

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

Université M'hamed Bougara Boumerdes

Faculté des Sciences



Mémoire Présenté

Pour L'Obtention Du Diplôme De Master

Domaine : Mathématique et Informatique

Département : Mathématiques

Option : Mathématique Financier

Par : BOUKHAMES Oussama

Et : ZAIDI Mohamed

Thème

***Estimation Non Paramétrique De La Densité
De Probabilité Et De La Fonction De
Régression***

Soutenu à l'UMBB, en juin 2016, devant le jury composé de :

Mr AKLIOUAT.K	Président à l'UMBB -Boumerdes.
Mme YASSA.M	Encadreur à l'UMBB - Boumerdes
Mr RAHMOUNE.A	Examineur à l'UMBB - Boumerdes.

Année Universitaire 2015-2016

Remerciements

Nous tenons à exprimer nos plus vifs remerciements à ALLAH pour nous avoir facilité ce mémoire.

Nous remercions sincèrement tous ceux qui ont contribué de près ou de loin, à la réalisation de ce Project, plus particulièrement :

Notre promotrice Mme YASSA Malika pour l'aide et le temps qu'il a voulu nous consacrer et qu'on ne remerciera jamais assez pour ses lectures, ses conseils et sa patience.

L'ensemble des membres de jury qui nous font le grand honneur d'accepté d'examiner et d'évaluer notre travail.

Une spéciale remerciement pour Mr MOUCI Aïssa pour son aide.

Finalement, nous remercions tous ceux qui nous aïdions de près ou de loin quel qu'ils soient et d'où qu'ils soient.



Dédicaces

Je dédie mon mémoire :

*A mon père Youcef & ma mère Hassiba vous êtes pour moi une source de vie car sans vos sacrifices, votre tendresse et votre affection je ne pourrais arriver jusqu'au bout. Que *allah* vous garde afin que votre regard puisse suivre ma destinée.*

A mon frère Abdou & ma sœur Meriem sui été toujours présent pour moi

A mes amis avec qui j'ai partagé des moments des plus agréables.

Particulièrement à mon binôme Mohamed

Oussama



Dédicaces

Je dédie mon mémoire :

A mon père Mouloud & ma mère Khdauidj vous êtes pour moi une source de vie car sans vos sacrifices, votre tendresse et votre affection je ne pourrais arriver jusqu'au bout. Que Allah vous garde afin que votre regard puisse suivre ma destinée.

A mon frère Omar & mes sœurs sui été toujours présent pour moi.

A mes amis avec qui j'ai partagé des moments des plus agréables.

Particulièrement à mon binôme Oussama.

Mohammed



Table des matières	1
Liste des figures	3
Introduction général	4
I-Estimation de la densité	7
Introduction.....	8
1-Notions préliminaires	8
1-1-La convergence en moyenne quadratique.....	8
1-2-L'espérance d'une fonction de la variable aléatoire	9
1-3-La densité de probabilité	9
1-4-L'espérance conditionnelle	11
1-5-convergence en loi	12
1-6-Convergence presque sûre	12
2-Estimation par les histogrammes.....	13
2-1-Histogramme	13
2-2-Relation entre densité de probabilité et histogramme des données	14
2-3-Estimateur de densité par histogramme	15
2-4-Propriétés statistiques de l'estimateur par histogramme	16
2-5-Utilisation de l'estimateur de densité par histogrammes	18
3-Estimateur simple de la densité	21
4-méthode du noyau	23
4-1-Définitions et propriétés élémentaires	23
4-2-Choix de paramètre de lissage h	30
4-3-Cross validation	32
Conclusion	36
II-Estimation de la fonction de régression	37
Introduction	38
1-Estimateurs à noyau de la fonction de régression	39
1-1-L'estimateur de Nadaraya-Watson	39
1-2-L'estimation par polynômes locaux	40
1-3-Propriétés asymptotiques de l'estimateur de Nadaraya-Watson	42
Conclusion	43

III-Simulation	45
Introduction	46
1-Le choix de langage de programmation	46
2-Plan de simulation de fonction de la densité	47
2-1-Algorithmme de simulation	47
2-2-Résultat de simulation	48
2-3-Interpritation	50
3-Plan de simulation de fonction de régression	50
3-1-Algorithmme de simulation	51
3-2-Résultat de simulation	52
3-3-Interpritation	54
Conclusion générale	55
Annexe	58
Bibliographie	60

Liste Des Figures

[1]	Figure 1.1: fonction de la densité(1)	10
[2]	Figure 1.2 : fonction de la densité(2)	10
[3]	La Figure 2.1 : un histogramme de 100 observations ; sur un intervalle $\Omega = [-5, 5]$; $h = 0.8$.	14
[4]	La Figure 2.2 : un histogramme de 100000observations ; sur un intervalle $\Omega = [-5, 5]$; $h = 0.2$.	15
[5]	Figure 2.3.a : Histogramme de 300 observations construit avec une granularité $h = 2.5$.	19
[6]	Figure 2.3.b : Histogramme de 300 observations construit avec une granularité $h = 0.1$.	19
[7]	Figure 2.4.a : Histogramme de 300 observations construit sur la portion $\Omega = [-5, 5]$.	20
[8]	Figure 2.4.b: Histogramme de 300 observations construit sur la portion $\Omega = [-4, 6]$.	20
[9]	Figure 2.1: Noyau Uniform (Rosenblatt).	24
[10]	Figure 2.2: Noyau Triangulaire.	24
[11]	Figure 2.3: Noyau d'Epanechnikov ou parabolique.	25
[12]	Figure 2.4: Noyau gaussien.	25
[13]	Figure 3.1: $N = 50$, $\mu=0$, $\sigma=0.5$, $h=0.242371$.	48
[14]	Figure 3.2: $N = 500$, $\mu=0$, $\sigma=0.5$, $h=0.152926$.	49
[15]	Figure 3.3: $N = 5000$, $\mu=0$, $\sigma=0.5$, $h=0.096489$.	49
[16]	Figure 3.4: $N = 50000$, $\mu=0$, $\sigma=0.5$, $h=0.060881$.	50
[17]	Figure 3.5: $N = 10$, $\mu=0$, $\sigma=1$, $a=10, b=1$, $h=0.969486$.	52
[18]	Figure 3.6: $N = 50$, $\mu=0$, $\sigma=1$, $a=10, b=1$, $h=0.611704$.	52
[19]	Figure 3.7: $N = 500$, $\mu=0$, $\sigma=1$, $a=10, b=1$, $h=0.385961$.	52
[20]	Figure 3.8: $N = 1000$, $\mu=0$, $\sigma=1$, $a=10, b=1$, $h=0.243524$.	53



***Introduction
générale***

Introduction générale

La densité est une notion essentielle en statistique et probabilité. Même si la fonction de densité et de répartition, toutes les deux, caractérisent entièrement la loi de probabilité d'une variable aléatoire, la densité de probabilité à un net avantage sur le plan visuel. En fait, la fonction de densité est beaucoup plus facile à interpréter que celle de la fonction de répartition.

Pour la régression, la représentation graphique de deux variables aléatoires n'est pas suffisante pour établir une relation interprétable. Le but de l'analyse de régression est alors d'approximer la moyenne conditionnelle $m(x) = E(Y/X = x)$ qui s'appelle la fonction de régression

Pour estimer ces deux fonctions la densité et la fonction de régression, on ne dispose généralement que d'un ensemble fini d'observations issu d'une même variable aléatoire. Pour construire des estimateurs des deux fonctions. Il existe deux approches très complémentaires pour réaliser ces estimations : l'approche paramétrique et l'approche non-paramétrique.

Dans le cas paramétrique, on a une information sur la loi de probabilité, par exemple on suppose que la densité est connue mais fonction de paramètres inconnus et on désire estimer ces paramètres ou que la fonction de régression a une forme particulière (linéaire, polynomiale...) et on estime les paramètres du modèle sur la base de l'information que l'on a sur la loi de probabilité.

La particularité des méthodes d'estimation non paramétrique est que le paramètre inconnu qu'on cherche à estimer n'est pas supposé d'appartenir à une famille indiquée par un petit nombre de paramètres réels. En général, dans la théorie non paramétrique, on suppose que le nombre de paramètres qui décrivent la loi des observations est infini

Par exemple dans le cas de la fonction de densité et la fonction de la régression on ne les suppose aucune hypothèse à priori sur la forme de la densité et la forme de la relation entre les variables pour la régression

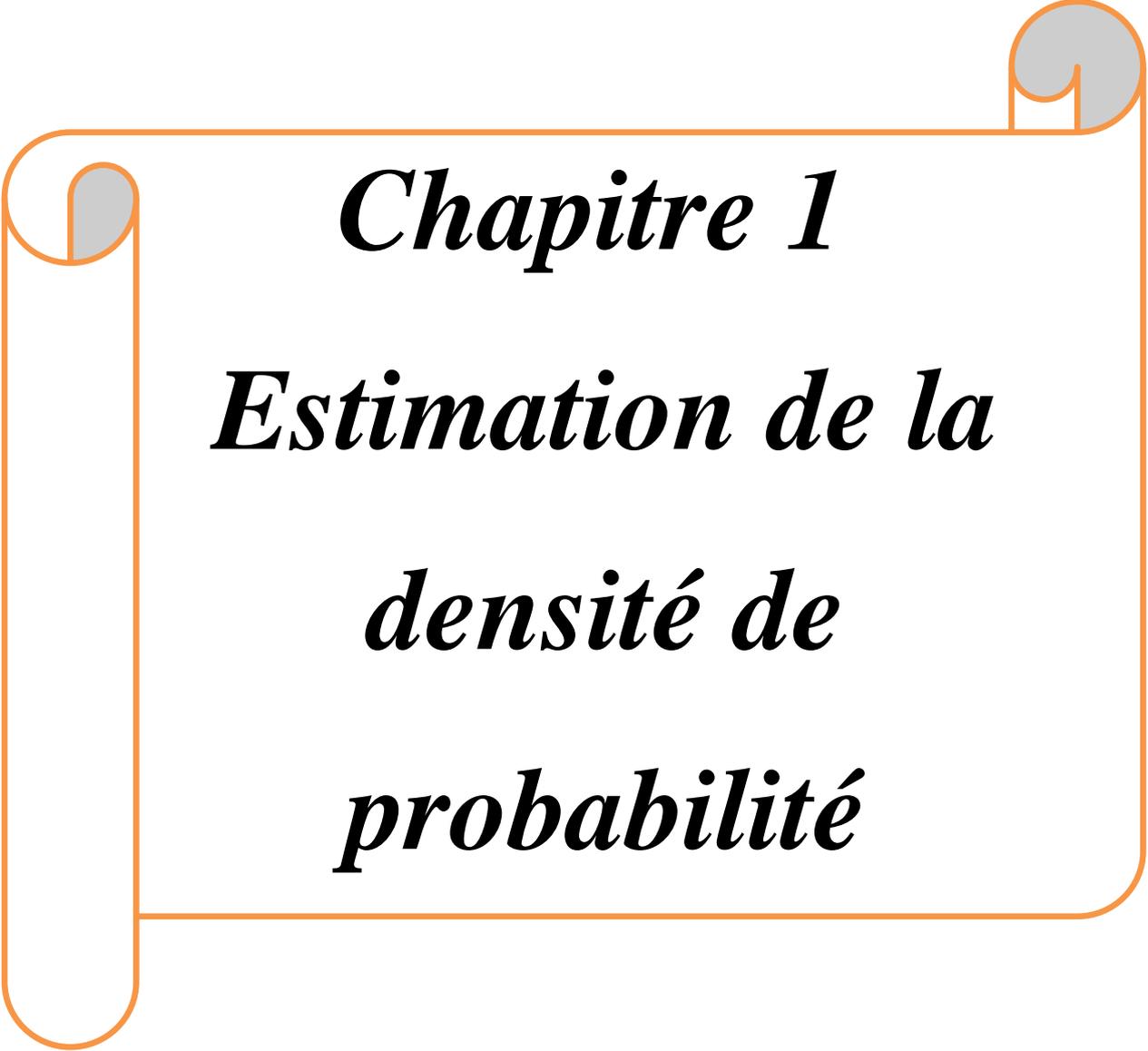
Dans ce mémoire nous étudierons quelques méthodes d'estimation non paramétrique.

Dans le chapitre un nous concentrons sur les méthodes d'estimation de la densité :

La méthode d'estimation par histogramme, la méthode d'estimation simple de la densité 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. Nous présentons également, les propriétés statistiques de chaque méthode d'estimation.

Dans le chapitre deux, nous présentons la méthode de Nadaraya Watson et celle de polynômes locaux qui est une généralisation de la méthode de Nadaraya Watson. Nous présentons aussi, les propriétés statistiques des estimateurs obtenus.

Dans le chapitre trois nous illustrons notre étude par des simulations afin de la performance de ces estimateurs.



Chapitre 1
Estimation de la
densité de
probabilité

Introduction

Pour estimer la fonction de densité que l'on suppose connue par l'estimation paramétrique, nous utilisons par exemple la méthode de l'estimation de vraisemblance ou bien la méthode de moindre carré.....etc., mais si cette fonction de densité est inconnue nous utilisons l'estimation non paramétrique

L'avantage principal de l'estimation non-paramétrique de la densité de probabilité sous-jacente à un ensemble fini d'observations est de ne pas nécessiter d'hypothèses a priori sur l'appartenance de cette densité à une famille de lois connues. L'estimation ne concerne pas les paramètres permettant de sélectionner une loi, mais directement la fonction elle-même (d'où le terme non-paramétrique).

Soit x_1, x_2, \dots, x_n .n observation équipondérées issues d'une variable aléatoire réelle X de densité de probabilité réelle $f(x)$ inconnue. Comment obtenir une estimation de $f(x)$ à partir de la seule information contenue dans l'échantillon ?

Ce problème, que l'on désigne généralement par estimation non paramétrique de la densité de probabilité a fait l'objet de multiples travaux par des méthodes diverses, citons :

- L'estimateur par histogramme.
- L'estimateur par la méthode du noyau.

Dans ce chapitre, nous allons présenter une étude détaillée de l'estimateur par la méthode du noyau ainsi que ses propriétés statistiques.

1-Notions préliminaires

1-1-La convergence en moyenne quadratique

1-1-1-Convergence en moyenne ou dans $L^1[a, b]$

La suite de fonctions $\{f_n\}$ converge vers f en moyenne quadratique ou dans $L^1[a, b]$ si :

- pour tout n les fonctions f_n sont dans $L^1[a, b]$,
- si f est aussi dans $L^1[a, b]$,
- et si de plus $\|f_n - f\|_{L^1[a, b]} \xrightarrow{n \rightarrow +\infty} 0$.

On note alors :

$$f_n \xrightarrow{L^1[a, b]} f$$

Remarque : attention, on peut avoir $\|f_n - f\|_{L^1[a, b]} \xrightarrow{n \rightarrow +\infty} 0$ sans que les fonctions f_n et f soient dans $L^1[a, b]$.

1-1-2-Convergence en moyenne quadratique ou dans $L^2(I)$

La suite de fonctions $\{f_n\}$ converge vers f en moyenne quadratique ou dans $L^2[a, b]$ ssi :

- pour tout n les fonctions f_n sont dans $L^2[a, b]$.
- si f est aussi dans $L^2[a, b]$.
- et si de plus $\|f_n - f\|_{L^2[a, b]} \xrightarrow{n \rightarrow +\infty} 0$.

On note alors :

$$f_n \xrightarrow{L^2[a, b]} f$$

On dit aussi, en électronique, que les signaux f_n convergent au sens de l'énergie vers le signal f .

1-2-L'espérance d'une fonction de la variable aléatoire

Définition : On considère une variable aléatoire X et une fonction h . On aimerait connaître l'espérance de la nouvelle variable $Y = h(X)$.

Proposition : Soit X une variable aléatoire, h une fonction définie sur \mathbb{R} . Alors :

1-Si X est une variable discrète à valeurs dans $D = \{x_1 \dots \dots x_n\}$. on a

$$E[h(X)] = h(x_1)P(X = x_1) + \dots + h(x_n)P(X = x_n) = \sum_{i=1}^n h(x_i)P(X = x_i)$$

2-Si X est une variable discrète à valeur dans l'ensemble infini $D = \{x_i : i \geq 1\}$. lorsque la somme est bien définie, on a :

$$E[h(X)] = \sum_{i=1}^{+\infty} h(x_i)P(X = x_i)$$

3-si X est une variable à densité f , lorsque l'intégrale est bien définie, on a :

$$E[h(X)] = \int_{\mathbb{R}} h(x)f(x)dx.$$

1-3-La densité de probabilité

Définition 1 : On appelle densité de probabilité d'une variable aléatoire continue X , toute fonction f continue et positive sur un intervalle $I \in \mathbb{R}$ telle que :

- $P(x \in I) = \int_I f(t)dt = 1$
- Pour tout intervalle $J = [\alpha, \beta]$ inclus dans I , on a : $P(X \in J) = \int_{\alpha}^{\beta} f(t)dt$.

D'autre part la fonction F définie par : $F(x) = P(X \leq x)$ est appelée la fonction de répartition de la variable X .

$$F(x) = \int_a^x f(t)dt \quad \text{ou} \quad \lim_{a \rightarrow -\infty} \int_a^x f(t)dt.$$

Remarques :

- Comme la fonction f est continue et positive, la probabilité $P(X \in I)$ correspond à l'aire sous la courbe Cf . Elle vaut alors 1 u.a.
- La probabilité $P(X \in J)$, avec $J = [\alpha; \beta]$, correspond à l'aire du domaine délimité par Cf , l'axe des abscisses et les droites d'équation $x = \alpha$ et $y = \beta$

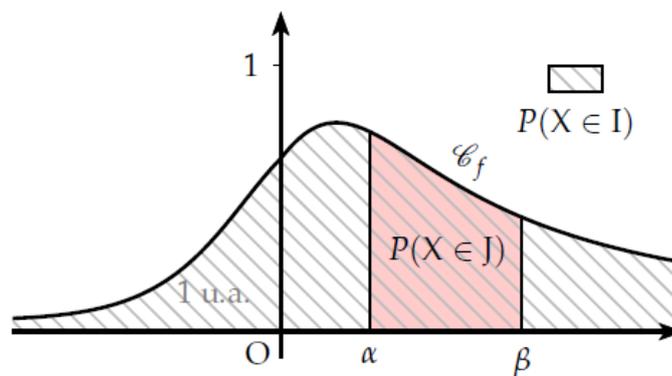


Figure 1.1: fonction de la densité (1)

- Comme la probabilité que X prenne une valeur isolée est nulle, que l'intervalle J soit ouvert ou fermé importe peu. Ainsi :

$$\begin{aligned} P(X \in [\alpha, \beta]) &= P(X \in]\alpha, \beta]) \\ &= P(X \in]\alpha, \beta]) \\ &= P(X \in]\alpha, \beta]) \end{aligned}$$

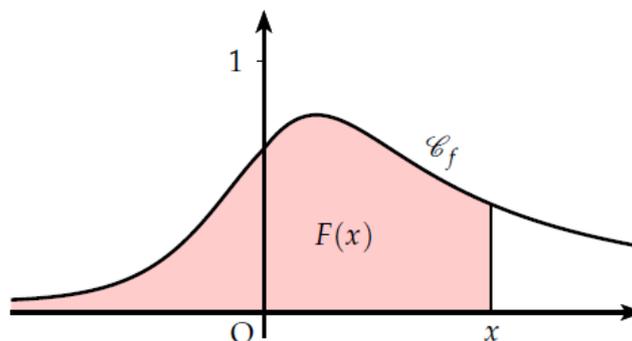


Figure 1.2 : fonction de la densité (2)

- L'écriture $(X \in I)$ est une notation abusive car X n'est pas un nombre, mais la fonction qui associe une issue à un nombre. Elle prolonge la notation déjà utilisée pour des variables discrètes ($X = \alpha$).

