

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

UNIVERSITÉ MOHAMED KHIDER, BISKRA

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

DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

TOUR Temader

Titre :

Sur l'estimation de la fonction de la densité

Membres du Comité d'Examen :

Fateh BENATIA	Prof	UMKB	Président
Abdallah SAYAH	MCA	UMKB	Encadreur
Souraya KHEIREDDINE	M.A.A	UMKB	Examineur

Septembre 2020

DÉDICACE

C'est avec un immense honneur et une grande modestie que je dédie ce modeste travail :

- A deux personnes qui m'ont donné tout ce que je veux : mes chers père et mère. Pour tout ce qu'ils ont fait pour moi pour que je sois ce que je suis aujourd'hui.

- A mes frères et sœurs.

- A toute ma famille, grands et petits.

- A tous les enseignants qui ont contribué à mes études.

- A mes meilleurs amis qui m'ont apporté leur support moral.

TOUR Temader.

REMERCIEMENTS

Tout d'abord je tiens à remercier Dieu de m'avoir donné le courage, la morale et la santé pour mener à bien ce travail.

Je tiens à exprimer mes remerciements et mon plus grand respect et appréciation à mon encadreur Dr. SAYAH Abdallah. Pour ses commentaires, remarques et suggestions qui ont donné une autre dimension à notre travail.

Mes remerciements vont aussi aux membres du jury, Dr. BENATIA Fateh et Melle KHEIREDDINE Souraya, qui ont bien accepté de lire et évaluer ce modeste travail.

Merci pour vos corrections et remarques constructives.

Enfin, je tiens à remercier ma famille pour leurs encouragements, soutien et sacrifice apportés tout au long de mon cheminement scolaire et universitaire. Merci à toute personne ayant contribué directement ou indirectement à ce travail.

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
List of Figures	v
Liste des tableaux	vii
Introduction	1
1 Définition de la variable aléatoire et l'estimateur	3
1.1 Espace probabilisé	3
1.2 Variable aléatoire	4
1.2.1 Variables aléatoires réelles discrètes	5
1.2.2 Variables aléatoires réelles continues	6
1.3 Convergences des suites des variables aléatoires	7
1.3.1 Estimateur	9
2 Estimation de la fonction de densité	11
2.0.2 Estimation paramétrique de la fonction de densité	11
2.1 Estimation non-paramétrique de la fonction de densité	13

2.2	Estimation par l'histogramme	13
2.2.1	Histogramme	13
2.3	Relation entre densité de probabilité et l'histogramme des données	15
2.4	Estimateur de densité par histogramme	15
2.4.1	Propriétés statistiques de l'estimateur par histogramme	16
2.4.2	Estimation à noyau de la densité	18
3	Simulation	32
	Conclusion	42
	Annexe : Abréviations et Notations	45

Table des figures

2.1	histogramme	14
2.2	histogramme de 100000 observations construit avec $h = 0.2$	15
2.3	Les courbes des noyau :Triangulaire,Gaussien, Epanechnikov, Biwiegh	22
3.1	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 100$	33
3.2	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 200$	34
3.3	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 1000$	34
3.4	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 100$	35
3.5	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 200$	35
3.6	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 1000$	36
3.7	Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov et un h "optimale" et pour n "varié"	36
3.8	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 100$	37

3.9	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 200$	38
3.10	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 1000$	38
3.11	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 100$	39
3.12	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 200$	39
3.13	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 1000$	40
3.14	Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov et un h "optimale" et pour n "varié"	40

Liste des tableaux

2.1	Table du quelque noyau existant	21
2.2	Table du quelque noyau avec l'efficacité	31

Introduction

La théorie de l'estimation est l'une des branches les plus basiques de la statistique mathématique. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. Depuis l'apparition de cette théorie, l'estimation de la densité de probabilité est l'un de ses principaux problèmes.

L'approche paramétrique consiste à supposer que la densité de probabilité f appartient à une famille de densités qui peuvent être décrites par un petit nombre (connu) de paramètres réels. Le statisticien qui opte pour une telle approche possède une bonne connaissance a priori du phénomène aléatoire. Il sait, par intuition ou par expérience, que la variable aléatoire X suit une loi f , tout en ignorant la valeur de son espérance ou de sa variance. Dans un tel contexte, l'estimation de la densité se réduit alors à un problème d'estimation de paramètres.

Ainsi, à moins d'avoir sur le phénomène aléatoire étudié des informations a priori très précises et indiscutables, le champ d'application d'un modèle paramétrique n'est satisfaisant que lorsque l'inflation du nombre de paramètres est telle que les méthodes d'estimation du modèle deviennent tout à fait inefficaces.

Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique propose de laisser parler les données, sans spécifier au préalable de forme sur f . L'outil d'estimation non paramétrique nous est fourni par l'histogramme : une fois les données regroupées en classes de valeurs, les fréquences empiriques

sont représentées par des aires rectangulaires dont les bases correspondent aux classes elles mêmes. L'histogramme convient bien pour des analyses relativement grossières, un problème vient du fait que l'histogramme donne une fonction qui n'est pas continue.

Il existe d'autres méthodes non paramétriques plus efficaces que la méthode de l'histogramme est celle de la méthode du noyau consiste à retrouver la continuité et qui est la plus utilisée vu sa simplicité et la qualité de l'estimation qu'elle assure. Rosenblatt (1956), et Parzen (1962) sont les premiers à proposer un estimateur à noyau d'une densité univariée. Cet estimateur est basé sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support.

Dans ce mémoire nous étudions l'estimation de la densité de probabilité à partir d'un échantillon indépendant et identiquement distribué (i.i.d). Il est divisé en trois chapitres.

Chapitre 1 : "Définition de la variable aléatoire et l'estimateur " : Pour le premier chapitre on pose un petit rappel sur les variables aléatoires par les deux approches discrètes et continues avec quelques types de convergence et ensuite on pose la définition de l'estimateur et leurs propriétés.

Chapitre 2 : "Estimation de la fonction de densité" : Dans le deuxième chapitre on fait le point sur l'estimation paramétrique de la fonction de densité. Ensuite, on fait une étude sur estimation non-paramétrique de la fonction de densité telle qu'on parle sur la méthode de l'histogramme et par la méthode de noyau, avec quelques méthodes sur le choix du paramètre de lissage.

Chapitre 3 : "Simulation" : Le troisième chapitre est une simulation par logiciel R.

Chapitre 1

Définition de la variable aléatoire et l'estimateur

Dans des domaines très différents comme les domaines scientifiques, sociologiques ou médicaux, on s'intéresse à de nombreux phénomènes dans lesquels apparaît l'effet du hasard. Ces phénomènes sont caractérisés par le fait que les résultats des observations varient d'une expérience à l'autre.

Le but est de rechercher d'estimer, à partir d'un échantillon, (la ou les valeurs numériques d'un ou plusieurs paramètres de la population considérée).

Ce chapitre permet de reprendre certaines notions de base des variables aléatoires, l'estimateur et leurs propriétés.

1.1 Espace probabilisé

Une expérience est appelée "aléatoire" s'il est impossible de prévoir à l'avance son résultat et si, répétée dans des conditions identiques. On appelle ensemble associé à une expérience aléatoire l'ensemble fondamental $\Omega = \{\text{tous les résultats possibles de cette expérience}\}$.

Un espace de probabilité(s) ou espace probabilisé est construit à partir d'un espace probabilisable en le complétant par une mesure de probabilité : il permet la modélisation

quantitative de l'expérience aléatoire étudiée en associant une probabilité numérique à tout événement lié à l'expérience. Formellement, c'est un triplet (Ω, \mathcal{F}, P) , \mathcal{F} est un ensemble des événements ou tribu sur Ω , et P est une probabilité sur \mathcal{F} , comme définie ci-dessous, (pour une étude plus détaillée voir, par exemple, Dusart [4] et Veysseyre [17]).

