}}\right] < \infty.$$

On note :

$$A = \sum_{n=1}^{\infty} P \left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}} \right].$$

On a :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(X_i) \quad \text{et} \quad F(x) = E[F_n(x)].$$

Alors l'expression finale est donnée par

$$\begin{aligned} A &= \sum_{n=1}^{\infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n I_{]-\infty, x]}(X_i) - \frac{1}{n} E \left[\sum_{i=1}^n I_{]-\infty, x]}(X_i) \right] \right| > \varepsilon \sqrt{\frac{\log n}{n}} \right], \\ &= \sum_{n=1}^{\infty} P \left[\left| \sum_{i=1}^n I_{]-\infty, x]}(X_i) - E \left[\sum_{i=1}^n I_{]-\infty, x]}(X_i) \right] \right| > \varepsilon \sqrt{n \log n} \right]. \end{aligned}$$

On utilise l'inégalité de Bernshtien Fréchet.

On sait que :

$$0 \leq I_{]-\infty, x]}(X_i) \leq 1,$$

en notant $t = \varepsilon \sqrt{n \log n}$. donc , par identification $\alpha_i = 0$ et $\beta_i = 1$.

nous trouvons :

$$\begin{aligned} A &= \sum_{n=0}^{\infty} P \left(\sum_{i=0}^n |I_{]-\infty, x]}(X_i) - E[I_{]-\infty, x]}(X_i)]| > t \right) \leq \sum_{n=0}^{\infty} 2 \exp \left(\frac{-2\varepsilon^2 n \log n}{n} \right), \\ &\leq \sum_{n=0}^{\infty} 2 \exp \left(\log n^{-2\varepsilon^2} \right), \\ &\leq \sum_{n=0}^{\infty} 2n^{-2\varepsilon^2}, \\ &\leq \sum_{n=0}^{\infty} \frac{2}{n^{2\varepsilon^2}} < \infty. \end{aligned}$$

Donc, $\exists \varepsilon > 0$ tel que :

$$2\varepsilon^2 > 1 \Rightarrow \varepsilon > \frac{1}{\sqrt{2}}.$$

Pour que la série converge.

D'où

$$\sum_{n=0}^{\infty} P \left[|F_n(x) - F(x)| > \varepsilon \sqrt{\frac{\log n}{n}} \right] < \infty.$$

De cela, nous concluons que

$$F_n(x) - F(x) = o \left(\sqrt{\frac{\log n}{n}} \right) \text{ P.C.O.}$$

3 Estimation non-paramétrique par la méthode de noyaux

Le concept de noyau a d'abord été introduit par ROSENBLATT(1956), mais c'est CACOULOS(1966) qui a été le premier à utiliser le terme " noyau" pour désigner la fonction que l'on utilise dans les méthodes non paramétriques.

Supposons que nous observons n variables aléatoires indépendantes identiquement distribuées (i.i.d). (X_1, X_2, \dots, X_n) de fonction de répartition inconnue F . On appelle estimateur à noyau pour F noté

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right).$$

Où H est une fonction de répartition et h_n est une suite des nombres réels positifs.(le paramètre de lissage).

Proposition 3.1. Soit X_1, X_2, \dots, X_n un n -échantillon de X de fonction de répartition F . Soit $\tilde{F}_n(x)$ un estimateur à noyau H vérifiant $\int_{\mathbb{R}} y H'(y) dy < \infty$. Alors : $\tilde{F}_n(x)$ est un estimateur asymptotiquement sans biais de $F(x)$

Preuve 3.1. Il suffit de montrer que

$$E \left[\tilde{F}_n(x) \right] \rightarrow F(x), \quad \text{Si } n \rightarrow \infty.$$

$$\begin{aligned}
E \left[\tilde{F}_n(x) \right] &= E \left[\frac{1}{n} \sum_{i=1}^n H \left(\frac{x - X_i}{h_n} \right) \right], \\
&= E \left[H \left(\frac{x - X_1}{h_n} \right) \right], \\
&= \int_{\mathbb{R}} H \left(\frac{x - z}{h_n} \right) f(z) dz, \\
&= \left[H \left(\frac{x - z}{h_n} \right) F(z) \right]_{-\infty}^{+\infty} + \frac{1}{h_n} \int_{-\infty}^{+\infty} H' \left(\frac{x - z}{h_n} \right) F(z) dz,
\end{aligned}$$

Par un changement de variables, on pose :

$$\begin{aligned}
\frac{x - z}{h_n} &= y \implies z = x - yh_n, \\
\implies dz &= -h_n dy.
\end{aligned}$$

On voit que le premier terme est nul car il est composé de deux fonctions de répartition.

Alors, $E \left[\tilde{F}_n(x) \right]$ devient :

$$E \left[\tilde{F}_n(x) \right] = \int_{-\infty}^{+\infty} H'(y) F(x - yh_n) dy,$$

D'après le développement limité de $F(x - yh_n)$ on a :

$$F(x - yh_n) = F(x) - h_n y F'(x) + o(h_n^2),$$

Donc ;

$$\begin{aligned}
E \left[\tilde{F}_n(x) \right] &= \int_{-\infty}^{+\infty} H'(y) (F(x) - h_n y F'(x) + o(h_n)), \\
&= F(x) \int_{-\infty}^{+\infty} H'(y) dy - h_n f(x) \int_{-\infty}^{+\infty} y H'(y) dy + o(h_n) \int_{-\infty}^{+\infty} H'(y) dy, \\
&= F(x) + h_n \int_{-\infty}^{+\infty} y H'(y) dy + o(h_n^2),
\end{aligned}$$

On pose :

$$M = \int_{-\infty}^{+\infty} yH'(y) dy.$$

Si $M < +\infty$ Alors ;

$$E \left[\tilde{F}_n(x) \right] - F(x) = -h_n M + o(h_n^2) \longrightarrow 0.$$

3.1 Etude asymptotique du biais et de la variance

Considérons l'estimateur à noyau

$$\tilde{F}_n(x) = \frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right),$$

La variance de l'estimateur à noyau est donnée par :

$$\begin{aligned} \text{Var} \left[\tilde{F}_n(x) \right] &= \text{Var} \left[\frac{1}{n} \sum_{i=1}^n H\left(\frac{x - X_i}{h_n}\right) \right], \\ &= \frac{1}{n} \sum_{i=1}^n \text{Var} \left[H\left(\frac{x - X_i}{h_n}\right) \right], \\ &= \frac{1}{n} \text{Var} \left[H\left(\frac{x - X_1}{h_n}\right) \right] \quad \text{car les } X_i \text{ sont identiquement distribuées} \\ &= \frac{1}{n} \left\{ E \left[H^2\left(\frac{x - X_1}{h_n}\right) \right] - \left(E \left[H\left(\frac{x - X_1}{h_n}\right) \right] \right)^2 \right\}, \\ &= \frac{1}{n} \left\{ \int_{\mathbb{R}} H^2\left(\frac{x - z}{h_n}\right) \cdot f(z) dz - F^2(x) \right\} \quad \text{car } E \left[H\left(\frac{x - X_1}{h_n}\right) \right] \rightarrow F(x), \\ &= \frac{1}{n} \left\{ \left[H^2\left(\frac{x - z}{h_n}\right) \cdot F(z) \right]_{-\infty}^{+\infty} + \int_{\mathbb{R}} \left(H^2\left(\frac{x - z}{h_n}\right) \right)' F(z) dz - F^2(x) \right\}. \end{aligned}$$

On pose $y = \frac{x - z}{h_n}$, alors on déduit

$$\text{Var} \left[\tilde{F}_n(x) \right] = \frac{1}{n} \left\{ F(x) \int_{\mathbb{R}} y (H^2(y))' dy - F^2(x) \right\}.$$

On sait que

$$\int_{\mathbb{R}} y H'(y) dy < \infty \implies \int_{\mathbb{R}} y (H^2(y))' dy < \infty.$$

D'où, $\text{Var} \left[\tilde{F}_n(x) \right] \rightarrow 0$ quand $n \rightarrow \infty$.