Définition 2 : L'espérance mathématique d'une variable aléatoire continue X de la densité est :

$$E(X) = \int_{(I)} t f(t) dt$$

Définition 3 : L'espérance d'une variable aléatoire est, lorsqu'elle existe, la moyenne des valeurs de cette variable, pondérées par leurs probabilités de réalisation. On voit bien comment traduire cette définition informelle dans le cas d'une variable aléatoire X discrète en posant :

$$E(X) = \sum_{x \in X(\Omega)} x P(X = x).$$

Cette formule n'a de sens que si la famille de réels $\{xP(X = x); x \in X(\Omega)\}$ est sommable, ce qui se traduit par la condition suivante pour l'existence de l'espérance de la v.a. discrète X :

$$\sum_{x \in X(\Omega)} |x| P(X = x) < +\infty.$$

1-4-L'espérance conditionnelle

Lorsqu'on travaille avec des variables aléatoires (v.a.s) discrètes on introduit la notion de probabilité conditionnelle par la formule :

$$P(Y = y|X = x) := \frac{P(Y = y, X = x)}{P(X = x)} \text{ si } P(X = x) > 0$$

D'où la définition d'espérance conditionnelle de $f(Y)$ sachant que $X = x$ par :

$$\sum_y f(y) P(Y = y|X = x) = u_f(x)$$

Pour toute fonction f à valeurs réelles ou complexes mesurable bornée. Souvent on utilise la notation $E(f(Y)|X = x) := u_f(x)$. Cette définition pose des problèmes lorsque on travaille avec des v.a.s continues où avec des modèles aléatoires plus complexes (suites infinies des v.a.s, espaces des fonctions, ...etc). Par exemple, si X est continue, alors $P(X = x) = 0$ pour tout $x \in \mathbb{R}$, donc la définition ci dessous est inutilisable. La façon plus efficace de contourner cette difficulté est de modifier la définition de probabilité conditionnelle, en effet,

il sera plus facile de définir d'abord l'espérance conditionnelle et après la probabilité conditionnelle comme un sous-produit.

1-5-convergence en loi

Définition 1 : (Convergence étroite des mesures.) Soit (μ_n) une suite de mesures de probabilité sur \mathbb{R}^d et μ une mesure sur \mathbb{R}^d . On dit que la suite (μ_n) converge étroitement vers la mesure de probabilité μ si pour toute fonction continue bornée $f: \mathbb{R}^d \rightarrow \mathbb{R}$.

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d \mu_n = \int_{\mathbb{R}^d} f d \mu$$

On note souvent $\mu_n \Rightarrow \mu$.

Définition 2 : On dit qu'une suite de v.a. $(X_n)_{n \geq 1}$ à valeurs dans \mathbb{R}^d converge en loi vers une v.a. X si les lois μ_{X_n} convergent étroitement vers μ_X , la loi de X . Autrement dit, si :

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)), \forall f \in C_b^0(\mathbb{R}^d, \mathbb{R})$$

On le note : $X_n \xrightarrow{\text{Loi}} X$.

N.B : On le note $C_b^0(\mathbb{R}^d, \mathbb{R})$ l'ensemble des fonctions continues bornées de \mathbb{R}^d dans \mathbb{R} .

Remarque : On voit que la convergence en loi ne fait référence qu'à la loi des v.a. (contrairement aux convergences p.s ; en probabilité, et L^p). On peut très bien définir la convergence en loi d'une suite de v.a. qui ne seraient pas définies sur le même espace de probabilité.

1-6-Convergence presque sûre

La convergence de variables aléatoires correspond à la convergence de fonctions. La convergence la plus faible est la convergence simple qui s'énoncerait : pour tout $\omega \in \Omega$. $X_n(\omega) \rightarrow X(\omega)$ quand $n \rightarrow +\infty$. En probabilité (et plus généralement en théorie de la mesure), il est trop restrictif de demander la convergence pour tous les $\omega \in \Omega$. À la place, on définit la convergence presque sûre :

Définition : Une suite $(X_n)_n$ de variables aléatoires converge presque sûrement vers X si la convergence est vraie avec une probabilité 1.

$\mathbb{P}(\omega \in \Omega ; \lim X_n(\omega) = X(\omega)) = 1$. Lorsque $n \rightarrow +\infty$.

On note la convergence presque sûre : $X_n \rightarrow X$.p.s.

D'un point de vue analyse, c'est une convergence plus faible que la convergence simple (Déjà très faible). Mais d'un point de vue probabiliste, ce sera une des convergences les plus fortes que nous manipulerons.

2-Estimateur par l'histogramme

L'histogramme est le plus ancien des estimateurs non-paramétriques de densité, l'origine des histogrammes est attribuée à John Graunt au XVIIe siècle répondant à l'objectif d'une représentation de la distribution de données. À ce titre, il peut être considéré comme un estimateur de la densité de probabilité sous-jacente à un ensemble fini d'observations. Nous voyons, dans ce chapitre, comment obtenir cette dernière propriété. Nous discutons ensuite des propriétés statistiques de l'estimateur de densité par histogramme ainsi que des problèmes engendrés par son utilisation pour représenter la densité de probabilité.

Dans la suite de ce manuscrit, l'ensemble de référence Ω est un intervalle et e_{min} (rsp. e_{max}) est le plus petit (rsp. plus grand) élément de Ω .

2-1-Histogramme

Construire un histogramme à partir d'un ensemble d'observations (précises) x_1, \dots, x_n consiste à partitionner l'intervalle de référence $\Omega = [e_{min}, e_{max}]$ en $p \in \mathbb{N}$ classes $A_k, k \in \{1, \dots, p\}$, et à compter le nombre d'observations appartenant à chaque classe A_k . Si toutes les classes de l'histogramme ont la même largeur, on dit que l'histogramme est régulier. On note $h \in \mathbb{R}^+$, la largeur des classes qui est alors appelée le « pas » ou la fenêtre de l'histogramme. La valeur de h est donnée par :

$$h = \frac{e_{max} - e_{min}}{p} \quad (\text{I.1})$$

Le nombre d'observations appartenant à chaque classe A_k est noté Acc_k . Il est défini :

$$Acc_k = \sum_{i=1}^n 1_{A_k}(x_i) \quad (\text{I.2})$$

Où $1_{A_k}(x_i)$ est la fonction caractéristique de l'ensemble A_k définie par :

$$1_{A_k}(x) = \begin{cases} 1 & \text{si } x \in A_k \\ 0 & \text{sinon} \end{cases} \quad (\text{I.3})$$

La Figure 2.1 présente un histogramme de 100 observations tirées aléatoirement d'une loi normale centrée réduite $N(0, 1)$. Ces observations sont réparties sur un intervalle de référence $\Omega = [-5, 5]$. La fenêtre de l'histogramme est fixée à $h = 0.8$.

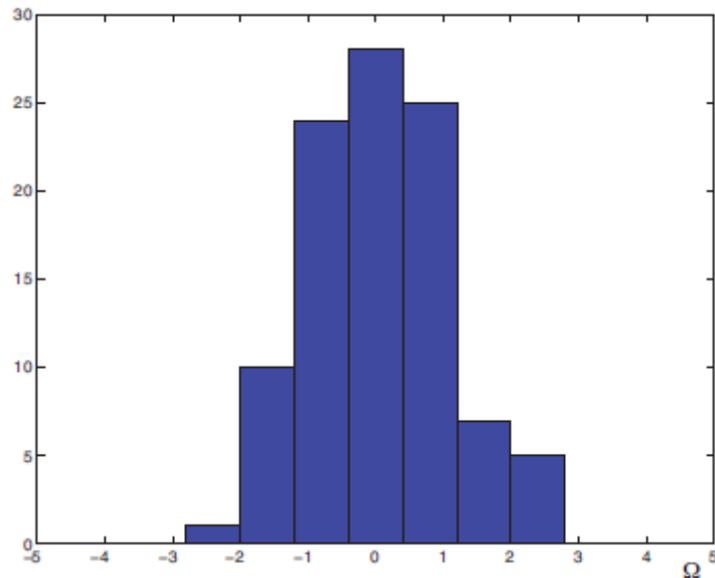


Figure 2.1

2-2-Relation entre densité de probabilité et histogramme des données

Histogramme et densité de probabilité sont liés par des conditions aux limites : une densité de probabilité peut être vue comme la limite d'un histogramme lorsque le nombre d'observations tend vers l'infini et que la valeur h de l'histogramme tend vers zéro.

La Figure 2.2 illustre, cette relation, en considérant le même processus d'observations précédent avec un grand nombre d'observations ($n = 100000$) et un plus faible pas ($h = 0.2$). En comparant la Figure 2.1 avec la Figure 2.2, on voit clairement que l'allure de cet histogramme se rapproche de l'allure de la densité de probabilité ($N(0, 1)$) quand le nombre d'observations n augmente et la valeur de h diminue.

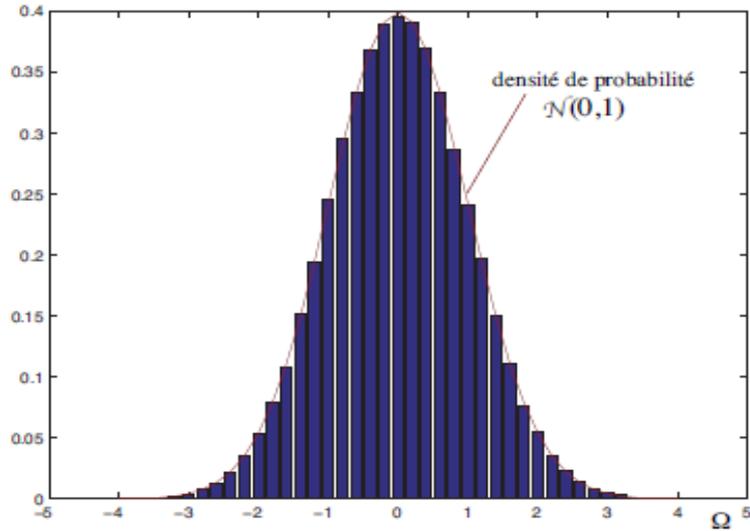


Figure 2.2

2-3-Estimateur de densité par histogramme

Soit un ensemble fini d'observations (x_1, \dots, x_n) de n variables aléatoires (X_1, \dots, X_n) indépendantes (au sens de probabilité) et identiquement distribuées (de même loi de probabilité) (i. i. d) de densité de probabilité commune f . Pour tout $k \in \{1, \dots, p\}$, soit Acc_k le nombre d'observation appartenant à la classe A_k , défini par l'expression (I.2), la probabilité de A_k (basée sur les observations x_i), notée $p(A_k)$, est donnée par :

$$p(A_k) = \frac{Acc_k}{n} \quad (I.4)$$

En émettant l'hypothèse, généralement irréaliste, que les observations se répartissent uniformément dans la classe A_k (de largeur h), on peut alors construire un estimateur de la densité f , pour tout $x \in \Omega$, par :

$$\hat{f}_h^n(x) = \frac{1}{h} \sum_{k=1}^p p(A_k) 1_{A_k}(x) \quad (I.5)$$

Cet estimateur peut aussi s'écrire

$$\hat{f}_h^n(x) = \frac{1}{nh} \sum_{k=1}^p Acc_k 1_{A_k}(x) \quad (I.6)$$

Dans la suite, nous émettons l'hypothèse que les classes $A_k, \forall k \in \{1, \dots, p\}$ forment une partition de Ω (c'est-à-dire $\Omega = \cup_{k=1}^p A_k$ et $\forall i, j \in \mathbb{N} : i \neq j, A_i \cap A_j = \emptyset$) et définissons pour chaque classe A_k , son centre A_k telles que :

$$\forall k \in \{1, \dots, p\}, Ak = [a_k - \frac{h}{2}, a_k + \frac{h}{2}] \text{ et } \forall k \in \{1, \dots, p - 1\}, a_{k+1} = a_k + h$$

2-4-Propriétés statistiques de l'estimateur par histogramme

Nous présentons, dans ce titre, les propriétés statistiques de l'estimateur par histogramme \hat{f}_h^n défini par l'expression (I.6) (pour une étude plus détaillée voir, par exemple, les livres de Bosq et Lecoutre [13] et de Simonoff [29]).

En statistiques, il est nécessaire de mesurer la qualité d'un estimateur. Pour cela, on évalue, d'une part, l'écart entre la moyenne de l'estimateur et la densité à estimer, ce critère d'évaluation est appelé biais¹, et d'autre part, la variance de l'estimateur (due au caractère aléatoire d'observations) qui caractérise la dispersion des valeurs de l'estimateur dans l'ensemble d'observations. On essaye généralement de réduire au mieux ces deux quantités.

2-4-1-Définition de biais

Le biais de l'estimateur \hat{f}_h^n (I.6) est donné [29], pour tout $x \in [Ak, Ak+1]$, $\forall k \in \{1, \dots, p\}$, par :

$$\begin{aligned} \text{Biais} \left(\hat{f}_h^n(x) \right) &= E \left(\hat{f}_h^n(x) \right) - f(x) \\ &= \frac{1}{2} \dot{f}(h - 2(x - a_k)) O(h^2) \end{aligned} \tag{I.7}$$

Où O est un terme résiduel et \dot{f} est la dérivée de f . \dot{f} doit être une fonction de $L(\Omega)$ absolument continue² et carrée intégrable³

2-4-2-Variance de l'estimateur

La variance de l'estimateur est donnée [29] pour tout $x \in \Omega$, par :

$$\begin{aligned} \text{Var} \hat{f}_h^n(x) &= E \left(\left(\hat{f}_h^n(x) \right)^2 \right) - \left(E \left(\hat{f}_h^n(x) \right) \right)^2 \\ &= \frac{f(x)}{nh} + O(n^{-1}) \end{aligned} \tag{I.8}$$

¹. Le biais d'un estimateur \hat{s}^n de $s \in L(\Omega)$ est donné par : $\text{biais}(\hat{s}^n) = E(\hat{s}^n) - s$ [68]

²Soit $A = [a, b]$ un intervalle de Ω et soit s une fonction de $L(\Omega)$. On dit que la fonction s est absolument continue sur A si, pour tout $\varrho > 0$, il existe un $\iota > 0$ tel que, pour toute suite (a_n, b_n) $n \in \mathbb{N}$ de sous-intervalle de A d'intérieurs disjoints : $\sum_{n \geq 0} (b_n - a_n) < \iota \Rightarrow \sum_{n \geq 0} (|s(b_n) - s(a_n)|) < \varrho$

³Une fonction $s \in L^1(\Omega)$ est dite carrée intégrable si : $\int_{-\infty}^{+\infty} |s(u)|^2 du < \infty$ ([62] page 79).

Cette variance tend vers zéro quand le produit nh tend vers l'infini quand le nombre d'observation n tend vers l'infini.

Afin d'apprécier la qualité de l'estimateur, il est usuel d'évaluer la distance entre l'estimateur et la densité à estimer. La distance la plus couramment utilisée est celle définie par la moyenne du carré de la valeur absolue de leur différence. Elle porte le nom d'erreur quadratique moyenne (MSE)⁴.

La convergence en moyenne quadratique de l'estimateur \hat{f}_h^n (I.6) a été établie par Lecoutre [28].

Nous avons, d'après [29] pour tout $x \in [a_k, a_k + 1]$:

$$\begin{aligned} MSE(\hat{f}_h^n(x)) &= Var \hat{f}_h^n(x) + Biases^2(\hat{f}_h^n(x)) \\ &= \frac{f(x)}{nh} + \frac{\hat{f}(x)^2}{4} (h - 2(x - a_k))^2 + O(n^{-1}) + O(h^3). \end{aligned} \quad (I.9)$$

Cette erreur quadratique moyenne tend vers zéro quand h tend vers zéro et nh tend vers l'infini quand n tend vers l'infini. Ce critère d'erreur quadratique moyenne est un critère local. On lui préfère généralement un critère plus global obtenu en l'intégrant sur tout le domaine $\Omega = [e_{min}, e_{max}]$. Ce critère porte le nom d'erreur quadratique moyenne intégrée (MISE)⁵.

La convergence en moyenne quadratique intégrée de l'estimateur \hat{f}_h^n défini par (I.6) a été établie par Lecoutre [28]. Nous avons, d'après [29] :

$$MISE = \frac{1}{nh} + \frac{h^2 R(f')}{12} + O(n^{-1}) + O(h^3) \quad (I.10)$$

Où $R(s) = \int_{\Omega} (s(x))^2 dx$, fonction sde carré intégrable.

Cette erreur quadratique moyenne intégrée tend vers zéro quand h tend vers zéro et nh tend vers l'infini quand n tend vers l'infini.

L'utilisation du critère MISE permet de définir une granularité optimale de l'histogramme notée h^* . Cette valeur optimale est la valeur qui minimise ce critère pour un nombre d'observations et une loi donnée. Cette valeur optimale est de la forme :

$$h^* = \left[\frac{6}{R(f')} \right]^{\frac{1}{3}} n^{-\frac{1}{3}} \quad (I.11)$$

⁴En anglais Mean Squared Error.

⁵En anglais Mean Integrated Squared Error

Notons que cette valeur est inutilisable en pratique car elle fait intervenir une connaissance à priori de densité inconnue f via l'intégrale du carré de sa dérivée ($R(f')$). Puisque, justement, cette technique est dédiée à l'estimation de densités dont la loi est inconnue, la valeur de h^* est généralement inconnue.