Définition 1.1 *Une mesure de probabilité, ou plus simplement une probabilité, sur Ω est une application $P : \Omega \rightarrow [0, 1]$ possédant les deux propriétés suivantes :*

1. $P(\Omega) = 1$.
2. Pour toute famille $A_1, A_2, \dots \in \mathcal{F}$ d'événements deux-à-deux disjoints,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

1.2 Variable aléatoire

Dans des nombreuses expériences aléatoires, on n'est pas intéressé directement par le résultat de l'expérience, mais par une certaine fonction de ce résultat. Considérons par exemple l'expérience qui consiste à observer, pour chacune des n pièces produites par une machine, si la pièce est défectueuse ou non. Nous attribuerons la valeur 1 à une pièce défectueuse et la valeur 0 à une pièce en bon état.

Définition 1.2 (variable aléatoire réelle) *Etant donné un ensemble fondamental Ω , une variable aléatoire réelle (v.a.r) est une application mesurable de (Ω, \mathcal{F}) dans $(\mathbb{R}, \mathcal{B})$, tel que \mathcal{B} est une tribu borélienne de \mathbb{R} :*

$$X : \omega \in \Omega \longmapsto X(\omega) \in \mathbb{R}.$$

Définition 1.3 (Loi de probabilité de v.a.r) *Soit Ω un ensemble fondamental muni d'une probabilité P , et soit X une v.a.r. On appelle **loi de probabilité de X** , notée P_X ,*

l'application qui à toute partie A de Ω associe

$$\begin{aligned} P_X(A) &= P(X \in A) \\ &= P(\{\omega \in \Omega : X(\omega) \in A\}). \end{aligned}$$

Fonction de répartition

Définition 1.4 *La fonction de répartition de la v.a.r X est définie par :*

$$F_X(x) = P(X \leq x), x \in \mathbb{R}.$$

Propriétés de la fonction de répartition :

1. $0 \leq F_X(x) \leq 1$.
2. $F_X(x)$ tend vers 0 en $-\infty$ et vers 1 en $+\infty$.
3. $F_X(x)$ est croissante et continue à droite.

Proposition 1.1 *Pour toutes les valeurs a et b appartenant à \mathbb{R} , on a l'identité :*

$$P(a < X \leq b) = F_X(b) - F_X(a), a < b.$$

1.2.1 Variables aléatoires réelles discrètes

Définition 1.5 *On dit qu'une v.a.r **discrète** si elle ne prend qu'un nombre fini ou infini dénombrable de valeurs :*

$$X \in \{x_i, i \in K \subset \mathbb{N}\}.$$

Dans ce cas, la loi de la variable aléatoire X est la loi de probabilité sur l'ensemble des valeurs possibles de X qui affecte la probabilité $P(X = x_i)$ au singleton $\{x_i\}$. En pratique, l'ensemble des valeurs que peut prendre X est \mathbb{N} ou une partie de \mathbb{N} .

Remarque 1.1

1. L'espérance mathématique (moment d'ordre 1) de la v.a.r discrète X , notée $E(X)$ est définie par (si la série $\sum_{i \in K} x_i P(X = x_i)$ est absolument convergente ou lorsque K est fini) :

$$E(X) = \sum_{i \in K} x_i P(X = x_i).$$

2. Le nombre :

$$Var(X) = E[(X - E(X))^2],$$

lorsqu'il existe, est appelé variance (moment d'ordre 2) de X , et l'écart type de X est :

$$\sigma(X) = \sqrt{Var(X)}.$$

1.2.2 Variables aléatoires réelles continues

Définition 1.6 Une v.a.r prend des valeurs sur un ensemble infini non dénombrable de points, est dite continue si elle existe une fonction f non négative, définie pour toute valeur x appartenant à \mathbb{R} et vérifiant, pour toute partie A de \mathbb{R} , la propriété :

$$P(X \in A) = \int_A f(x) dx,$$

et telle que :

$$\int_{\mathbb{R}} f(x) dx = 1$$

La fonction f est appelée la densité de probabilité de la variable aléatoire X .

La fonction de répartition de la variable aléatoire X est définie par :

$$F(x) = P(X < x) = \int_{-\infty}^x f(t) dt.$$

Remarque 1.2

1. L'espérance mathématique de la v.a.r continue X , définie sur l'espace probabilisé (Ω, \mathcal{F}, P) , est donnée par l'intégrale, si elle converge :

$$E(X) = \int_{\Omega} X dP = \int_{\Omega} x dP_x dx,$$

que l'on peut écrire, si f est la densité de probabilité de X

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

2. La variance, ou carré de l'écart-type σ , est donnée par l'intégrale si cette intégrale et la densité f existent

$$E[X - E(X)]^2 = Var(X) = \sigma^2 = \int_{\Omega} [x - E(X)]^2 f(x) dx.$$

1.3 Convergences des suites des variables aléatoires

Une suite (X_n) des variables aléatoires étant une suite des fonctions de Ω dans \mathbb{N} , il existe diverses façons de définir la convergence de (X_n) dont certaines jouent un grand rôle en calcul des probabilités.

1. Convergence en probabilité

La suite (X_n) converge en probabilité vers une v.a.r X et on note $X_n \xrightarrow{P} X$ si :

$$\forall \varepsilon > 0, \text{ on a } \lim_{n \rightarrow +\infty} P(\{|X_n - X| > \varepsilon\}) = 0.$$

2. La convergence presque sûre ou convergence forte

Définissons d'abord l'égalité presque sûre de deux variables aléatoires :

Définition 1.7 X et Y sont égales presque sûrement si

$$P(\{w : X(w) \neq Y(w)\}) = 0.$$

La convergence presque sûre se définit alors par :

Définition 1.8 La suite (X_n) converge presque sûrement (ou forte) vers X et on note $X_n \xrightarrow{p.s.} X$ si

$$P\left(\left\{w : \lim_{n \rightarrow +\infty} X_n(w) \neq X(w)\right\}\right) = 0$$

2. La convergence en loi

Bien que la plus faible, elle est très utilisée en pratique car elle permet d'approximer fonction de répartition de X_n par celle de X .

Définition 1.9 Soient (X_n) et X des variables aléatoires sur un même espace probabilisé (Ω, \mathcal{F}, P) , de fonctions de répartition respectives F_n et F ; on dit que les (X_n) convergent vers X en loi (et on note $X_n \xrightarrow{L} X$) si en tout point x où F est continue, les $F_n(x)$ convergent vers $F(x)$.

3. Convergence en moyenne quadratique

Une suite de v.a.r $(X_n)_{n \in \mathbb{N}}$ converge en moyenne quadratique vers une v.a.r X (et on note $X_n \xrightarrow{m.q.} X$) si :

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

– Lois des Grands Nombres

Ces lois décrivent le comportement asymptotique de la moyenne de l'échantillon. Elles sont de deux types : **lois faibles** mettant en jeu la convergence en probabilité et **lois fortes** relatives à la convergence presque sûre.

Théorème 1.1 Si (X_1, \dots, X_n) est un échantillon d'une v.a.r X tel que $E|X| < \infty$, alors

$$\text{Loi faible} \quad \bar{X}_n \xrightarrow{P} \mu \quad \text{quand } n \rightarrow \infty.$$

$$\text{Loi forte} \quad \bar{X}_n \xrightarrow{p.s} \mu \quad \text{quand } n \rightarrow \infty,$$

où $\mu := E(X)$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

– Théorème Central Limite

L'étude de somme $S_n = \sum_{i=1}^n X_i$ de variables indépendantes et de même loi joue un rôle capital en

statistique. Le théorème suivant connu sous le nom de **théorème central limite** établit la convergence vers la loi normale.

Théorème 1.2 Si X_1, X_2, \dots est une suite des v.a.r (i.i.d) de moyenne μ et une variance finie σ^2 , alors

$$\frac{(S_n - n\mu)}{\sigma\sqrt{n}} \xrightarrow{L} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

1.3.1 Estimateur

Définition 1.10 On appelle un estimateur de θ toute fonction mesurable T de $X = (X_1, X_2, \dots, X_n)$ dans Θ autrement dit :

$$T : X \rightarrow \Theta$$

$$X \rightarrow T(X)$$

$T(X)$ s'appelle l'estimateur de θ

Propriétés d'un estimateur :

Soit T un estimateur de θ , lorsque la taille n de l'échantillon intervient dans les propriétés, nous noterons T_n .