L'erreur quadratique moyenne (en anglais : " Mean squared error , MSE) de l'estimateur à noyau est donnée par $\text{MSE}(\tilde{F}_n) = E \left[\tilde{F}_n(x) - F(x) \right]^2$

$$\begin{aligned} \text{MSE}(x) &= E \left[\tilde{F}_n(x) - E \left[\tilde{F}_n(x) \right] + E \left[\tilde{F}_n(x) \right] - F(x) \right]^2, \\ &= \text{Var} \left[\tilde{F}_n(x) \right] + \left[\text{Biais} \left\{ \tilde{F}_n(x) \right\} \right]^2, \end{aligned}$$

D'après les résultats précédents on a montrer que :

$$\text{Var} \left\{ \tilde{F}_n(x) \right\} \rightarrow 0 \quad \text{et} \quad \text{Biais} \left\{ \tilde{F}_n(x) \right\} \rightarrow 0, \quad \text{si} \quad n \rightarrow \infty.$$

Alors ;

$$E \left[\tilde{F}_n(x) - F(x) \right]^2 \rightarrow 0.$$

Chapitre 2

Modèle de densité :

Tout au long de ce chapitre, on suppose que les observations X_1, \dots, X_n sont des variables indépendantes de même loi (i.i.d) de densité f . Pour simplifier, on suppose que les X_i sont à valeurs réelles et que f est la densité par rapport à la mesure de Lebesgue sur \mathbb{R} . Par conséquent, $P(X_i \in [a, b]) = \int_a^b f(x)dx$, $\forall a, b \in \mathbb{R}$. De plus on supposera que f est deux fois continûment différentiable.

1 Estimation par Histogramme

L'estimateur le plus rudimentaire pour estimer une densité est l'histogramme des fréquences. Supposons que l'on ait x_1, \dots, x_n , n observations issues d'une même loi de probabilité de densité f , où f est à support borné $[a, b[$. Pour estimer cette densité f par la méthode de l'histogramme, ce qui revient à approcher f par une fonction en escaliers, on découpe $[a, b[$ en k classes $[\alpha_i, \alpha_{i+1}[$ où $i = 1, \dots, k$, avec $a = \alpha_1$ et $b = \alpha_{k+1}$. L'estimateur histogramme s'écrit alors : $\forall t \in [a, b[$, $\exists i = 1, \dots, k$ tel que $t \in [\alpha_i, \alpha_{i+1}[$ et

$$\hat{f}_n(t) = \frac{f_i}{\alpha_{i+1} - \alpha_i},$$

où f_i est la fréquence du nombre de points de la classe correspondante. Ce que l'on peut encore écrire plus concisément : $\forall t \in [a, b[$,

$$\hat{f}_n(t) = \sum_{i=1}^k \frac{f_i}{\alpha_{i+1} - \alpha_i} \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(t),$$

où

$$f_i = \frac{1}{n} \sum_{j=1}^n \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(x_j),$$

il est facile de voir que l'estimateur du type histogramme de densité s'écrit

$$\hat{f}_n(t) = \sum_{i=1}^k \frac{1}{n(\alpha_{i+1} - \alpha_i)} \sum_{j=1}^n \mathbf{1}_{[\alpha_i, \alpha_{i+1}[}(x_j) (1),$$

Pour s'implifier les notations, on supposera maintenant les classes de même largeur, c'est-à-dire que pour tout $i = 1, \dots, k$; $\alpha_{i+1} - \alpha_i = b(n)$. Il est aisé de remarquer que \hat{f}_n est une densité de probabilité. Si on pense à la convergence de cet estimateur, il est clair que \hat{f}_n sera d'autant plus proche de la vraie densité f que les largeurs de classe seront plus étroites, d'où la nécessité d'imposer que $b(n) \rightarrow 0$ quand $n \rightarrow \infty$. En revanche, il ne faut pas que $b(n)$ tende trop vite vers 0, sinon on pourrait avoir des classes ne contenant aucun point, et donc une fonction en escalier \hat{f}_n avec des marches d'ordonnée nulle, très éloignée de la réalité. Il faut donc que, certes $b(n)$ tende vers 0 avec n , et que, malgré cela, il «tombe» de plus en plus de points dans chaque classe, ce que l'on peut résumer dans la condition :

$$nb(n) \rightarrow \infty \quad \text{quand} \quad n \rightarrow \infty.$$

L'erreur quadratique moyenne $E \left[\hat{f}_n(x) - f(x) \right]^2$ est de l'ordre de $n^{-2/3}$, pour un choix optimal de $b(n)$.

2 Estimateur à noyau

L'estimation de la densité par histogrammes est une méthode naturelle très répandue car elle est facilement implémentable. Cependant, l'estimateur de densité fourni par un histogramme ne peut pas être adapté à la situation assez courante où nous disposons d'une information a priori sur la régularité de la densité à estimer. Plus précisément, si l'on sait par avance que la densité de l'échantillon observé est, par

exemple, deux fois continûment différentiable, on aurait naturellement envie d'estimer cette densité par une fonction qui, elle aussi, est deux fois continûment différentiable. Or, les histogrammes sont des fonctions qui ne sont même pas continues. Il est naturel alors de vouloir "lisser" les histogrammes. On s'attend alors à ce que le résultat du lissage améliore non seulement l'aspect visuel de l'estimateur, mais produise de plus un estimateur plus proche de la vraie densité que l'estimateur par histogramme.

2.1 Définition et propriétés élémentaires

Soit $x \in \mathbb{R}$ et $h > 0$. Si l'on suppose que x est le centre d'une classe de l'histogramme et que h est la longueur des classes, l'estimateur de $f(x)$ par histogramme peut s'écrire comme

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \mathbf{1}(|X_i - x| \leq h/2) = \frac{1}{nh} \sum_{i=1}^n \mathbf{1}\left(\frac{|X_i - x|}{h} \leq \frac{1}{2}\right).$$

Une façon de généraliser les histogrammes consiste à utiliser la formule ci-dessus pour tout $x \in \mathbb{R}$ et pas seulement pour les centres des classes. Cette généralisation est certainement utile, car elle conduit vers un estimateur qui est constant par morceaux comme les histogrammes, mais à l'avantage d'avoir des plateaux de longueurs variables. Cependant, cela ne nous conduit pas vers un estimateur continu. On remarque aisément que la discontinuité de l'estimateur défini ci-dessus est une conséquence de la discontinuité de la fonction indicatrice. Par conséquent, en remplaçant $\mathbf{1}(|z| \leq \frac{1}{2})$ par une fonction K quelconque, on obtient l'estimateur

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

qui est continu et même l fois continûment différentiable.