En remplaçant, la valeur de h^* (I.11) dans l'expression (I.10), on obtient la valeur optimale de l'erreur quadratique moyenne intégrée asymptotique (AMISE)⁶ notée AMISE* :

$$AMISE^* = \left[\frac{9R(f')}{16} \right]^{\frac{1}{3}} n^{-\frac{2}{3}} \quad (I.12)$$

Geffroy a montré dans [27] la convergence uniforme⁷ et presque complète⁸ de l'estimateur \hat{f}_h^n (I.6). Tandis que la convergence en probabilité⁹ et presque complète en norme $L^1(\Omega)$ a été établie par S.Abou-Jaoudé [52].

2-5-Utilisation de l'estimateur de densité par histogrammes

Utiliser un histogramme pour estimer la densité présente l'intérêt de la simplicité.

Cette simplicité a cependant une contrepartie désavantageuse. En effet, comme expliqué dans [20], la représentation par histogramme de la densité sous-jacente à un ensemble fini d'observations n'est robuste au choix ni de la fenêtre ni de l'intervalle de référence de la partition sur laquelle l'histogramme est bâti.

Nous illustrons ce manque de robustesse sur un exemple de 300 observations tirées aléatoirement d'une distribution normale $N(0, 1)$. Sur la Figure 2.3, nous avons choisi deux valeurs de largeur de la fenêtre : $h = 2.5$ (Figure 2.3 (a)) et $h = 0.1$ (Figure 2.3 (b))

On peut, par exemple, s'intéresser à la localisation du mode principal (c'est-à-dire la valeur la plus fréquemment prise par ces observations) sur l'histogramme. Pour $h = 2.5$, la localisation du mode est claire dans l'intervalle $[0, 2]$. Tandis que pour $h = 0.1$, la localisation de ce mode est dans l'intervalle $[-0.14, -0.06]$.

⁶En anglais Asymptotic Mean Integrated Squared Error.

⁷ Un estimateur \hat{s}_n converge uniformément vers $s \in L(\Omega)$ si $\lim_{n \rightarrow \infty} \{ \sup_{u \in \Omega} |\hat{s}_n(u) - s(u)| \} = 0$.

⁸ Un estimateur \hat{s}_n converge presque complète vers une fonction $s \in L(\Omega)$ si pour tout $u \in \Omega$ et tout $\varphi > 0$, on a $\sum_n P(|\hat{s}_n(u) - s(u)| > \varphi) < \infty$.

⁹ Un estimateur \hat{s}_n converge en probabilité vers une fonction $s \in L(\Omega)$ si pour tout $u \in \Omega$ et tout $\varphi > 0$, on a $\lim_{n \rightarrow +\infty} P(|\hat{s}_n(u) - s(u)| \geq \varphi) = 0$.

Figure 2.3 : Histogramme de 300 observations construit avec une granularité $h = 2.5$ (a) et $h = 0.1$ (b).

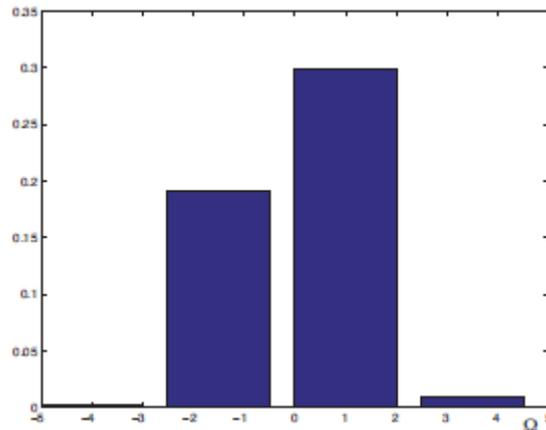
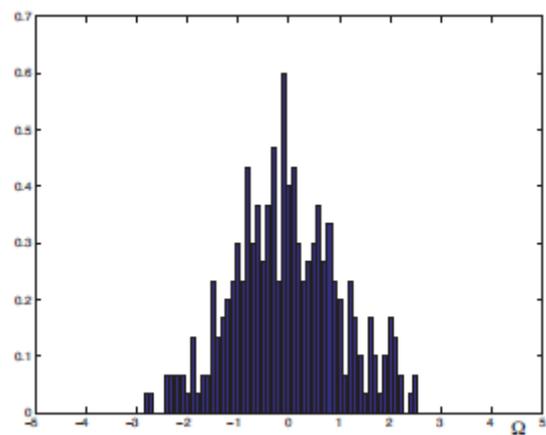


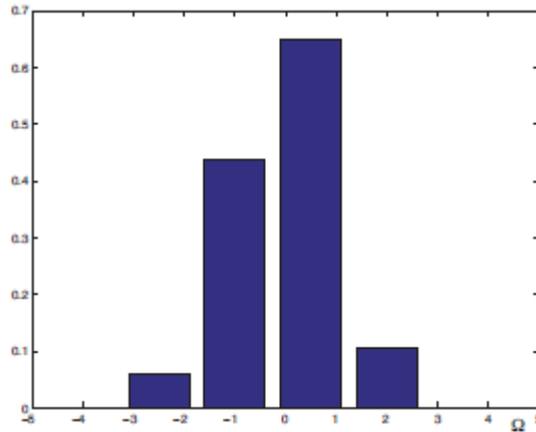
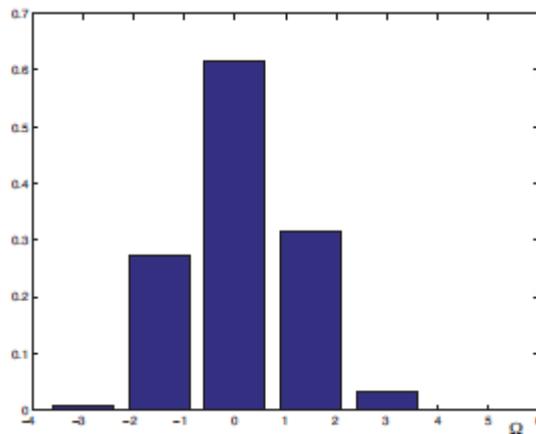
Figure 2.3.a $h = 2.5$



La Figure 2.3.b $h = 0.1$

Sur la Figure 2.4, nous avons choisi deux positions différentes de l'intervalle de référence Ω : $\Omega = [-5, 5]$ (Figure I.4 (a)) et $\Omega = [-4, 6]$ (Figure 2.4 (b)). La Figure 2.4 montre l'influence du positionnement de Ω , sur la représentation d'observations. Sur la Figure 2.4 (b), la distribution semble symétrique alors qu'elle semble ne pas l'être sur la Figure 2.4 (a).

Figure 2.4 : Histogramme de 300 observations construit sur la portion $\Omega = [-5, 5]$ (a) et $\Omega = [-4, 6]$ (b) .

Figure 2.4 (a). $\Omega = [-5, 5]$ Figure 2.4 (b). $\Omega = [-4, 6]$

Runkler [58], Van Den Berg [30] et Strauss [43] ont tenté de réduire l'influence du partitionnement arbitraire de l'intervalle de référence Ω en remplaçant le partitionnement binaire par un partitionnement flou de Ω . L'estimateur de densité par des histogrammes basé sur un partitionnement flou ainsi que ses propriétés statistiques ont été établies par L. Waltman [34] et Loquin et Strauss [31].

Du point de vue de l'estimation non-paramétrique de densité, cet estimateur est une fonction étagée, et donc discontinue. À cause de cette discontinuité, l'histogramme ne peut pas être adapté à la situation où nous disposons d'une information à priori sur la fenêtre de la densité à estimer. Par exemple, si l'on sait par avance que la densité de probabilité est deux fois continûment différentiable, l'estimation naturelle de cette densité doit être deux fois continûment différentiable. Les histogrammes, qui produisent des estimations discontinues, ne peuvent donc pas répondre à ce type de problématique.

3-Estimateur simple de la densité

L'estimateur simple de densité, appelé aussi la méthode d'estimation par les histogrammes mobiles, a été proposée par Fix et Hodges [17]. L'estimation de la densité en un point $x \in \Omega$, par cette méthode, consiste à construire autour de x un intervalle (ou fenêtre) de largeur h $\left(\left[x - \frac{h}{2}, x + \frac{h}{2}\right]\right)$ et à compter le nombre d'observations dans cet intervalle.

Partons du lien existant entre la densité de probabilité f et la fonction de répartition F :

$$\forall x \in \Omega, \quad F(x) = \int_{-\infty}^x f(u) du \quad (\text{I.13})$$

On peut écrire

$$\begin{aligned} f(x) &= \lim_{h \rightarrow 0} \frac{p\left(x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2}\right)}{h} \\ &= \lim_{h \rightarrow 0} \frac{F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right)}{h} \end{aligned}$$

L'estimateur simple de f , noté \hat{f}_h^n peut alors être défini, pour tout $x \in \Omega$, par :

$$\begin{aligned} \hat{f}_h^n(x) &= \frac{1}{h} \frac{|\{i : x - \frac{h}{2} \leq x_i \leq x + \frac{h}{2}\}|}{n} \\ &= \frac{1}{nh} \sum_{i=1}^n 1_{\left[x - \frac{h}{2}, x + \frac{h}{2}\right]}(x_i) \\ &= \frac{1}{nh} \sum_{i=1}^n 1_{\left[-\frac{1}{2}, \frac{1}{2}\right]}\left(\frac{x - x_i}{h}\right) \end{aligned} \quad (\text{I.14})$$

Où $|\cdot|$ est le cardinal d'un ensemble. Cet estimateur peut aussi s'écrire

$$\hat{f}_h^n(x) = \frac{1}{nh} \sum_{i=1}^n w\left(\frac{x - x_i}{h}\right) \quad (\text{I.15})$$

Où w est une fonction de poids définie par :

$$w(x) = \begin{cases} 1 & \text{si } x \in \left[-\frac{1}{2}, \frac{1}{2}\right] \\ 0 & \text{sinon} \end{cases} \quad (\text{I.16})$$

Cette fonction de poids n'est autre que la densité de probabilité uniforme sur l'intervalle $\left[-\frac{1}{2}, \frac{1}{2}\right]$

Le biais de l'estimateur $\hat{f}_h^n(x)$ (I.15) est donné [60], pour tout $x \in \Omega$, par :

$$\begin{aligned} \text{Biais}(\hat{f}_h^n(x)) &= E(\hat{f}_h^n(x)) - f(x) \\ &= \frac{1}{h} \left(F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right) \right) - f(x) \end{aligned} \quad (\text{I.17})$$

L'expression (I.17) montre que ce biais converge vers zéro quand la largeur de fenêtre h tend vers zéro.

La variance de cet estimateur est donnée [60], pour tout $x \in \Omega$, par :

$$\text{Var}(\hat{f}_h^n(x)) = \frac{1}{2nh^2} \left(F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right) \right) \left(1 - F\left(x + \frac{h}{2}\right) + F\left(x - \frac{h}{2}\right) \right) \quad (\text{I.18})$$

Cette variance tend vers zéro lorsque nh tend vers l'infini quand le nombre d'observations n tend vers l'infini. Ce qui donne [60] la convergence en moyenne quadratique :

$$\text{si } h \rightarrow 0 \text{ et } nh \rightarrow \infty \text{ quand } n \rightarrow \infty, \text{MSE}(\hat{f}_h^n(x)) = E(\hat{f}_h^n(x) - f(x))^2 \rightarrow 0 \quad (\text{I.19})$$

L'inconvénient majeur de l'estimateur simple défini par l'expression (I.15) est comme l'estimation par histogramme de fournir une estimation discontinue. Elle est cependant discontinue uniquement aux points $\left\{ \left(x_i - \frac{h}{2}, x_i + \frac{h}{2}\right) \right\}_{i \in \{1, \dots, n\}}$ contrairement à l'estimateur par histogramme qui est discontinue aux frontières de chaque classe et a une dérivée nulle partout ailleurs. Cette discontinuité est une conséquence de la discontinuité de la fonction de poids $w(x)$ (I.16).

Comme le souligne Silverman [4], cette méthode d'estimation est finalement comparable à celle de l'histogramme. En effet, considérons un histogramme construit à partir d'intervalles de largeur h . Si on suppose qu'aucune des observations ne soit aux bornes de ces intervalles et si $x \in \Omega$ est au centre de l'un de ces intervalles alors l'estimateur simple, au point x , $\hat{f}_h^n(x)$ (I.15)

n'est autre que l'estimateur par histogramme $\hat{f}_h^n(x)$ (I.6). L'estimateur simple peut être vu comme une manière de construire un histogramme où chaque point est le centre d'un intervalle de l'ensemble $\{A_k, k = 1, \dots, p\}$ de l'histogramme.

4-Méthode du noyau

L'estimation par 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 constitue 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) et une somme de fonctions continues reste continue.

4-1-Définitions 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

$$\hat{f}_h^n(x) = \frac{1}{nh} \sum_{i=1}^n 1(|X_i - x| \leq h/2) = \frac{1}{nh} \sum_{i=1}^n 1\left(\frac{|X_i - x|}{h} \leq \frac{1}{2}\right) \quad (\text{I.20})$$

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 certes utile, car elle conduit vers un estimateur qui est constant par morceaux comme les histogrammes, mais a 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 $1(|z| \leq \frac{1}{2})$ par une fonction K quelconque, on obtient l'estimateur :

$$\hat{f}_h^K(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (\text{I.21})$$

Qui est continu et même ℓ –fois continûment différentiable du moment où la fonction K l'est. On arrive ainsi à la définition suivante :

4-1-1-Définitions

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

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

On dit alors que K est le noyau de cet estimateur et h est la fenêtre.

Selon cette définition, toute fonction K peut servir comme noyau d'estimation d'une densité.

Les noyaux les plus couramment utilisés en pratique sont :

➤ **Le noyau uniforme (Rosenblatt)**

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

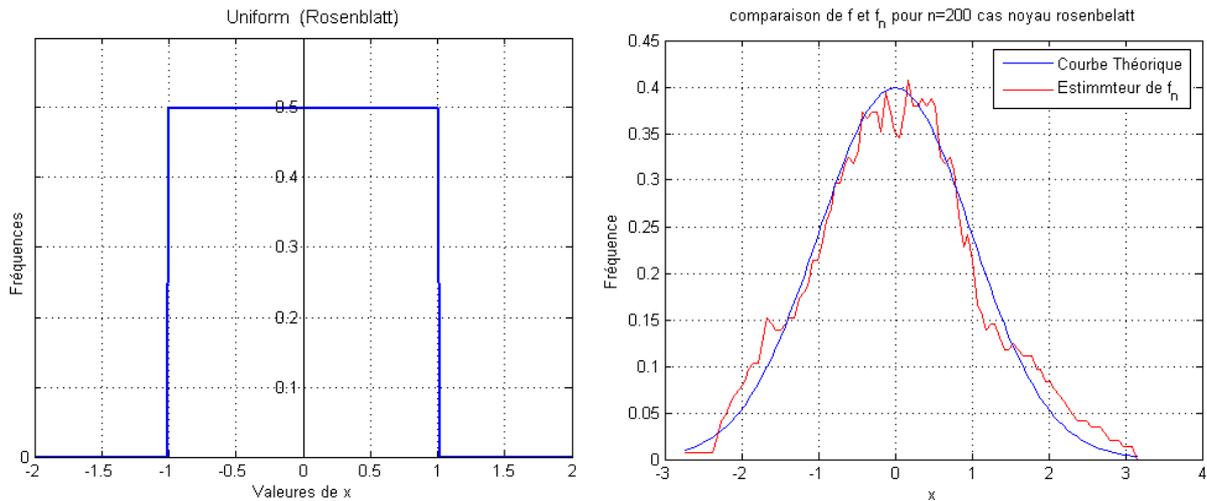


Figure 2.1: Noyau Uniform (Rosenblatt)

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

$$K(x) = (1 - |x|)1_{[-1,1]}(x)$$

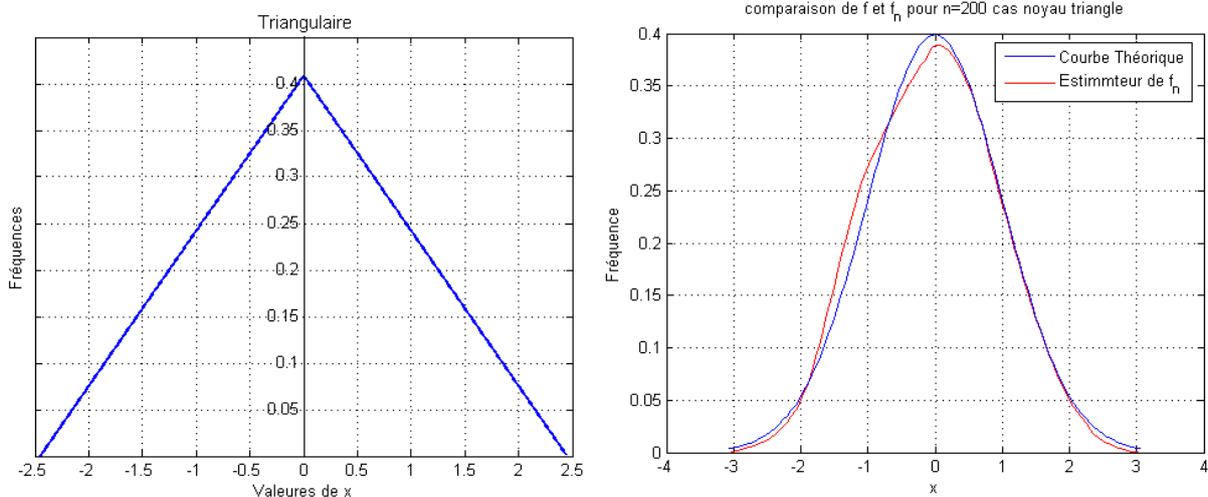


Figure 2.2: Noyau Triangulaire.

- **Le noyau d'Epanechnikov ou parabolique :** Il définit par la forme suivante :

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

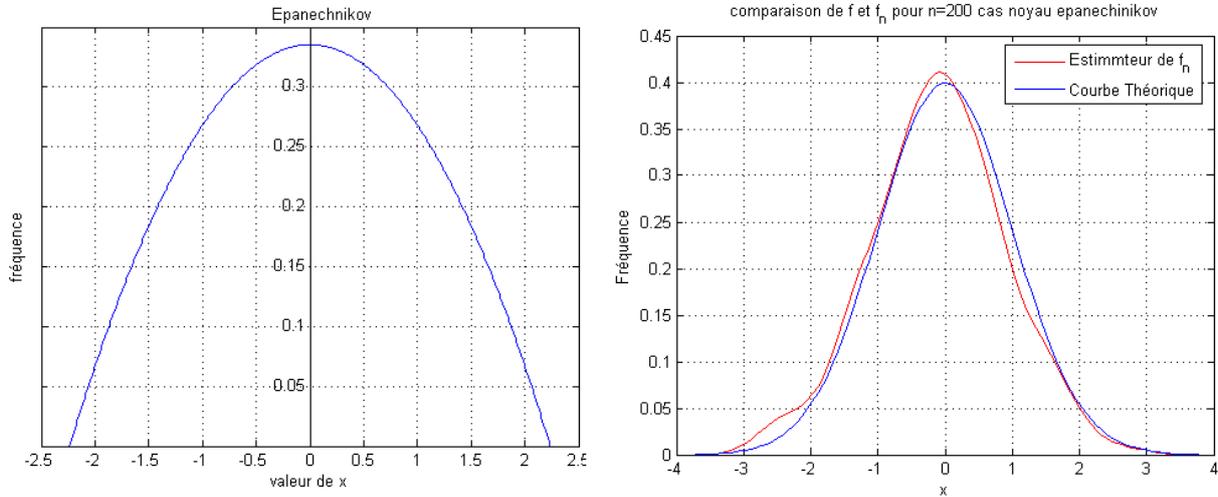


Figure 2.3: Noyau d'Epanechnikov ou parabolique.