Estimateur avec biais : Un estimateur T_n de θ est dit avec biais si pour tout $\theta \in \Theta$ (Θ ouvert de \mathbb{R}) et tout entier positif n

$$E(T_n) = \theta + \text{Biais}(T_n).$$

La quantité $\text{Biais}(T_n)$ est le biais de l'estimateur T_n .

Par exemple : $Y = \frac{1}{n+1} \sum_{i=1}^n X_i$, est un estimateur biaisé de $E(X)$ car : $E(Y) = \frac{1}{n+1} \sum_{i=1}^n E(X_i) = \frac{n}{n+1} E(X)$.

Estimateur sans biais : Un estimateur T_n de θ et $\theta \in \Theta$ est dit sans biais si :

$$\text{Biais}(T_n) = 0 \text{ alors : } E(T_n) = \theta.$$

Par exemple : la moyenne empirique ($\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$) est un estimateur sans biais de la moyenne μ .

Estimateur asymptotiquement sans biais : Un estimateur T_n de θ est dit asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} \text{Biais}(T_n) = 0 \text{ alors } \lim_{n \rightarrow \infty} E(T_n) = \theta.$$

Par exemple : $\tilde{S}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ est un estimateur asymptotiquement sans biais de σ^2 .

Définition 1.11 soit T_n un estimateur de θ . Le risque quadratique de T_n ou erreur quadratique moyenne (en anglais "Mean squared Error" et on notée (MSE)) est :

$$MSE(T_n) = E[(T_n - \theta)^2].$$

Remarque 1.3 L'erreur quadratique moyenne peut être écrite comme une somme de la variance et du carré du biais de l'estimateur

$$MSE(T_n) = \text{Var}(T_n) + [E(T_n) - \theta]^2.$$

Chapitre 2

Estimation de la fonction de densité

Il y a plusieurs approches pour estimer une densité à partir de données, Il y a l'approche paramétrique, où l'objectif est d'estimer les paramètres d'une distribution connue. Une autre approche est l'estimation non paramétrique, la méthode la plus connue étant l'estimation à noyau. C'est cette dernière méthode qui sera appliquée. Le principe d'estimation non paramétrique est de laisser les données elle mêmes, donc au lieu de supposer une loi paramétrique et d'utiliser un échantillon pour estimer les différents paramètres, on utilise directement l'échantillon comme estimateur de la densité.

2.0.2 Estimation paramétrique de la fonction de densité

Un estimateur T est une fonction $T_n : x \rightarrow T_n(x, X_1, \dots, X_n)$ mesurable par rapport à l'observation (X_1, \dots, X_n) . Si l'on sait à priori que T appartient à une famille paramétrique $\{T(x, \theta), \theta \in \Theta\}$ où $\Theta \in \mathbb{R}^d$ et $T(., .)$ est une fonction continue, on parle alors d'estimation paramétrique, car estimer T revient à estimer le paramètre fini dimensionnel θ .

Au phénomène étudié, nous associons maintenant un modèle statistique P_θ qui dépend d'un paramètre θ . Pour se faire une idée de la valeur inconnue du paramètre θ , à partir des observation (X_1, \dots, X_n) qui sont i.i.d, on calcule ensuite une certaine valeur numérique, que l'on considérera comme valeur approchée de θ qu'on appellera une estimateur de θ .

Dans ce cas où il n'y a pas d'estimateur évident, on cherche un estimateur par la méthode de vraisemblance, ou par la méthode des moments,...etc

1. Soit un échantillon issu d'une v.a.r X normale de fonction de densité f qui dépend de deux paramètres qui sont inconnus (μ, σ^2) :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Pour estimer la fonction de densité f il se fait d'estimer le paramètre inconnue $\theta = (\mu, \sigma^2)$, où μ est la moyenne de X et σ^2 est sa variance.

On a dans ce cas :

$$\hat{\mu} = \bar{X} \text{ et } \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2,$$

et l'estimateur de $\theta = (\mu, \sigma^2)$ est $\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2) = (\bar{X}, \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2)$.

Donc l'estimateur \bar{f} de f est donné :

$$\begin{aligned} \bar{f}(x) &= \frac{1}{\sqrt{2\pi}\sqrt{\hat{\sigma}^2}} \exp \left[\frac{-1}{2} \left(\frac{x - \hat{\mu}}{\sqrt{\hat{\sigma}^2}} \right)^2 \right] \\ &= \frac{1}{\sqrt{2\pi} \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} \exp \left[\frac{-1}{2} \left(\frac{x - \bar{X}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}} \right)^2 \right]. \end{aligned}$$

2. Soit X un v.a.r suit la loi exponentielle de fonction de densité $f(x) = \lambda e^{-\lambda x}$ avec $\lambda > 0$ paramètre inconnu.

Pour estimer la fonction de densité f il se fait d'estimer le paramètre inconnue λ .

L'estimateur de λ est $\hat{\lambda} = \frac{1}{\bar{X}}$.

Donc l'estimateur \bar{f} de f est donné :

$$\bar{f}(x) = \hat{\lambda} e^{-\hat{\lambda}x} = \frac{1}{\bar{X}} e^{-\frac{1}{\bar{X}}x}.$$

2.1 Estimation non-paramétrique de la fonction de densité

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 à 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).

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

2.2 Estimation par l'histogramme

En statistique, l'histogramme est une représentation graphique de la répartition de la v.a.r et le plus simple et 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. On va voir, dans cette section, comment obtenir cette dernière propriété et on va discuter ensuite les propriétés statistiques de l'estimateur de densité par histogramme.

2.2.1 Histogramme

Construire un histogramme à partir d'un ensemble d'observations (précises) x_1, \dots, x_n consiste à partitionner l'intervalle de référence $E = [e_{\min}, e_{\max}]$ en $p \in \mathbb{N}$ classes C_k , $k \in \{1, \dots, p\}$ et à compter le nombre d'observations appartenant à chaque classe C_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 :

$$h = \frac{e_{\max} - e_{\min}}{p}.$$

Le nombre d'observations appartenant à chaque classe C_k est noté ν_k . Il est défini :

$$\nu_k = \sum_{i=1}^n 1_{C_k}(x_i),$$

où 1_{C_k} est la fonction indicatrice de l'ensemble C_k définie par :

$$1_{C_k} = \begin{cases} 1 & \text{si } x \in C_k \\ 0 & \text{sinon} \end{cases},$$

Figure 2.1 présente un histogramme de 100 observations tirées aléatoirement d'une loi normale centrée réduite $\mathcal{N}(0, 1)$. La fenêtre de l'histogramme est fixée à $h = 0.7$.

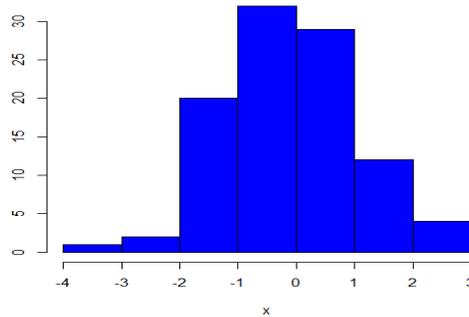


FIG. 2.1 – histogramme

2.3 Relation entre densité de probabilité et l'histogramme des données

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.

On illustre dans la Figure 2.2, cette relation, en considérant le même processus d'observations précédent avec un grand nombre d'observations ($n = 100000$) et un plus faible pas ($h = 0.02$). 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 ($\mathcal{N}(0, 1)$) quand le nombre d'observations n augmente et la valeur de h diminue.