Exemple fondamentale 2.1. *Dans cet exemple, nous donnons la simulation de l'estimateur à noyau d'une densité de loi normale $N(0;1)$*

Code R

```
n=100
X=rnorm(n) echantillon Noyau Gaussien
K=function(t)(1/sqrt(2*pi))*exp(-0.5*t^2)
h = 0.2parametredelissageh
s = 1000
a = min(X)
b = max(X)
x = seq(a,b,length = s)Intervalle[a,b]
w = numeric(n)
fn = numeric(s)
for(jin1 :s)
  for(iin1 :n)
    D < -(x[j] - X[i])w[i] = K(D/h)
  fn[j] = sum(w)/(n * h)
plot(x, fn, xlab = "x", ylab = "fn(x)", type=l, col=4, lwd= 2)
lines(x, dnorm(x), lwd = 2, lty = 2)
```

Dans ce dernier graphe, la ligne discontinue c'est la densité théorique $f(x)$ d'une loi normale centrée réduite et la ligne continue c'est la densité à noyau $f_n(x)$: Ici, le noyau K est une densité normale et le paramètre de lissage est $h = 0.2$ ($h \simeq n^{-1/5}$).

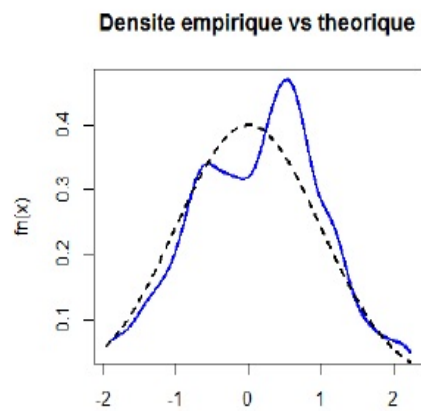


FIGURE 2.1 – Densités théorique et empirique.

Définition 2.1. Soit $K : \mathbb{R} \rightarrow \mathbb{R}$ une fonction quelconque et soit h un réel positif. On appelle estimateur à noyau la fonction

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

où h est un paramètre appelé paramètre de lissage il dépend de n et il vérifie :

$$h \rightarrow 0, \text{ lors que } n \rightarrow \infty,$$

et K est une densité de probabilité appelée noyau

Noyaux usuels :

-le noyau rectangulaire :

$$K(u) = \frac{1}{2} \mathbf{1}_{[-1,1]}(u),$$

-le noyau triangulaire :

$$K(u) = (1 - |u|) \mathbf{1}_{[-1,1]}(u),$$

-le noyau d'Epanechnikov :

$$K(u) = \frac{3}{4} (1 - u^2) \mathbf{1}_{[-1,1]}(u),$$

-Le noyau gaussien

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Les courbes de ces noyaux sont présentées ci-dessous :

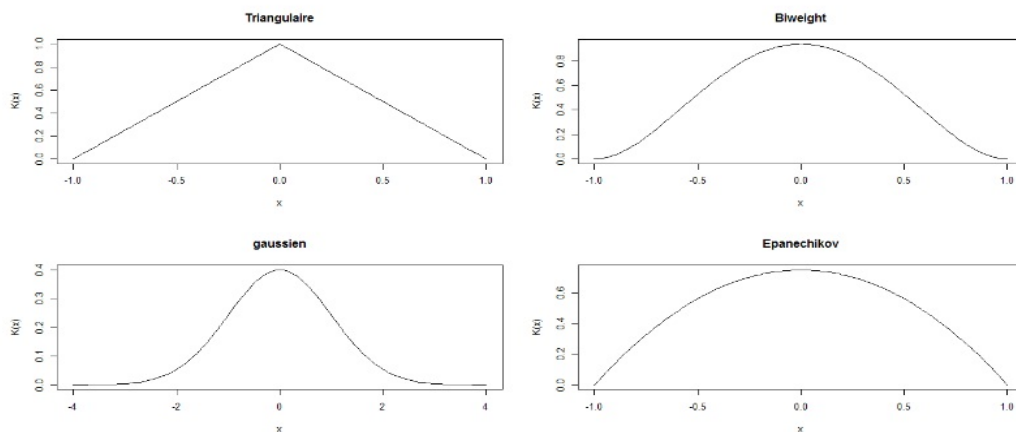


FIGURE 2.2 – Les courbes des noyaux : triangulaire, Biweight, Gaussien et Epanechnikov

Code R

```

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

```

Lemme 2.1. Si K est positive et $\int_{\mathbb{R}} K(u) du = 1$, alors $\hat{f}_h^K(\cdot)$ est une densité de probabilité. De plus \hat{f}_h^K est continue si K est continue.

Preuve 2.1. L'estimateur à noyau est positive et continue car la somme des fonctions positives et continues est elle-même une fonction positive et continue. Il faut donc vérifier que l'intégrale de $\hat{f}_h^K(\cdot)$ vaut un. En effet,

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}_h^K(x) dx &= \int_{\mathbb{R}} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) dx, \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K\left(\frac{X_i - x}{h}\right) dx. \end{aligned}$$

On pose : $u = \frac{X_i - x}{h}$,

noustrouvons $\frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K(u) h du = 1$.

On voit donc que, tout comme l'estimateur par histogramme, l'estimateur à noyau est une densité de probabilité. Il a de plus l'avantage d'être continu à condition que K le soit, ce qui n'était pas le cas pour les histogrammes. Par conséquent, lorsqu'on estime une densité continue, il est naturel de s'attendre que l'estimateur à noyau soit meilleur que l'estimateur par histogramme. Le but de la suite de ce chapitre est de donner des résultats quantitatives caractérisant le gain obtenu par l'utilisation de \hat{f}_h^K par rapport à \hat{f}_h^H .

2.2 Etude du biais et de la variance

Lorsqu'on définit un estimateur à noyau, on a non-seulement le choix de la fenêtre $h > 0$ mais aussi celui du noyau K . Il y a certain nombre de conditions qui sont considérées comme usuelles pour les noyaux et qui permettent d'analyser le risque de l'estimateur à noyau qui en résulte.