- **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 :

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

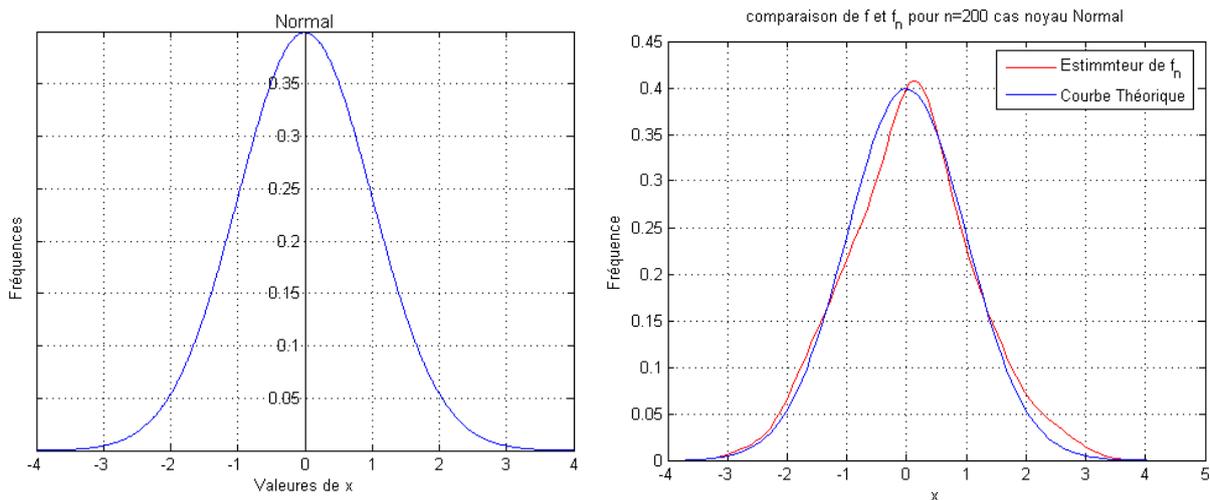


Figure 2.6: Noyau gaussien

Le noyau d'Epanechnikov est apprécié puisqu'il est celui qui a la meilleure efficacité (mais les autres noyaux présentés sont à peine moins efficaces).

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

Démonstration.

L'estimateur à noyau est positif 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,

$$\begin{aligned} \int_{-\infty}^{+\infty} \hat{f}_h^K(x)dx &= \int_{-\infty}^{+\infty} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) dx \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) dx \quad (u = (X_i - x) / h) \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} \sum_{i=1}^n K(u)h du = 1. \end{aligned} \tag{I.22}$$

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 quantitatifs caractérisant le gain obtenu par l'utilisation de \hat{f}_h^K par rapport à \hat{f}_h^n .

4-1-2-Propriétés statistiques de l'estimateur à noyau

Nous présentons, dans cette partie, les propriétés statistiques de l'estimateur de densité de Parzen-Rosenblatt défini par l'expression (I.22) (pour une étude plus détaillée, voir par exemple Bosq et Lecoutre [13], Tsybakov [2] ou Silverman [4]).

Nous considérons ici les noyaux sommatifs K centrés de variance finie de $K(\Omega)$ c'est-à-dire vérifiant :

$$\int_{\Omega} uk(u)du = 0 \quad \text{et} \quad \int_{\Omega} u^2k(u)du = \sigma_k^2 > 0 \tag{I.23}$$

σ_k^2 est appelée la variance du noyau K [1].

Le biais de l'estimateur \hat{f}_{kh}^n (I.22) est donné [29], pour tout $x \in \Omega$, par :

$$\text{biais} \left(\hat{f}_{kh}^n(x) \right) = \frac{h^2 \sigma_k^2 f''(x)}{2} + \mathcal{O}(h^4) \quad (\text{I.24})$$

Où f'' est la dérivée seconde de la fonction de densité f . f'' doit être une fonction absolument continue.

L'expression (I.24) montre que ce biais converge vers zéro quand la largeur de bande h tend vers zéro. Il découle de l'expression (I.24) que le biais de \hat{f}_{kh}^n ne dépend pas directement du nombre d'observations mais seulement de la fonction de noyau K et donc l'augmentation du nombre d'observations ne peut pas, à elle seule, diminuer le biais.

La variance de cet estimateur est donnée [29], pour tout $x \in \Omega$ par :

$$\text{Var} \left(\hat{f}_{kh}^n(x) \right) = \frac{f(x)R(k)}{nh} + \mathcal{O}(h^{-1}) \quad (\text{I.25})$$

$$\text{Où } R(k) = \int_{\Omega} (K(u))^2 du$$

Cette variance tend vers zéro quand nh tend vers l'infini quand n tend vers l'infini.

La convergence en moyenne quadratique de l'estimateur \hat{f}_{kh}^n (I.22) a été établie par Bosq et Lecoutre [13]. Nous avons, d'après [29], pour tout $x \in \Omega$:

$$\text{MSE} \left(\hat{f}_{kh}^n(x) \right) = \frac{f(x)R(k)}{nh} + \frac{h^4 \sigma_k^4 (f''(x))}{4} + \mathcal{O}(h^6) + \mathcal{O}(h^{-1}) \quad (\text{I.26})$$

En intégrant l'expression (I.26) sur tout le domaine, on obtient [29] l'erreur quadratique moyenne intégrée de \hat{f}_{kh}^n :

$$\text{MSIE} = \frac{R(k)}{nh} + \frac{h^4 \sigma_k^4 R(f'')}{4} + \mathcal{O}(h^6) + \mathcal{O}(h^{-1}) \quad (\text{I.27})$$

$$\text{Où } R(f'') = \int_{\Omega} (f''(u))^2 du$$

La largeur de bande optimale de \hat{f}_{kh}^n notée h^* , est donnée [29] par :

$$h^* = \left(\frac{R(k)}{\sigma_k^4 R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} \quad (\text{I.28})$$

Notons que h^* est une quantité déterministe qui dépend du nombre d'observations n .

Elle dépend aussi de la densité inconnue f via l'intégrale du carré de sa dérivée seconde ($R(f'')$) et ne peut pas donc être utilisée telle qu'elle dans les calculs. Une façon classique de corriger ce défaut consiste à remplacer la quantité $R(f'')$, dans l'expression (I.30), par un estimateur approprié [1]. Cette correction a fait l'objet de beaucoup de travaux (voir, par exemple, Hall et Marron [45], Jones et Sheather [38]).

En remplaçant la valeur de h^* (I.30) dans l'expression (I.27), on obtient [29] la valeur du $AMISE^*$:

$$AMISE^* = \frac{5}{4} (\sigma_k R(k))^{\frac{4}{5}} (R(f''))^{\frac{1}{5}} n^{-\frac{4}{5}} \quad (I.29)$$

Epanechnikov a proposé dans [59] un noyau optimal de l'estimateur $\hat{f}_{kh}^n(x)$ (I.22) qui minimise le critère $AMISE^*$ (I.31) par rapport au noyau K . La solution k_{opt} obtenue est donnée par :

$$K_{pot}(u) = \begin{cases} \frac{3}{4} (1 - u^2) & \text{si } |u| \leq 1 \\ 0 & \text{sinon} \end{cases} \quad (I.30)$$

Ce noyau est appelé le noyau d'Epanechnikov. Nous notons, dans la suite, e et non k_{opt} le noyau d'Epanechnikov défini par (I.32).

Le couple largeur de bande h et forme de noyau K joue un rôle important dans la méthode d'estimation à noyau. Nous discutons, dans la section suivante, de l'influence de ce choix sur l'estimation de la densité f sous-jacente à un ensemble fini d'observations.

La convergence ponctuelle presque sûre de l'estimateur \hat{f}_{kh}^n (I.22) a été établie par Devroyet Wagner [33]. Bertrand Retali a montré dans [37] la convergence uniforme presque sûre de cet estimateur, tandis que la convergence en $L^1(\Omega)$ a été établie par Gilck [39] et Devroy [32].

Un résultat important issu de [4] montre que si f'' la dérivée seconde de f , existe et que la largeur de bande $h = qn^{-\frac{1}{5}}$ (avec q une constante), alors, pour tout $x \in \Omega$, l'estimateur à noyau $\hat{f}_{kh}^n(x)$ est asymptotiquement normal (c'est-à-dire converge en loi vers la loi normale)

$$n^{\frac{2}{5}} \{ \hat{f}_{kh}^n(x) - f(x) \} \xrightarrow[n \rightarrow \infty]{L} N \left(\frac{q^2}{2} f''(x) \sigma_k^2, \frac{1}{q} f(x) R(x) \right) \quad (I.31)$$

Où σ_k^2 est la variance de K définie dans (I.23).

L'expression (I.33) est équivalente, par une transformation linéaire de la loi¹⁰ normale,

¹⁰ Une suite des variables aléatoires $(X_n)_{n \in \mathbb{N}}$ converge en loi vers la variable aléatoire X

à

$$\{\hat{f}_{kh}^n(x) - f(x)\} \xrightarrow[n \rightarrow \infty]{L} N\left(\underbrace{\frac{h^2}{2} f''(x) \sigma_k^2}_{\text{biais de } \hat{f}_{kh}^n}, \underbrace{\frac{1}{nh} f(x) R(x)}_{\text{variance de } \hat{f}_{kh}^n}\right) \quad (\text{I.32})$$

Sous l'hypothèse que le terme de biais (I.24) de \hat{f}_{kh}^n est négligeable par rapport au terme de la variance (I.25) de \hat{f}_{kh}^n , l'intervalle de confiance au seuil α sur la valeur de la densité $f(x)$,

Pour tout $x \in \Omega$, est donné [60] par :

$$\left[\hat{f}_{kh}^n(x) - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{f}_{kh}^n(x) R(k)}{nh}}, \hat{f}_{kh}^n(x) + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{f}_{kh}^n(x) R(x)}{nh}} \right] \quad (\text{I.33})$$

Où $Z_{1-\frac{\alpha}{2}}$ est le $(1 - \frac{\alpha}{2})$ quantile¹¹ de la loi normale centrée réduite.

Notons que cette méthode donne un intervalle de confiance à $(1 - \frac{\alpha}{2})\%$ des valeurs de la densité f et non pas pour toutes les valeurs de f . Une autre technique de calcul proposée par Bickel et Rosenblatt [44] sous le nom de bandes de confiance¹² permet de donner un autre intervalle d'estimation de la densité f . Sous l'hypothèse que f est une densité sur $[0, 1]$ et que $h = n^{-l}, l \in (\frac{1}{5}, \frac{1}{2})$ Bickel et Rosenblatt ont démontré que, pour tout $x \in [0, 1]$, nous avons le résultat asymptotique suivant :

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\hat{f}_{kh}^n(x) - Z_{n,\alpha} \sqrt{\frac{\hat{f}_{kh}^n(x) R(k)}{nh}} \leq f(x) \leq \hat{f}_{kh}^n(x) + Z_{n,\alpha} \sqrt{\frac{\hat{f}_{kh}^n(x) R(x)}{nh}}\right) \\ = 1 - \alpha \end{aligned} \quad (\text{I.36})$$

$$Z_{n,\alpha} = \left(-\frac{\log\left\{-\frac{1}{2} \log(1 - \alpha)\right\}}{(2l \log(n))^{\frac{1}{2}}} + d_n \right)^{\frac{1}{2}} \quad (\text{I.35})$$

$$d_n = (2l \log(n))^{\frac{1}{2}} + (2l \log(n))^{-\frac{1}{2}} \log\left(\frac{1}{2\pi} \frac{R(k')}{R(k)}\right)^{\frac{1}{2}} \quad (\text{I.36})$$

¹¹ Si X est une variable aléatoire suivant la loi normale centrée réduite, alors la valeur du quantile d'ordre $1 - \frac{\alpha}{2}$

¹² En anglais confidence bands.

Où k' est la dérivée de noyau k et $R(k') = \int_{\Omega} (k'(u))^2 du$

Comme le fait remarquer Härdle [60] les intervalles d'estimation de f donnés par les bandes de confiance (I.36) sont plus étroits que ceux donnés par les intervalles de confiance (I.35).

4-2-Choix de paramètre de lissage h

La mise en œuvre pratique de l'estimateur à noyau de la densité de probabilité nécessite la spécification de paramètre de lissage h . Ce choix est très important, il existe de nombreuses situations où il est acceptable de choisir le paramètre de lissage subjectivement par œil. Il s'agirait de regarder les estimations de densité serval sur une plage de paramètre de lissage et en sélectionnant la densité qui est le "plus agréable" dans un certain sens. Une telle stratégie est de commencer par un large paramètre de lissage et de diminuer la quantité de lissage des fluctuations Unutil qui sont plus " aléatoire " que " structurelle " commencent à apparaître. Cette approche est plus viable lorsque l'utilisateur a des raisons de croire qu'il existe une certaine structure dans les données, une telle connaissance de la position des modes. Cependant, il y'a aussi de nombreux cas où il est très bénéficial d'avoir le paramètre de lissage sélectionnée automatiquement à partir des données. Une des raisons est qu'il peut prendre beaucoup de temps pour choisir le paramètre de lissage à l'œil, il y'a beaucoup d'estimations de densité nécessaires pour un problème donné. Une autre est que, dans de nombreux cas, l'utilisateur n'a pas de connaissance préalable de la structure des données et n'aurait aucun sentiment pour lequel le paramètre de lissage donne une estimation plus proche de la densité réelle. Lorsque estimateurs à noyau sont utilisés comme composants de procédures statistiques plus grandes, la sélection automatique du paramètre de lissage est généralement nécessaire [62].

Le problème de sélection du paramètre de lissage [61] est présent dans tous les types d'estimation du noyau, Actuellement disponibles sélecteurs de bande passante peuvent être divisés en deux classes, La première classe se composent de formules facilement calculables simples, mais sans aucune garantie sur mathématiques d'être à proximité du paramètre de lissage optimale. Nous appellerons ces sélecteurs de bande passante rapide et simple, Une méthode qui utilise les données X_1, \dots, X_n pour produire une largeur de bande h est appelé un sélecteur du paramètre de lissage, bien que l'estimation de densité du noyau offre un cadre pratique pour le développement de nombreuses idées clés.

Le deuxième type de sélecteur paramètre de lissage sera étiqueté comme salut technologie car de telles procédures de sélection sont basés sur des arguments mathématiques plus complexes et nécessitent beaucoup plus d'effort de calcul, mais visent à donner une bonne réponse pour

les classes très générales de fonctions sous-jacentes. Elle est motivée par visant à minimiser $MISE (f_h(x))$.

4-2-1-Méthode de Rule of thumb

Une approche très facile et naturel consiste à utiliser une famille \hat{A} au niveau de la distribution d'attribuer une valeur à $\int (f''(x))^2 dx$ terme de f'' dans l'expression h_{AMIS} pour la fenêtre idéale par exemple Si x_1, x_2, \dots, x_n . Est un échantillon i.i.d. d'une variable aléatoire, alors l'estimateur non-paramétrique par la méthode du noyau de la densité est :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (I.37)$$

Où K est un noyau (kernel en anglais) et h un paramètre nommé fenêtre, qui régit le degré de lissage de l'estimation. Bien souvent, K est choisi comme la densité d'une fonction gaussienne standard (espérance nulle et variance unitaire) :

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (I.38)$$

Si on choisit f comme étant la distribution de loi normale de moyenne 0 et de variance σ^2 ($N(0; \sigma^2)$) On aura alors:

$$\int (f''(x))^2 dx = \sigma^{-5} \int (\phi''(x))^2 dx = \frac{3}{8} \prod^{-\frac{1}{2}} \sigma^{-5} \simeq 0.212\sigma^{-5} \quad (I.39)$$

De plus, si K est un noyau gaussien, alors la valeur pour le h_{opt} notée dans ce cas par h_{rot} est obtenue en substituant ce noyau et la valeur $R (f'')$ obtenue

$$\begin{aligned} h_{opt} &= (4\pi)^{-\frac{1}{10}} \left[\frac{3}{8} \pi^{-\frac{1}{2}} \sigma \right] n^{-\frac{1}{5}} \\ &= \left(\frac{4}{3} \right)^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \\ &= 1.06\sigma n^{-\frac{1}{5}}. \end{aligned} \quad (I.40)$$

Remarques

- Évidemment, h_{rot} fonctionne bien si la densité réelle est gaussienne
- Forcément bien si la densité réelle n'est pas gaussienne, il ne fonctionne pas.

4-3-Cross validation

4-3-1-Validation croisée non biaisée

Cette méthode appelée Validation Croisée non Biaisée a été proposée par Rudemo [51] en 1982 et Bowman [8] en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$\begin{aligned}
 UCV(h) &= \int_{\mathbb{R}} |f_h(x) - f(x)|^2 dx - \int_{\mathbb{R}} f^2(x) dx \\
 &= \int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx.
 \end{aligned}
 \tag{I.41}$$

Puisque $\int_{\mathbb{R}} f^2(x) dx$ ne dépend pas du paramètre de lissage h . On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$\int_{\mathbb{R}} f_h^2(x) dx - 2 \int_{\mathbb{R}} f_h(x) f(x) dx.
 \tag{I.42}$$

On veut premièrement trouver un estimateur de $\int_{\mathbb{R}} f_h(x) f(x) dx$. Remarquons que

$$\int_{\mathbb{R}} f_h(x) f(x) dx = \mathbb{E}(f_h(x))$$

L'estimateur empirique de $\int_{\mathbb{R}} f_h(x) f(x) dx$, est alors :

$$\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i).$$

Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{R}} f_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n f_{h,i}(x_i)$$

Où $f_{h,i}(x_i) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i}^n K\left(\frac{x_i - x_j}{h}\right)$ est l'estimateur de la densité construit à partir de l'ensemble de points sauf le point x_i .