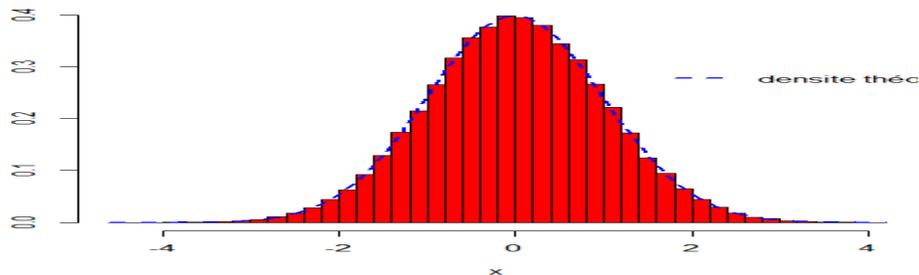


FIG. 2.2 – histogramme de 100000 observations construit avec $h = 0.2$

2.4 Estimateur de densité par histogramme

Soit un ensemble fini d'observations (x_1, \dots, x_n) de n (v.a.r) (X_1, \dots, X_n) i.i.d de densité de probabilité commune f . Pour tout $k \in \{1, \dots, p\}$, soit ν_k le nombre d'observation appartenant à la classe C_k , la probabilité de C_k (basée sur les observations x_i), notée $P(C_k)$, est donnée par :

$$P(C_k) = \frac{\nu_k}{n}.$$

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

$$\tilde{f}_h(x) = \frac{1}{h} \sum_{k=1}^p P(C_k) 1_{C_k}(x).$$

Cet estimateur peut aussi s'écrire

$$\tilde{f}_h(x) = \frac{1}{nh} \sum_{k=1}^p \nu_k 1_{C_k}(x).$$

2.4.1 Propriétés statistiques de l'estimateur par histogramme

On présente, les propriétés statistiques de l'estimateur par histogramme \tilde{f}_h . 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 biais, 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 essaie généralement de réduire au mieux ces deux quantités.

1. Le biais de l'estimateur \tilde{f}_h est donné par :

$$\text{Biais}(\tilde{f}_h(x)) = E(\tilde{f}_h(x)) - f(x) = \frac{h}{2} f'(x) + O(h^2),$$

2. La variance de l'estimateur \tilde{f}_h est donnée par :

$$\text{Var}(\tilde{f}_h(x)) = E\left[\left(\tilde{f}_h(x)\right)^2\right] - \left[E\left(\tilde{f}_h(x)\right)\right]^2 = \frac{f(x)}{nh} + O(n^{-1}).$$

Cette variance tend vers zéro quand le produit $nh \rightarrow \infty$ quand le nombre d'observation $n \rightarrow \infty$.

- L'erreur quadratique moyenne (MSE) de l'estimateur \tilde{f}_h est donné par

$$\begin{aligned} MSE\left(\tilde{f}_h(x)\right) &= \left[\text{Biais}\left(\tilde{f}_h(x)\right) \right]^2 + \text{Var}\left(\tilde{f}_h(x)\right) \\ &= \frac{f(x)}{nh} + \frac{h^2}{4}(f'(x))^2 + O(n^{-1}) + O(h^4). \end{aligned}$$

Cette erreur quadratique moyenne tend vers zéro quand h tend vers zéro et $nh \rightarrow \infty$ quand $n \rightarrow \infty$, ce critère d'erreur quadratique moyenne est un critère local.

L'approximation asymptotique de la MSE est donnée par :

$$AMSE(\tilde{f}_h(x)) = \frac{f(x)}{nh} + \frac{h^2}{4}(f'(x))^2.$$

La fenêtre optimale h^{**} qui minimise l' $AMSE(\tilde{f}_h(x))$ est donnée par :

$$h^{**} = \left(\frac{2f(x)}{(f'(x))^2} \right)^{\frac{1}{3}} n^{-\frac{1}{3}},$$

l'erreur quadratique moyenne asymptotique associée à h^{**} est

$$AMSE^{**}\left(\tilde{f}_h(x)\right) = \frac{3}{(4)^{\frac{2}{3}}}(f(x)f'(x))^{\frac{2}{3}} n^{-\frac{2}{3}}.$$

- L'erreur quadratique moyenne intégrée ($MISE$) de l'estimateur \tilde{f}_h est donné par :

$$MISE\left(\tilde{f}_h(x)\right) = \frac{h^2}{12} \int (f'(x))^2 dx + \frac{1}{nh} + O(h^3) + O(n^{-1}).$$

Cette erreur quadratique moyenne intégrée tend vers zéro quand h tend vers zéro et $nh \rightarrow \infty$ quand $n \rightarrow \infty$, ce critère d'erreur quadratique moyenne est un critère global.

L'approximation asymptotique de la $MISE$ est donné par :

$$AMISE\left(\tilde{f}_h(x)\right) = \frac{h^2}{12} \int (f'(x))^2 dx + \frac{1}{nh}. \quad (2.1)$$

L'utilisation du critère *AMISE* permet de définir une fenêtre 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 :

$$h^* = \left[\frac{6}{\int (f'(x))^2 dx} \right]^{\frac{1}{3}} n^{-\frac{1}{3}}. \quad (2.2)$$

Cette technique est dédiée à l'estimation de densités dont la loi est inconnue, la valeur de h^* est généralement inconnue. (pour une étude plus détaillée voir, par exemple, Simonoff [15]).

En remplaçant, la valeur de h^* (2.2) dans l'expression (2.1), on obtient la valeur optimale de l'erreur quadratique moyenne intégrée asymptotique (*AMISE*) notée $AMISE^*$:

$$AMISE^*(\tilde{f}_h(x)) = \left[\frac{9 \int (f'(x))^2 dx}{16} \right]^{\frac{1}{3}} n^{-\frac{2}{3}}.$$

2.4.2 Estimation à noyau de la densité

En statistique, l'estimation par noyau (ou encore méthode de Parzen-Rosenblatt) est une méthode non-paramétrique d'estimation de la densité de probabilité d'une variable aléatoire. Elle se base sur un échantillon d'une population statistique et permet d'estimer la densité en tout point du support. En ce sens, cette méthode généralise astucieusement la méthode d'estimation par un histogramme.

La construction d'un estimateur à noyaux

Supposons que nous observons n (v.a.r) i.i.d X_1, \dots, X_n de densité de probabilité inconnue f de \mathbb{R} dans $[0, \infty[$.

Soit $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X , la fonction de répartition

empirique est donnée par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x)}.$$

Rappelons que la densité de probabilité f est égale à la dérivée de la fonction de répartition F (si cette dérivée existe). On peut donc écrire

$$\begin{aligned} f(x) &= \frac{dF(x)}{dx} \\ &= \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}, \quad h > 0 \\ &= \lim_{h \rightarrow 0} \frac{P(x-h \leq X \leq x+h)}{2h}. \end{aligned}$$

La loi forte des grands nombres nous donne

$$\forall x \in \mathbb{R} : F_n(x) \xrightarrow{p.s} F(x) \text{ si } n \rightarrow +\infty.$$

Rosenblatt (1956) est le premier qui a donné un exemple d'estimateur à partir de $F_n(x)$ pour $h > 0$ est petit.

Un estimateur de $f(x)$ est alors :

$$\begin{aligned} \widehat{f}_n(x) &= \frac{F_n(x+h) - F_n(x-h)}{2h} \\ &= \frac{1}{2h} \left\{ \frac{1}{n} \sum_{i=1}^n \left(1_{(X_i \leq x+h)} - 1_{(X_i \leq x-h)} \right) \right\} \\ &= \frac{1}{2nh} \sum_{i=1}^n \left(1_{(x-h \leq X_i \leq x+h)} \right) \\ &= \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \left(1_{(-1 \leq \frac{X_i - x}{h} \leq +1)} \right) \\ &= \frac{1}{nh} \sum_{i=1}^n k_0 \left(\frac{X_i - x}{h} \right), \end{aligned}$$

où

$$k_0(y) = \begin{cases} \frac{1}{2} & \text{si } y \in [-1, 1] \\ 0 & \text{sinon} \end{cases}$$

Avec $k_0(\cdot)$ est une densité de probabilité uniforme sur l'intervalle $[-1, 1]$ est appelée noyau de Rosenblatt. Cet estimateur peut être généralisé en remplaçant la fonction de poids $k_0(\cdot)$ par une fonction de poids plus générale k .

Notion de noyau

Nous définissons maintenant plus généralement la notion d'estimateur à noyau.

Définition 2.1 Soit $k : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction positive et intégrable, tel que $\int_{\mathbb{R}} k(t)dt = 1$, k est dit noyau.

