



Faculté des Sciences Exactes et Informatique
Département de Mathématiques

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Spécialité : Mathématiques fondamentales.

Option : Probabilités et Statistique.

Thème

Inégalités exponentielles dans l'estimation fonctionnelle non-paramétrique

Présenté par :

Belkhelfa Selma.

Devant le jury :

Président : M^{me} Laoudj Ckekraoui Farida (M.C.A) Université de Jijel
Encadreur : M^{me} Abdi Zeyneb (M.A.A) Université de Jijel
Examineur : M^{me} Djeridi Zohra (M.C.B) Université de Jijel

Promotion **2020/2021**

Remerciements

Avant toute considération, je remercie le Grand Dieu le tout puissant qui, m'a aidé pour achever ce travail.

Je tiens tout à remercier en premier lieu mon encadreur *M^{me}* **Abdi Zeyneb** d'avoir accepté de m'encadrer et pour sa continuité à me soutenir et à m'encourager. Je voudrai aussi le remercier pour sa gentillesse, sa disponibilité et du temps consacré à mon travail.

Je veux également remercier *M^{me}* **Laoudj Ckekraoui Farida**, pour nous avoir fait l'honneur de présider le jury de notre mémoire Nos remerciements vont également aux *M^{me}* **Djeridi Zohra** honorer de leur examinateur dans ce jury.

Je remercie également les membres du département de Mathématique et Informatique de m'avoir permis de travailler dans de bonnes conditions pendant la réalisation de mon travail. Merci également a tous les enseignants qui m'ont aidé pendant mon cursus, sans oublier leurs conseils précieux.

Je remercie aussi toute personne de prés ou de loin a contribué à la finalisation de ce travail.

Dédicace

Grâce à Allah le tout puissant, qui m'a donné la volonté, la force, et qui m'a éclairé vers le bon chemin, que nul ne peut se faire sans son désir.

À mon père, l'homme qui a fait beaucoup pour moi.

À ma mère, qui a toujours été une source de tendresse d'encouragement.

À mes soeurs :(Oumaiama, Ikram, Rahma)

À mes amies

À mon fiancé, Je le souhaite tout le bonheur du monde.

Table des matières

1	L'estimation fonctionnelle non-paramétrique	10
1.1	Estimation par la méthode des noyaux	10
1.2	Noyaux	11
1.2.1	Quelques formes de noyaux	11
1.2.2	Propriétés d'un estimateur à noyau	12
1.3	Principe de la méthode de noyau	12
1.3.1	Estimateur à noyau :	12
1.4	Quelques définitions	14
1.5	La représentation graphique de l'estimateur sous différent type de noyaux	15
1.6	Propriétés de l'estimateur à noyau	17
1.6.1	Étude du Biais	17
1.7	Expression d'erreur quadratique moyenne (MSE) et moyenne intégrée (MISE)	20
1.8	Lecture Asymptotique de MSE et MISE	21
1.9	Choix du paramètre de lissage h	21
1.9.1	Méthode de la validation croisée par les moindres carrés	22
2	Inégalités exponentielles : Cas de variables aléatoires i.i.d	23
2.1	Les modes de convergence aléatoires	23
2.1.1	Convergence presque complète	23
2.1.2	Convergence presque sûrement	24
2.1.3	Convergence en moyenne quadratique	24
2.1.4	Convergence en loi	25
2.1.5	Convergence en probabilité	25
2.2	Inégalités exponentielles : Cas de variables aléatoires i.i.d	26
2.2.1	Inégalité de Markov	26

2.2.2	Inégalité de Bienaymé-Tchebychev (1853-1884)	26
2.2.3	Inégalité de Bernstein (1924)	27
2.2.4	Inégalité de Kolmogorov (1929)	29
2.2.5	Inégalité de Hoeffding (1963)	30
2.2.6	Inégalité de Talagrand (1996)	32
3	Simulation	35
3.1	Modèle de régression	35
3.2	Simulation de la fonction de régression sous différent type de noyaux	36
3.2.1	Plan de simulation	36
3.2.2	Résultats de simulation	36
3.3	Simulation de la fonction de régression suivant différents valeurs de h	40
3.3.1	Plans de simulation	40
3.4	Étude de la convergence presque complète par l'utilisation de l'inégalité de Berstein	43
4	Conclusion	45
	Bibliographie	47

Table des figures

1.1	Estimateur à noyau basé sur 7 observations ($h=4$)	14
1.2	Noyau uniforme (Rosenblatt)	15
1.3	Noyau Triangulaire	16
1.4	Noyau d'Epanechnikov ou parabolique	16
1.5	Noyau gaussien	17
1.6	Le "trad-off" biais-variance en fonction de h	19
3.1	La représentation graphique de θ et $\hat{\theta}$ sous le noyau Triangulaire	37
3.2	La représentation graphique de θ et $\hat{\theta}$ sous le noyau Biweigh	38
3.3	La représentation graphique de θ et $\hat{\theta}$ sous le noyau Gaussien	39
3.4	Représentation graphique de $\hat{\theta}$ et θ pour($h=0.1000, 0.3625, 0.8875, 1.1500$)respectivement	41
3.5	Représentation graphique de $\hat{\theta}$ et θ pour($h=0.1000, 0.3625, 0.8875, 1.1500$) respectivement	42

Notation

- $\mathbb{1}$:Indicatrice
- $K(\cdot)$: Le noyau
- $\hat{f}_n(X)$: L'estimateur
- X, Y : Les variable aléatoire
- i.i.d : Indépendantes et identiquement distribuées.
- \sum : La somme
- \prod : produit
- E : L'espérons mathématique
- h : Le paramètre de lissage
- \int :l'intégrale
- $V(\cdot)$:La variance
- \lim : La limite
- CV :Validation croise
- $|\cdot|$:La valeur absolue

Introduction générale

Dans la statistique on trouve deux approches d'estimation dans la littérature, la première est l'approche paramétrique qui se résume en l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon. Dans ce cas, on construit les estimateurs en utilisant les différentes méthodes d'estimation (méthode des moments, de maximum de vraisemblance,...)

Cependant, l'approche non-paramétrique consiste généralement à estimer à partir des observations une fonction inconnue élément d'une certaine classe fonctionnelle. Cette branche de la statistique a subi un intérêt croissant autant sur le plan théorique que pratique.

la statistique non-paramétrique regroupe l'ensemble des méthodes statistiques qui permettent de tirer de l'information pertinente de données sans faire l'hypothèse que la loi de probabilité de ces observations appartient à une famille connue.

L'estimation non paramétrique remonte à **Rosenblatt** (1956, [20]) et **Parzen**(1962,[19]) où la méthode du noyau de convolution est utilisée pour estimer la fonction de densité de probabilité. Qui ont proposé une classe d'estimation à noyau d'une densité unimodale.

Les estimateurs à noyau sont des fonctions de deux paramètres K appelé noyau, et h dit paramètre de lissage (largeur de fenêtre), les propriétés de convergence de l'estimateur à noyau ont été établies par **parzen** (1962), **silverman** et **Nadaraya** (1964,[17]), les théorèmes relatifs à l'erreur quadratique moyenne et l'erreur quadratique moyenne intégrée ont été obtenus par **parzen** (1962).

Malgré leur intérêt apporté au domaine de la statistique, les estimateurs à noyaux connaissent quelques inconvénients parmi lesquels se situent les problèmes de support. Le plus connu est l'effet du bord dans l'estimation des fonctions à support borné au moins d'un côté. Plusieurs solutions à ce problème ont été proposées dans la littérature.

De multiples inégalités exponentielles existent dans l'estimation non-paramétrique fonctionnelle, elles permettent de déterminer la vitesse de convergence des différents estimateurs.

La construction des inégalités exponentielles s'inspire principalement du théorème central li-

mite et de la loi fort des grands nombres. Cette construction requiert l'utilisation de l'inégalité exponentielle de Markov, il s'agit de définir une borne supérieure idéale de la fonction génératrice des moments.

Les auteurs ont énoncé plusieurs types d'inégalités exponentielles selon le modèle utilisé. Dans le cas où les v.a sont i.i.d, **Bernstein** (1924) propose une inégalité trouvant un usage particulier pour la démonstration de la convergence fort en estimation non-paramétrique.

Ce travail s'articule autour de 3 chapitres :

Le premier chapitre est d'établir certaines propriétés de l'estimateur de la fonction de répartition basé sur la méthode de noyau. Nous établissons ces propriétés statistiques (biais, variance, erreur quadratique moyenne, l'erreur quadratique moyenne intégrée, choix de paramètre de lissage et la méthode de la validation croisée).

Dans le deuxième chapitre nous présentons les différents modes de convergence et quelque inégalités exponentielles dans le cas ou les variables aléatoires sont i.i.d et quelque remarques sont données concernant leurs preuve.

Et on termine notre travail par une application de la convergence presque complète par l'utilisation de l'inégalité de Bernstein, ainsi une étude de simulation concernant la simulation de la fonction de régression et son estimateur(f et \hat{f}) suivant les différents type de noyaux, puis une représentations graphiques de (f et \hat{f}) suivant les valeurs de h (le paramétriser de lissage).

L'estimation fonctionnelle non-paramétrique

Introduction

Dans ce chapitre nous allons présentée une méthode d'estimation de la densité, c'est à dire comment estimer non- paramétriquement la densité de probabilité f ?

Pour répondre à cette question on considère les observations x_1, x_2, \dots, x_n , il existe plusieurs méthodes d'estimation non- paramétrique d'une densité (la méthode par histogramme, la méthode de noyaux,...etc), l'objectif de notre étude dans ce chapitre est de la construction d'un estimateur de f_n noté \hat{f}_n par la méthode de noyau avec $\hat{f}_n = f(x, X_1, X_2, \dots, X_n)$.

Pour l'estimation non- paramétrique, ils existent plusieurs méthodes d'estimation mais, dans ce chapitre on base seulement sur la méthode de noyau.

1.1 Estimation par la méthode des noyaux

L'estimateur à noyau est probablement l'estimateur le plus utilisé et certainement le plus étudier mathématiquement car il possède des propriétés qui le rendent fort intéressant. Supposons que nous observons n variables aléatoires i.i.d, x_1, x_2, \dots, x_n de densité f, l'objectif de notre étude est la construction d'un estimateur de f en un point fixe x, notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 , la densité est la dérivé de la fonction de répartition, ce qui permet d'écrire pour tout x :

$$f(x) = F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$$

Donc pour un $h > 0$ fixe « petite » on peut penser à estimer f(x) par :

$$\hat{f}_n(x) = \frac{1}{2h} (F_n(x+h) - F_n(x-h)) = \frac{1}{2nh} \sum_{i=1}^{+\infty} \mathbb{1}_{]x-h, x+h]}(x_i)$$

On a Alors :

$$E(\hat{f}_n(x)) = \frac{1}{2h}(E[F_n(x+h)] - E[F_n(x-h)]) = \frac{1}{2h}[F(x+h) - F(x-h)] \xrightarrow{h \rightarrow \infty} f(x)$$

Il faut donc faire dépendre h de la taille de l'échantillon, et le faire tendre vers 0 quand $n \rightarrow \infty$, de sorte que $\hat{f}_n(x)$ soit un estimateur asymptotiquement sans biais de $f(x)$, h sera donc dorénavant noté h_n . l'estimateur \hat{f}_n reste une fonction en escalier, pour obtenir quelque chose de plus lisse, on peut remarquer que :

$$\begin{aligned} \hat{f}_n(x) &= \frac{1}{2nh_n} \sum_{i=1}^n \mathbf{1}_{]x-h_n, x+h_n]}(x_i) \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \mathbf{1}_{]x-h_n, x+h_n]} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} \mathbf{1}_{]-1,1]} \left(\frac{x - X_i}{h_n} \right) \\ &= \frac{1}{nh_n} \sum_{i=1}^n K \left(\frac{x - X_i}{h_n} \right) \end{aligned}$$

Cet estimateur appelé estimateur de Rosenblatt, est le premier exemple d'estimateur à noyau construit à l'aide du noyau $K = \frac{1}{2} \mathbf{1}_{]-1,1]}$

1.2 Noyaux

Définition 1.2.1. [10] Soit $k : \mathbb{R} \rightarrow \mathbb{R}$, on dit que K est un noyau si et seulement si :

$$\int_{-\infty}^{+\infty} K(u) du = 1$$

- K est dit positive si $K(u) \geq 0 \forall u$.
- K est dit symétrique si $K(u) = K(-u) \forall u$.