HYPOTHESE : On suppose que K vérifie les 4 conditions suivantes :

1. (H.1) : $\int_{\mathbb{R}} K(u) du = 1$.
2. (H.2) : K est une fonction paire ou, plus généralement, $\int_{\mathbb{R}} u K(u) du = 0$.

$$3. (H.3) : \int_{\mathbb{R}} u^2 |K(u)| du < \infty.$$

$$4. (H.4) : \int_{\mathbb{R}} K(u)^2 du < \infty.$$

Proposition 2.1. *Si les trois premières conditions de l'hypothèse K sont remplies et f est une densité bornée dont la dérivée seconde est bornée, alors*

$$\text{Biais} \left(\hat{f}_h^K(x) \right) \leq C_1 h^2,$$

où $C_1 = \frac{1}{2} \sup_{z \in \mathbb{R}} |f''(z)| \int_{\mathbb{R}} u^2 |K(u)| du$, et sous les hypothèses (H.1)-(H.4), on a :

$$\text{Var}(\hat{f}_h^K(x)) \leq \frac{C_2}{nh},$$

avec, $C_2 = \sup_{z \in \mathbb{R}} f(z) \int_{\mathbb{R}} u^2 |K(u)| du$.

Preuve 2.2. *Commençons par calculer le biais :*

$$\begin{aligned} E_f \left[\hat{f}_h^K(x) \right] &= \frac{1}{nh} \sum_{i=1}^n E_f \left[K \left(\frac{X_i - x}{h} \right) \right], \\ &= \frac{1}{nh} \sum_{i=1}^n \int_{\mathbb{R}} K \left(\frac{y - x}{h} \right) f(y) dy, \\ &= \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{y - x}{h} \right) f(y) dy, \\ &= \int_{\mathbb{R}} K(u) f(x + uh) du. \end{aligned}$$

En effectuant un développement limité à l'ordre 2, il vient

$$\begin{aligned} E_f \left[\hat{f}_h^K(x) \right] &= \int_{\mathbb{R}} K(u) f(x + uh) du \\ &= \int_{\mathbb{R}} K(u) \left[f(x) + (uh) f'(x) + \frac{(uh)^2}{2} f''(\zeta_u) \right] du \quad (\zeta_u \in [x, x + uh]) \\ &= f(x) \underbrace{\int_{\mathbb{R}} K(u) du}_{=1} + h f'(x) \underbrace{\int_{\mathbb{R}} u K(u) du}_{=0} + \frac{h^2}{2} \int_{\mathbb{R}} u^2 K(u) f''(\zeta_u) du. \end{aligned}$$

Il en résulte que d'où la première assertion de la proposition. Pour prouver la seconde assertion, on utilise le fait que les variables aléatoires $Y_i = K((X_i - x)/h)$, $i = 1, \dots, n$ sont i.i.d. et que la variance de la somme de variables indépendantes coïncide avec la somme des variances :

$$\begin{aligned}
 \text{Var}_f \left[\hat{f}_h^K(x) \right] &= \frac{1}{(nh)^2} \text{Var}_f \left[\sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) \right], \\
 &= \frac{1}{(nh)^2} \sum_{i=1}^n \text{Var}_f \left[K \left(\frac{X_i - x}{h} \right) \right], \\
 &= \frac{1}{(nh)^2} \times n \times \text{Var}_f \left[K \left(\frac{X_1 - x}{h} \right) \right], \\
 &\leq \frac{1}{nh^2} E_f \left[K \left(\frac{X_1 - x}{h} \right)^2 \right], \\
 &= \frac{1}{nh^2} \int_{\mathbb{R}} K \left(\frac{y - x}{h} \right)^2 f(y) dy, \\
 &= \frac{1}{nh} \int_{\mathbb{R}} K(u)^2 f(x + uh) du, \\
 &\leq \frac{1}{nh} \underbrace{\sum_z f(z) \int_{\mathbb{R}} K(u)^2 du}_{\text{}}.
 \end{aligned}$$

C'est exactement ce qu'il fallait démontrer.

2.3 Quelques remarques :

Les évaluations du biais et de la variance que l'on vient de démontrer ont un certain nombre de conséquences résumées ci-dessous.

1. **Vitesse de convergence :** On déduit de la Proposition 2.2 que le risque MISE de $\hat{f}_h^K(x)$ admet la majoration suivante :

$$\text{MISE} \left(\hat{f}_h^K(x) \right) \leq C_1^2 h^4 + \frac{C_2}{nh}$$

On vérifie aisément que la valeur de la fenêtre h qui minimise le majorant du $MISE$ est $h_{opt} = (C_2/4C_1^2)^{1/5} n^{-4/5}$. En injectant cette valeur dans l'expression du MSE on obtient :

$$MSE \left(\hat{f}_{h_{opt}}^K(x) \right) \leq Const.n^{-4/5}.$$

Cela montre que la vitesse de convergence de l'estimateur à noyau est de $n^{-4/5}$. Elle est donc meilleure que la vitesse $n^{-2/3}$ obtenue pour les histogrammes. Par conséquent, les estimateurs à noyau sont préférables aux histogrammes lorsqu'il s'agit d'estimer une densité deux fois continument différentiable.

2. **Optimalité de la vitesse** : On peut démontrer qu'il est possible d'estimer f à une vitesse meilleure que $n^{-4/5}$ sans imposer des hypothèses supplémentaires (de régularité ou de structure) sur la densité inconnue f .
3. **Sur-lissage et sous-lissage** : Lorsque la fenêtre h est très petit, le biais de l'estimateur à noyau est très petit face à sa variance et c'est cette dernière qui détermine la vitesse de convergence du risque quadratique. Dans ce type de situation, l'estimateur est très volatile et on parle de sous-lissage (under-smoothing, en anglais). En revanche, lorsque h grandit, la variance devient petite et c'est le biais qui devient dominant. L'estimateur est alors très peu variable et est de moins à moins influencé par les données. On parle alors d'un effet de sur-lissage (over-smoothing en anglais). En pratique, il est primordial de trouver la bonne dose de lissage qui permet d'éviter le sous-lissage et le sur-lissage.
4. **Comparaison avec le cadre paramétrique** : Dans la théorie statistique paramétrique classique, la vitesse de convergence usuelle pour le risque quadratique est de $n^{-4/5}$ obtenue pour l'estimateur à noyau est meilleur que $n^{-2/3}$ obtenu pour l'estimateur par histogramme mais reste quand-même inférieure à la vitesse paramétrique. Ceci est tout à fait naturel et traduit la complexité de l'estimation non-paramétrique comparée à l'estimation paramétrique. On peut remarquer également que lorsque la régularité de la densité tend vers l'infini ($\beta \rightarrow \infty$), la vitesse de convergence se rapproche de plus en plus de la "vitesse paramétrique"

Chapitre 3

Régression non paramétrique

Soient les observations de couples $(x_i; y_i)$; $i = 1, \dots, n$, où les x_i représentent les valeurs observées de la variable explicative X et les y_i représentent celles de la variable dépendante Y . La méthode la plus communément utilisée pour étudier la relation entre ces deux variables est la régression linéaire simple, qui suppose un modèle de la forme

$$Y_i = \alpha + \alpha_1 X_i + \epsilon_i; i = 1, \dots, n$$

où les erreurs aléatoires ϵ_i sont non corrélées, de moyenne nulle et de variance σ^2 . Cette méthode possède l'avantage d'être facile à interpréter et, lorsque les postulats sur les résidus ϵ_i sont vérifiés, elle permet de faire des tests d'hypothèses statistiques formels sur les paramètres. Par contre, il arrive que la linéarité de la relation ne soit pas toujours respectée. Dans ce cas, il est préférable de choisir un modèle plus flexible qui reflète mieux la relation entre X et Y . Le modèle de régression non paramétrique suivant peut alors être employé :

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

où $r(x_i)$ représente la moyenne conditionnelle de la courbe de régression, c'est-à-dire

$$r(x) = E(Y|X = x).$$

L'objectif est d'estimer la fonction r pour expliquer et prédire Y à partir de X . Pour cela on dispose des réalisations de n couples de variables $(X_1, Y_1), \dots, (X_n, Y_n)$. On va supposer que les (X_i, Y_i) sont indépendants.