Pour $n \in \mathbb{N}^*$ on appelle $h = h_n > 0$ la fenêtre ou paramètre de lissage et $\hat{f}_n(x)$ l'estimateur à noyau de la densité de probabilité f (ou l'estimateur de Parzen -Rosenblatt) défini pour tout $x \in \mathbb{R}$ par :

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

Lemme 2.1 Si k est une densité de probabilité, alors \hat{f}_n est aussi une densité de probabilité, en effet

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

en utilisant un changement de variable $t = \frac{y - x}{h} \Rightarrow dt = -\frac{1}{h} dx$, on obtient :

$$\int_{\mathbb{R}} \widehat{f}_n(x) dx = \int_{\mathbb{R}} k(t) dt = 1.$$

Définition 2.2 *Un noyau est dit symétrique si, pour tout t dans son ensemble de définition*

$$k(t) = k(-t).$$

On peut dire que tout noyau symétrique vérifie les conditions suivantes :

1. $k(t) \geq 0, \forall t \in \mathbb{R}$.
2. $\int_{\mathbb{R}} tk(t)dt = 0$.
3. $0 < \int_{\mathbb{R}} t^2k(t)dt < \infty$.
4. $\int_{\mathbb{R}} k^2(t)dt < \infty$.
5. $\int_{\mathbb{R}} k(t)dt = 1$.
6. $\int_{\mathbb{R}} |k(t)|dt < \infty, \sup_{-\infty < t < +\infty} |k(t)| < \infty, \lim_{t \rightarrow \infty} |tk(t)| < \infty$.

Le tableau suivant présente un exemple de quelques noyaux continus symétriques et leurs formes sont présentées dans la figure 2.3

noyau	$k(t)$
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \text{ si } t \in \mathbb{R}$
Epanechnikov	$\frac{3}{4} (1 - t^2) \mathbf{1}_{(t \leq 1)}$
Quartique (Biweight)	$\frac{15}{16} (1 - t^2)^2 \mathbf{1}_{(t \leq 1)}$
Triangulaire	$(1 - t) \mathbf{1}_{(t \leq 1)}$

TAB. 2.1 – Table du quelque noyau existant

Consistance

La consistance de l'estimateur est basée sur le lemme suivant (**Lemme de Bochner**)

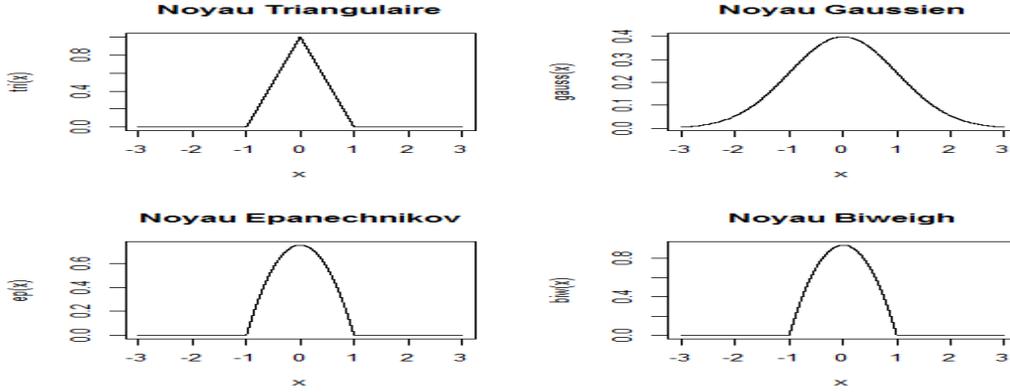


FIG. 2.3 – Les courbes des noyau :Triangulaire,Gaussien, Epaechnikov, Biwiegh

Lemme 2.2 Soit k une fonction vérifie les conditions précédentes, et soit $g \in L^1(\mathbb{R})$ i.e $(\int |g(t)|dt < \infty)$, et $g_n(x) = \frac{1}{h_n} \int k(\frac{1}{h_n})g(x - y)dy$, où (h_n) est la suite des constantes positives ayant $\lim_{n \rightarrow \infty} h_n = 0$, et si x est un point de continuité de g alors :

$$\lim_{n \rightarrow \infty} g_n(x) = g(x) \int_{\mathbb{R}} k(y)dy.$$

Théorème 2.1 L'estimateur à noyau \hat{f}_n dans les conditions précédentes est asymptotiquement sans biais si $h \rightarrow 0$ quand $n \rightarrow \infty$ i.e

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

Théorème 2.2 L'estimateur \hat{f}_n dans les conditions précédentes est cohérent si l'on ajoute la contrainte supplémentaire $\lim_{n \rightarrow \infty} nh \rightarrow \infty$ alors

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

et

$$MSE(\hat{f}_n(x)) = \left[Biais(\hat{f}_n(x)) \right]^2 + Var(\hat{f}_n(x)).$$

Nous allons démontrer que la variance tend vers zéro si $\lim_{n \rightarrow \infty} nh \rightarrow \infty$, que le biais tend vers zéro a été prouvé dans le théorème précédent ainsi, $MSE(\hat{f}_n(x)) \rightarrow 0$, c'est à dire $\hat{f}_n(x)$ est un estimateur consistant.(pour une étude plus détaillée voir, par exemple, Achour [1]).

Propriétés

Biais de \hat{f}_n

Soit x fixe dans \mathbb{R} . Le biais de l'estimateur à noyau $\hat{f}_n(x)$ est :

$$\begin{aligned} \text{Biais}(\hat{f}_n(x)) &= E(\hat{f}_n(x)) - f(x) \\ &= \frac{h^2}{2} f''(x) \int_{\mathbb{R}} t^2 k(t) dt + o(h^2) \\ &= \frac{h^2}{2} \mu_2(k) f''(x) + o(h^2) \quad \text{avec } \mu_2(k) = \int_{\mathbb{R}} t^2 k(t) dt. \end{aligned}$$

Variance de \hat{f}_n

Soit x fixe dans \mathbb{R} . La variance de l'estimateur $\hat{f}_n(x)$ est :

$$\begin{aligned} \text{Var}(\hat{f}_n(x)) &= \frac{1}{nh} f(x) \int_{\mathbb{R}} k^2(t) dt + o\left(\frac{1}{nh}\right) \\ &= \frac{1}{nh} R(k) f(x) + o\left(\frac{1}{nh}\right) \quad \text{avec } R(k) = \int_{\mathbb{R}} k^2(t) dt. \end{aligned}$$

Preuve. Si on a n -échantillon (X_1, \dots, X_n) des variables aléatoires i.i.d nous avons donc

Pour le biais

$$\begin{aligned} E(\hat{f}_n(x)) &= \frac{1}{nh} E \left(\sum_{i=1}^n k \left(\frac{X_i - x}{h} \right) \right) \\ &= \frac{1}{h} E \left(k \left(\frac{y - x}{h} \right) \right) \\ &= \frac{1}{h} \int_{\mathbb{R}} k \left(\frac{y - x}{h} \right) f(y) dy, \end{aligned}$$

on utilise un changement de variable $t = \frac{y-x}{h} \Rightarrow dt = \frac{1}{h} dy$, on obtient :

$$E(\widehat{f}_n(x)) = \int_{\mathbb{R}} k(t) f(x+th) dt;$$

En utilisant le développement de Taylor d'ordre 2 de f au voisinage de x on obtient :

$$f(x+th) = f(x) + thf'(x) + \frac{(th)^2}{2} f''(x) + o(h^2).$$

Alors

$$\begin{aligned} E(\widehat{f}_n(x)) &= \int_{\mathbb{R}} k(t) \left[f(x) + thf'(x) + \frac{(th)^2}{2} f''(x) + o(h^2) \right] dt \\ &= f(x) \int_{\mathbb{R}} k(t) dt + hf'(x) \int_{\mathbb{R}} tk(t) dt + \frac{h^2}{2} f''(x) \int_{\mathbb{R}} t^2 k(t) dt + o(h^2). \end{aligned}$$

Sous les conditions précédentes on a :

$$E(\widehat{f}_n(x)) = f(x) + \frac{h^2}{2} f''(x) \int_{\mathbb{R}} t^2 k(t) dt + o(h^2).$$