1.2.1 Quelques formes de noyaux

Les noyaux les plus couramment utilisés en pratique sont :

-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 gaussien :

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, u \in \mathbb{R}$$

-Le noyau d'Epanechnikov :

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

-Le noyau : Biweight (quadratique)

$$K(u) = \frac{15}{16}(1 - u^2)\mathbf{1}_{[-1,1]}$$

1.2.2 Propriétés d'un estimateur à noyau

- $K(u) \geq 0$ (positive)

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

- $\int_{\mathbb{R}} uK(u)du = 0$ (symétrie)

L'estimateur à noyau est une fonction de densité \hat{f}_n qui a les même propriétés de continuité et de différentiable que K :

- * Si K est continue, \hat{f}_n sera une fonction continue
- * Si K est différentiable, \hat{f}_n sera une fonction différentiable.
- * Si K peut prendre des valeurs négative, alors \hat{f}_n a aussi des valeurs négative. [30]

1.3 Principe de la méthode de noyau

L'estimation par la méthode de noyau (ou méthode de Parzen-Rozenblatt)est une méthode non-paramétrique d'estimation de la densité d'une variable aléatoire. Cette méthode permet d'obtenir une densité continue et de construire en ce sens une généralisation de la méthode de l'histogramme, en effet la fonction indicatrice utilisée pour l'histogramme est ici remplacée par une fonction continue (le noyau) est une somme des fonctions continue qui reste continues .

1.3.1 Estimateur à noyau :

Définition 1.3.1. *Un estimateur à noyau noté \hat{f}_n de la fonction f est défini par :*

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \quad (1.1)$$

Où $(h_n)_{n \geq 1}$ est une suite de réels positive appelée paramètre de lissage ou largeur de la fenêtre qui tend vers 0 quand n tend vers l'infini et K appelé fonction de poids ou noyau.

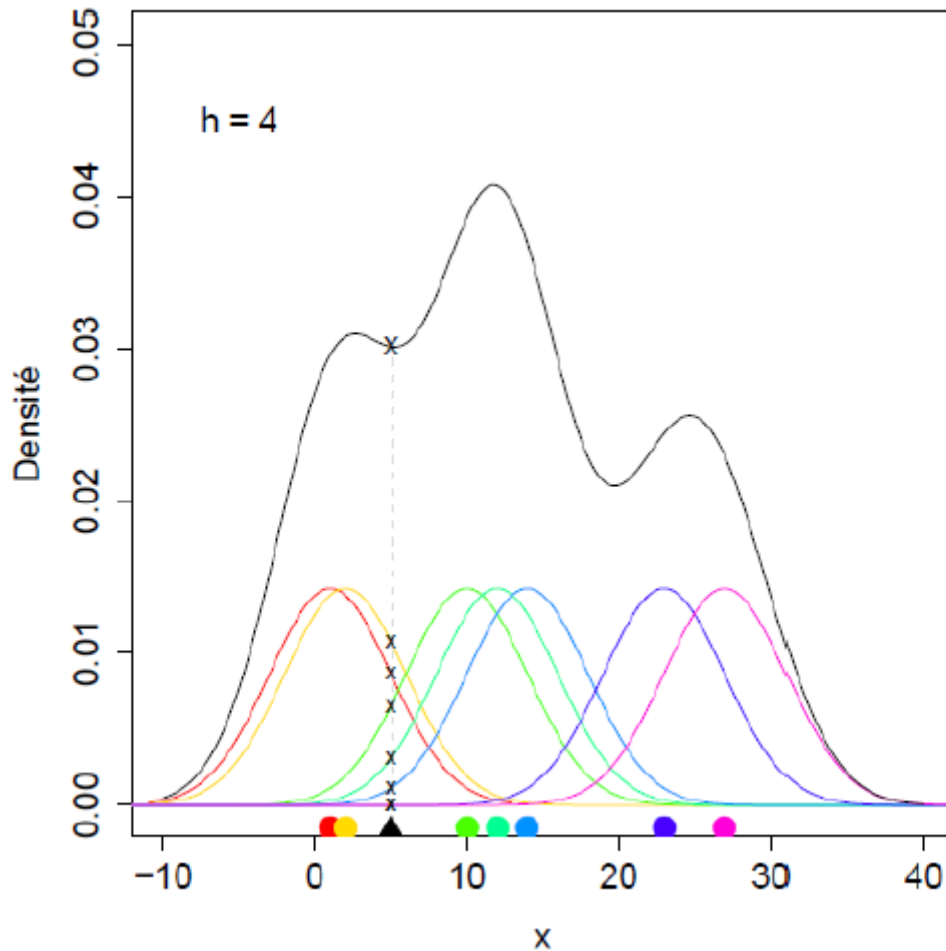
Propriété 1.3.1. *Un estimateur à noyau est une densité*

Preuve. [10]

$$\begin{aligned}
 \int_{-\infty}^{+\infty} \hat{f}_n(x) dx &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x - X_i}{h_n}\right) dx \\
 &= \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) h_n du \quad \text{changement de variable } \left(u = \frac{x - X_i}{h_n}\right) \\
 &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u) du \\
 &= \frac{1}{n} n = 1
 \end{aligned}$$

□

Pour mieux saisir l'intuition derrière l'estimateur à noyau, nous avons construit cet estimateur à partir de l'équation (1.1) en utilisant un ensemble de données constitué seulement de 7 observations. Le noyau K a été choisi comme étant la densité d'une loi normale de moyenne 0 et de variance 1 et le paramètre de lissage h égale à 4. On centre d'abord un noyau individuel sur chacune des 7 observations et la valeur de l'estimateur à noyau $\hat{f}(z)$ au point z est simplement la somme des ordonnées de chacun des 7 noyaux individuels à ce point x comme représenté à la figure (1.1).

FIGURE 1.1 – Estimateur à noyau basé sur 7 observations ($h=4$)

Par exemple si $x = 5$ on a $\hat{f}(x) = 0.03$ qui est égale à la somme des densités des 7 noyaux gaussiens au même point $x = 5$

1.4 Quelques définitions

Définition 1.4.1. [1] On dit qu'un estimateur \hat{f} de f est sans biais si :

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

Définition 1.4.2. [1] On dit qu'un estimateur \hat{f} de f est asymptotiquement sans biais si :

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

Définition 1.4.3. [1] Un estimateur \hat{f} de f dit asymptotiquement uniformément sans biais si :

$$\lim_{n \rightarrow \infty} \sup \left| E \left[\hat{f}_n(x) - f(x) \right] \right| = 0$$

1.5 La représentation graphique de l'estimateur sous différent type de noyaux

[2] • Le noyau uniforme (Rosenblatt)

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

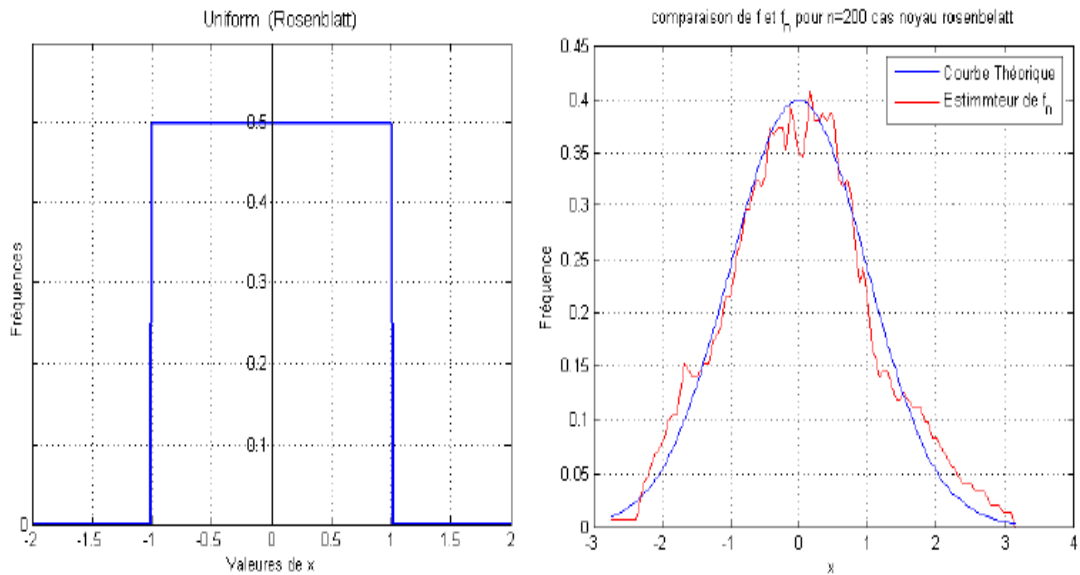


FIGURE 1.2 – Noyau uniforme (Rosenblatt)

• **Le noyau triangulaire** : Ce noyau a un avantage par rapport au noyau de Rosenblatt, c'est qu'il est continu partout, ce qui conduit à une estimation de f_n continue. Ce noyau s'écrit sous la forme :

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

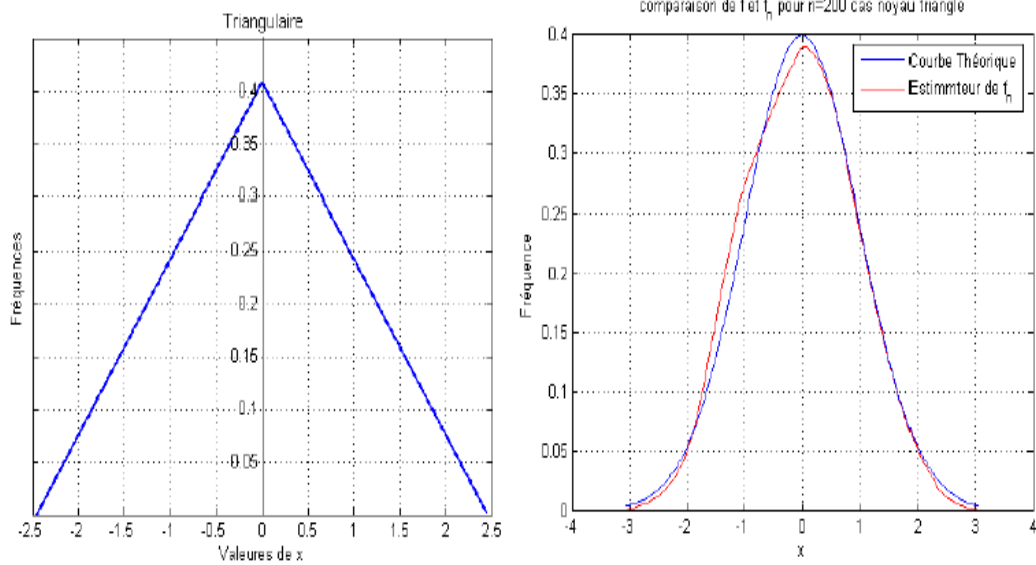


FIGURE 1.3 – Noyau Triangulaire

- **Le noyau d'Epanechnikov ou parabolique** : est défini par la forme suivante :

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

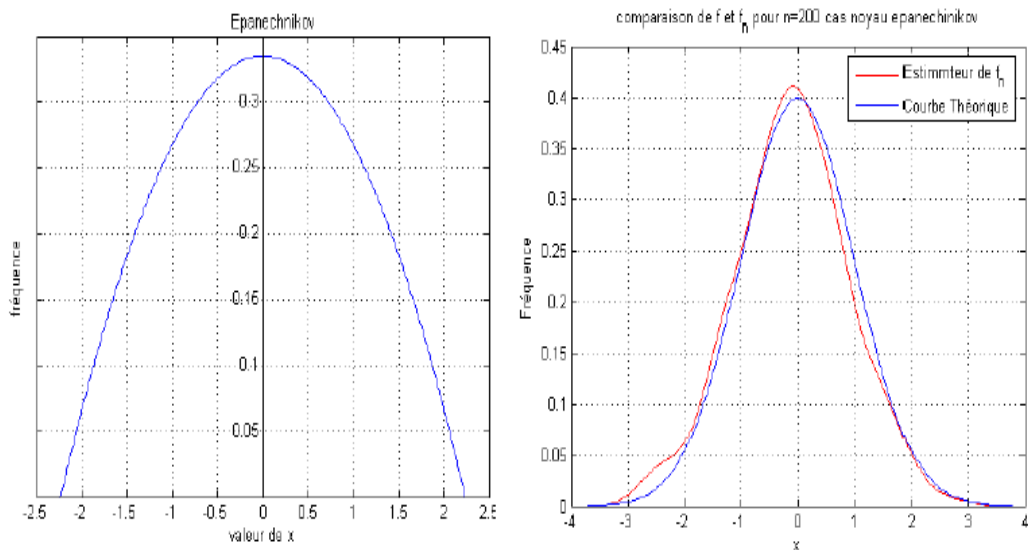


FIGURE 1.4 – Noyau d'Epanechnikov ou parabolique

- **Le noyau gaussien** : L'avantage du noyau gaussien est que plus la valeur de h est élevée plus on élargit la fenêtre, ce qui a un effet de lissage globale important, mais le coût de calcul dans le cas de ce noyau est très élevé du fait de son support infini. Ce noyau s'écrit sous la

forme :