Montrons maintenant que $UCV(h)$ est un estimateur sans biais de $MISE(h) - R(f)$.

On a :

$$MISE(h) - R(f) = \mathbb{E} \left[\int f_h^2(x) dx - 2 \int f_h(x) f(x) dx \right] \quad (I.43)$$

Il suffit de montrer que $\int f_h^2 dx$ et $\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$ sont des estimateurs sans biais de $\mathbb{E}[\int f_h^2(x) dx]$ et $\mathbb{E}[\int f_h(x) f(x) dx]$ respectivement. Or $\mathbb{E}[\int f_h^2(x) dx]$ admet l'estimateur sans biais trivial $\int f_h^2 dx$. Il reste donc à montrer que $\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i)$ est un estimateur sans biais de $\mathbb{E}[\int f_h(x) f(x) dx]$

On a d'une part,

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i) \right] &= \mathbb{E} \left[\frac{1}{n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n K \left(\frac{x_i - x_j}{h} \right) \right] \\ &= \mathbb{E} \left[\frac{1}{n(n-1)h} \sum_{j \neq 1}^n \int K \left(\frac{x_i - x_j}{h} \right) f(x) dx \right] \\ &= \frac{1}{n} \int f(z) \int K \left(\frac{x - z}{h} \right) f(z) dx dz. \end{aligned} \quad (I.44)$$

D'autre part,

$$\begin{aligned} \mathbb{E} \left[\int f_h(x) f(x) \right] &= \mathbb{E} \left[\frac{1}{nh} \sum_{i=1}^n \int K \left(\frac{x_i - x_j}{h} \right) f(x) dx \right] \\ &= \frac{1}{n} \int f(z) \int K \left(\frac{x - z}{h} \right) f(x) dx dz. \end{aligned} \quad (I.45)$$

Ce qui implique que :

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i) \right] = \mathbb{E} \left[\int f_h(x) f(x) \right]$$

Finalement, un estimateur sans biais de $MISE(h) - R(f)$ est donnée donc par $UCV(h)$.

En utilisant l'équation précédente, le critère $UCV(h)$ devient :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{i \neq j, j=1}^n \left[\int \frac{1}{n^2 h^2} K\left(\frac{x-x_i}{h}\right) K\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i-x_j}{h}\right) \right]. \quad (I.46)$$

Nous noterons h_{UCV} l'estimateur de h qui minimise $UCV(h)$

La popularité de cette méthode est due à la motivation intuitive et au fait que cet estimateur est asymptotiquement optimal sous de faibles conditions. L'optimalité asymptotique de la validation croisée non biaisée a été obtenue par Stone [56].

Soit x_1, x_2, \dots, x_n un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f . Utilisant le noyau gaussien on obtient :

$$UCV(h) = \frac{1}{2n^2 h \sqrt{\pi}} \left(n + 2 \sum_{i=1}^n \sum_{i \neq j, j=1}^n \exp\left(-\left(\frac{x_i - x_j}{2h}\right)^2\right) \right) - \frac{2}{\sqrt{2\pi} n(n-1)h} \sum_{i=1}^n \sum_{i \neq j, j=1}^n \exp\left(-\frac{(x_i - x_j)^2}{2h^2}\right) \quad (I.47)$$

Cette méthode présente deux problèmes majeurs (ou points faibles) : d'une part son manque de robustesse par rapport aux changements de taille de l'échantillon c'est-à-dire le résultat de simulation peut se révéler extrêmement variable d'un échantillon à l'autre, d'autre part, la fonctionnelle à minimiser a souvent tendance à présenter plusieurs minimums locaux [22]. Pour d'autres études, voir Hall [23], Burman [9], Scott et Terrell [54].

4-3-2-Validation croisée biaisée

Un critère de validation croisée biaisée, a été introduit par Scott et Terrell [54] en 1987 pour remédier aux problèmes de validation croisée non biaisée. Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance.

L'Erreur Quadratique Intégrée Moyenne Asymptotique s'écrit sous la forme :

$$AMISE = \frac{h^4}{4} \sigma_K^4 R(f'') + \frac{R(K)}{nh}. \quad (I.48)$$

Le paramètre de lissage basé sur la méthode de validation croisée biaisée est la valeur h qui minimise un estimateur du AMISE. On peut estimer le AMISE si l'on estime $R(f'')$. Un estimateur naturel de ce terme est donné par $R(f''_h)$ où f_h est l'estimateur de la densité qui utilise la méthode du noyau. Scott et Terrell [54]

Supposant que le noyau K satisfait aux conditions suivantes :

$$\int K''(u)du = 0, \mu_1(K'') = \int uK''(u) = 0, \mu_2(K'') = \int u^2K''(u) = 2 \quad (\text{I.49})$$

On obtient le développement asymptotique :

$$\mathbb{E}[R(f_h'')] = R(f'') + \frac{R(K'')}{nh^5} + O(h^2). \quad (\text{I.50})$$

Proposition 1 : (Scott et Trelle) [54]

Soit x_1, x_2, \dots, x_n un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f .

Pour un noyau K on obtient :

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{\mu_2^2(K)}{4n^2} \sum_i \sum_{j, j \neq i} K_h^{(2)} K_h^{(2)}(x_i - x_j). \quad (\text{I.51})$$

Proposition2 : Soit x_1, x_2, \dots, x_n un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f . En choisissant le noyau gaussien on obtient :

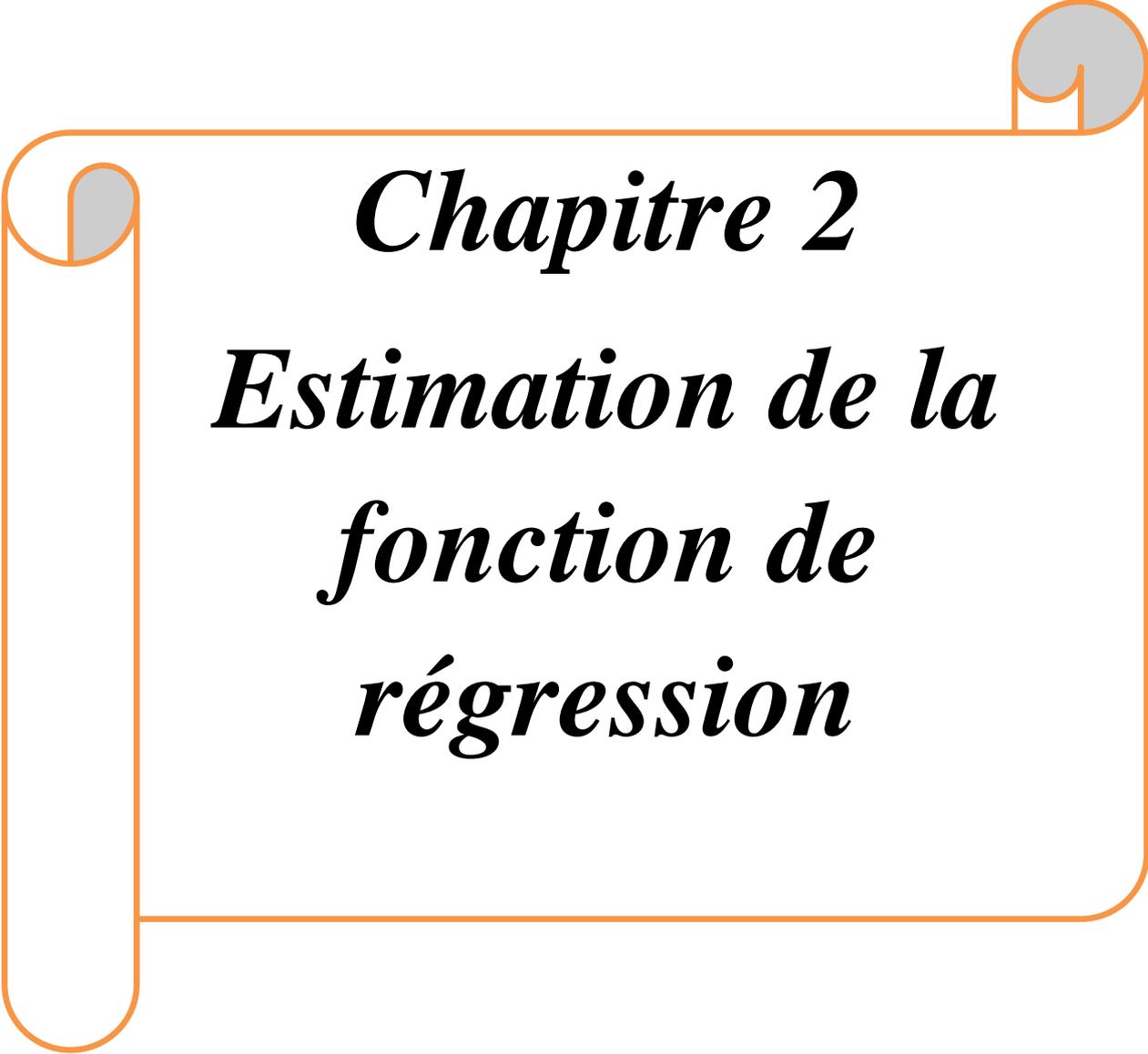
$$\begin{aligned} BCV(h) = & \\ & \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2h\sqrt{\pi}} \sum_{i=1}^n \sum_{j=1, i \neq j}^n \left[\left(\frac{x_i - x_j}{h} \right)^4 - 12 \left(\frac{x_i - x_j}{h} \right)^2 \right. \\ & \left. + 12 \right] \exp \left[- \frac{(x_i - x_j)^2}{4h^2} \right]. \end{aligned} \quad (\text{I.52})$$

Des résultats de simulations ont été obtenus pour la méthode de validation croisée biaisée dans le travail de Par k et Marron [46]. Les auteurs ont constaté que la méthode validation croisée biaisée présente le même point faible que celui de la méthode validation croisée non biaisée. Cette méthode nous donne plusieurs minimums locaux pour la fonctionnelle cible à minimiser. Cependant, d'après plusieurs simulations, les auteurs proposent de choisir la valeur inférieure parmi les minimums locaux.

Conclusion

Dans ce chapitre, Nous avons étudié les méthodes d'estimation non paramétriques de la densité de probabilité (méthode d'estimation par histogramme, simple et la méthode d'estimation par, noyau et aussi estimateur de densité Parzen-Rosenblatt qui peut être vue comme une extension de la méthode d'estimation par histogramme. Nous avons présenté également, les propriétés statistiques de chaque méthode d'estimation.

Nous avons étudié aussi les méthodes de choix de paramètre de lissage (méthode de Rule of thumb et la méthode de validation croisée biaisée et non biaisée).



Chapitre 2
Estimation de la
fonction de
régression

Introduction

La méthode la plus communément utilisée pour étudier la relation entre deux variables est

La régression linéaire simple, qui suppose un modèle de la forme :

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{II.1}$$

Où les erreurs aléatoires ε_i sont non corrélées, de moyennes nulles et de variances σ^2 indépendantes de X . Cette méthode paramétrique 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 lorsque la linéarité de la relation est mise en défaut, on préfère choisir un modèle plus flexible qui reflète mieux la relation entre X et Y . On utilise le modèle de régression non paramétrique

$$Y_i = m(X_i) + \varepsilon_i \tag{II.2}$$

Où ε_i représente la variation de Y_i autour de X_i et obéit aux mêmes hypothèses que dans le cas linéaire et a part certaines conditions de continuités et de lissages il n'y a habituellement aucune contrainte associée à (X) . On cherchera, dans une famille fixée de fonctions quelle est celle pour laquelle les Y sont les plus proches de $m(X)$. Cette proximité se mesure en général par un risque utilisant l'erreur quadratique moyenne (MSE en anglais), et on essaiera alors de déterminer la fonction $m^*(X_i)$ qui rendra cette erreur la plus petite possible, c'est à dire à trouver une fonction $m^*(X_i)$ qui minimise l'erreur quadratique moyenne :

$$\mathbb{E}|m^*(X) - Y|^2 = \min_m \mathbb{E}|m(X) - Y|^2 \tag{II.3}$$

Il est connu que ce minimum est donné par l'Espérance conditionnelle :

$$r(x) = \mathbb{E}(Y|X = x) \tag{II.4}$$

En effet, déterminer $m^*(x)$ revient à calculer $\underset{m}{\operatorname{argmin}} \mathbb{E}((m(X) - Y)^2 | X = x)$. En différenciant l'espérance $\mathbb{E}((m(X) - Y)^2 | X = x)$ par rapport à m , en égalant le résultat à 0, et finalement en isolant m , on obtient :

$$\begin{aligned} \frac{\partial}{\partial m} \mathbb{E}(m(X) - Y)^2 | X = x &= 2\mathbb{E}(m(X) - Y | X = x) \\ &= 2m(X) - \mathbb{E}(Y | X = x) \\ &= 0 \end{aligned} \tag{II.5}$$

Ce qui implique :

$$m^*(x) = \mathbb{E}(Y|X = x) =: r(x). \quad (\text{II.6})$$

Le fait que la dérivée seconde, qui est égale à 2, soit positive nous permet de conclure que c'est bien un minimum.

1-Estimateurs à noyau de la fonction de régression

1-1-L'estimateur de Nadaraya-Watson

Notons que la fonction : $r(x) = E(Y | X = x)$ peut s'écrire sous la forme :

$$r(x) = \frac{\int y f_{X,Y}(x, y) dy}{f(x)} =: \frac{\bar{r}_1(x)}{f(x)} \quad (\text{II.7})$$

Un estimateur à noyau de la densité conjointe \hat{f}_{XY} , peut être obtenu en suivant les mêmes étapes que pour $\hat{f}_{PR}(x)$, et on a alors :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right) \quad (\text{II.8})$$

Et on obtient l'estimateur de Nadaraya- Watson (1964) (NW)

$$\hat{r}_{NW,n}(x) = \frac{\frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} =: \frac{\hat{r}_{1,n}(x)}{\hat{f}_{PR}(x)} \quad (\text{II.9})$$

En effet si on remplace la densité conjointe $f_{X,Y}$ par son estimateur défini dans (II.8) dans le numérateur de (II.7) on a :

$$\begin{aligned} \int y \hat{f}_{X,Y}(x, y) dy &= \int \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) y K\left(\frac{Y_i - y}{h}\right) dy \\ &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int \frac{1}{h} y K\left(\frac{Y_i - y}{h}\right) dy. \end{aligned} \quad (\text{II.10})$$

Si on suppose de plus que le noyau K est symétrique on montre en faisant le changement de variables $u = \frac{Y_i - y}{h}$ que $\int \frac{1}{h} y K\left(\frac{Y_i - y}{h}\right) dy = X_i$, d'où l'expression (II.9). Le noyau K détermine la forme du voisinage autour du point x alors que la fenêtre h contrôle la taille de ce voisinage, c'est à dire le nombre d'observations prises pour effectuer la moyenne locale : il

est alors naturel que la fenêtre h soit prépondérante pour la consistance de l'estimateur $\hat{r}_{NW,n}$. En fait, h détermine le degré de lissage de l'estimateur $\hat{r}_{NW,n}$. Supposons que l'estimateur soit seulement évalué aux points d'observations $\{X_i, i = 1, \dots, n\}$, alors, lorsque K est à support compact, on obtient :

$$\lim_{h \rightarrow 0} \hat{r}_{NW,n}(X_i) = \frac{K(0)Y_i}{K(0)} = Y_i \quad (\text{II.11})$$

Plus précisément on a :

$$\lim_{h \rightarrow 0} \hat{r}_{NW,n}(X_i) = \begin{cases} Y_i & \text{si } x = X_i \quad 1 \leq i \leq n \\ 0 & \text{ailleurs} \end{cases} \quad (\text{II.12})$$

Ce qui veut dire que lorsque h tend vers 0. L'estimateur $\hat{r}_{NW,n}$ a tendance à reproduire les données, la courbe obtenue est proche d'une interpolation des points $\{(X_i, Y_i), i = 1, \dots, n\}$, c'est un phénomène de sous-lissage (la variance de l'estimateur est trop grande). Inversement,

$$\lim_{h \rightarrow +\infty} \hat{r}_{NW,n}(x) = \frac{\sum_{i=1}^n K(0)Y_i}{\sum_{i=1}^n K(0)} = \frac{1}{n} \sum_{i=1}^n Y_i \quad (\text{II.13})$$

On voit alors que l'estimateur $\hat{r}_{NW,n}$ ne dépend plus de x : lorsque h tend vers l'infini, on a un phénomène de sur-lissage (le biais est trop grand). Ces deux comportements nous conduisent à penser que le choix de h est en fait un arbitrage biais-variance, que nous étudierons plus en détails dans le comportement asymptotique de l'estimateur NW.

1-2-L'estimation par polynômes locaux

L'estimation de la fonction de régression par la méthode des polynômes locaux (LP) est une généralisation de l'estimateur de Nadaraya-Watson. L'idée maitresse de cette approche est de considérer le problème de la régression sous l'angle des moindres carrés (M.C.). Nous rappelons que la fonction de régression est elle-même solution d'un problème des M.C. Si on reprend l'estimateur de Nadaraya-Watson, alors on a d'après (II.9) :

$$\hat{r}_{1,n}(x) - \hat{r}_{NW,n}(x) \hat{f}_{PR}(x) = 0. \quad (\text{II.14})$$

L'estimateur de la régression $\hat{r}_{NW,n}$ peut donc être regardé comme la solution du problème des moindres carrés (M.C.) pondérés suivant :

$$\underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n (Y_i - \theta)^2 K\left(\frac{X_i - x}{h}\right). \quad (\text{II.15})$$

C'est à dire que l'estimateur $\hat{r}_{NW,n}$ est obtenu par une approximation des M.C. localement constante. Le principe de l'estimation localement polynômiale consiste en l'ajustement local d'un polynôme de degré p aux données $(X_i, Y_i)_{i=1, \dots, n}$. Pour $p \in \mathbb{N}$, nous cherchons à ajuster le polynôme

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

Aux données $(X_i, Y_i)_{i=1, \dots, n}$ par la méthode des Moindres Carrés. Pour cela, on doit supposer l'existence de la $(p + 1)$ ème dérivée de la fonction de régression $r(\cdot)$ au point x pour pouvoir faire un développement de Taylor et on a alors :

$$\begin{aligned} r(u) &\approx r'(x)(u-x) + \frac{r''(x)}{2}(u-x)^2 + \dots + \frac{r^{(p)}(x)}{p!}(u-x)^p \\ &\approx \sum_{j=0}^p \frac{r^{(j)}(x)}{j!} (u-x)^j =: \sum_{j=0}^p a_j (u-x)^j \end{aligned} \quad (\text{II.16})$$

Pour u appartenant à un voisinage de x .