Ainsi on a

$$\begin{aligned} \text{Biais}(\widehat{f}_n(x)) &= E(\widehat{f}_n(x)) - f(x) \\ &= \frac{h^2}{2} f''(x) \int_{\mathbb{R}} t^2 k(t) dt + o(h^2), \end{aligned}$$

si on pose $\mu_2(k) = \int_{\mathbb{R}} t^2 k(t) dt$ alors

$$\text{Biais}(\widehat{f}_n(x)) = \frac{h^2}{2} \mu_2(k) f''(x) + o(h^2).$$

Pour la variance

$$\begin{aligned} \text{Var}(\widehat{f}_n(x)) &= \text{Var} \left[\frac{1}{nh} \sum_{i=1}^n k \left(\frac{X_i - x}{h} \right) \right] \\ &= \frac{1}{n} \left\{ \frac{1}{h^2} E \left(\left[k \left(\frac{y - x}{h} \right) \right]^2 \right) - \frac{1}{h^2} \left[E \left(k \left(\frac{y - x}{h} \right) \right) \right]^2 \right\} \\ &= \frac{1}{nh^2} \int k^2 \left(\frac{y - x}{h} \right) f(y) dy - \frac{1}{n} \left[\frac{1}{h} \int k^2 \left(\frac{y - x}{h} \right) f(y) dy \right]^2, \end{aligned}$$

on utilise un changement de variable $t = \frac{y - x}{h} \Rightarrow dt = \frac{1}{h} dy$, et on obtient :

$$\text{Var}(\widehat{f}_n(x)) = \frac{1}{nh} \int_{\mathbb{R}} k^2(t) f(x + th) dt - \frac{1}{n} \left(\int_{\mathbb{R}} k(t) f(x + th) dt \right)^2.$$

Donc d'après le développement de Taylor de f au voisinage de x

$$\text{Var}(\widehat{f}_n(x)) = \frac{1}{nh} \int_{\mathbb{R}} k^2(t) (f(x) + o(1)) dt - \frac{1}{n} \left(\int_{\mathbb{R}} k(t) (f(x) + o(1)) dt \right)^2.$$

Sous les conditions précédentes on a :

$$\text{Var}(\widehat{f}_n(x)) = \frac{1}{nh} f(x) \int_{\mathbb{R}} k^2(t) dt + o\left(\frac{1}{nh}\right),$$

si on pose $R(k) = \int_{\mathbb{R}} k^2(t) dt$ alors

$$\text{Var}(\widehat{f}_n(x)) = \frac{1}{nh} R(k) f(x) + o\left(\frac{1}{nh}\right).$$

■

Remarque 2.1

1. Si $h \rightarrow 0$; quand $n \rightarrow \infty$, alors $Biais(\hat{f}_n(x)) \rightarrow 0$.
2. Si $h \rightarrow 0$ et $nh \rightarrow \infty$; quand $n \rightarrow \infty$, alors $Var(\hat{f}_n(x)) \rightarrow 0$.
3. $\hat{f}_n \xrightarrow{P} f$ si $h \rightarrow 0$ et $nh \rightarrow \infty$.

Erreur quadratique moyenne(MSE) de \hat{f}_n

L'erreur quadratique moyenne est donne par :

$$\begin{aligned} MSE(\hat{f}_n(x)) &= \left[Biais(\hat{f}_n(x)) \right]^2 + Var(\hat{f}_n(x)) \\ &= \frac{1}{4}h^4\mu_2^2(k)(f''(x))^2 + \frac{1}{nh}f(x)R(k) + o(h^4) + o\left(\frac{1}{nh}\right). \end{aligned}$$

- $h \rightarrow 0$ et $nh \rightarrow \infty$; quand $n \rightarrow \infty$ Alors $MSE(x) \rightarrow 0 \Rightarrow \hat{f}_n \xrightarrow{m.g} f$.

L'approximation asymptotique de la MSE est donnée par :

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

Facilement on trouve que h_{opt}^* (le paramètre de lissage optimal locale, qui varie en fonction de x) qui minimise asymptotique de MSE

$$\begin{aligned} h_{opt}^* &= \arg \min AMSE(\hat{f}_n(x)) \\ &= \left(\frac{f(x)R(k)}{\mu_2^2(k)(f''(x))^2} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \end{aligned}$$

et l'erreur quadratique moyenne asymptotique associée à h_{opt}^* est :

$$AMSE_{opt}^*(\hat{f}_n(x)) = 5 \left(\frac{1}{4n}f(x)R(k) \right)^{\frac{4}{5}} \left(\left[\frac{1}{2}\mu_2(k)f''(x) \right]^2 \right)^{\frac{1}{5}}.$$

Consistance forte (La convergence presque sûre)

Dans le même article, Parzen a établi la normalité asymptotique, ainsi que la convergence uniforme en probabilité. Son travail est un outil important et a été largement développé

par plusieurs chercheurs (Devroye et Györfi (1985), Silverman (1986), Izenman (1991), Scott (1992)). Pour une étude plus détaillée, voir par exemple Bosq et Lecoutre (1987) et Tsybakov (2009). En (1976) Nadaraya a énoncé le théorème suivant sur la consistance forte de l'estimateur.

Théorème 2.3 *Si $k(\cdot)$ est à variation bornée et si pour tout $\gamma > 0$; la série $\sum_{n \in \mathbb{N}} e^{(-\gamma nh^2)}$ converge, alors*

$$\lim_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |\widehat{f}_n(t) - f(t)| = 0 \quad p.s.,$$

si et seulement si la densité f est uniformément continue. (pour une étude plus détaillée voir, Achour 2.1)

Erreur quadratique moyenne intégrée ($MISE$) :

L'erreur quadratique moyenne intégrée de l'estimateur \widehat{f}_n est donnée par :

$$\begin{aligned} MISE(\widehat{f}_n(x)) &= \int_{\mathbb{R}} MSE(\widehat{f}_n(x)) dx \\ &= \frac{1}{4} h^4 \mu_2^2(k) R(f''(x)) + \frac{1}{nh} R(k) + o(h^4) + o\left(\frac{1}{nh}\right). \end{aligned}$$

L'approximation asymptotique de la $MISE$ est donnée par :

$$AMISE(\widehat{f}_n(x)) = \frac{1}{4} h^4 \mu_2^2(k) R(f''(x)) + \frac{1}{nh} R(k).$$

La fenêtre optimale globale de $AMISE(\widehat{f}_n(x))$, notée h_{opt} , est donnée par :

$$\begin{aligned} h_{opt} &= \arg \min MISE(\widehat{f}_n(x)) \\ &= \left(\frac{R(k)}{\mu_2^2(k) R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \end{aligned} \tag{2.3}$$

L'erreur quadratique moyenne intégrée asymptotique associée à h_{opt} est :

$$AMISE_{opt} = \frac{5}{4}(\mu_2)^{\frac{2}{5}}(R(k))^{\frac{4}{5}}(R(f''))^{\frac{1}{5}}n^{-\frac{4}{5}}.$$

Choix du paramètre de lissage

1. La règle simple de référence

L'idée de cette méthode revient à Deheuvels (1977) avant d'être publiée par Silvermann (1986). Le choix du paramètre de lissage par cette méthode consiste à remplacer la partie inconnue $R(f'')$ dans l'expression de l'estimateur optimal h_{opt} , donné par l'équation (2.3), par une distribution classique afin d'obtenir un estimateur pour h . Si on choisit f comme étant la distribution normale de moyenne 0 et de variance σ^2 , on obtient

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}, \text{ avec } \sigma \in \mathbb{R}^+ \text{ et } x \in \mathbb{R}$$

$$\text{alors } f''(x) = \frac{1}{\sigma^3\sqrt{2\pi}}\left(\left(\frac{x}{\sigma}\right)^2 - 1\right)e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}.$$

La quantité inconnue $R(f'')$ s'écrit alors

$$R(f'') = \int_{\mathbb{R}} [f''(x)]^2 dx = \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}}.$$

Donc, en faisant référence à une densité de probabilité normale, l'expression du paramètre de lissage optimal asymptotique devient

$$h_{opt} = \left(\frac{8\sqrt{\pi}R(k)}{3\mu_2^2(k)}\right)^{\frac{1}{5}} \hat{\sigma}n^{-\frac{1}{5}},$$

où $\hat{\sigma}$ est un estimateur de σ (d'après Silverman(1986)), tel que

$$\hat{\sigma} = \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right]^{\frac{1}{2}}, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

1. Si k noyau Gaussien $\Rightarrow h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$.
2. Si k noyau Epanchinkov $\Rightarrow h_{opt} = 2.34\hat{\sigma}n^{-\frac{1}{5}}$.
3. Si k noyau Quartique $\Rightarrow h_{opt} = 2.78\hat{\sigma}n^{-\frac{1}{5}}$.