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

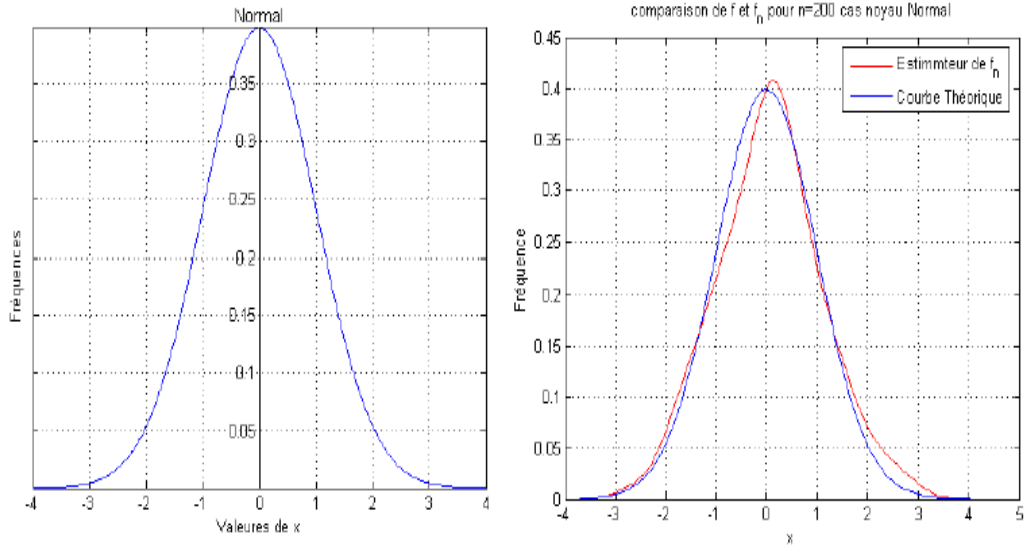


FIGURE 1.5 – Noyau gaussien

1.6 Propriétés de l'estimateur à noyau

Nous allons maintenant donner quelques propriétés statistiques élémentaires de l'estimateur de la densité f .

1.6.1 Étude du Biais

Supposons que l'on dispose d'un échantillon d'observations x_1, x_2, \dots, x_n , issu d'une variable aléatoire possédant pour fonction de densité, f que l'on désire estimer. On suppose que \hat{f}_n est l'estimateur à noyau obtenu en utilisant le noyau K et le paramètre de lissage h_n et \hat{f}_n défini par l'équation (1.1)

HYPOTHÈSES

On suppose que K vérifie les 4 conditions suivantes :

- 1- $\int_{\mathbb{R}} K(u) du = 1$
- 2- K est une fonction paire ou, plus généralement $\int_{\mathbb{R}} uK(u) du = 0$
- 3- $\int_{\mathbb{R}} u^2 |K(u)| du < \infty$
- 4- $\int_{\mathbb{R}} K(u)^2 du < \infty$

Proposition 1.6.1. *Si les trois premières conditions de l'hypothèse sur K sont remplies, alors :*

$$\text{Biais}[\hat{f}_n(x)] \simeq \frac{1}{2}f''(x)\mu_2h^2 + O(h^2)$$

Où $\mu_2 = \int_R K(u)u^2 du$

Si, de plus la condition 4 de l'hypothèse est satisfaite alors :

$$v[\hat{f}_n(x)] \simeq \frac{1}{nh}f(x)R(x) + o\left(\frac{1}{nh}\right)$$

Où $R(x) = \int_R K^2(u) du$

Preuve. [14]

Commençons par le calcul de Biais :

Considérons l'estimateur à noyau :

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

Où nous avons introduit la notation :

$$K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$$

Pour une version transformée de K :

$$\begin{aligned} E[\hat{f}_n(x)] &= E[K_h(x - X_i)] \quad \text{car les } X_i \text{ sont i.id} \\ &= \int_R K_h(x - y)f(y)dy \\ &= \int_R K(u)f(x - uh)du \quad \text{avec } u = \frac{x - y}{h} \quad \text{et } du = -\frac{1}{h}dy \\ &= \int_R K(u)[f(x) - f'(x)uh + \frac{1}{2}f''(x)u^2h^2 + O(h^2)]du \quad \text{par le développement de Taylor d'ordre} \\ &= f(x) \int_R K(u)du - f'(x)h \int_R K(u)u + \frac{1}{2}f''(x)h^2 \int_R K(u)u^2du + O(h^2) \end{aligned}$$

Alors :

$$\begin{aligned} \text{Biais}[\hat{f}_n(x)] &= E[\hat{f}_n(x)] - f(x) \\ &= \frac{1}{2}f''(x)h^2 \int_R K(u)u^2du + O(h^2) \\ &= \frac{1}{2}f''(x)\mu_2h^2 + O(h^2) \end{aligned} \tag{1.2}$$

Avec $\mu_2 = \int_R K(u)u^2du$

D'où la première assertion de la proposition.

Pour la variance on calcule :

$$\begin{aligned}
 v[\hat{f}_n(x)] &= E[\hat{f}_n^2(x)] - E[\hat{f}_n(x)]^2 \\
 &= \frac{1}{n} [E(K_h^2(x-X))] - [E(K_h(x-X))]^2 \\
 &= \frac{2}{nh^2} \int_R K^2\left(\frac{x-y}{h}\right) f(y) dy - \left[\int_R K\left(\frac{x-y}{h}\right) f(y) dy \right]^2 \\
 &= \frac{1}{nh} \int_R K^2 f(x-uh) du \quad \text{avec } u = \frac{x-y}{h} \\
 &= \frac{1}{nh} \int_R K^2(u) [f(x) - f'(x)hu + \dots] du \\
 &= \frac{1}{nh} f(x) \int_R K^2(u) du - f'(x) \int_R K^2(u) du + o\left(\frac{1}{nh}\right)
 \end{aligned} \tag{1.3}$$

Donc nous trouvons que :

$$\begin{aligned}
 v[\hat{f}_n(x)] &= \frac{1}{nh} f(x) \int_R K^2(u) du + o\left(\frac{1}{nh}\right) \\
 &= \frac{1}{nh} f(x) R(x) + o\left(\frac{1}{nh}\right)
 \end{aligned} \tag{1.4}$$

Avec $R(x) = \int_R K^2(u) du$ □

La figure suivante nous permet de mieux voir le comportement du biais et de la variance.

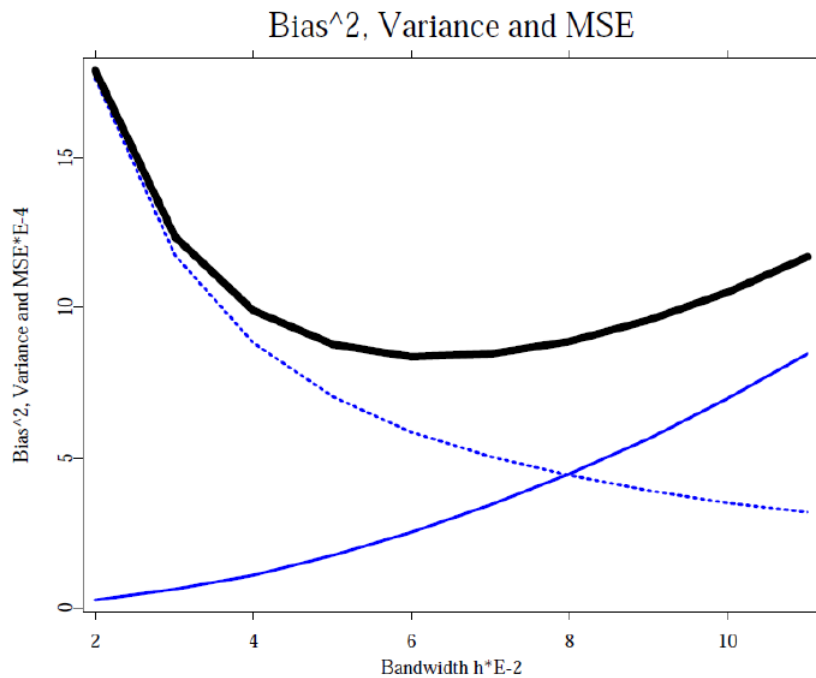


FIGURE 1.6 – Le "trad-off" biais-variance en fonction de h

La variance est représentée par la courbe en pointillé et le biais par la courbe fine, la courbe en gras représente le MSE.

Discussion du comportement du Biais et de la Variance

- * Le biais décroît si h diminue mais la variance augmente.
- * La variance diminue si h augmente mais le biais augmente.
- * Pour que la variance tende vers zéro, Il faut que $nh \rightarrow \infty$.
- * La variance est plus grande pour des valeurs, Plus grandes de la densité.

1.7 Expression d'erreur quadratique moyenne (MSE) et moyenne intégrée (MISE)

Définition 1.7.1. *L'erreur moyenne quadratique $\hat{f}_n(x)$ de $f(x)$ au point x est donnée par :*

$$MSE(\hat{f}_n(x)) = E[\hat{f}_n(x) - f(x)]^2 = var[\hat{f}_n(x)] + [Biais\hat{f}_n(x)]^2$$

Preuve. [1]

$$\begin{aligned} MSE(\hat{f}_n(x)) &= E(f(x) - \hat{f}_n(x))^2 \\ &= E(f(x)^2) - 2E(f(x)\hat{f}_n(x)) + E(\hat{f}_n(x)^2) \\ &= E(f(x)^2) - 2E f(x)\hat{f}_n(x) + E(\hat{f}_n(x)^2) + [E\hat{f}_n(x)]^2 - [E\hat{f}_n(x)]^2 \\ &= f(x)^2 - 2f(x)E(\hat{f}_n(x)) + E(\hat{f}_n(x)^2) + [E(\hat{f}_n)]^2 - [E(\hat{f}_n(x))]^2 \\ &= [E(\hat{f}_n(x)) - f(x)]^2 + E(\hat{f}_n(x)^2) - [E(\hat{f}_n(x))]^2 \\ &= [Biais(\hat{f}_n(x))]^2 + v(\hat{f}_n(x)) \end{aligned}$$

□

Le MSE est donné par :

$$MSE(\hat{f}_n(x)) = [Biais(\hat{f}_n(x))]^2 + var(\hat{f}_n(x))$$

Donc si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, On a que $E[\hat{f}_n(x) - f(x)]^2 \rightarrow 0$ pour tout point x , L'estimateur simple $\hat{f}_n(x)$ est alors un estimateur consistant de $f(x)$.

Définition 1.7.2. *L'erreur moyenne quadratique intégrée MISE donnée par :*

$$MISE(\hat{f}_n(x)) = \int MSE(\hat{f}_n(x))dx$$

1.8 Lecture Asymptotique de MSE et MISE

Les expressions asymptotiques du biais et de la variance nous permettant de trouver l'expression asymptotique pour la (MSE) et l'erreur quadratique moyenne intégrée (MISE), Ces expressions ont été obtenues sous certaines conditions sur K et en supposant que la densité de probabilité f avait toutes les dérivées (continues) nécessaires. A partir de (1.2) et (1.3) on peut obtenir facilement l'expression suivante pour la MSE et la MISE :

$$MSE[\hat{f}_n(x)] = \frac{1}{4}h^4\mu_2^2(f''(x))^2 + \frac{1}{nh}f(x)R(x) + o(h^4 + \frac{1}{nh}) \quad (1.5)$$

$$MISE[\hat{f}_n(x)] = \frac{1}{4}h^4\mu_2^2 \int_R (f''(x))^2 dx + \frac{1}{nh}f(x)R(x) + o(h^4 + \frac{1}{nh}) \quad (1.6)$$

Sous des conditions appropriées d'intégrabilité de f et ses dérivées. On note l'approximation asymptotique de la MSE par :

$$AMSE[\hat{f}_n(x)] = \frac{1}{4}h^4\mu_2^2(f''(x))^2 + \frac{1}{nh}f(x)R(x)$$

Et l'approximation asymptotique de la MISE par :

$$AMISE[\hat{f}_n(x)] = \frac{1}{4}h^4\mu_2^2R(f'') + \frac{1}{nh}f(x)R(x)$$

1.9 Choix du paramètre de lissage h

Comme nous l'avons précédemment signalé, pour avoir une bonne estimation par la méthode des noyaux, il faut bien choisir le paramètre de lissage h_n puisque celui-ci joue un rôle essentiel dans le processus. Les méthodes existantes pour le choix de h_n peuvent être classées en deux catégories La première catégorie est purement théorique. Elle consiste à minimiser l'erreur quadratique moyenne intégrée (MISE). En effet, la valeur optimale théorique de h_n notée h_{opt} s'obtient en minimisant la MISE asymptotique donnée en (1.6). Ainsi, pour un échantillon de taille n donné et pour un noyau K fixé, Cette valeur optimale de h_n est donnée par :

$$h_{opt} = \left\{ \frac{\int_R K^2(u)du}{\mu_2^4 \int_R (f''(x))^2 dx} \right\}^{\frac{1}{5}} n^{-\frac{1}{5}}$$

Où $\mu_2^4 = (\int_R K(u)u^2 du)^2$

On ne peut pas utiliser ce paramètre de lissage optimal h_{opt} car il dépend encore de la quantité inconnue $\int_R (f''(x))^2 dx$.