Pour ajuster localement le polynôme (II.16) aux données $(X_i, Y_i)_{i=1, \dots, n}$ par la méthode des M.C. pondérés par $K\left(\frac{-x}{h}\right)$ on doit minimiser par rapport au vecteur $a =$

$(a_0, a_1, \dots, a_p) \in \mathbb{R}^{p+1}$ la quantité

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p a_j (X_i - x)^j \right\}^2 K\left(\frac{X_i - x}{h}\right). \quad (\text{II.17})$$

Comme pour l'estimateur de Nadaraya-Watson, les paramètres K et h déterminent la forme et la taille du voisinage autour de x .

1-3-Propriétés asymptotiques de l'estimateur de Nadaraya-Watson

L'estimateur à noyau de la régression est donc dépendant du choix des deux paramètres que sont la fenêtre h et le noyau K . Nous verrons par la suite que le paramètre crucial est la fenêtre pour obtenir de bonnes propriétés asymptotiques. Toute fois le noyau ne doit pas être néglige, il permet de réduire le biais de l'estimateur en s'appuyant sur des propriétés de régularité de la courbe de régression.

1-3-1-propriété de l'estimateur de Nadaraya Watson

Dans ce qui suit nous allons déterminer les conditions sur la fenêtre et le noyau, nécessaires à la consistance de l'estimateur $\hat{r}_{NW;n}$. Nous utiliserons pour cela la décomposition biais variance suivante :

$$\mathbb{E}[(\hat{r}_{NW;n}(x) - r(x))^2] = \text{Var}[\hat{r}_{NW;n}] + [\mathbb{E}(\hat{r}_{NW;n}(x)) - r(x)]^2. \quad (\text{II.18})$$

On dénote par $\xrightarrow{L_2}$ (resp. \xrightarrow{p}) la convergence en norme L_2 (resp. en probabilité). Lorsque (II.18) tend vers 0, on déduit que $\hat{r}_{NW;n}(x) \xrightarrow{L_2} r(x)$ ce qui implique que $\hat{r}_{NW;n} \xrightarrow{p} r(x)$.

1-3-1-1- Calcul de la variance

Pour calculer la variance de l'estimateur $\hat{r}_{NW;n}$ ainsi que son expression asymptotique, Posons :

$$\bar{\sigma}^2(x) := \text{Var}(Y|X = x) = \frac{1}{f(x)} \int y^2 f_{X,Y}(x, y) dy - [r(x)]^2. \quad (\text{II.19})$$

Et $K =: \int k^2(t) dt$, lorsque ces deux expressions sont bien définies. Nous avons besoin aussi du lemme de Bochner que nous rappelons ci-dessous.

Lemme 2.1. (Bochner) Supposons que g soit une fonction bornée sur \mathbb{R} , continue dans un voisinage du point $x_0 \in \mathbb{R}$ et que Q soit une fonction sur \mathbb{R} telle que $\int |Q(t)| dt < +\infty$, alors :

$$\lim_{h \rightarrow 0} \frac{1}{h} \int Q\left(\frac{z - x_0}{h}\right) g(z) dz = g(x_0) \int Q(t) dt. \quad (\text{II.20})$$

Preuve : Posons : $t = \frac{z-x_0}{h}$, alors pour tout $h > 0$

$$\begin{aligned} & \left| \frac{1}{h} \int Q\left(\frac{z-x_0}{h}\right) g(z) dz - g(x_0) \int Q(t) dt \right| = \left| \int [g(x_0 + th) - g(x_0)] Q(t) dt \right| \\ & \leq \sup_{|t| < \frac{1}{2}} |g(x_0 + th) - g(x_0)| \int |Q(t)| dt + \int_{|t| \geq \frac{1}{2}} |g(x_0 + th) - g(x_0)| |Q(t)| dt \\ & \leq \sup_{|s| < \frac{1}{2}} |g(x_0 + th) - g(x_0)| \int |Q(t)| dt + 2 \sup_t |g(t)| \int_{|t| \geq \frac{1}{2}} |Q(t)| dt. \end{aligned}$$

On obtient le résultat en faisant tendre h vers 0.

Nous sommes maintenant en mesure d'évaluer la variance de l'estimateur de Nadaraya-Watson grâce à la proposition suivante :

Proposition 2.2. On suppose que $E(Y^2) < +\infty$. A chaque point de continuité des fonctions $r(x)$, $f(x)$ et $\sigma^2(x)$ tels que $f(x) > 0$, on a :

$$\text{Var}[\hat{r}_{NW;n}(x)] = \frac{1}{nh} \left(\frac{\bar{\sigma}^2(x)}{f(x)} \kappa \right) (1 + o(1)). \tag{II.21}$$

Où le terme $o(1)$ tend vers 0 lorsque h tend vers 0.

Remarque 2.2. Dans le calcul des variances et covariances précédentes, nous retrouvons invariablement le terme $\kappa =: \int K^2(t) dt$. Un choix approprié du noyau K assure la finitude de cette intégrale. On peut prendre par exemple $K(\cdot)$ à variation bornée sur \mathbb{R} et à support compact, suppositions qui implique évidemment les hypothèses $K(1 - 3)$. Si on veut obtenir une variance minimale, on peut minimiser κ suivant K dans une certaine classe de noyaux fixée.

1-3-1-2-Calcul du biais

Le calcul du biais repose essentiellement sur des développements de Taylor, ce qui nous conduit à poser certaines conditions de régularités sur les fonctions $r(\cdot)$ et $f(\cdot)$ qui déterminerons l'ordre du biais asymptotique en fonction du paramètre de lissage h . Afin de faciliter les calculs et du fait que l'estimateur $\hat{r}_{NW;n}(\cdot)$ est lui même sous forme d'un rapport aléatoire, nous allons introduire un terme de centrage sous forme de rapport, défini par :

$$\tilde{E}(\hat{r}_{NW;n}(x)) = \frac{\mathbb{E}(\hat{r}_{1;n}(x))}{\mathbb{E}(\hat{f}_{PR}(x))}. \quad (\text{II.22})$$

Pour évaluer le biais de l'estimateur, nous commençons d'abord par donner une proposition due à Nadaraya (1989).

Proposition 2.3. (Nadaraya 1989).

- Si Y est bornée et si nh tend vers $+\infty$, lorsque n tend vers $+\infty$, on a :

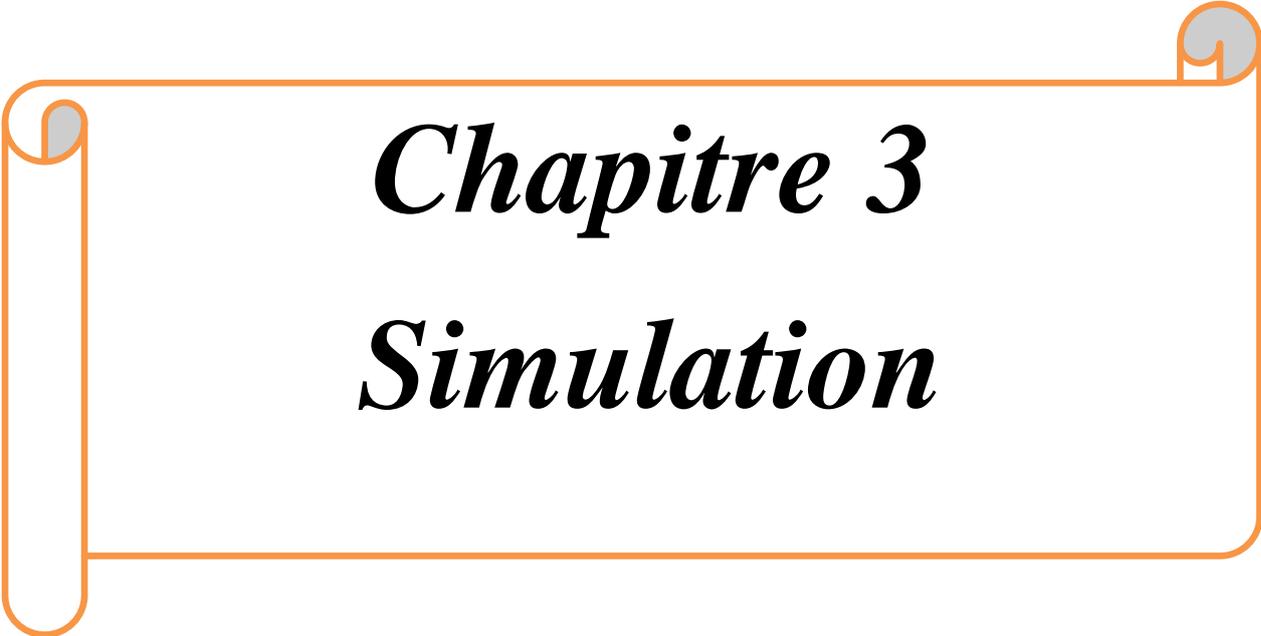
$$\mathbb{E}(\hat{r}_{NW;n}(x)) = \bar{\mathbb{E}}(\hat{r}_{NW;n}(x)) + o((nh)^{-1}). \quad (\text{II.23})$$

- Si $E(Y^2) < +\infty$ et si nh^2n tend vers $+\infty$, lorsque n tend vers $+\infty$, on a :

$$\mathbb{E}(\hat{r}_{NW;n}(x)) = \bar{\mathbb{E}}(\hat{r}_{NW;n}(x)) + o((\sqrt{nh})^{-1}). \quad (\text{II.24})$$

Conclusion

Dans ce chapitre, nous avons présenté l'estimateur à noyau de la fonction de régression par la méthode Nadaraya-Watson ainsi que ses différentes propriétés asymptotiques et l'estimateur par polynôme locaux.



Chapitre 3
Simulation

Introduction

Dans les chapitres (1) et (2), nous avons donné les résultats théoriques de l'estimateur non paramétrique de la densité de probabilité et la fonction de régression à savoir l'analyse exacte de l'estimateur.

Ensuite nous avons défini l'estimateur à noyau (Parzen-Rosenblatt) ainsi que ses propriétés de convergences dans le cas où la taille de l'échantillon est fixée ou aléatoire.

Dans ce chapitre, nous nous intéressons à l'estimation de la densité de probabilité et l'estimateur de la fonction de la régression dans le cas où la taille de l'échantillon est fixée

1-Le choix de langage de programmation

Dans ce mémoire, on a utilisé le langage de programmation Matlab version 7.9.0 dans la simulation des différents systèmes.

MATLAB (matrix Labor tory) est un langage de programmation de quatrième génération émulé par un environnement de développement du même nom ; il est utilisé à des fins de calcul numérique. Développé par la société The Math Works, MATLAB permet de manipuler des matrices, d'afficher des courbes et des données, de mettre en œuvre des algorithmes, de créer des interfaces utilisateur, et peut s'interfacer avec d'autres langages comme le C, C++, Java, et Fortran. Les utilisateurs de MATLAB (environ un million en 2004) sont des milieux très différents comme l'ingénierie, les sciences et l'économie dans un contexte aussi bien industriel que pour la recherche. Matlab peut s'utiliser seul ou bien avec des toolbox (boîte à outils).

Le langage MATLAB a été conçu par Cleve Moler à la fin des années 1970 à partir des bibliothèques Fortran, LINPACK et EISPACK, le logiciel MATLAB est construit autour du langage MATLAB. Une interface en ligne de commande, qui est un des éléments du bureau MATLAB, permet d'exécuter des commandes simples. Des séquences de commandes peuvent être sauvegardées dans un fichier texte, typiquement avec l'éditeur MATLAB, sous la forme d'un "script" ou encapsulée dans une fonction.

2-Plan de simulation de fonction de la densité

Nous considérons un échantillon $X(i) \ i = 1, \dots, n$. Est une suite de variables aléatoires indépendantes et identiquement distribuées de densité de probabilité f suit loi normale $(X_i \rightarrow N(\mu, \sigma^2))$ avec N d'observations.

Dans le but d'estimer f dans un intervalle donné. On suppose que F représente la fonction de répartition et f leur fonction de densité de probabilité avec la forme :

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

Où :

K : est le noyau qui on a choisi

h : est un paramètre de fenêtre

Dans ce chapitre nous avons utilisons le noyau gaussien sur la forme :

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

d'espérance nulle et variance unitaire.

Le choix le plus important c'est le choix de h car tous les noyaux on donne un bon estimateur mais c'est on choisit le mauvaise h on tombe dans un estimateur mauvais

De plus si K est un noyau gaussien alors la valeur de h optimale noté h_{opt}

Avec : $h_{opt} = 1.06\sigma n^{-\frac{1}{5}}$

2-1-Algorithmme de simulation

Afin de simuler cet échantillon décrit plus haut et d'évaluer ses performances durant un intervalle donné, on suit les étapes suivantes :

- Générer l'échantillon X_i selon la loi normale.
- Donner le nombre d'observation N de cette simulation.
- Donner l'intervalle de l'espace simulé.
- Choisir le noyau $K(\cdot)$ (Dans ce cas on a travaillé avec le noyau gaussien)
- Choisir la fenêtre de lissage h .
- Estimer $f(x)$ avec leurs estimateurs.
- Tracer le graphe des densités estimées.