2. La méthode de validation croisée

La méthode de validation croisée du type moindres carrés a été introduite par Rudemo (1982) et Bowman (1984). Cette méthode permet d'obtenir un paramètre de lissage simple et attrayant. La méthode est motivée par la décomposition de l'erreur quadratique moyenne intégrée $MISE\{\hat{f}_n(\cdot)\}$ de l'estimateur à noyau.

On a

$$\hat{f}_{-i}(x) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n k_h(x - X_j),$$

est l'estimateur de la densité construit à partir de l'ensemble de points sauf le point X_i .

La quantité

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

est appelée la quantité de "validation croisée".

Le paramètre de lissage du type "validation croisée" est la valeur de h qui minimise cette quantité de validation croisée, c'est-à-dire

$$\hat{h}LSCV = \arg \min_h LSCV(h).$$

La méthode de ré-injection (plug-in)

En adoptant le critère de l'erreur quadratique moyenne Intégrée ($MISE$), Scott, Tapia et Thomson [13] choisissent d'estimer la fonction $R(f'')$ dans l'expression de h_{opt} donnée par l'équation (2.3) à l'aide de l'estimateur naturel $\hat{R}(f'')$ défini comme suit :

$$\widehat{R}(f'') = R(\widehat{f}_n''),$$

où \widehat{f}_n'' désigne la dérivée seconde de l'estimateur à noyau \widehat{f}_n . alors avec un noyau k deux fois dérivable, on a :

$$\widehat{f}_n''(x) = \frac{1}{nh^3} \sum_{i=1}^n k''\left(\frac{X_i - x}{h}\right)$$

En choisissant par exemple le noyau gaussien :

$$k(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}, \text{ si } t \in \mathbb{R}$$

L'estimateur $\widehat{R}(f'')$ s'écrit comme suit :

$$\widehat{R}(f'') = \frac{3}{8n^2 h^9 \sqrt{\pi}} \sum_{i=1}^n \sum_{j=1}^n \left[h^4 - (x_i - x_j)^2 h^2 + \frac{1}{12} (x_i - x_j)^4 \right] \exp \left[\frac{(x_i - x_j)^2}{4h^2} \right].$$

Scott, Tapia et Thomson [13] proposent finalement d'injecter l'estimateur $\widehat{R}(f'')$ dans l'expression de h_{opt} , on obtient l'estimateur de h :

$$h_{opt} = \left(\frac{R(k)}{\mu_2^2(k) \widehat{R}(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Le choix optimal du noyau

Pour mesurer l'efficacité d'un noyau symétrique on peut calculer le rapport de *AMISE* des 2 noyaux

$$0 < eff(k_1, k_2) = \frac{AMISE(k_1, n, h)}{AMISE(k_2, n, h)} \leq 1$$

Le choix des noyaux n'influe pas trop dans le cas des noyaux symétrique.

Le tableau suivant présente quelques noyaux avec l'efficacité.

noyau	Equation	$R(k)$	$\mu_2(k)$	$eff(k)$
Epanechnikov	$\frac{3}{4} (1 - t^2) 1_{(t \leq 1)}$	$\frac{3}{5}$	$\frac{1}{5}$	1
Quartique	$\frac{15}{16} (1 - t^2)^2 1_{(t \leq 1)}$	$\frac{5}{7}$	$\frac{1}{7}$	0.994
Triangulaire	$(1 - t) 1_{(t \leq 1)}$	$\frac{350}{429}$	$\frac{1}{9}$	0.986
Gaussien	$\frac{1}{\sqrt{2\pi}} e^{-t^2/2}$, si $t \in \mathbb{R}$	$\frac{1}{2\sqrt{\pi}}$	1	0.951
Uniforme	$(\frac{1}{2}) 1_{(t \leq 1)}$	$\frac{1}{2}$	$\frac{1}{3}$	0.930

TAB. 2.2 – Table du quelque noyau avec l'efficacité

Chapitre 3

Simulation

L'objectif de ce chapitre, est de vérifier par simulation l'influence du choix de paramètre de lissage et du noyau pour l'estimateur à noyau de la densité. Les simulations sont réalisées par le logiciel R. On a vu que l'estimateur à noyau de la densité dépend de la taille de l'échantillon n mais aussi de la fenêtre h et du noyau k , nous allons donc étudier les cas suivants : "choix du noyau, choix de paramètre de lissage".

Plan de simulation

Premièrement, pour l'estimation de la densité nous avons choisi de simuler les densités de probabilité de deux types de lois, la loi normale qui a une fonction de densité ($D1$) définie sur \mathbb{R} et la deuxième est la loi exponentielle (de paramètre $\lambda = 1$) dont la densité ($D2$) est définie sur un support positif $[0, +\infty[$. Les deux fonctions de densité sont définies comme suit :

$D1$. La loi normale $\mathcal{N}(0, 1)$:

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

$D2$. La loi exponentielle $\xi(1)$:

$$f(x) = \lambda \exp(-\lambda x), x \in [0, +\infty[.$$

En utilisant l'estimateur à noyau de la densité

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

Algorithme de simulation

L'algorithme de simulation que nous avons utilisé :

- Simuler un échantillon de taille n pour une densité qui appartient à la famille paramétrique comme : $D1, D2$.
- Construire l'estimateur par la méthode du noyau à partir des observations.
- Tracer les deux courbes : densité test et la densité estimée.

1. Etude du premier type de loi (La loi normale $\mathcal{N}(0, 1)$)

Choix de noyau

En choisissant $h = h_{opt} = 1.06\sigma(n^{-\frac{1}{5}})$ (avec $\sigma = 1$ est l'écart type) fixé et on utilisera quelques noyaux k (Triangulaire, Gaussien, Epanechnikov, Quartique.) et la taille de l'échantillons $n = 100, n = 200, n = 1000$.

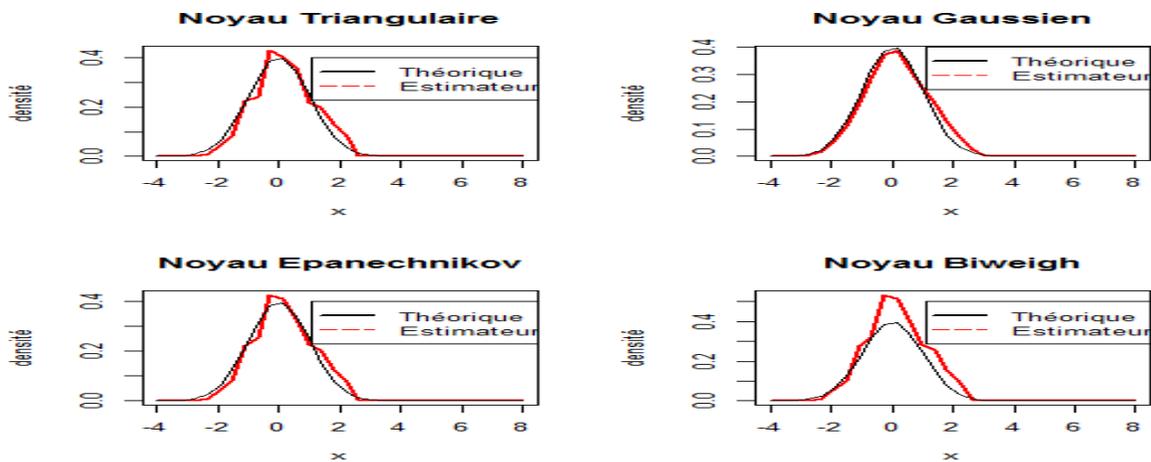


FIG. 3.1 – Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 100$.