La deuxième catégorie est celle dite des méthodes pratiques. Elle utilise seulement les observations. Plusieurs auteurs ont travaillé sur ces méthodes parmi lesquels on peut citer Scott et al. [25], Rudemo [22], Stone [25], Bowman [5], Marron [16], Berlinet et Devroye [4], Park et Marron [23], Sarda et Vieu [24], Cuevas et al. [7] ainsi que Youndjé et al. [28, 29].

On s'intéresse dans la suite à une seule méthode (la méthode de la validation croisée) parmi les deux méthodes en pratique : La méthode de réinjections ("Plug-in") et la méthode de validation croisée par les moindres carrés ("Least Squares Cross Validation")

1.9.1 Méthode de la validation croisée par les moindres carrés

L'idée principale de cette méthode de validation croisée est celle de la minimisation par rapport à h_n d'un estimateur de la MISE. Pour plus de détails, on peut se référer à de nombreux auteurs tels que Bowman [5], Marron [16], Rudemo [22] et Stone [25]. L'approche de validation croisée par les moindres carrés consiste à minimiser un estimateur convenable de la MISE qui s'écrit :

$$\begin{aligned} MISE(\hat{f}_n) &= E \left[\int_R (\hat{f}_n(x) - f(x))^2 dx \right] \\ &= E \left[\int_R \hat{f}_n^2(x) dx - 2 \int_R \hat{f}_n(x) f(x) dx \right] + \int_R f^2(x) dx \end{aligned} \quad (1.7)$$

Comme la dernière intégrale $\int_R f^2(x) dx$ ne dépend pas de h_n , La valeur h_{opt} qui minimise la MISE est aussi celle qui minimise la fonction

$$J(h_n) = E \left[\int_R \hat{f}_n^2(x) dx - 2 \int_R \hat{f}_n(x) f(x) dx \right]$$

Pour trouver h_{opt} , il suffit d'estimer sans biais $J(h_n)$ par :

$$CV(h_n) = \int_R \hat{f}_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,i}(X_i)$$

Avec

$$\hat{f}_{n,i}(x) = \frac{1}{(n-1)h_n} \sum_{j \neq i} K\left(\frac{x - X_j}{h_n}\right)$$

Est l'estimateur calculé à partir de l'échantillon privé de l'observation X_i . La valeur optimale de h_n est donnée par :

$$h_{opt} = \operatorname{argmin}_{h_n > 0} CV(h_n)$$

Inégalités exponentielles : Cas de variables aléatoires i.i.d

Dans ce chapitre nous présentons quelques inégalités exponentielles dans le cas des variables aléatoires i.i.d qui permettent de donner une majoration de l'erreur de l'estimation qui quantifie la qualité de l'estimateur et ainsi la convergence presque complète de l'estimateur, des précisions et des remarques sont données concernant leurs preuves

2.1 Les modes de convergence aléatoires

On considère ici, une suite infinie de v.a. X_1, X_2, \dots, X_n notée en bref X_n . On peut définir plusieurs modes de convergence pour une telle suite. On notera F_{X_n} la fonction de répartition de X_n .

2.1.1 Convergence presque complète

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 $\lim_{n \rightarrow \infty} X_n = X$), si et seulement si :

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

Définition 2.1.1. [8] *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 $X_n = O_{p.co}(U_n)$, si et seulement si :*

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

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

2.1.2 Convergence presque sûrement

Définition 2.1.2. [13] On dit que X_n converge presque sûrement (ou converge avec probabilité 1, ou converge fortement) vers la v.a. X si, quelque soit $\varepsilon > 0$ donné,

$$\lim_{n \rightarrow \infty} P \left(\sup_{m \geq n} |X_m - X| < \varepsilon \right) = 1$$

Et l'on note $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$

Cette définition est complexe mais on peut voir qu'elle équivaut à dire que la suite M_n où $M_n = \sup_{m \geq n} |X_m - X|$ converge vers 0 en probabilité. Comme pour tout n $\sup_{m \geq n} |X_m - X| \geq |X_n - X|$, il est clair que la convergence presque sûre entraîne la convergence en probabilité (d'où les qualificatifs de convergence forte et convergence faible). On admettra les propositions ci-après, qui pourront nous être utiles par la suite.

Proposition 2.1.1. soit X_n elle que $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ et g une fonction continue

$$g(X_n) \xrightarrow[n \rightarrow \infty]{p.s.} g(x).$$

Proposition 2.1.2. soit X_n elle que $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ et Y_n telle que $Y_n \xrightarrow[n \rightarrow \infty]{p.s.} Y$. Si f est continue dans R^2 alors :

$$f(X_n, Y_n) \xrightarrow[n \rightarrow \infty]{p.s.} f(X, Y).$$

Ces deux propositions sont également vraies pour la convergence en probabilité. Elles s'étendent également à des fonctions de k variables aléatoires où $k > 0$

2.1.3 Convergence en moyenne quadratique

Définition 2.1.3. [13] On dit que X_n converge en moyenne quadratique vers la v.a. X si les v.a. X_1, X_2, \dots, X_n ont un moment d'ordre 2 et si :

$$\lim_{n \rightarrow \infty} E[(X_n - X)^2] = 0$$

Et l'on note $X_n \xrightarrow[n \rightarrow \infty]{m.q.} X$

La convergence m.q. est particulièrement facile à manipuler car elle repose sur la convergence usuelle d'une suite de nombres $E[(X_n - X)^2]$. Nous y recourrons abondamment, d'autant plus

qu'elle implique la convergence en probabilité. On admettra la hiérarchie d'implications suivantes entre les différents modes de convergence :

$$c.p.c \implies c.p.s \implies p_{\uparrow m.q} \implies L$$

En outre $P \iff L$ dans le cas de la convergence vers une constante. Notons que, dans le cas général, il n'y a pas, entre convergence m.q. et convergence p.s de domination de l'une sur l'autre.

2.1.4 Convergence en loi

Définition 2.1.4. [13] On dit que X_n converge en loi vers la v.a. X si l'on a, en tout x où sa fonction de répartition F_X est continue.

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

Et l'on note $X_n \xrightarrow[n \rightarrow \infty]{L} X$

On dira aussi que la loi de X est la loi limite ou asymptotique de la suite x_n . En pratique la loi limite sera utile pour donner une approximation pour le calcul de la probabilité d'un événement sur X_n quand n sera assez grand :

$$P(X_n \in A) = P(X \in A)$$

Pour la convergence en loi comme pour les autres modes de convergence un cas particulier important est celui où X est une v.a, c'est-à-dire, que la suite converge vers une constante c .

2.1.5 Convergence en probabilité

Définition 2.1.5. [13] On dit que X_n converge en probabilité (ou converge faiblement) vers la v.a X si, quelque soit $\varepsilon > 0$ donné :

$$\lim_{n \rightarrow \infty} P(|X_n - X| < \varepsilon) = 1$$

Et l'on note $X_n \xrightarrow[n \rightarrow \infty]{P} X$

Pour ce mode de convergence comme pour les suivants la convergence vers une constante c s'explique naturellement en remplaçant X par c .

2.2 Inégalités exponentielles : Cas de variables aléatoires i.i.d

La plus part de ces résultats ont été obtenus par une approche basée sur l'utilisation de l'inégalité de Markov exponentielle. Le rôle de cette méthode est de trouver la meilleure borne supérieure de la fonction génératrice des moments.

2.2.1 Inégalité de Markov

Soit X une variable aléatoire réelle finie. On suppose que X est presque sûrement positive ou nulle, alors pour tout réel strictement positif, on a

$$P(X \geq a) \leq \frac{E(X)}{a} \quad (\text{preuve voir [9]})$$

2.2.2 Inégalité de Bienaymé-Tchebychev (1853-1884)

Théorème 2.2.1. [15] *Soit $r > 0$ un nombre réel, et X une variable aléatoire réelle définie sur un espace probablisable. On suppose que $|X|^r$ est fini (on dira que X admet un moment d'ordre r ou encore que X est r -intégrable). alors on a :*

$$P(|X| \geq t) \leq \frac{E|X|^r}{t^r}$$

Pour $r = 1$ (resp. $r = 2$), ce résultat est connu sous le nom d'inégalité de Markov (resp. d'inégalité de Bienaymé-Tchebychev). Il s'agit des premières inégalités pouvant servir à évaluer des déviations par rapport à la moyenne.

Preuve.

1-Inégalité de Markov :

Soit la fonction de densité de x , $f(x)$

Pour $t > 0$

$$\begin{aligned} E(X) &= \int_0^{\infty} xf(x)dx \geq \int_t^{\infty} xf(x)dx \\ &\geq t \int_t^{\infty} f(x)dx \\ &\geq tP(X \geq t) \\ &\implies P(X \geq t) \leq \frac{E(X)}{t} \end{aligned}$$

2- Inégalité de Bienaymé-Tchebychev

$\forall t > 0 : P(|X - E(X)| \geq t) = P(|X - E(X)|^2 \geq t^2)$ en appliquant l'inégalité de Markov à la

variable aléatoire $[X - E(X)]^2$ nous obtenons :

$$P(|X - E(X)| \geq t) \leq \frac{E([X - E(X)]^2)}{t^2} = \frac{v(X)}{t^2} = \frac{\sigma^2}{t^2}$$

Dans le cas où $E(X) = 0$

$$P(|X| \geq t) \leq \frac{\sigma^2}{t^2}$$

□

2.2.3 Inégalité de Bernstein (1924)

Théorème 2.2.2. [15] Soient X_1, \dots, X_n n variables aléatoires indépendantes identiquement distribuées d'espérance nulle. Posons $s_n = \sum_{i=1}^n X_i$ Supposons $\forall i$, il existe une constant $M > 0$ telle que :

$$E|X_i|^q \leq q! M^{q-2} E(X_i^2) < +\infty, q = 3, 4, \dots$$

Et posons

$$v_n^2 = \sum_{i=1}^n E(X_i)^2$$

Alors $\forall t > 0$ on a :

$$P(|S_n| \geq t) \leq 2 \exp\left(-\frac{t^2}{2(tM + 2v_n^2)}\right)$$

Remarque 2.2.1. Cette inégalité est plus utilisée pour montrer la convergence presque sûr et la convergence presque complète dans l'estimation non paramétrique.

Preuve.