1 Estimateur de Nadaraya-Watson

On suppose que les (X_i, Y_i) admettent une densité $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ et on suppose que pour tout $x > 0$, $f_X(x) = \int f(x, y) dy > 0$, (f_X est la densité de X). On peut alors écrire

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

Un estimateur naturel de $r(x)$ est donné par

$$\hat{r}_n(x) = \begin{cases} \int y \frac{\hat{f}(x, y)}{\hat{f}_X(x)} dy & \text{si } \hat{f}_X(x) \neq 0 \\ 0 & \text{si } \hat{f}_X(x) = 0 \end{cases}$$

Soit $h \rightarrow 0$ quand $n \rightarrow \infty$, alors les estimateurs de $f(x, y)$ et $f(x)$ est donné par

$$\begin{aligned} \hat{f}_{n,X}(x) &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \\ \hat{f}_n(x, y) &= \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right), \end{aligned}$$

Définition 1.1. Si K est un noyau d'ordre 1 alors $\forall x \in \mathbb{R}$, estimateur de la fonction de régression est donné par :

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

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

Plus généralement, des estimateurs de $r(x)$ peuvent s'écrire :

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

$$\text{où } \omega_{n,i}(x) = \begin{cases} \frac{K\left(\frac{X_i-x}{h}\right)}{\sum_{i=0}^n K\left(\frac{X_i-x}{h}\right)} & \text{si } \sum_{i=0}^n K\left(\frac{X_i-x}{h}\right) \neq 0 \\ 0 & \text{sinon} \end{cases}.$$

$\omega_{n,i}(x)$ sont des poids dépendant de x et de (X_1, \dots, X_n) .

Remarquons aussi que, si $\sum_{i=0}^n K\left(\frac{X_i-x}{h}\right) = 0$ i.e. si x se trouve dans une zone où il n'y a pas de X_i , alors $\hat{r}(x) = 0$. Et sinon, comme $\sum_{i=0}^n \omega_{n,i}(x) = 1$, alors Y_i est une moyenne pondérée des Y_i qui correspondent aux points X_i proches de x .

Dans la pratique, comme K est en général symétrique et décroissant sur \mathbb{R}^+ , le poids associé à Y_i dans cette moyenne pondérée est d'autant plus grand que X_i est proche de x . Les Y_i associés à des points X_i qui sont loin de x n'ont pas ou peu d'impact sur l'estimation de $r(x)$. C'est en cela que la méthode est locale, au contraire de l'EMC non paramétrique.

Preuve 1.1. Supposons donc que $\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) \neq 0$. alors

$$\begin{aligned} \hat{r}_n(x) &= \int \frac{y \hat{f}_n(x, y)}{\hat{f}_{n,X}(x)} dy, \\ &= \frac{1}{\hat{f}_{n,X}(x)} \int y \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) K\left(\frac{Y_i-y}{h}\right) dy, \\ &= \frac{nh}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)} \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) \int y K\left(\frac{Y_i-y}{h}\right) dy, \\ &= \frac{1}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) \frac{1}{h} \int y K\left(\frac{Y_i-y}{h}\right) dy, \\ &= \frac{1}{\sum_{i=1}^n K\left(\frac{X_i-x}{h}\right)} \sum_{i=1}^n K\left(\frac{X_i-x}{h}\right) Y_i. \end{aligned}$$

Pour la dernière ligne, on a utilisé le fait que

$$\frac{1}{h} \int y K\left(\frac{Y_i - y}{h}\right) dy = \frac{1}{h} \int (Y_i - uh) K(u) h du = Y_i K(u) du - h \int K(u) du = Y_i$$

Remarque 1.1. Si K est continu, positif et à support sur \mathbb{R} (par exemple le noyau gaussien) alors $\hat{r}_n(x)$ est continu.

Remarque 1.2. Si la densité f_X est connue. Dans ce cas, il est préférable d'utiliser

$$\hat{r}_n(x) = \begin{cases} \int y \frac{\hat{f}(x, y)}{f_X(x)} dy & \text{si } f_X(x) \neq 0 \\ 0 & \text{si } f_X(x) = 0 \end{cases}$$

Proposition 1.1. On suppose f_X connue. On s'intéresse à l'estimation de $r(x)$ pour x fixé. Soit K un noyau d'ordre 1. On suppose de plus les hypothèses suivantes

- (H1) $f_X(x) > 0$.
- (H2) Il existe $\epsilon > 0$ tel que les fonctions f_X et r sont continument dérivables sur $[x - \epsilon, x + \epsilon]$
- (H3) Pour tout $y \in \mathbb{R}$ et si $|u| \leq \epsilon$ on a :

$$|f(x + u, y) - f(x, y)| \leq M(x, y)\epsilon,$$

où

$$\int y^2 M(x, y) dy < \infty \quad \text{et} \quad \int y^2 f(x, y) dy < \infty$$

- (H4) K est un noyau à support dans $[-1, 1]$ et de carré intégrable.

Proposition 1.2. Sous les hypothèses (H1)-(H4), il existe une constante $C(x)$ (dépendant de x) telle que

$$E[(\tilde{r}_n(x) - r(x))^2] \leq C(x)(h^2 + \frac{1}{nh}),$$

Remarque 1.3. Si on choisit une fenêtre h telle que $h \asymp n^{-1/3}$ (le signe signifie “de l'ordre de ”), il existe une constante $C(x)$ telle que

$$E[(\tilde{r}_n(x) - r(x))^2] \leq C(x)n^{-2/3}$$

Preuve 1.2. On utilise la décomposition biais/variance :

$$E[(\tilde{r}_n(x) - r(x))^2] = \text{Biais}^2 + \text{variance}$$

Etude asymptotique du biais

L'étude asymptotique du biais basée sur

1. le théorème du transfert et le théorème de Fubini .
2. les hypothèses (H1)-(H4).
3. l'inégalité des accroissements finis.

$$\begin{aligned} E[(\tilde{r}_n(x))] &= \mathbf{E} \left(\frac{1}{nhf_X(x)} \sum_{i=1}^n Y_i K \left(\frac{X_i - x}{h} \right) \right), \\ &= \frac{1}{hf_X(x)} \mathbf{E} \left[Y_1 K \left(\frac{X_1 - x}{h} \right) \right], \\ &= \frac{1}{hf_X(x)} \int \int y K \left(\frac{t - x}{h} \right) f(t, y) dt dy, \\ &= \frac{1}{f_X(x)} \int \int y K(v) f(x + vh, y) dv dy. \end{aligned}$$