2-2-Résultat de simulation

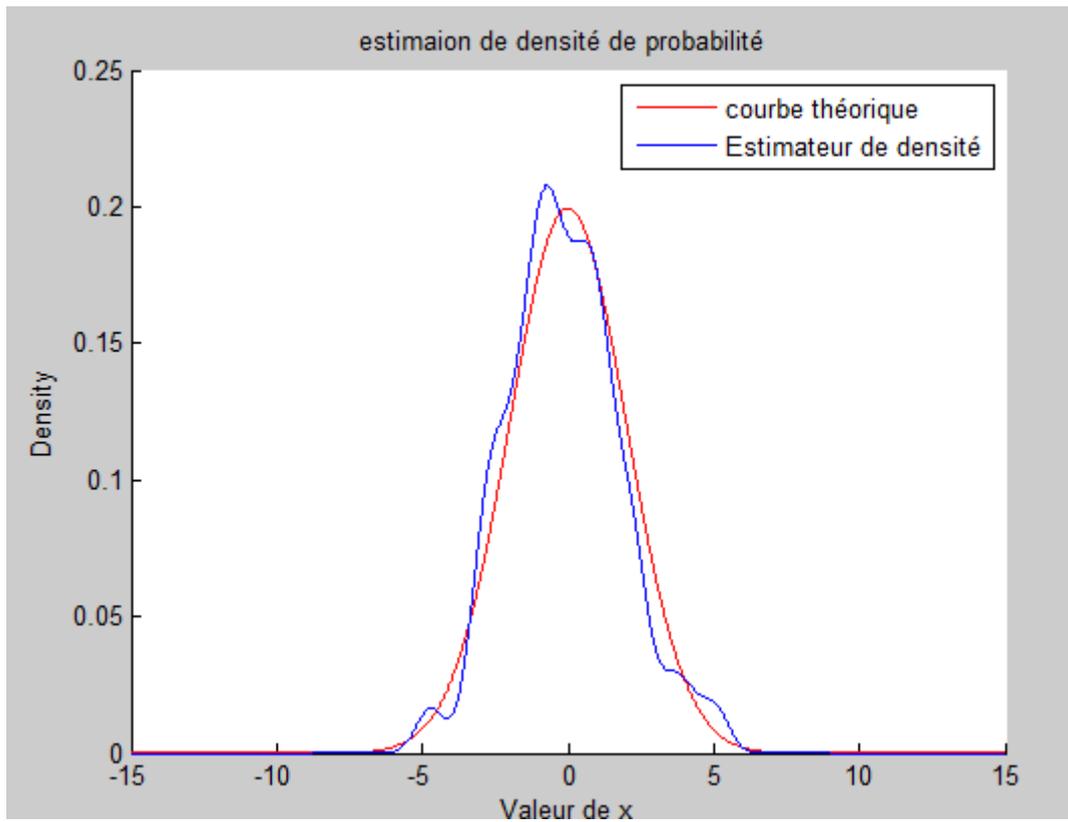


Figure 3.1: $N = 50$, $\mu=0$, $\sigma=0.5$, $h=0.242371$

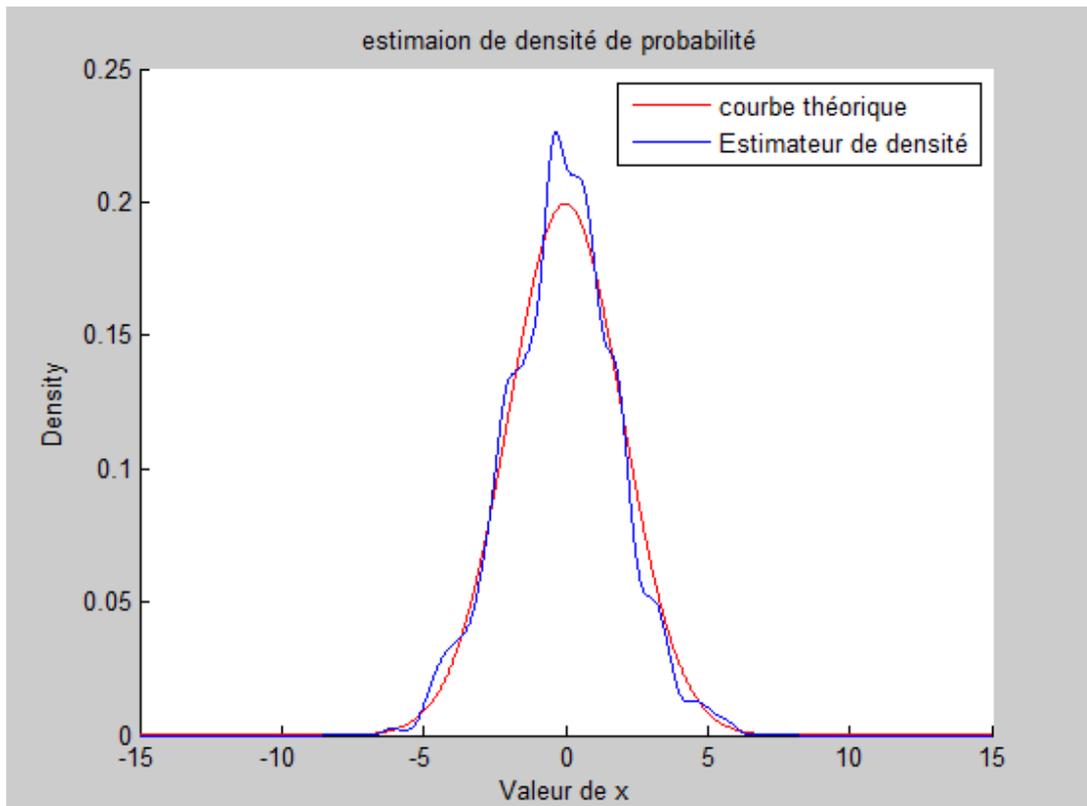


Figure 3.2: $N = 500$, $\mu=0$, $\sigma=0.5$, $h=0.152926$

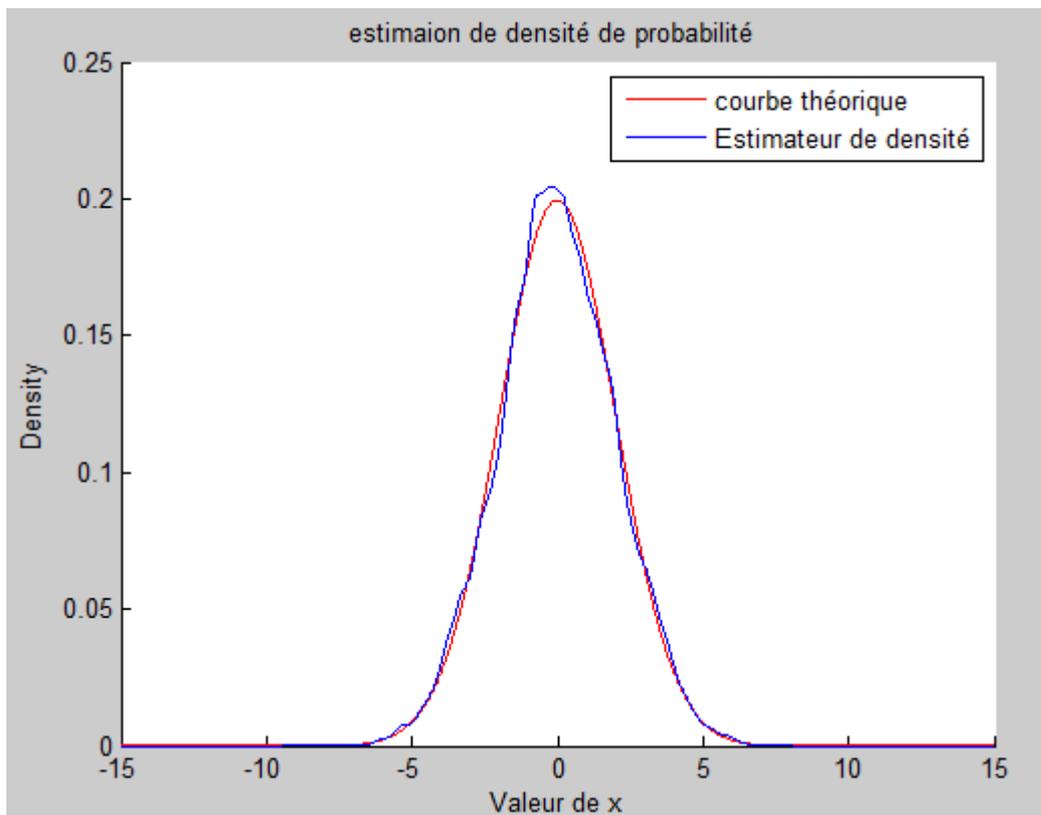


Figure 3.3: $N = 5000$, $\mu=0$, $\sigma=0.5$, $h=0.096489$

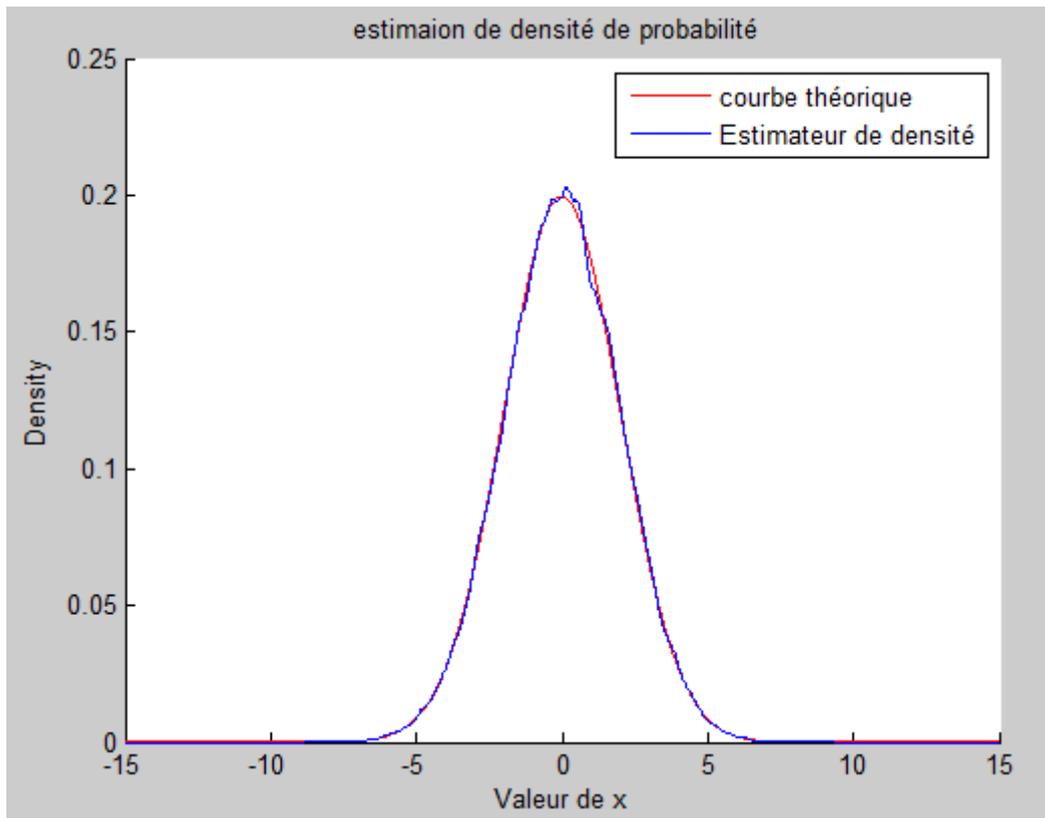


Figure 3.4: $N = 50000$, $\mu=0$, $\sigma=0.5$, $h=0.060881$

2-3-Interpritation

Quand on augmente le nombre d'observations N . On remarque que l'erreur diminue et l'information de l'estimation est presque là même avec l'information théorique. Les figures illustres cette diminution de l'erreur. En comparant les figures 3.1, 3.2, 3.3, 3.4 ; on voit clairement que l'allure de l'estimateur Parzen-Rosenblatt se rapproche de l'allure de la densité de probabilité théorique quand le nombre d'observations N augmente et la fenêtre h diminue. En générale, les caractéristiques de performances obtenues dans les différentes observations de cet échantillon avec l'estimateur de Parzen-Rosenblatt sont très proches de celles théoriques. Plus N est grand, ce qui donne une bonne estimation de densités.

3-Plan de simulation de fonction de régression

On Considère un échantillon $X(i) = 1, \dots, n$. Est une suite de variables aléatoires indépendantes et identiquement distribuées. Suivent une loi normale $(X_i \rightarrow N(\mu, \sigma^2))$ avec N d'observations.

La méthode la plus communément utilisée pour étudier la relation entre deux variables est

La régression linéaire simple, qui suppose un modelé de la forme :

$$Y_i = aX_i + b + \varepsilon_i$$

Tel que :

- ε_i est un bruit blanc ($\varepsilon_t \rightarrow N(\mathbf{0}, \mathbf{1})$).
- a et b sont des constants à choisir.

Le modèle de régression utiliser pour cette simulation est sur la forme :

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

Avec :

- ε_i représente la variation de Z_i autour de X_i .
- $r(X_i)$ est l'estimateur de Nadarata-Watson, sa forme est :

$$\hat{r}_{NW,n}(x) = \frac{\frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}$$

Où :

- K : est le noyau gaussien
- h : est un paramètre de fenêtre

3-1-Algorithmme de simulation

Nous allons organiser les étapes de simulation comme suite :

- Générer l'échantillon X_i selon la loi normale.
- Donner le nombre d'observation N et les constants a et b de cette simulation.
- Choisir un modèle de régression
- Donner l'intervalle de l'espace simulé.
- Choisir le noyau $K(\cdot)$ (Dans ce cas on a travaillé avec le noyau gaussien)
- Choisir la fenêtre de lissage h .
- Estimer Z_i .
- Tracer le graphe des régressions estimées.

3-2-Résultat de simulation

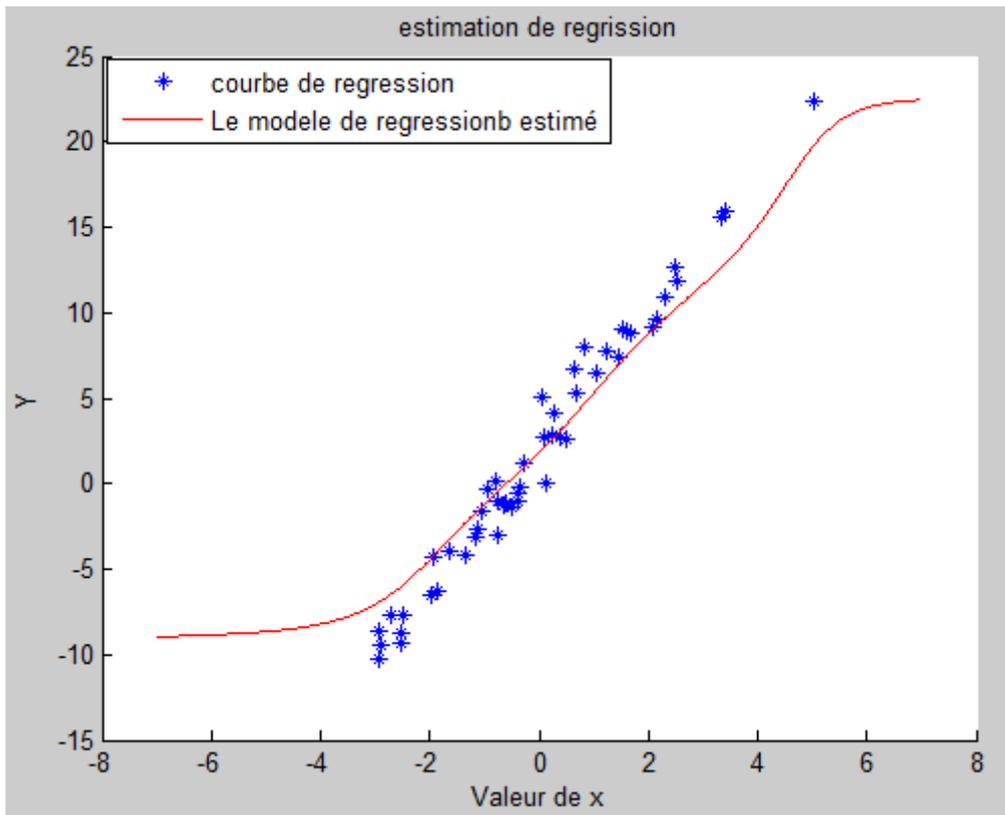


Figure 3.5: $N = 50$, $\mu = 0$, $\sigma = 2$, $a = 4$, $b = 2$, $h = 0.969486$.

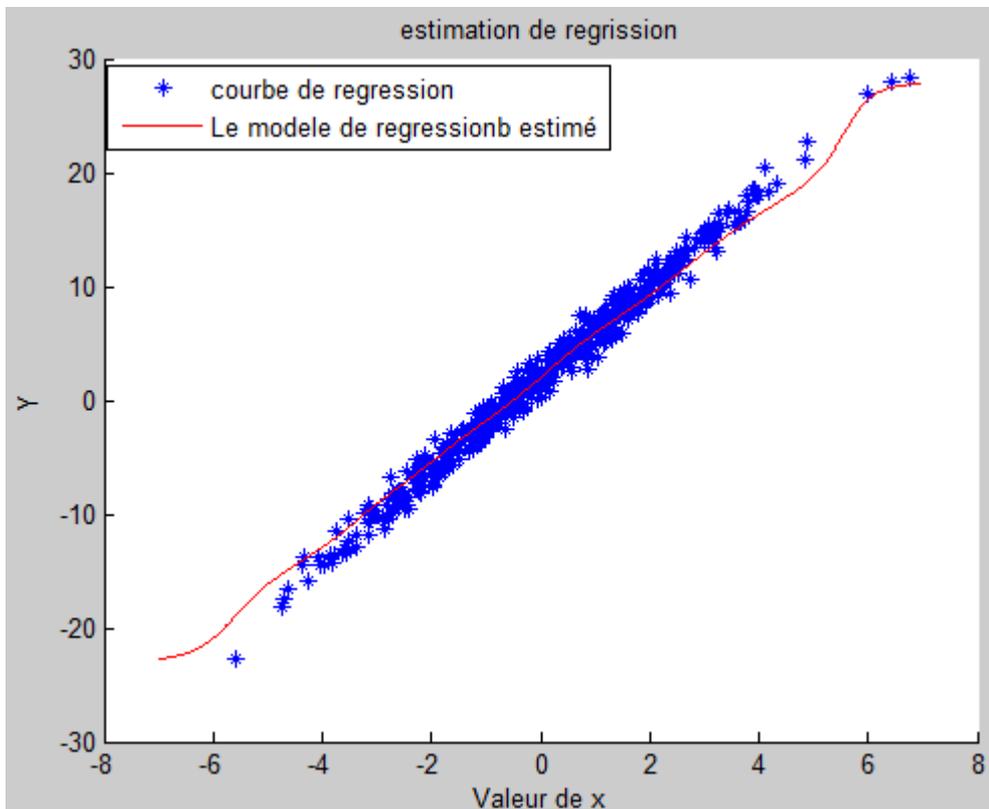


Figure 3.6: $N = 500$, $\mu = 0$, $\sigma = 2$, $a = 4$, $b = 2$, $h = 0.611704$.

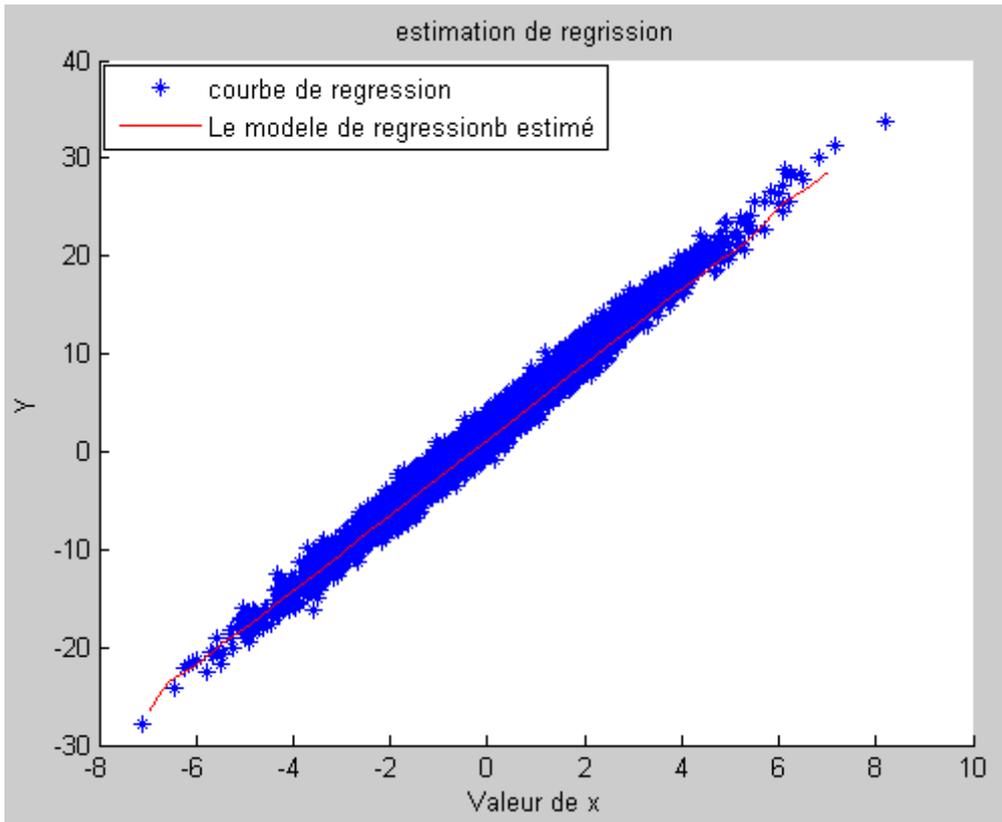


Figure 3.7: $N = 5000$, $\mu=0$, $\sigma=2$, $a=4$, $b=2$, $h=0.385961$.

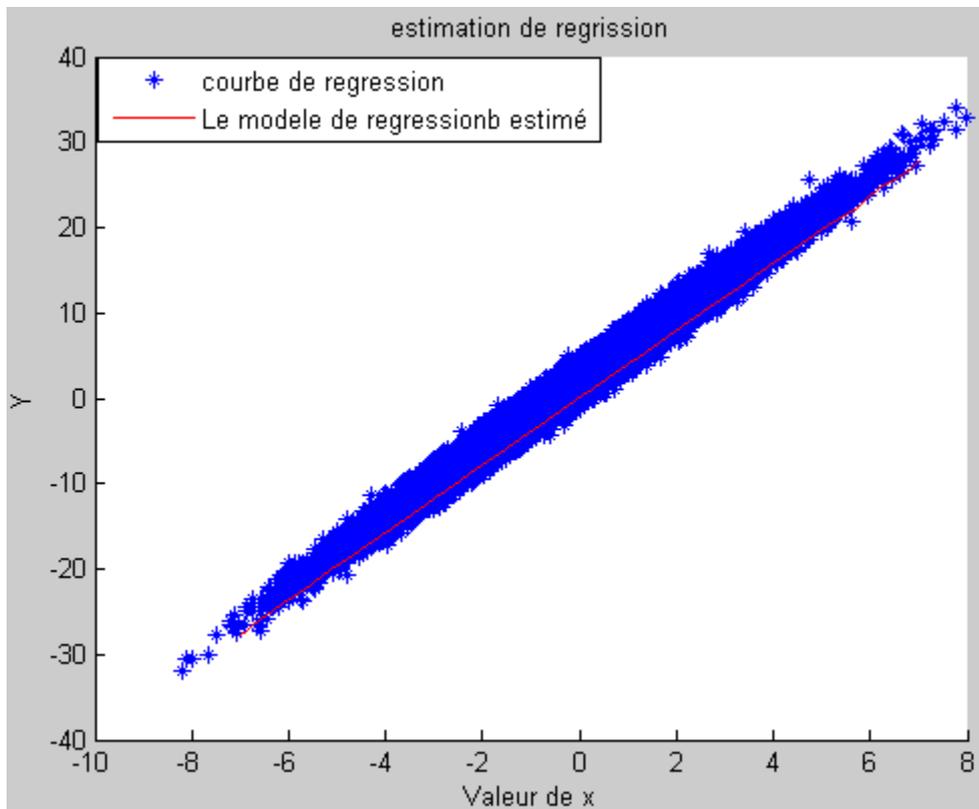


Figure 3.8: $N = 50000$, $\mu=0$, $\sigma=2$, $a=4$, $b=2$, $h=0.243524$.