Soit $0 \leq s < \frac{1}{M}$

$$\begin{aligned} E[e^{sX_i}] &\leq 1 + \sum_{q=2}^{\infty} \frac{1}{q!} s^q E|X_i|^q \\ &\leq 1 + s^2 E(X_i^2) \sum_{q=2}^{\infty} (sM)^{q-2} \quad (\text{par la condition de Gramer}) \\ &= 1 + \frac{s^2 E(X_i^2)}{1 - sM} \\ &\leq e^{\left(\frac{s^2 E(X_i^2)}{1 - sM}\right)} \end{aligned}$$

On applique l'inégalité de Markov à s_n , on arrive à :

$$\begin{aligned} P(s_n \geq t) &\leq e^{(-st)} \prod_{i=1}^n E[e^{(\lambda X_i)}] \\ &\leq e^{(-st)} e^{\left(\frac{s^2 v_n^2}{1-sM}\right)} \\ &\leq e^{\left(\frac{s^2 v_n^2}{1-sM} - st\right)} \end{aligned}$$

Optimiser ce majorant conduit à :

$$\frac{d}{ds} \left(\frac{s^2 v_n^2}{1-sM} - st \right) = 0$$

C'est-à-dire

$$s = s_0 = \frac{t}{Mt + 2v_n^2} > 0$$

On arrive à :

$$\begin{aligned} \left(\frac{s_0^2 v_n^2}{1-s_0 M - s_0 t} \right) &= \left(\frac{\left(\frac{t}{Mt + 2v_n^2} \right)^2 v_n^2}{1 - \left(\frac{t}{Mt + 2v_n^2} \right) M} - \left(\frac{t}{Mt + 2v_n^2} \right) t \right) \\ &= \frac{t^2}{2(Mt + 2v_n^2)} - \frac{t^2}{Mt + 2v_n^2} \\ &= -\frac{t^2}{2(Mt + 2v_n^2)} \end{aligned}$$

Conduit à l'inégalité de Bernstein :

Donc :

$$P(s_n \geq t) \leq e^{\left(-\frac{t^2}{2(tM + 2v_n^2)}\right)}$$

Les mêmes calculs seront faits pour la variable aléatoire $(-X_i)$ car

$$p(s_n \leq -t) = P(-s_n \geq t)$$

Et

$$P(|s_n| \geq t) \leq P(s_n \geq t) + P(s_n \leq -t)$$

Conduit à l'inégalité de Bernstein

$$P(|s_n| \geq t) \leq 2e^{\left(-\frac{t^2}{2(tM + 2v_n^2)}\right)}$$

□

Lemme 2.2.1. [3] Soit X_1, X_2, \dots, X_n des v.a.r. centrées, indépendantes et de même loi, telles qu'il existe deux réels positifs d et σ^2 vérifiant :

$$|X_1| \leq d \quad \text{et} \quad E(\Delta_1^2) \leq \sigma^2$$

Alors, pour tout $\varepsilon \in]0, \frac{\sigma^2}{d}[$ on a :

$$P\left(\frac{1}{n} \left| \sum_{i=1}^n \xi_i \right| > \varepsilon\right) \leq 2 \exp\left\{\frac{-n\varepsilon^2}{4\sigma^2}\right\}$$

2.2.4 Inégalité de Kolmogorov (1929)

Théorème 2.2.3. [15] Soient X_1, X_2, \dots, X_n des variables aléatoires i.i.d d'espérance nulle, on note $v_n^2 = \sum_{i=1}^n E(X_i)^2$ et $s_n = \sum_{i=1}^n X_i$, telles que $|X_i| \leq Mv_n$ p.s pour $1 \leq i \leq n$ où M est une constante positive, Pour tout $t > 0$

(a). si $Mt \leq 1$ alors :

$$P(s_n/v_n \geq t) \leq \exp\left(\frac{-t^2}{2} \left(1 - \frac{Mt}{2}\right)\right)$$

(b). si $Mt \geq 1$, alors :

$$P(s_n/v_n \geq t) \leq \exp\left(\frac{-t}{4M}\right)$$

Preuve.

Soit $s > 0$, pour $sMt \leq t$:

$$\begin{aligned} E[\exp(sX_i)] &= 1 + \sum_{m=2}^{\infty} \frac{1}{m!} s^m E(X_i^m) \\ &\leq 1 + \frac{s^2 E(X_i^2)}{2} \left(1 + \frac{Msv_n}{3} + \frac{M^2 s^2 v_n^2}{4.3} + \dots\right) \\ &\leq \exp\left(\frac{s^2 E(X_i^2)}{2} \left(\frac{1 + Msv_n}{2}\right)\right) \end{aligned}$$

Par l'inégalité de Markov nous obtenons :

$$\exp(-sv_n t) E(\exp-ss_n) \leq \exp\left\{-sv_n t + \frac{t^2 v_n^2}{2} \left(\frac{1 + Msv_n}{2}\right)\right\}$$

En considérant :

$s = \frac{t}{v_n}$ où $s = \frac{1}{Mv_n}$ pour les cas $Mt \leq 1$ et $Mt \geq 1$ respectivement on optimise le membre de droit de la dernière inégalité et on conclut la preuve. \square

2.2.5 Inégalité de Hoeffding (1963)

L'inégalité de Hoeffding est une inégalité de concentration concernant les sommes de variables aléatoires indépendantes et bornées.[15]

Théorème 2.2.4. Soient Y_1, Y_2, \dots, Y_n des variables aléatoires indépendantes telles que :

$$a_i \leq Y_i \leq b_i \text{ et } E(Y_i) = 0$$

On note que $\Delta_i = b_i - a_i$

o na : Pour $t > 0$

$$P \left(\sum_{i=1}^n Y_i \geq t \right) \leq \exp \left(\frac{-2t^2}{\sum_{i=1}^n \Delta_i^2} \right) \quad (2.1)$$

Pour la preuve de cette inégalité, Et afin d'obtenir une majoration de la fonction génératrice des moments, Hoeffding utilise la convexité de la fonction exponentielle.

Preuve.

Par la majoration de Chernoff (i.e, une application de l'inégalité de Markov à l'exponentielle de la somme de variable aléatoire d'intérêt), nous obtenons pour tout $s > 0$

$\forall t > 0,$

$$\begin{aligned} P \left(\sum_{i=1}^n Y_i \geq t \right) &= P \left(s \sum_{i=1}^n Y_i \geq st \right) \\ &= P \left(\exp(s \sum_{i=1}^n Y_i) \geq \exp(st) \right) \\ &\leq \frac{E(\exp(s \sum_{i=1}^n Y_i))}{\exp(st)} \\ &= \exp(st) \prod_{i=1}^n E(\exp(sY_i)) \end{aligned}$$

Le but est d'optimiser cette inégalité, Avant cela Hoeffding donne une majoration de $E(\exp(sY))$, Dans un lemme dit de Hoeffding. □

Lemme 2.2.2. si $a \leq Y \leq b, E(Y) = 0$ et $\Delta = b - a$, Alors :

$$E(e^{sY}) \leq e^{\left(\frac{s^2}{8}\Delta^2\right)}$$

Preuve.

Pour i fixé, on décompose Y en une combinaison convexe $Y = \lambda a + (1 - \lambda)b$, avec $0 < \lambda < 1$

On pose $\lambda = \frac{b-Y}{b-a}$ et $Y = \frac{b-Y}{b-a}a + \frac{Y-a}{b-a}b = \lambda a + (1 - \lambda)b$

La convexité de la fonction exponentielle conduit à :

$$e^{(sY)} = e^{s(\lambda a + (1-\lambda)b)} \leq \lambda e^{sa} + (1 - \lambda)e^{sb}$$

Par passage à l'espérance, Comme $E(Y) = 0$, on a

$$E(e^{sY}) \leq \frac{b}{b-a}e^{sa} + \frac{-a}{b-a}e^{sb}$$

On pose

$$a = \frac{-a}{b-a} \implies a = -(b-a)a = -\Delta a$$

On obtient :

$$\begin{aligned} E(e^{sY}) &\leq (1-a)e^{as\Delta} + (1-a)s\Delta \\ \log(E(e^{sy})) &\leq -as\Delta + \log(1-a + as\Delta) \\ E(e^{sY}) &\leq e^{(\psi(u))} \end{aligned}$$

Avec $\psi(u) = -au + \log((1-a) + ae^u)$ et $u = s\Delta$

Nous avons $\psi(0) = 0$

$$\begin{aligned} \psi'(u) &= -a + \frac{ae^u}{(1-a) + ae^u} \\ \psi''(u) &= \frac{ae^u((1-a) + ae^u) - (ae^u)^2}{((1-a) + ae^u)^2} \\ &= \frac{(1-a)ae^u}{((1-a) + ae^u)^2} \\ \frac{1}{\psi''(u)} &= \frac{(1-a)^2 + 2a(1-a)e^u + a^2e^{2u}}{(1-a)ae^u} \\ &= \frac{(1-a)}{ae^u} + 2 + \frac{ae^u}{(1-a)} \\ &= x + 2 + \frac{1}{x} \end{aligned}$$

La fonction $x + 2 + \frac{1}{x}$ un minimum en $x = 1$, par conséquent $\psi''(u) \leq \frac{1}{4}$ comme $\psi(u) = 0$ et $\psi'(0) = 0$, on a par la formule de Taylor qu'il existe une constante $\Theta \in R$, avec $0 \leq |\Theta| \leq u$ telle que :

$$\begin{aligned} \psi(u) &= \psi(0) + \frac{u}{1!}\psi'(0) + \frac{u^2}{2!}\psi''(0) \\ &\leq \frac{1}{8}u^2 \\ &\leq \frac{1}{8}(s\Delta)^2 \end{aligned}$$

O na :

$$\begin{aligned} E(\exp(sY)) &\leq e^{(\psi(u))} \\ &\leq \exp\left(\frac{s^2}{8}\Delta^2\right) \end{aligned}$$

$$E(e^{s\sum_{i=1}^n Y_i}) = \prod_{i=1}^n E(e^{sY_i}) \leq \prod_{i=1}^n e^{\frac{s^2}{8}\Delta_i^2}$$

Ainsi, pour tout $t > 0, \forall s \geq 0$

$$P\left(\sum_{i=1}^n Y_i \geq t\right) \leq e^{\left(-st + \frac{s^2}{8}\sum_{i=1}^n \Delta_i^2\right)}$$

En considérant : $s = \frac{4t}{\sum_{i=1}^n \Delta_i^2}$

On optimise le membre de droite de la dernière inégalité et on conclut la preuve. \square

Exemple 2.2.1.

Si X_1, \dots, X_n sont des variables aléatoires de Rademacher (i.e. $P(X_1 = 1) = P(X_1 \neq 1) = \frac{1}{2}$) i.i.d. Alors, d'après l'inégalité de Hoeffding, nous avons :

$$P(S \geq t) \leq e^{-\frac{t^2}{2n}} \quad \text{pour tout } t > 0$$

D'où, en pose $r = \frac{t}{\sqrt{n}}$, nous avons

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \geq r\right) \leq e^{-\frac{r^2}{2}} \quad \text{pour tout } r > 0$$

2.2.6 Inégalité de Talagrand (1996)

Les inégalités de concentration de type Talagrand sont une extension aux fonctions de variables indépendantes des inégalités exponentielles de Bennett ou de Bernstein pour les sommes de variables indépendantes. Ce type d'inégalités est apparu la première fois dans les travaux de Hoeffding (1963,[11]). Talagrand ([27], théorème 4.2.) donne une extension du résultat de Bennett aux fonctionnelles de variables aléatoires indépendantes. Les inégalités de Talagrand s'appliquent en particulier à la fonctionnelle $Z_n = \sup S_n(f)$, $f \in \mathfrak{F}$, où $S_n(f) = f(X_1) + \dots + f(X_n)$ et \mathfrak{F} une famille dénombrable de fonctions mesurables à valeurs dans \mathbb{R} . On obtient alors :

Théorème 2.2.5. Soient X_1, \dots, X_n des variables aléatoires indépendantes à valeurs dans un espace mesurable (\mathcal{X}, A) . Soit \mathfrak{F} une famille dénombrable de fonctions mesurables à valeurs dans \mathbb{R} , vérifiant $\|f\|_\infty \leq b < \infty$ pour tout $f \in \mathfrak{F}$. Soient $Z = \sup_{f \in \mathfrak{F}} \sum_{i=1}^n f(X_i)$ et $v = E(\sup_{f \in \mathfrak{F}} \sum_{i=1}^n f^2(X_i))$. Alors pour tout x strictement positif

$$P(Z \geq E(Z) + x) \leq K \exp\left(-\frac{1}{K'} \frac{x}{b} \log\left(1 + \frac{xb}{v}\right)\right)$$

Et

$$P(Z \geq E(Z) + x) \leq K \exp\left(-\frac{x^2}{2(c_1v + c_2bx)}\right) \quad (2.2)$$

Où K, K', c_1 et c_2 sont des constantes universelles strictement positives. De plus, Les mêmes inégalités sont valables si l'on remplace Z par $-Z$ La preuve de Talagrand repose sur des inégalités isopérimétriques pour les mesures produits. Ledoux (1996) retrouve ces inégalités par des méthodes entropiques en utilisant la méthode dite de Herbst. Il obtient (2.2) avec $K = 2$, $c_1 = 42$ et $c_2 = 8$. Mais il prend v de la forme $E(\sup_{i=1}^n f^2(X_i) : f \in \mathfrak{F}) + CbE(Z)$

Exemple 2.2.2. [9]

De la loi binomiale : On réalise n expériences indépendantes et on suppose que lors de chacune de ces expériences, La probabilité d'un événement appelé "succès" est p . Soit S_n le nombre de succès obtenus lors de ces n expériences. La variable aléatoire S_n , somme de n variables de Bernoulli indépendantes, de même paramètre p , Suit une loi binomiale : $S_n \rightarrow B(n; p)$. On s'intéresse alors à la variable aléatoire $\frac{S_n}{n}$, proportion de succès sur n expériences, a donc pour espérance $E(\frac{S_n}{n}) = p$ et pour variance $V(\frac{S_n}{n}) = \frac{1}{n^2}V(S_n) = \frac{p(1-p)}{n}$. Comme $p(1-p)$ atteint son maximum lorsque $p = \frac{1}{2}$, on a ainsi $p(1-p) \leq \frac{1}{4}$. En appliquant l'inégalité de Bienaymé-Chebyshev, il vient

$$P(|S_n/n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}.$$

Ainsi pour tout $\varepsilon > 0$, il existe $n > 0$ (plus précisément $n > \frac{1}{4n\varepsilon^2}$ tel que $P(|S_n/n - p| \geq \varepsilon) \leq n$ ou encore $\lim_{n \rightarrow \infty} P(|S_n/n - p| \geq \varepsilon) = 0$. La variable aléatoire $\frac{S_n}{n}$ converge en probabilité vers p .