En outre on a :

$$\begin{aligned} r(x) &= \mathbf{E} [Y | X = x] \\ &= \int y f_Y(y | X = x) dy \\ &= \int y \frac{f(x, y)}{f_X(x)} dy. \end{aligned}$$

Donc :

$$r(x)f_X(x) = \int yf(x, y)dy.$$

Ensuite, nous pouvons d eduire que

$$r(x + vh)f_X(x + vh) = \int yf(x + vh, y)dy,$$

Alors :

$$\begin{aligned} E[\tilde{r}_n(x)] - r(x) &= \frac{1}{f_X(x)} \left[\int \int yK(v)f(x + vh, y)dvd y - \int yf(x, y)dy \right], \\ &= \frac{1}{f_X(x)} \left[\int \int yK(v)f(x + vh, y)dvd y - \int \int yK(v)f(x, y)dvd y \right], \\ &= \frac{1}{f_X(x)} \left[\int K(v)f(x + vh)r(x + vh)dv - \int K(v)r(x)f_X(x)dv \right], \\ &= \frac{1}{f_X(x)} [(v)[f_X(x + vh) - f_X(x) + f_X(x)]r(x + vh)dv], \\ &- \left[\int_{-1}^1 K(v)f_X(x)[r(x + vh) - r(x)]dv \right]. \end{aligned}$$

On applique l'in egalit e des accroissements finis sur r et f_X et sous l'hypoth ese (H4), il existe une constante $C(x)$ telle que, pour tout $|u| \leq \epsilon$,

$$\begin{aligned} |r(x + u) - r(x)| &\leq C(x)u, \\ |f_X(x + u) - f_X(x)| &\leq C(x)u. \end{aligned}$$

On applique ces in egalit es avec $u = vh$ pour $|v| \leq 1$ et $|h| \leq \epsilon$, nous obtenons :

$$\begin{aligned} E[\tilde{r}_n(x)] - r(x) &\leq \frac{1}{f_X(x)} \left[\int_{-1}^1 |K(v)||f_X(x + vh) - f_X(x)||r(x + vh)|dv \right], \\ &+ \int_{-1}^1 |K(v)||r(x + vh) - r(x)|dv, \\ &\leq \frac{C(x)}{f_X(x)} \left[\int_{-1}^1 |K(v)||hv||r(x + vh)|dv \right] + C(x) \int_{-1}^1 |K(v)||hv|dv. \end{aligned}$$

Ensuite, d'après l'hypothèse (H2), il existe une constante $c(x)$ telle que $|r(x + hv)| \leq c(x)$ pour tout $|h| \leq \epsilon$ et tout $|v| \leq 1$. On a :

$$|E[\tilde{r}_n(x)] - r(x)| \leq C_1(x)h,$$

avec,

$$C_1(x) = C(x) \left(\frac{c(x)}{f_X(x)} + 1 \right) \int |K(v)| dv.$$

Etude asymptotique de la variance

$$\begin{aligned} \text{Var}[(\tilde{r}_n(x))] &= \text{Var} \left(\frac{1}{nhf_X(x)} \sum_{i=1}^n Y_i K \left(\frac{X_i - x}{h} \right) \right), \\ &= n \text{Var} \left(\frac{1}{nhf_X(x)} Y_1 K \left(\frac{X_1 - x}{h} \right) \right), \\ &= n \frac{1}{n^2 h^2 f_X^2(x)} \text{ar} \left(Y_1 K \left(\frac{X_1 - x}{h} \right) \right), \\ &\leq \frac{1}{nh^2 f_X^2(x)} E \left(Y_1^2 K^2 \left(\frac{X_1 - x}{h} \right) \right), \\ &= \frac{1}{nh^2 f_X^2(x)} \int y^2 K^2 \left(\frac{t - x}{h} \right) f(t, y) dt dy, \\ &= \frac{1}{nh f_X^2(x)} \int y^2 K^2(v) f(x + vh, y) dv dy. \end{aligned}$$

Sous l'hypothèse (H.3), on obtient : l'expression finale est donnée par :

$$\begin{aligned} \text{Var}[(\tilde{r}_n(x))] &\leq \frac{1}{nh f_X^2(x)} \left(\int y^2 K^2(v) M(x, y) \epsilon dv dy + \int y^2 K^2(v) f(x, y) \epsilon dv dy \right), \\ &= \frac{\int K^2(v)}{nh f_X^2(x)} \left(\epsilon \int y^2 M(x, y) dv dy + \int y^2 f(x, y) \epsilon dv dy \right) \end{aligned}$$

Finalement, si $|h| \leq \epsilon$, la moajoration de la partie de variance est donnée par :

$$\text{Var}[(\tilde{r}_n(x))] \leq \frac{C_2(x)}{nh}$$

où

$$C_2(x) = \frac{\int K^2(v)}{nhf_X^2(x)} \left(\epsilon \int y^2 M(x, y) dv dy + \int y^2 f(x, y) \epsilon dv dy \right).$$

Cette quantité est finie d'après les hypothèses (H.3) et (H.4)

Calcul du risque quadratique

$$E[(\tilde{r}_n(x) - r(x))^2] \leq C_1^2(x)h^2 + \frac{C_2(x)}{nh}.$$

On équilibre les deux termes

$$h^2 \approx \frac{1}{nh} \iff h \approx n^{-\frac{1}{3}},$$

et si on choisit une fenêtre $h^* = cn^{-\frac{1}{3}}$ avec c une constante positive, on a

$$E[(\tilde{r}_n(x) - r(x))^2] \leq C_s(x)n^{-2/3}.$$

2 L'estimation par la méthode des polynômes locaux dans le cas réel

2.1 Le principe de la méthode

L'estimation de la fonction de régression par la méthode des polynômes locaux est fondée sur une simple généralisation de l'estimateur de Nadarya-Watson appelé à noyaux. L'idée principale de cette approche est de considérer le problème de la régression sous l'angle des moindres carrés en considérant la fonction de régression $r(\cdot)$ comme solution d'un problème de moindres carrés. Par convenance, nous rappelons la définition de l'estimateur à noyau :

lorsque $K \geq 0$,

$$\hat{r}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} = \frac{\hat{g}_n(x, y)}{\hat{f}_n(x)}.$$

où K est un noyau et $h = h_n$ est le paramètre de lissage. Nous avons, lorsque $K \geq 0$,

$$\{\widehat{g}_n(x, y) - \widehat{r}_n(x)\widehat{f}_n(x)\} = 0.$$

L'estimateur de la régression $\widehat{r}_n(x)$ peut donc être regardé comme la solution du problème de moindres carrés pondérés suivant :

$$\min_{a_0 \in \mathbb{R}} \sum_{i=1}^n (Y_i - a_0)^2 K\left(\frac{X_i - x}{h}\right). \quad (3.1)$$

En d'autres termes, l'estimateur $\widehat{r}_n(x)$ est obtenu par une approximation des moindres carrés localement constante des valeurs Y_i .