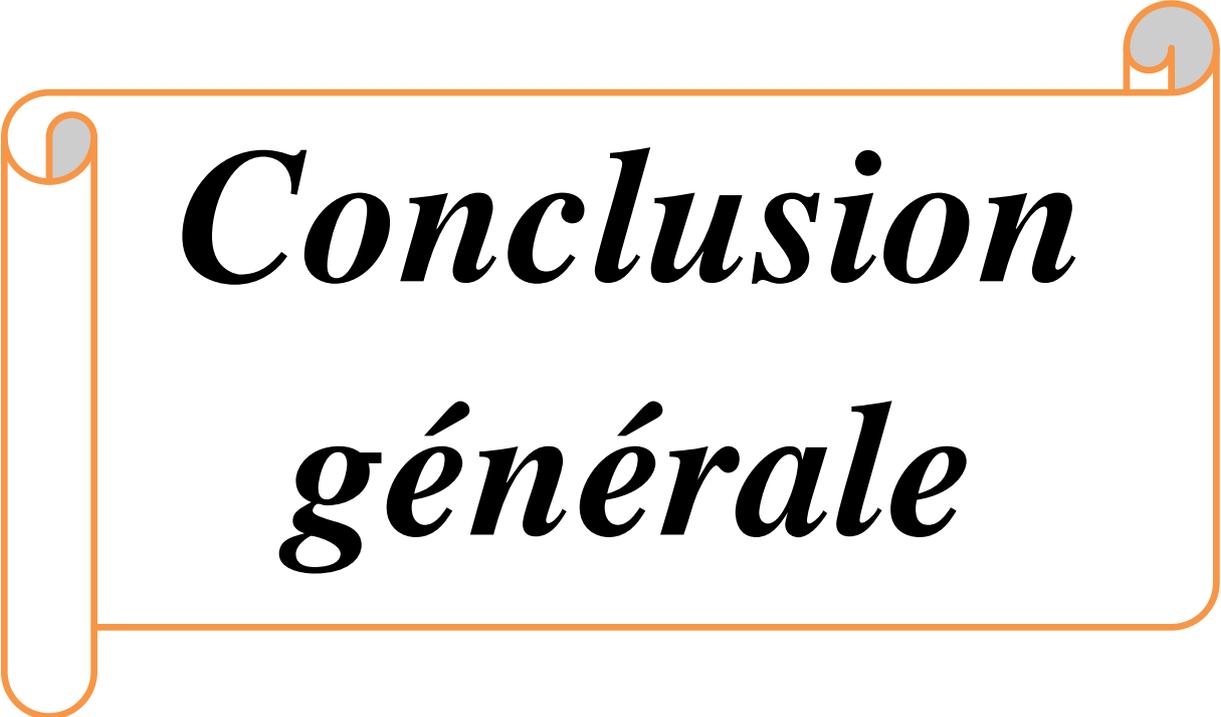
3-3-Interpritation

Les résultats obtenus montrent que l'information de l'estimateur par la méthode Nadaraya-Watson dépend de la taille de l'échantillon N .

Lorsque N augmente et h tend vers zéro, la courbe de l'estimateur de Nadaraya-Watson a obtenu est proche d'une interpolation des points $\{(X_i, Y_i), i = 1, \dots, n\}$

Lorsque N augmente l'estimateur de Nadaraya-Watson diminue l'erreur quadratique.

Lorsque N augmente la courbe de l'estimateur de Nadaraya-Watson est presque une droite.



*Conclusion
générale*

Conclusion générale

Au cours de notre mémoire de fin d'études, 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.

Ce mémoire se compose de deux parties essentielles.

La première partie est théorique, elle correspond aux différents résultats que nous avons cités Dans le chapitre un nous sommes concentrés sur les méthodes d'estimation de la densité :

La méthode d'estimation par histogramme, la méthode d'estimation simple de la densité 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. Nous avons présenté également, les propriétés statistiques de chaque méthode d'estimation.

Dans le chapitre deux, nous avons présenté la méthode de Nadaraya Watson et celle de polynômes locaux qui est une généralisation de la méthode de Nadaraya Watson. Nous avons présenté aussi, les propriétés statistiques des estimateurs obtenus.

La deuxième partie est pratique, nous avons fait des simulations pour illustrer la performance de notre estimateur.

Dans le chapitre trois, nous avons vu que la courbe de l'estimateur de Parzen-Rosenblatt se rapproche de la courbe de la densité de probabilité théorique quand le nombre d'observations N augmente et la fenêtre h diminue.

En générale, les caractéristiques de performances obtenues dans les différentes observations de l'échantillon par l'estimateur de Parzen-Rosenblatt sont très proches de celles théoriques. Plus N augmente, plus l'estimateur de densité est performant.

Pour l'estimateur de Nadaraya Watson nous avons obtenu que :

Lorsque N augmente et h tend vers zéro, la courbe de l'estimateur de Nadaraya-Watson a obtenu est proche d'une interpolation des points $\{(X_i, Y_i), i = 1, \dots, n\}$

Lorsque N augmente l'estimateur de Nadaraya-Watson diminue l'erreur quadratique.

Lorsque N augmente la courbe de l'estimateur de Nadaraya-Watson est presque une droite.

Résumé

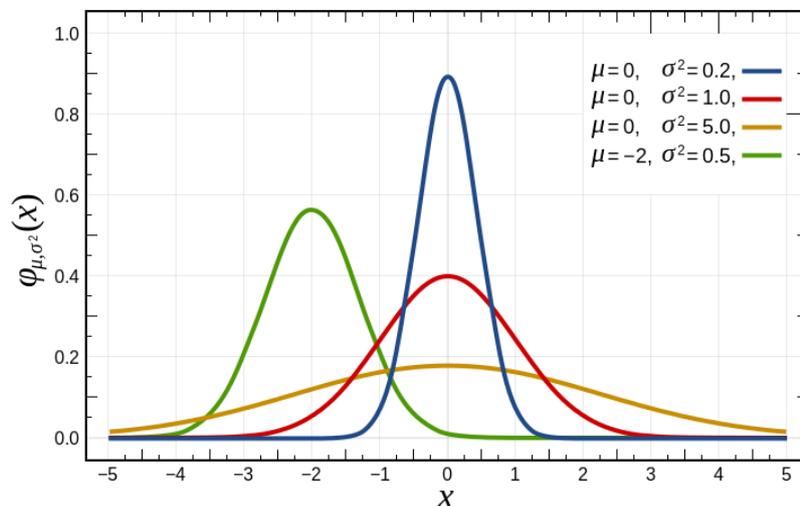
Dans ce mémoire, nous nous intéressons à l'estimation non paramétrique de la densité de probabilité et la fonction de la régression. Dans un premier temps, nous rappelons les méthodes et les propriétés de l'estimateur à noyau de la densité de probabilité. Puis nous nous pencherons sur l'estimation de la fonction de régression. Notre étude est illustrée par des simulations afin de montrer la performance des estimateurs étudiés.

Loi normale

En théorie des probabilités et en statistique, la loi normale est l'une des lois de probabilité les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires. Elle est en lien avec de nombreux objets mathématiques dont le mouvement, le bruit blanc gaussien ou d'autres lois de probabilité. Elle est également appelée loi gaussienne, loi de Gauss ou loi de Laplace-Gauss des noms de Laplace (1749-1827) et Gauss (1777-1855), deux mathématiciens, astronomes et physiciens qui l'ont étudiée. Plus formellement, c'est une loi de probabilité absolument continue qui dépend de deux paramètres : son espérance, un nombre réel noté μ , et son écart type, un nombre réel positif noté σ .

La densité de probabilité de la loi normale est donnée par :

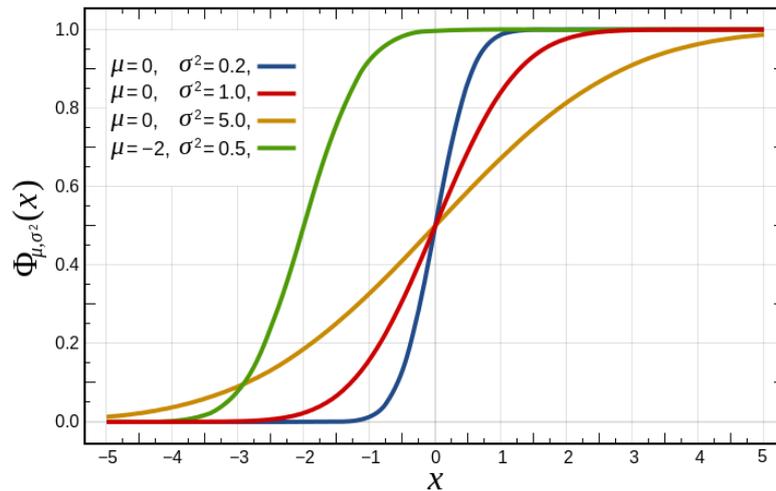
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} . \text{ pour tout } x \in \mathbb{R}$$



Densité de probabilité (ou fonction de masse) La courbe rouge représente la fonction φ , densité de probabilité de la loi normale centrée réduite.

La fonction de répartition de la loi normale est donnée par :

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt . \text{ pour tout } x \in \mathbb{R}$$



Fonction de répartition La courbe rouge représente la fonction Φ , fonction de répartition de la loi normale centrée réduite.

Lorsqu'une variable aléatoire X suit la loi normale, elle est dite gaussienne ou normale et il est habituel d'utiliser la notation avec la variance σ^2 :

$$X \sim N(\mu, \sigma^2)$$

L'espérance mathématique :

$$\mathbb{E}(X) = \mu.$$

Variance en intégrant par parties :

$$V(X) = \sigma^2.$$

Bibliographies

- [1]. A.Berlinet and G. Biau. Estimation de densité et prise de décision. In *Décision et Reconnaissance de Formes en Signal*, ed. Lengellé, R., pages 141–179. Hermès, Paris, 2002.
- [2]. A.Tsybakov. *Introduction to Nonparametric Estimation*. Springer science and business media, New York, 2009.
- [3]. Arnak S. DALALYAN. *STATISTIQUE AVANCÉE : MÉTHODES NON - PARAMÉTRIQUES* Ecole Centrale de Paris
- [4]. B. Silvermann. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, 1986.
- [5]. BARDET, J.-M., BERTRAND, P., AND BILLAT, V. Estimation non paramétrique de la densité spectrale d'un processus gaussien échantillon aléatoirement. arxiv preprint arxiv:0802.1388 (2008).
- [6]. Bosq, D. (1969). Estimation de la densité conditionnelle et de la régression. *C. R. Acad. Sci. Paris Sér. A-B*, 269, A661-A664.
- [7]. BOUEZMARNI, T., AND SCAILLET, O. Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* 21, 390-412 (2003).
- [8]. BOWMAN, A. W. An alternative method of cross-validation for the smoothing density estimates.
- [9]. BURMAN, P. A date dependent approach to density estimation. *Zeitschrift Für Wahrscheinlichkeitstheorie and Verwandte Gebiete*, 69, 609-628 (1985).
- [10]. CARBON, M., AND FRANCO, C. Estimation non paramétrique de la densité et de la régression-Prévision non paramétrique. *La Revue de Modula*, ISSN 1145-895X, 15 1 25 (1995).
- [11]. COLLOMB, G. Estimation non paramétrique de la régression: revue bibliographique.
- [12]. CYBAKOV, A. B. *Introduction à l'estimation non paramétrique*, vol. 41. Springer Science & Business Media, 2003.
- [13]. D. Bosq and J. Lecoutre. *Théorie de l'estimation fonctionnelle*. Paris Economica, 1987.
- [14]. Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés, *Revue de Statistique Appliquée*, XXV, p.5-42.
- [15]. DEHEUVELS, P. Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée* 25, 3 (1977), 5–42.

- [16]. DEHEUVELS,P., AND HOMINAL, P. Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Revue de Statistique Appliquée*, 27, 47,68 (1979).
- [17]. E.Fix and J.Hodges. Discriminatory analysis, nonparametric discrimination : consistency properties. Report No. 4, Project No. 21-49-004, USAF School of Aviation Medicine, 1951.
- [18]. E.Miranda,A. Bugarín, S. Li, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, editors, *Soft Methods in Probability and Statistics*, pages 45–52. Springer, 2006.
- [19]. ElFassi,K. (2008). Asymptotic theory of local polynomial estimate equipercntile equating function. European seminar, Bordeaux, France.
- [20]. F.Comby. Estimation du mouvement apparent majoritaire dans une séquence d'images vidéo par accumulation de votes bimodaux sur un histogramme approché.
- [21]. Ferraty,F.et Vieu, P. (2002/2003). *Statistique fonctionnelle : Modèles Non paramétrique de Régression*. Cours de DEA.
- [22]. HALL,P.AND MARRON, J. S. Local minima in cross validation function. *Journal of the royal statistical society*, 90,149-173 (1991).
- [23]. HALL,P.Cross ; validation in density estimation. *Biometrika*, 69, 383-390 (1982).
- [24]. Härdle,W. (1990). *Applied Nonparametric Regression*. Cambridge University Press, Cambridge.
- [25]. Härdle,W. et Kelly, G. (1987). Nonparametric kernel regression estimation –optimal choice of bandwidth. *Statistics*, 18.1, 21-35.
- [26]. I.Gannaz. D'épartement IF Cours de quatrième année INSA de Lyon Introduction `a la statistique *International Statistical Review/Revue Internationale de Statistique* (1981), 75–93.
- [27]. J.Geoffrey. Sur l'estimation d'une densité dans un espace métrique. *C.R. Acad. Paris Sér. A 278* :1449–1452, 1974.
- [28]. J.Lecoutre. Contribution à l'estimation non paramétrique de la régression. PhD thesis, Université de Pierre et Marie Curie-ParisVI-France, 1982.
- [29]. J.S.Simonoff. *Smoothing Methods in Statistics*. Springer-Verlag, 1996.
- [30]. J.van den berg, U. Kaymak, and W.-M. van den bergh. Financial markets analysis by using a probabilistic fuzzy modelling approach. *Int. J. Approx. Reasoning*,35(3) :291–305, 2004.
- [31]. K.Loquin and O. Strauss. Fuzzy histograms and density estimation. In J. Lawry,
- [32]. L.Devroye. *A Course in Density Estimation*. Birkhauser, Boston, 1987.

- [33]. L.Devrye and T. Wagner. Nonparametric discrimination and density estimation. Technical Report no. 183, University of Texas, Austin, 1976.
- [34]. L.Waltman, U. Kaymak, and J. Van Den Berg. Fuzzy histograms: A statistical analysis. In EUSFLAT-LFA 2005 Joint 4th EUSFLAT 11th LFA Conference, pages 605–610, Barcelona, September 2005.
- [35]. Lecoutre,J. (1982). Contribution à l'estimation non paramétrique de la régression. PhD thesis, Université de Pierre et Marie Curie-Paris VI-France.
- [36]. Lois de probabilité à densité Christophe ROSSIGNOL Année scolaire 2012/2013
- [37]. M.B.Retali. Convergence uniforme d'un estimateur de la densité par la méthode de noyau. *Revue Roumaine Math. Pure et Appliquées*, 23 :361–385, 1978.
- [38]. M.C. Jones and S. J. Sheather. Using non-stochastic terms to advantage in kernelbased estimation of integrated squared density derivatives. *Statistics &Probability Letters*, 11(6) 511–514, 1991.
- [39]. N.Glick. Consistency conditions for probability estimators and integrals of density estimators. *Utilitas Math.*, 23 :61–74, 1974.
- [40]. Nadaraya, E. A. (1989). *Nonparametric Estimation of Probability Densities and Regression Curves*. Kluwer, Dordrecht.
- [41]. NADARAYA,E. On nonparametric estimation density function and regression. *Theory Probab P.P.L*, 10,186-190 (1965).
- [42]. Nadaraya,E.A. (1964) On estimating regression. *Theory of Probability and its Applications* 10,186-190.
- [43]. O.Strauss, F. Comby, and M.-J. Aldon. Rough histograms for robust statistics. In ICPR, pages 2684–2687, 2000.
- [44]. P.Bickel and M. Rosenblatt. On some global measures of the deviations of density function estimators. *Annals of Statistics*, 3(6) :1071–1095, 1973.
- [45]. P.Hall and J. S. Marron. Estimation of integrated squared density derivatives. *Statistics &Probability Letters*, 6(2) :109–115, 1987.
- [46]. PARK,B.U. AND MARRON, S. J. Comparison of data driven bandwidth selectors. *Journal of the American Statistical Association*, 85, 66-72 (1990).
- [47]. Parzen,E.(1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, 1065-1076.
- [48]. PhD thesis, LIRMM-Université Montpellier II-France, 2001.
- [49]. Prakasa,Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.

- [50]. ROSENBLATT, M. Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist.*, 27, 832-837 (1956).
- [51]. RUDEMO, M. Empirical choice of histogram and kernel density estimators.
- [52]. S. Abou-Jaoudé. Sur une condition nécessaire et suffisante de la convergence presque complète de l'estimateur de la partition fixe pour une densité. *C.R. Acad. Paris Sér. A-B*, 283(16 :AII) :A1107–A1110, 1976.
- [53]. S. Tisserant – *Éléments de Statistique* – 2009
- [54]. SCOTT, D. W., AND TERRELL, G. Biased and unbiased cross validation in density estimation. *Journal of the American Statistical Association*, 82, P1131-1146 (1987).
- [55]. SILVERMAN, B. W. *Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability*, London (1986).
- [56]. STONE, C. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 12, P1285-1297 (1984).
- [57]. Stute, W. (1982). A law of the iterated logarithm for kernel density estimators. *Ann. Probab* 10.2, 414-422.
- [58]. T. A. Runkler. Fuzzy histograms and fuzzy chi-squared tests for independence. In *IEEE international conference on fuzzy systems*, pages 1361–1366, 2004.
- [59]. V. A. Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theoretical probability, applications*, 14 :153–158, 1969.
- [60]. W. Härdle. *Smoothing Techniques, With Implementations in S*. Springer, 1991.
- [61]. WAND, M. P., AND JONES, M. C. Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association* 88, 422 (1993) 520–528.
- [62]. WAND, M. P., AND JONES, M. C. *Kernel smoothing*. Crc Press, 1994.
- [63]. www.Les-Mathematiques.net.