Théorème 2.2.6. Soit (X_n) une suite de variables aléatoires sur le même espace probabilisé $(\Omega; P)$ admettant des espérances et des variances vérifiant

$$\lim_{n \rightarrow +\infty} E(X_n) = l \text{ et } \lim_{n \rightarrow +\infty} V(X_n) = 0$$

alors les (X_n) convergent en probabilité vers X

Preuve.

Soit $\varepsilon > 0$

Posons $E(X_n) = l + u_n$ avec $\lim u_n = 0$, Alors il existe $N \in \mathbb{N}$ tel que :

$$n \geq N \Rightarrow |u_n| < \varepsilon/2$$

et donc à partir du rang N ,

$$|X_n - E(X_n)| < \varepsilon/2 \implies |X_n - l| < \varepsilon, \quad (2.3)$$

Car $|X_n - l| = |X_n - E(X_n) + E(X_n) - l| \leq |X_n - E(X_n)| + |E(X_n) - l|$

L'implication (2.3) peut être encore écrite sous la forme

$$|X_n - l| \geq \varepsilon \implies |X_n - E(X_n)| \geq \varepsilon/2$$

Par conséquent, en utilisant l'inégalité de Bienaymé-Chebyshev,

$$P(|X_n - l| \geq \varepsilon) \leq P(|X_n - E(X_n)| \geq \varepsilon/2) \leq \frac{V(X_n)}{(\varepsilon/2)^2}$$

qui tend vers 0 quand n tend vers l'infini

Conséquence : Pour que (X_n) converge en probabilité vers X , il suffit que $E(X_n - X) \rightarrow 0$ et $V(X_n - X) \rightarrow 0$ lorsque $n \rightarrow \infty$ (la démonstration passe par l'inégalité de Bienaymé-Chebyshev). \square

Simulation

Dans les deux chapitres précédents, nous avons donné des résultats théoriques sur l'estimation fonctionnelles non- paramétrique ainsi quelques inégalités exponentielles.

Et dans cette section nous proposons une simulation effectuée pour confirmer les différents aspects théoriques abordés dans notre étude.

On commence notre travail dans ce chapitre par une présentation d'un modèle de régression, puis une simulation à l'aide de logiciel R.

3.1 Modèle de régression

Définition 3.1.1. [12] *Le modèle de la régression est l'un des modèles les plus fréquemment rencontrés en statistique paramétrique et non paramétrique.*

On considère $(X_1, Y_1), \dots, (X_n, Y_n)$ des couples de variables aléatoires indépendantes réelles et de même loi que (X, Y) . Dans le modèle de régression non paramétrique on suppose l'existence d'une fonction f qui exprime la valeur moyenne de la variable à expliquer Y en fonction de la variable explicative X , c'est à dire :

$$Y = \theta(X) + \varepsilon \tag{3.1}$$

Telle que ε est un bruit blanc indépendante de X .

On commence notre simulation par :

Algorithme de simulation

Les étapes de l'algorithme pour effectuer des bons estimateurs (courbe de régression) sont comme suite :

- Générer un échantillon de couple aléatoire (X_i, Y_i) de taille n .
- Estimer la courbe de régression $Y = f(X) + \varepsilon$ par la méthode du noyau.

L'estimateur de la courbe de régression pour le modèle $Y = f(X) + \varepsilon$ par la méthode du noyau est :

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

Le paramètre de lissage est choisi par la méthode de validation croisée (Voir le chapitre 1).

3.2 Simulation de la fonction de régression sous différent type de noyaux

Ici, les noyaux considérées sont les suivants :

1. **Le noyau triangulaire** : $K_1(u) = (1 - |u|) \mathbb{1}_{[-1,1]}(u)$.
2. **Le noyau biweigh** : $K_2 = \frac{15}{16}(1 - u^2) \mathbb{1}_{[-1,1]}(u)$.
3. **Le noyau gaussien** : $K_3(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, u \in R$

3.2.1 Plan de simulation

On définit le modèle de régression suivant :

$$Y = X + 1 + \varepsilon \tag{3.2}$$

- Générer l'échantillon $X \sim N(0, 1)$ et $\varepsilon \sim N(0, 1)$.
- Donner le nombre d'observation N de cette simulation.
- Choisir le noyau $K(\cdot)$
- Choisir la fenêtre de lissage h (dans ce cas $h = \log(n^{\frac{-1}{5}})$)

3.2.2 Résultats de simulation

Les résultats obtenus sont des représentations graphiques de f et \hat{f} selon différents types de noyaux (K_1, K_2, K_3) pour différents taille d'échantillon n ($n=150$; $n=200$; $n=250$; $n=300$)

Cas de noyau K_1

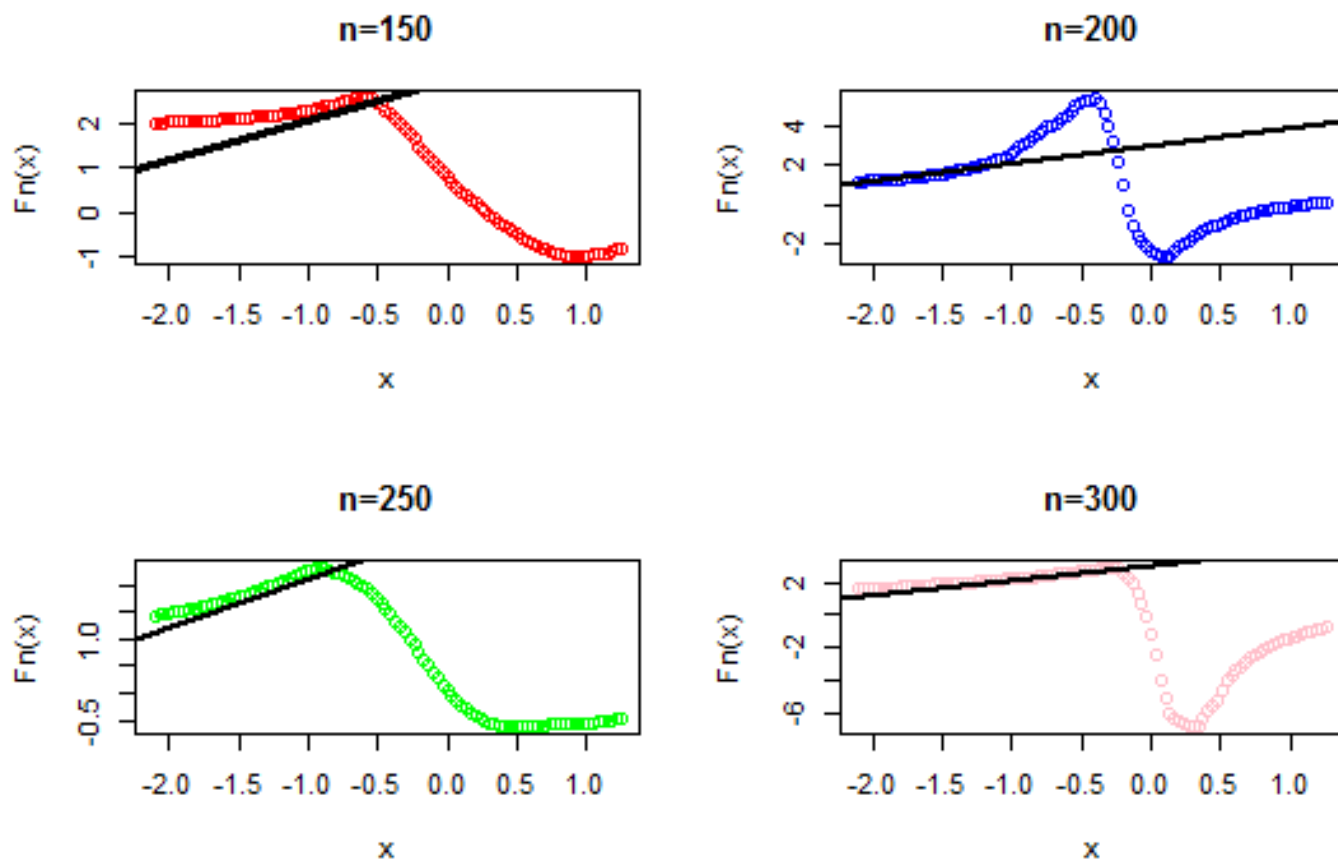


FIGURE 3.1 – La représentation graphique de θ et $\hat{\theta}$ sous le noyau Triangulaire

Interprétation

Les résultats obtenus dans la figure (3.1) montrent que pour la taille de l'échantillon " n " augmente, la courbe de $\hat{\theta}$ (représenté par la courbe en pointillé coloré) converge que la courbe de θ (représenté par la courbe noir fine) sous le noyau K_1 (le noyau triangulaire).

Cas de noyau K_2

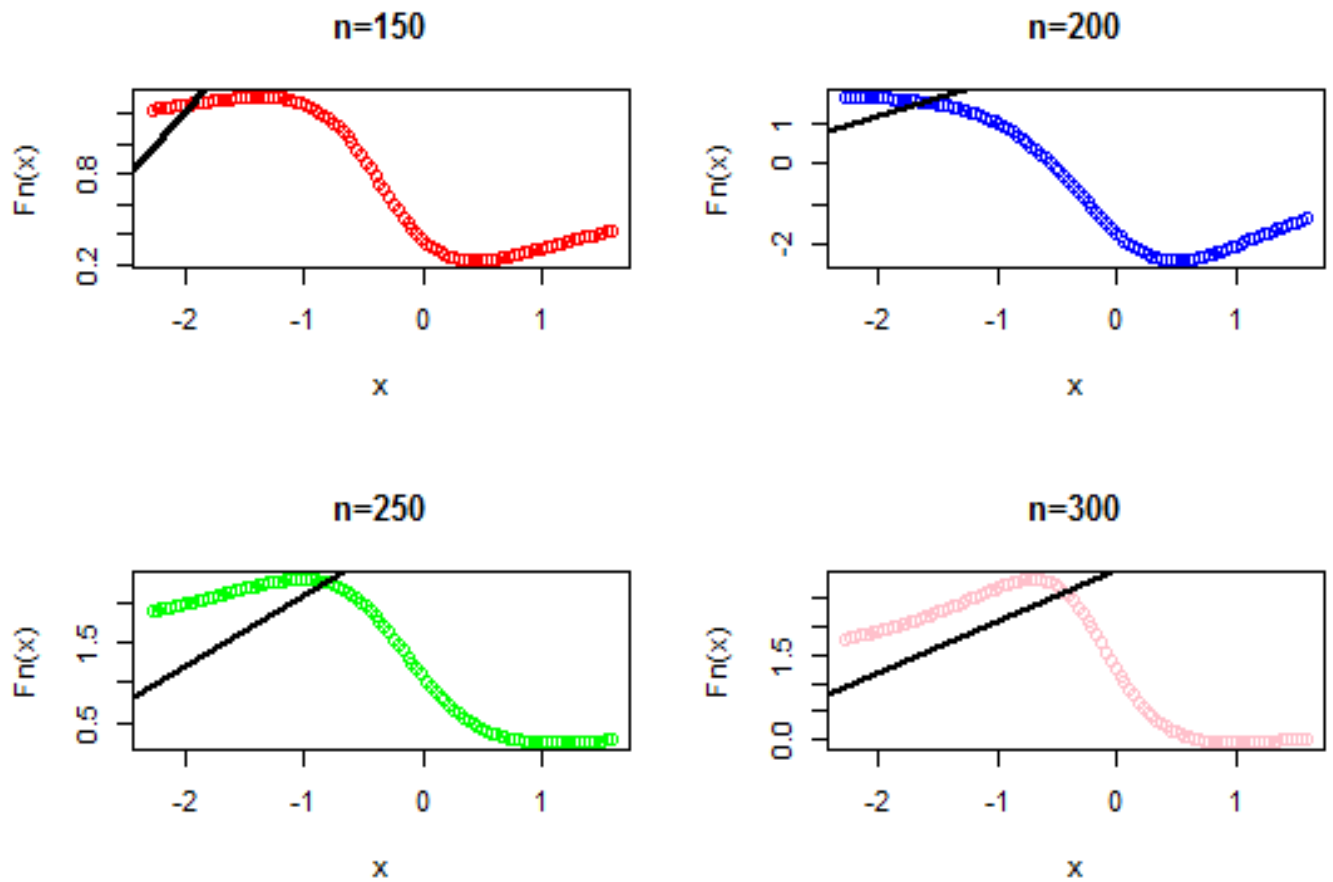


FIGURE 3.2 – La représentation graphique de θ et $\hat{\theta}$ sous le noyau Biweigh

Interprétation

Les résultats du figure (3.2) montrent que lorsque la taille de l'échantillon "n" augmente la courbe de $\hat{\theta}$ (représenté par la courbe en pointillé coloré) diverge vers la courbe de θ (représenté par la courbe noir fine) sous le noyau K_2 .