Plus généralement, définissons l'approximation des moindres carrés localement polynomiale de la façon suivante :

l'estimation localement polynomiale consiste en l'ajustement local d'un polynôme de degré p aux données $\{(X_i, Y_i) : 1 \leq i \leq n\}$.

En effet, soit p un entier naturel fixé. Nous cherchons à ajuster le polynôme

$$a_0 + a_1(\cdot - x) + a_2(\cdot - x)^2 + \dots + a_p(\cdot - x)^p$$

aux données (X_i, Y_i) , via la méthode des moindres carrés pondérés.

Premièrement, on suppose l'existence de la $(p + 1)$ -ième dérivée de la fonction de régression $r(\cdot)$ au point x . Nous pouvons alors approximer localement la fonction de régression $r(x)$ par un polynôme d'ordre p .

Il s'ensuit, via le développement de Taylor autour du point x ,

$$\begin{aligned} r(z) &\approx r(x) + r'(x)(z - x) + \frac{r''(x)}{2}(z - x)^2 + \dots + \frac{r^{(p)}(x)}{p!}(z - x)^p \\ &\approx \sum_{j=0}^p \frac{r^{(j)}(x)}{j!}(z - x)^j =: \sum_{j=0}^p a_j(z - x)^j. \end{aligned} \quad (3.2)$$

Lorsque z est situé dans un voisinage du point x .

A présent, nous ajustons localement le polynôme (3.2) aux données

$\{(X_i, Y_i) : 1 \leq i \leq n\}$ par la méthode des moindres carrés pondérés avec comme fonction de poids $K\left(\frac{X_i - x}{h}\right)$.

Il faut minimiser par rapport au vecteur $a = (a_0, \dots, a_p)^T \in \mathbb{R}^{p+1}$ la quantité suivante :

$$\sum_{i=1}^n (Y_i - r(z))^2 K \left(\frac{X_i - x}{h} \right). \quad (3.3)$$

On remplace $r(X_i)$ par son développement de Taylor :

$$\sum_{i=1}^n \left(Y_i - \sum_{j=0}^p a_j (X_i - x)^j \right)^2 K \left(\frac{X_i - x}{h} \right). \quad (3.4)$$

La solution de ce problème de minimisation peut-être donnée en utilisant les notations matricielles, le problème (3.4) peut se résumer ainsi

$$\min (Y - Qa)^t K (Y - Qa). \quad (3.5)$$

où Q la matrice définie par :

$$Q = \begin{pmatrix} 1 & (X_1 - x) & (X_1 - x)^2 & \cdots & (X_1 - x)^p \\ 1 & (X_2 - x) & (X_2 - x)^2 & \cdots & (X_2 - x)^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (X_n - x) & (X_n - x)^2 & \cdots & (X_n - x)^p \end{pmatrix}$$

et

$$Y = (Y_1, \dots, Y_n)^t \quad \text{et} \quad a = (a_0, \dots, a_p)^t$$

On désigne par K la matrice diagonale de poids :

$$K = \begin{pmatrix} K \left(\frac{X_1 - x}{h_n} \right) & 0 & \cdots & 0 \\ 0 & K \left(\frac{X_2 - x}{h_n} \right) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & K \left(\frac{X_n - x}{h_n} \right) \end{pmatrix}$$

En supposant l'inversibilité de $Q^t K Q$, et d'après la théorie des moindres carrés, le vecteur de solution est donné par

$$\hat{a} = (Q^t K Q)^{-1} Q^t K Y$$

où $\hat{a} = (\hat{a}_0, \dots, \hat{a}_p)^T \in \mathbb{R}^{p+1}$, le vecteur qui minimise l'expression (3.5).

D'après (3.2), la dérivée k -ième $r^{(k)}(x)$ peut-être donc estimer par $\hat{a}_k \times k!$ pour $k = 0, 1, \dots, p$. Il s'ensuit la définition suivante : La statistique

$$r_n^{(k)}(x) = \hat{a}_k \times k!, \quad 0 \leq k \leq p, \quad (3.6)$$

est l'estimateur localement polynomial d'ordre p de la dérivée k -ième de la régression $r^k(x)$.

Remarque 2.1. Lorsque $k = p = 0$, on retrouve bien l'estimateur à noyau $\hat{r}_n(x)$.

2.2 La méthode locale linéaire

Un exemple particulièrement intéressant est le cas $p = 1$ et $k = 0$.

L'estimateur $\hat{r}_n^{LL}(x)$ de la fonction de régression est appelé l'estimateur localement linéaire.

D'après (3.4) et (3.6), il est égale à \hat{a}_0 lorsque $\hat{a} = (\hat{a}_0, \hat{a}_1)$ désigne le vecteur solution de l'équation des moindres carrés suivant :

$$\min_{a_0, a_1 \in \mathbb{R}} \sum_{i=1}^n \{Y_i - a_0 - a_1(X_i - x)\}^2 K \left(\frac{X_i - x}{h_n} \right).$$

Plus explicitement, l'estimateur localement linéaire est défini par :

$$\hat{r}_n^{LL}(x) =: \frac{\hat{g}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{g}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{f}_{n,1}^2(x)}, \quad (3.7)$$

où

$$\hat{f}_{n,j}(x) =: \frac{1}{h_n} \sum_{i=1}^n \left\{ \frac{X_i - x}{h_n} \right\}^j K \left(\frac{X_i - x}{h_n} \right), \quad j = 0, 1, 2,$$

$$\widehat{g}_{n,j}(x) =: \frac{1}{h_n} \sum_{i=1}^n \left\{ \frac{X_i - x}{h_n} \right\}^j K \left(\frac{X_i - x}{h_n} \right), \quad j = 0, 1.$$

Chapitre 4

Application sur des données simulées

La fonction de régression est un outil fondamental pour décrire la relation entre deux variables aléatoires. Dans ce chapitre, nous allons déterminer ce lien par les deux méthodes d'estimation non paramétrique : locale linéaire (L.L) et la méthode de noyaux (L.C). L'objectif principal est de montrer à l'aide de données simulées réelles la supériorité de la méthode d'estimation par polynômes locaux sur la méthode à noyau.

Simulée les données

: on suppose que l'on dispose de l'observation de n variables $Y_i, i = 1, \dots, n$ indépendantes, et l'on se place dans le cadre du modèle de régression standard :

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

avec, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ (i.i.d) et $E(\varepsilon_i) = 0$, $Var(\varepsilon_i) = \sigma^2$. où les prédicteurs x_i sont déterministes et à valeurs dans \mathcal{R} et r est une fonction inconnue que l'on cherche à estimer. On va estimer la fonctions de régression suivante à l'aide d' un estimateur de Nadaraya-Watson et locale linéaire.