Cas de noyau K_3

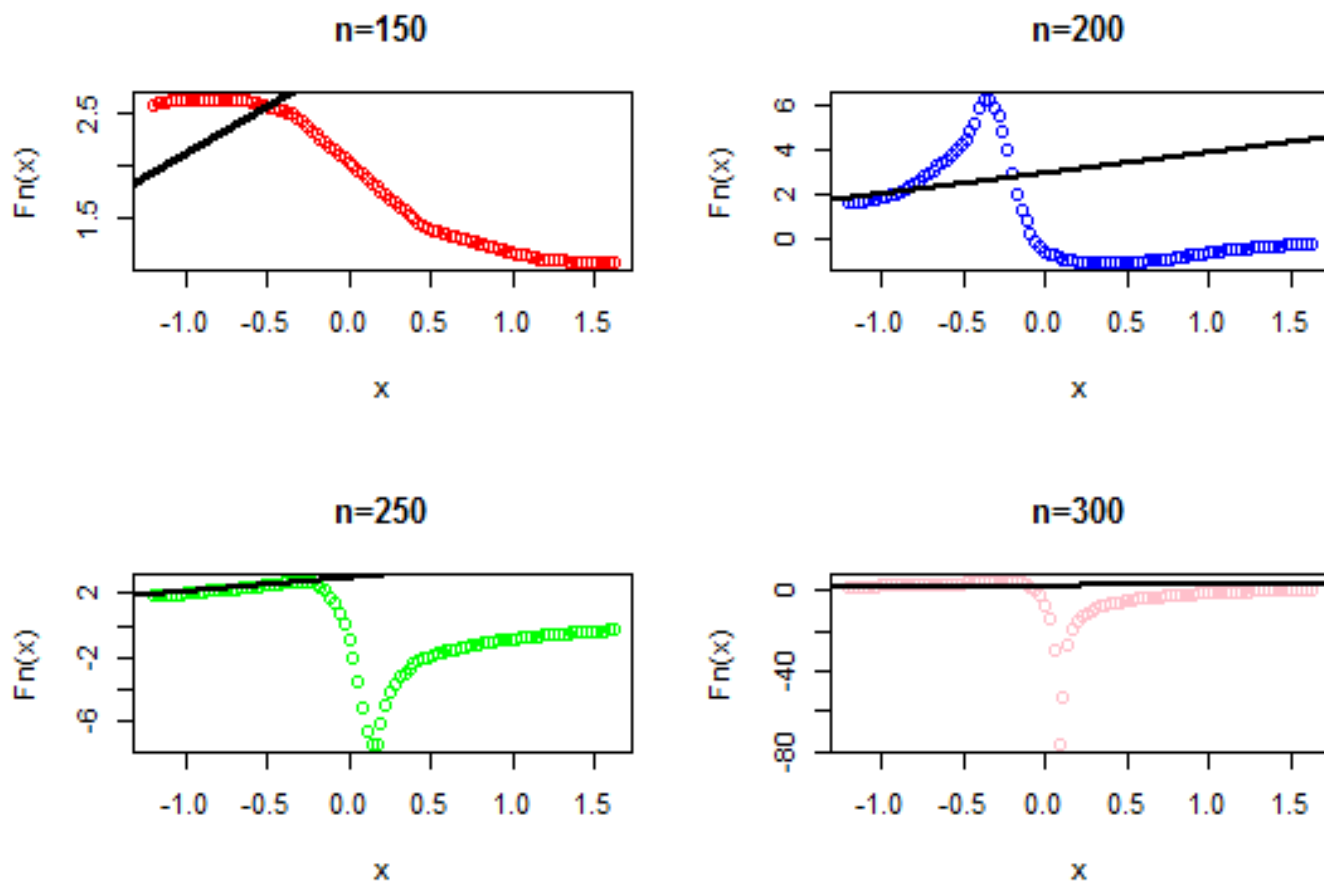


FIGURE 3.3 – La représentation graphique de θ et $\hat{\theta}$ sous le noyau Gaussien

Interprétation

Les résultats du figure (3.3) montrent que lorsque la taille de l'échantillon "n" augmente, la courbe de $\hat{\theta}$ (représenté par la courbe en pointillé coloré) converge vers la courbe de θ (représenté par la courbe noir fine) sous le noyau gaussien.

Comparaison

D'après les trois figures (figure 3.1, figure 3.2, figure 3.3) on remarque que le meilleur noyau qui nous donne le bon estimateur pour le modèle(3.2) c'est le noyau K_3 (le noyau gaussien).

3.3 Simulation de la fonction de régression suivant différents valeurs de h

3.3.1 Plans de simulation

On définit le modèle de régression suivant :

$$Y = X^2 + \varepsilon \tag{3.3}$$

- Générer l'échantillon $X \sim N(0, 1)$ et $\varepsilon \sim N(0, 1)$.
- Choisir le noyau $K(\cdot)$
- Choisir la fenêtre de lissage h (dans ce cas ($h_1=0.1000$; $h_2=0.3625$; $h_3=0.8875$; $h_4=1.1500$))

Les noyaux utilisés dans cette partie sont les suivants :

Le noyau gaussien : $K_3(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, $u \in R$

Le noyau triangulaire : $k_1(u) = (1 - |u|) \mathbb{1}_{[-1,1]}(u)$.

Cas de noyau K_1

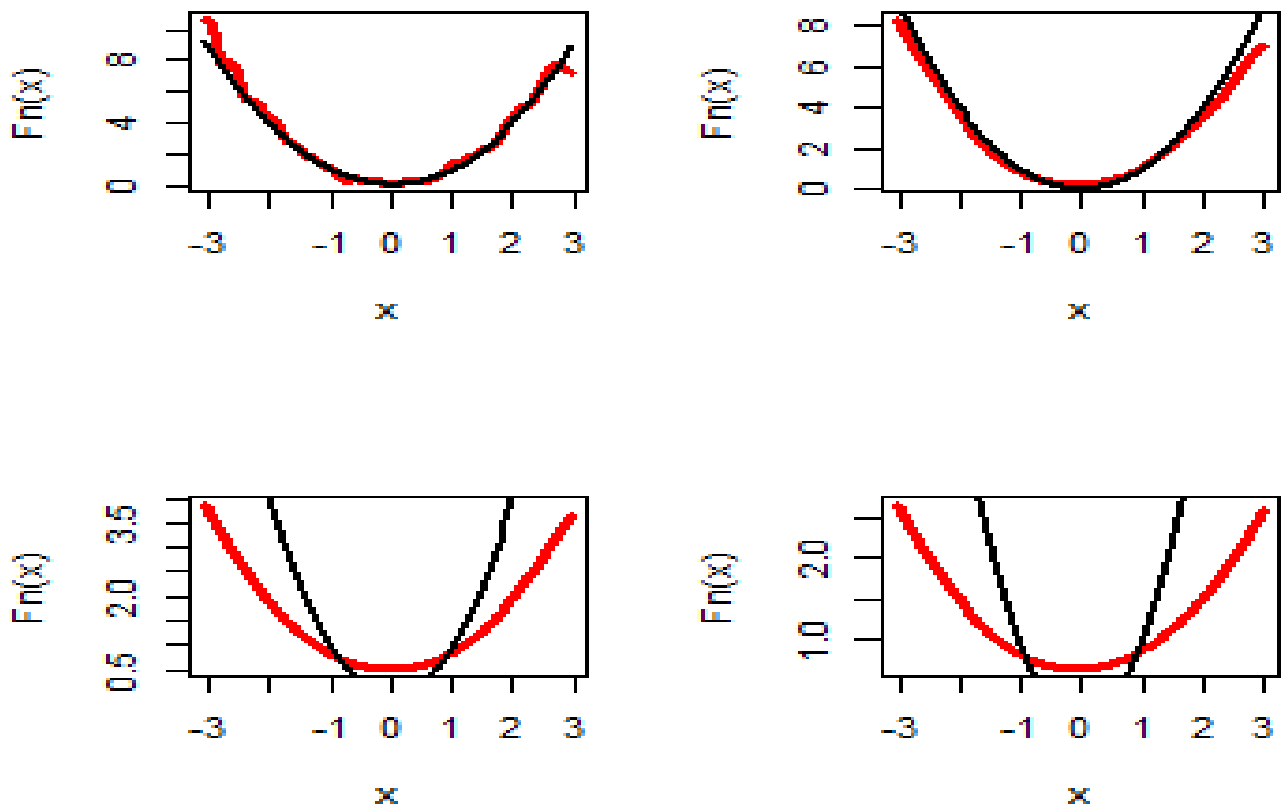


FIGURE 3.4 – Représentation graphique de $\hat{\theta}$ et θ pour ($h=0.1000, 0.3625, 0.8875, 1.1500$) respectivement

Interprétation

Les résultats du figure (3.4) montrent que le meilleur h qui nous donne le bon estimateur c'est le $h=0.3625$ pour le noyau gaussien.

Cas de noyau K_3

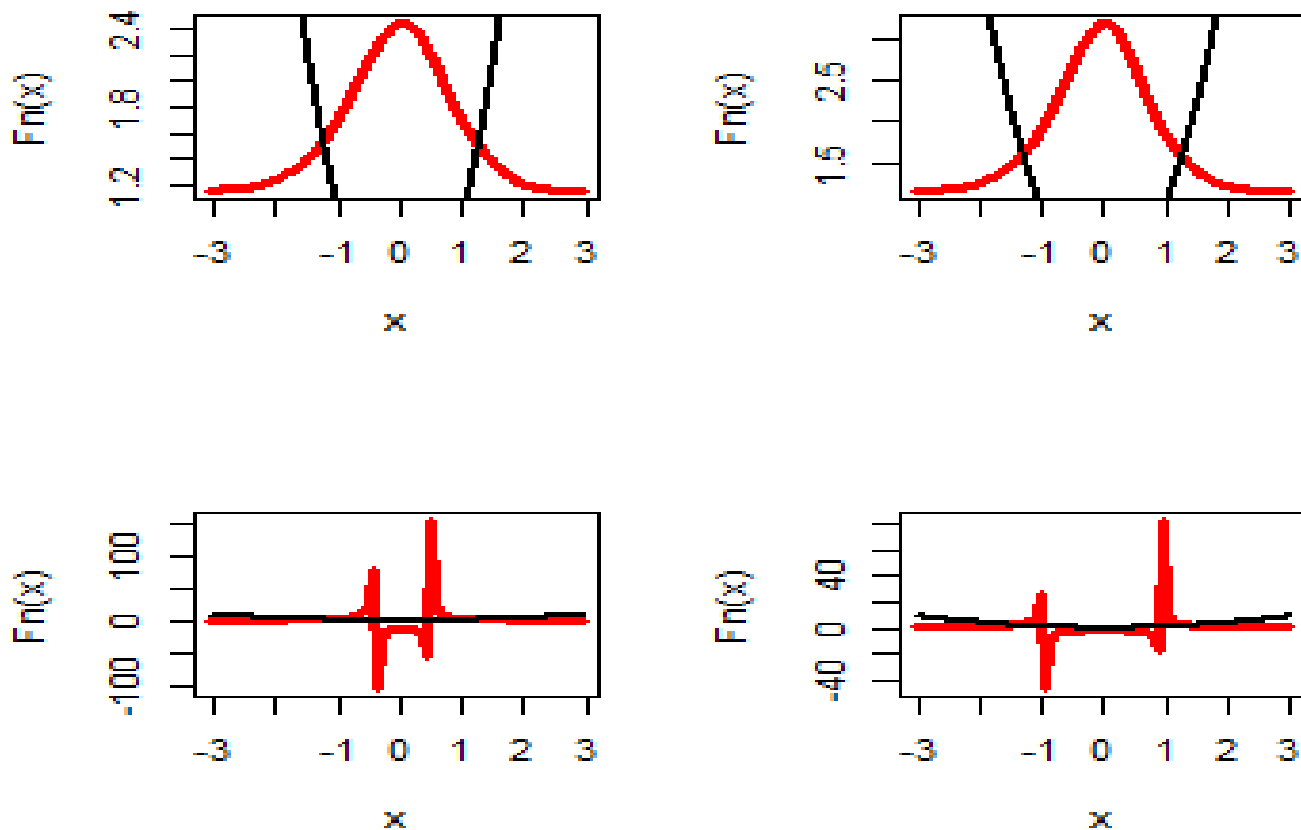


FIGURE 3.5 – Représentation graphique de $\hat{\theta}$ et θ pour ($h=0.1000, 0.3625, 0.8875, 1.1500$) respectivement

Interprétation

Les résultats du figure (3.5) montrent que le meilleur h qui nous donne le bon estimateur c'est le $h=0.8875$ pour le noyau triangulaire.

Comparaison

D'après les figures (figure 3.4, figure 3.5) on remarque que le meilleur noyau qui nous donne le bon h ($h=0.3625$) pour le modèle(3.3) c'est le noyau K_3 (le noyau gaussien).

3.4 Étude de la convergence presque complète par l'utilisation de l'inégalité de Bernstein

Dans cette partie, on va appliquer la convergence presque complète par l'utilisation du lemme(2.2.1) indiquée dans le chapitre (2), pour cela on considère le modèle $Y = \theta(X) + \varepsilon$ et on va essayer de montrer la convergence p.c de $\hat{\theta}(X)$ vers $\theta(X)$ C-à-d :

$$\left[\hat{\theta}_n(X) \xrightarrow{c.p.c} \theta_n(X) \right] \iff \forall \varepsilon_1 > 0 : \\ \lim_{n \rightarrow \infty} P \left(|\hat{\theta}_n(X) - \theta_n(X)| > \varepsilon_1 \right) = 0$$

Preuve.

On a :

$$\begin{aligned} |\hat{\theta}_n(X) - \theta_n(X)| &= |\hat{\theta}_n(X) - E(\hat{\theta}_n(X)) + E(\hat{\theta}_n(X)) - \theta_n(X)| \\ &\leq |\hat{\theta}_n(X) - E(\hat{\theta}_n(X))| + |E(\hat{\theta}_n(X)) - \theta_n(X)| \end{aligned}$$

On a : $E(\hat{\theta}_n(X)) - \theta_n(X) = 0$ (car $\hat{\theta}_n$ est sans biais voir[18])

Donc :

$$P \left(|\hat{\theta}_n(X) - \theta_n(X)| > \varepsilon_1 \right) \leq P \left(|\hat{\theta}_n(X) - E(\hat{\theta}_n(X))| > \varepsilon_1 \right)$$

On va montrer que :

$$\lim_{n \rightarrow \infty} P \left(|\hat{\theta}_n(X) - E(\hat{\theta}_n(X))| > \varepsilon_1 \right) = 0$$

Par l'utilisation de l'inégalité de Bernstein :

On a :

$$\hat{\theta}_n(X) = \frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{X - X_i}{h} \right)$$

Alors :

$$\begin{aligned} \left| \hat{\theta}_n(X) - E \left(\hat{\theta}_n(X) \right) \right| &= \frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{X - X_i}{h} \right) - E \left[\frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{X - X_i}{h} \right) \right] \\ &= \frac{1}{nh} \sum_{i=1}^n (Y_i - E(Y_i)) K \left(\frac{X - X_i}{h} \right) \\ &= \frac{1}{nh} \sum_{i=1}^n \varepsilon_i K \left(\frac{X - X_i}{h} \right) \end{aligned}$$

3.4. ÉTUDE DE LA CONVERGENCE PRÉSCQUE COMPLÈTE PAR L'UTILISATION DE
L'INÉGALITÉ DE BERSTEIN

On pose : $\varepsilon_i K \left(\frac{X-X_i}{h} \right) = \xi_i$

$E(\xi_i) = 0$ car $E(\varepsilon_i) = o$ (ε est une bruit blanc)

O na :

$$P \left(\frac{1}{n} \left| \sum_{i=1}^n \xi_i \right| > \varepsilon_1 \right) \leq 2exp \left\{ \frac{-n\varepsilon_1^2}{4\sigma^2} \right\}$$

$$P \left(\frac{1}{h} \frac{1}{n} \left| \sum_{i=1}^n \xi_i \right| > \frac{1}{h} \varepsilon_1 \right) \leq 2exp \left\{ \frac{-n\varepsilon_1^2}{4\sigma^2} \right\}$$

D'après l'inégalité de Berstein :

$$P \left(\frac{1}{nh} \left| \sum_{i=1}^n \xi_i \right| > \varepsilon' \right) \leq 2exp \left\{ \frac{-n\varepsilon'^2}{4\sigma^2} \right\}$$

Avec : $\varepsilon' = \frac{\varepsilon_1}{h}$

$$P \left(\frac{1}{nh} \left| \sum_{i=1}^n \xi_i \right| > \varepsilon' \right) \leq 2exp \{-g_n(X)\} \quad \text{avec } g_n(X) = \left\{ \frac{-n\varepsilon'^2}{4\sigma^2} \right\}$$

Donc :

$$P \left(\left| \hat{\theta}_n(X) - E(\hat{\theta}_n(X)) \right| > \varepsilon' \right) \leq 2exp \{-g_n(X)\}$$

$$\lim_{n \rightarrow \infty} P \left(\left| \hat{\theta}_n(X) - E(\hat{\theta}_n(X)) \right| > \varepsilon' \right) \leq \lim_{n \rightarrow \infty} 2exp \{-g_n(X)\}$$

$$\lim_{n \rightarrow \infty} P \left(\left| \hat{\theta}_n(X) - E(\hat{\theta}_n(X)) \right| > \varepsilon' \right) \leq 0$$

Alors :

$$\lim_{n \rightarrow \infty} P \left(\left| \hat{\theta}_n(X) - \theta_n(X) \right| > \varepsilon_1 \right) = 0$$

Donc : La convergence présque complète de $\hat{\theta}_n(X)$ vers $\hat{\theta}_n(x)$.

Et on écrit :

$$\hat{\theta}_n(X) \xrightarrow{c.p.c} \theta_n(X)$$

□

Conclusion

La littérature sur l'estimation non paramétrique et les inégalité exponentielles reste encore d'actualité, notamment ce qui concerne l'estimation non paramétrique pour la fonction de régression.

Les simulations que nous avons réalisées, nous permettent de montrer la performance de notre estimateur et compte tenu aussi des résultats des simulations, le choix de paramètre de lissage est crucial dans le comportement asymptotique de l'estimateur, ainsi le choix du noyau, nous avons alors sélectionné la valeur de paramètre de lissage h qui fournisse le meilleur estimateur de la fonction de régression.

Par ailleurs nous avons montré que la méthode du noyau est un outil très efficace pour estimer la fonction de régression.

Dans ce travail aussi, nous avons établi un résultat de la convergence presque complète par l'utilisation de l'inégalité de Bernstein dans le cas des variable aléatoire i.i.d.

Perspectives de recherche :

- Estimer la fonction de régression dans le cas multidimensionnel.
- Estimer la fonction de régression dans le cas des variables aléatoires associées mélangeants.

Résumé

Dans ce mémoire, nous étudions l'estimation non paramétrique du modèle de régression pour les variables fonctionnelles i.i.d ainsi que quelques illustrations graphiques concernant le meilleur noyau qui nous donne le bon estimateur sous différents modèles de régression.

Nous avons donné la convergence presque complète de l'estimateur non paramétrique pour la fonction de régression par l'utilisation de l'inégalité exponentielle et construisons l'estimateur par la méthode du noyau pour la fonction de régression .

Mots clés : Estimation non paramétrique, la convergence presque complète, inégalité exponentielle, la méthode du noyau.

Bibliographie

- [1] **AMROUN, S**, (2011), *Sur L'Estimation De La Courbe De Régression De La Moyenne*, Mémoire De Magister, Université A. Mira de Béjai.
- [2] **Boukhames, O** and **ZAIDI, M**, (2015-2016), *Estimation Non Paramétrique De La Densité De Probabilité Et De La Fonction De Régression* ,Mémoire de Master,université de Boumerdes.
- [3] **Belguerna, A**, *Propriétés asymptotiques pour des estimateurs non paramétriques Des processus autorégressifs*, thèse de DOCTEUR , Université de Sidi Bel-Abbès
- [4] **Berlinet, A.** and **Devroye, L.** (1989), 'A comparison of kernel density estimates',*Publications de l'Institut de Statistique de l'Université de Paris*.
- [5] **Bowman, A**, (1984), 'An alternative method of cross-validation for the smoothing density estimate',*Biometrika*, 71 , 352–360.
- [6] **Boukhames,S.** and **ZAIDI ,M.**(2016),*Estimation non paramétrique de la densité de probabilité et de la fonction de régression*,Mémoire de Master,université de Boumerdes.
- [7] **Cuevas, A. , Cao, R,** and **Gonzalez-Manteiga, W.** (1994), 'A comparative study of several smoothing methods in density estimation', *Computational Statistics and Data Analysis*, 17, 235–239.
- [8] **Djemoi,G** , (2017),*Modélisation non paramétrique pour les variables aléatoires fonctionnelles cas de données indépendantes*,Mémoire de Master,université ouargla.
- [9] **Dusart,P**,(2018)*Cours de Statistiques inférentielles*.
- [10] **Ferhat, I**, (2012),*Estimation d'une fonction de densité par la méthode des noyaux et application à la var*,Mémoire de Master,université de Tizi-ouzou.
- [11] **Hoeffding ,W**,(1963), *Probability inequalities for sums of bounded random variables*, **J, Amer**, *Statist. Assoc*, 58,13-30.

-
- [12] **Jemai, A.** (2018), *Estimation fonctionnelle non paramétrique au voisinage du bord*, Thèse de docteur de l'université de poitiers, Université de Carthage, Tunisie.
- [13] **Lejeune, M.** (2010), *Statistique La théorie et ses applications*, Springer-Verlag France, Paris.
- [14] **Refai, M.** (2018), *Estimation Fonctionnelle de la Densité Conditionnelle*, Mémoire de Master, université de Mostaganem.
- [15] **Mousli, A.** (2015), *Sur Les Inégalités Exponentielles dans l'Estimation Fonctionnelle*, mémoire de Magister université de Houari Boumediene.
- [16] **Marron, J.S.** (1987), 'A comparison of cross-validation techniques in density estimation', *The Annals of Statistics*, 15, 152–162.
- [17] **Nadarya, E.** *Some new estimates for distribution function*, *Theory of Probability and its Application*, 9, 497-500.
- [18] **Mouhou, M.** (2015), *Prévision non paramétrique via la fonction de régression*, Memoire de Master, Université de Bejaia.
- [19] **Parzen, E.** *On the estimation of a probability density function and mode*, *Annals of Mathematical Statistics*, 33 :1065-1027, 1962.
- [20] **Rosenblatt, M** *Remarks on some non parametric estimates of a density function*, *Annals of Mathematical Statistics*, 27 :8332-837, 1956.
- [21] **Ruppert, D.** (1982), 'Almost sure approximations to the Robbins-Monro and Kiefer-Wolfowitz processes with dependent noise', **Annals of Probability**, 10, 178–187.
- [22] **Rudemo, M.** (1982), 'Empirical choice of histograms and kernel density estimators', *Scandinavian Journal of Statistics*, 9, 65–78.
- [23] **Park, B.U.** and **Marron, J. S.** (1992), 'On the use of pilot estimators in bandwidth selection', *Journal of Non parametric Statistics*, 3 , 231–240.
- [24] **Sarda, P.** and **Vieu, P.** (1991), 'Smoothing parameter selection in hazard estimation', *Statistics and Probability Letters*, 5, 429–434.
- [25] **Stone, C. J.** (1984), 'An asymptotically optimal window selection rule for kernel density estimates', *The Annals of Statistics* , 12, 1285–1297.
- [26] **Scott, D.W., Tapia, R.A.** and **Thompson, J.R.** (1977), 'Kernel density estimation revisited', *Nonlinear Analysis*, 1, 339–372.
- [27] **Talagrand** ,(1996), *New concentration inequalities in product spaces*, *Invent, Math*, 126 :503-563.

- [28] **Youndjé, E., Sarda, P. and Vieu, P.** (1996a), '*Optimal smooth hazard estimates*', *Test*, 2, 379–394.
- [29] **Youndjé, E., Sarda, P. and Vieu, P.** (1996b) '*Choix de paramètres de lissage en estimation de densité conditionnelle*', *Publications de l'Institut de Statistique de Paris*, 38, 57–80.
- [30] **Zernoun ,K** ,(2016-2017),*Estimation non paramétrique de la densité de probabilité et de la fonction de répartition par des séries orthogonales*,*Mémoire de Master,université de Bejaia*