Code R

```
n <- 1000
```

```
x <- runif(n)
```

```
Data 1
```

```
rsnr = 5 rapport signal sur bruit
```



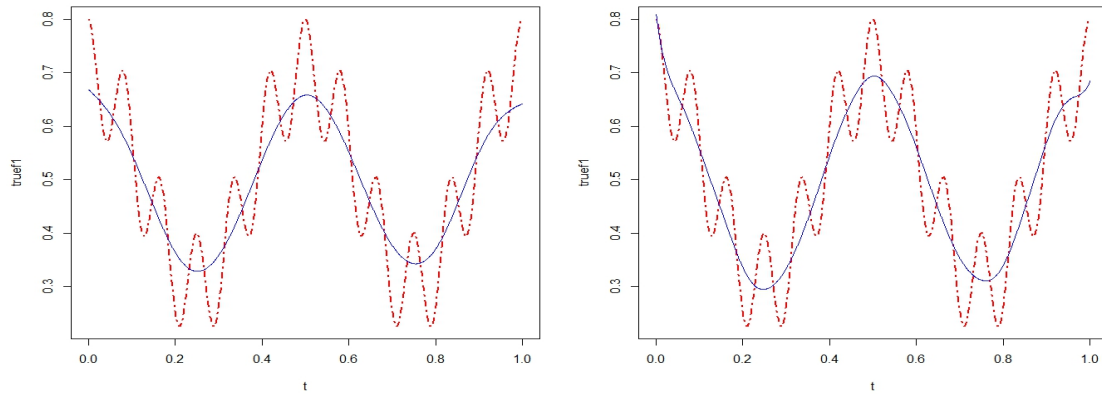
```
r = 0.5 + (0.2 * cos(4 * pi * x)) + (0.1 * cos(24 * pi * x))
sigma1 <- sd(f1)/rsnr Niveau de bruit
Y1 <- r + rnorm(n,mean=0,sd=sigma1)
Visualisation des donnees avec la fonction de regression
dev.new()
plot(x, Y1,type="l",pch=19)
library(KernSmooth)
Estimation a noyau
h = 0.2
hatf1 = locpoly(x, Y1,degree=0,bandwidth=h,gridsize=T,range.x=c(0,1))y
lines(t, r,type="l",col="blue", pch=19)
on choisit un ensemble de valeurs pour h
h.grid = seq(0.001,0.05,length.out = 8)

erreur = rep(0,length(h.grid))

for (j in 1 :length(h.grid))
r= locpoly(x, Y1,degree=1,bandwidth=h.grid[j],gridsize=T,range.x=c(0,1))y
erreur[j] = sum((truef1-hatf1)^2)
```

Résultats de simulation

Avec le programme du logiciel R , on obtient la courbe dans la figure qui représente la fonction théorique comme régression et son estimateur par la méthode de noyaux (figure à gauche) et par la méthode locale linéaire (figure à droite) avec le paramètre de lissage $h = 0.05$.



Commentaires

La comparaison des deux diagrammes indique que l'approche locale linéaire (graphique à droite) donne de meilleurs résultats que la méthode classique du noyau (graphique à gauche). Ceci est confirmé par l'erreur quadratique moyenne (MSE) pour différentes valeurs de paramètre de lissage h comme l'indique le tableau suivant :

Le paramètre (h)	0.010	0.014	0.018	0.023	0.027	0.032	0.036	0.041
MSE(L.C)	0.607	1.238	2.252	3.320	4.187	4.781	5.163	5.428
MSE(L.L)	0.669	1.214	2.146	3.151	3.992	4.586	4.969	5.217

Remarque 0.2. les estimateurs localement polynomiaux sont supérieurs aux estimateurs à noyaux dans le cadre du dispositif expérimental aléatoire, car la variance de cette méthode et celle de la méthode du noyau sont les mêmes, alors que d'après Fan [11], l'estimateur localement linéaire ou localement polynomial a un meilleur biais que l'estimateur à noyau. Il est d'ordre $O(h^{p+1})$, alors qu'il était d'ordre $O(h)$ pour la méthode du noyau.

Conclusion

Dans ce mémoire, nous avons traité les techniques de l'estimation non paramétrique de la fonction de densité de probabilité et la fonction de régression. Nous avons utilisé méthode de noyau (Parzen-Rosenblatt) pour estimer la densité et celle Nadaraya-watson pour la régression. Ces méthodes sont particulièrement utilisées pour évaluer la performance de l'estimateur.

Les principaux résultats que nous avons établi la vitesse de convergence presque-complète de la fonction de répartition empirique en précisant sa vitesse de convergence, lorsque les observations sont indépendantes. Ensuite, nous avons concentrés sur les méthodes d'estimation de la densité : La méthode d'estimation par histogramme, la méthode d'estimation par noyau de la densité et la régression. Nous avons présenté également, les propriétés statistiques de chaque modèle fonctionnels. De plus nous avons construit un estimateur local linéaire de la fonction de régression et nous réalisation des simulations par utilisant le logiciel R qui nous permet d'observer l'influence le paramétré de lissage sur la méthode de noyaux et local linéaire.

Bibliographie

- [1] *S. Bernshtein . Sur le théorème limite du calcul des probabilités. Mathematische Annalen* 85, 237-241. (1922)
- [2] *A. Berlinet. Hierarchies of higher order kernels. Probability Theory and Related Fields*, 94, 489-504, (1993).
- [3] *S. Bianconcini. A reproducing kernel perspective of smoothing spline estimators. Department of Statistics, University of Bologna, Via Belle Arti, 41 - 40126 Bologna, Italy, (2008).*
- [4] *I.J. Bienaymé. Considérations à l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés. C.R. Acad. Sci. Paris* 37, 309-324. (1853).
- [5] *N. Breaz and M. Aldea. On the smoothing spline regression models. Acta Universitatis Apulensis. No 15, (2008).*
- [6] *D. Blondin. Lois limites uniformes et estimation non-paramétrique de la régression. Thèse Doctorat, Université Paris 6, 1 - 26., (2004).*
- [7] *D. Bosq. Nonparametric Statistics for Stochastic Processes. Université Pierre et Marie Curie . Institut de Statistique. 1996*
- [8] *R. Cao, M. A. Delgado, and W. Gonzalez-Manteiga. Non parametric curve estimation : An overview. Investigaciones Economicas, vol. XXI (2), 209-252, (1997).*
- [9] *Y. Cao. Inégalités d'oracle pour l'estimation de la régression. Thèse Doctorat, Université de Provence, (2008).*
- [10] *L. Devroye, and L. Györfi, (1985) Nonparametric Density Estimation : The L1 View. John Wiley Sons, New York.*

-
- [11] J. Fan. *Design-adaptive nonparametric regression*. *J. Amer. Statist. Assoc.*, **87**, 998-1004. (1992)
- [12] F. Ferraty, and P. Vieu. *Nonparametric functional data analysis. Theory and Practice*. Springer Verlag. (2006)
- [13] D. Foata, and A. Fuchs, *Calcul des probabilités*, Masson. (1996)
- [14] E. Nadaraya, *On non parametric estimation density function and regression*. *Theory Probab P.P.L* (1965).
- [15] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*. *Annals of Mathematical Statistics* 27, 832837. (1956)
- [16] XI. SONG CHEN (*Probability density function estimation using gamma kernels*. *Annals of the Institute of Statistical Mathematics* 54, 471480. (2000).
- [17] E. Parzen, . "On estimation of a probability density function and mode", *The Annals of Mathematical Statistics*, vol.33, no.3, p. pp. 1065-1076. (1962)