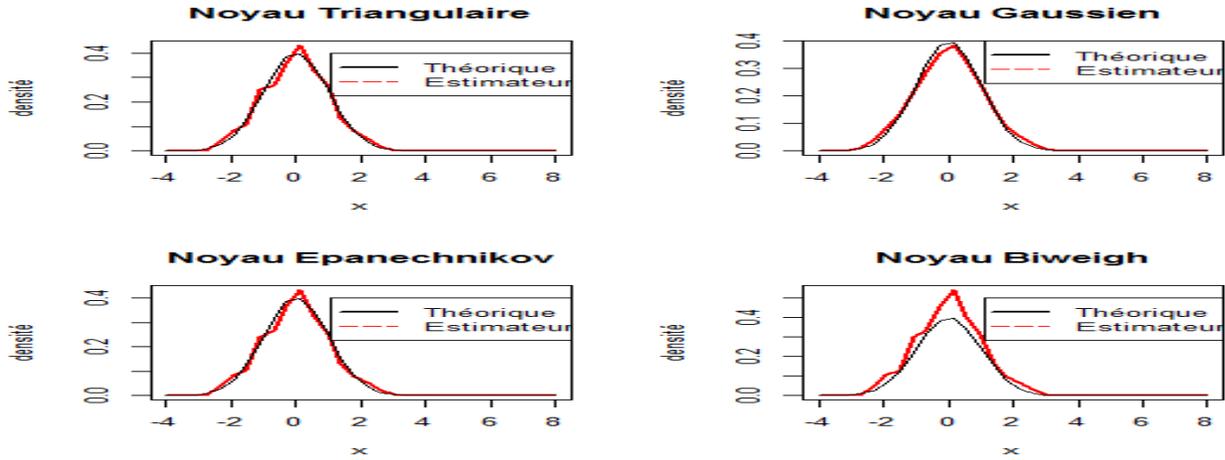


FIG. 3.2 – Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 200$.

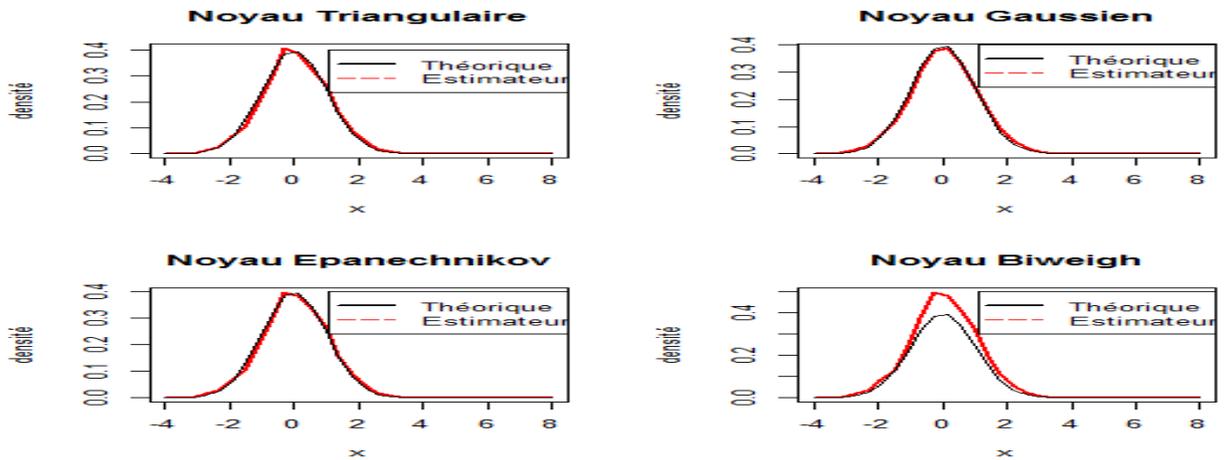


FIG. 3.3 – Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 1000$.

Remarque 3.1

1. Nous remarquons ici que l'estimateur lisse dès que n est grand pour chaque noyau utilisé.
2. Nous remarquons sur le graphe ci-dessus que quand n est grand l'estimateur \hat{f}_n est proche de la densité théorique.

Choix de paramètre de lissage

1. En choisissant $h = 0.1, h = 0.3, h = 0.5, h = 0.6$ ("h" varié) et on utilisera le noyau optimale k de Epanechnikov et la taille de l'échantillons $n = 100, n = 200, n = 1000$.

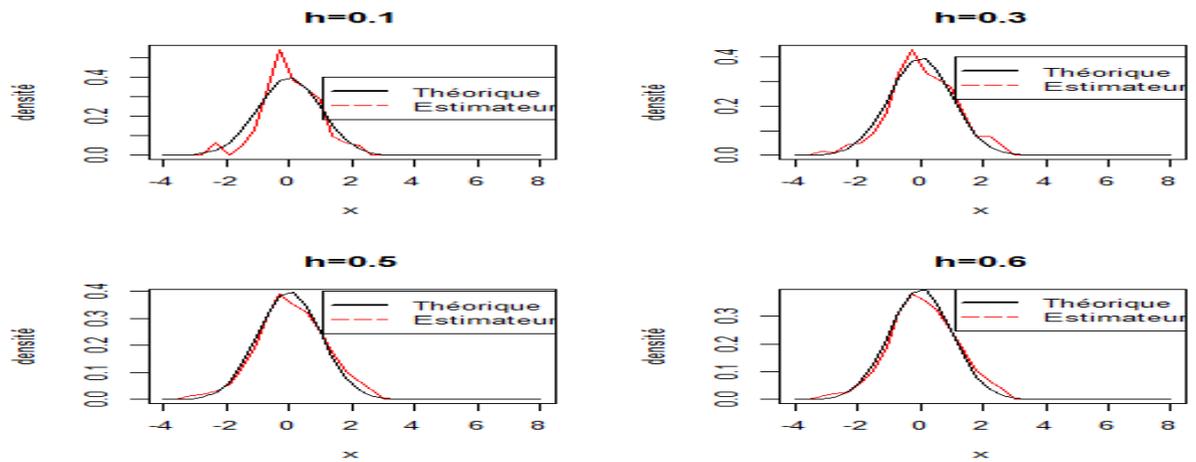


FIG. 3.4 – Comparaison entre la densité théorique $\mathcal{N}(0,1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 100$.

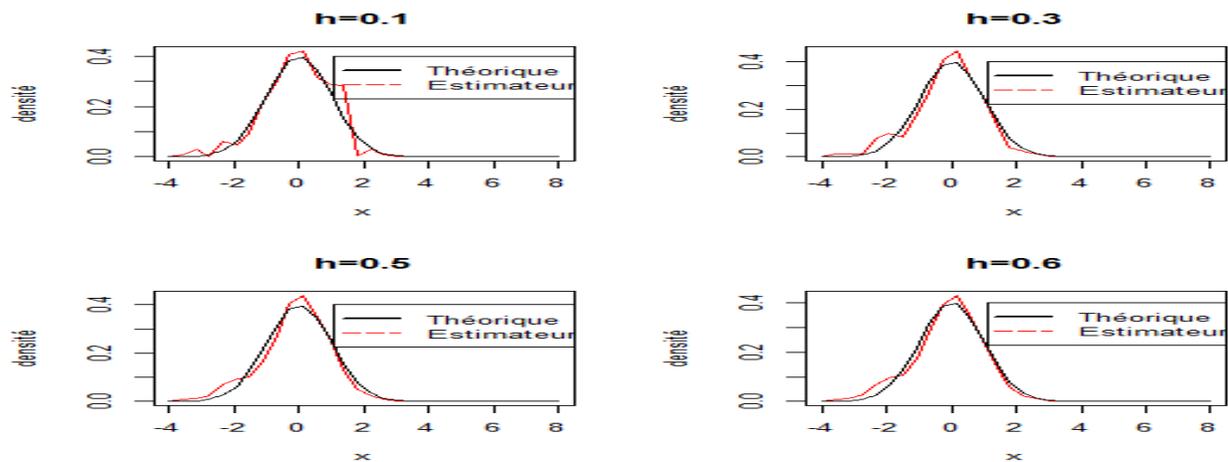


FIG. 3.5 – Comparaison entre la densité théorique $\mathcal{N}(0,1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 200$.

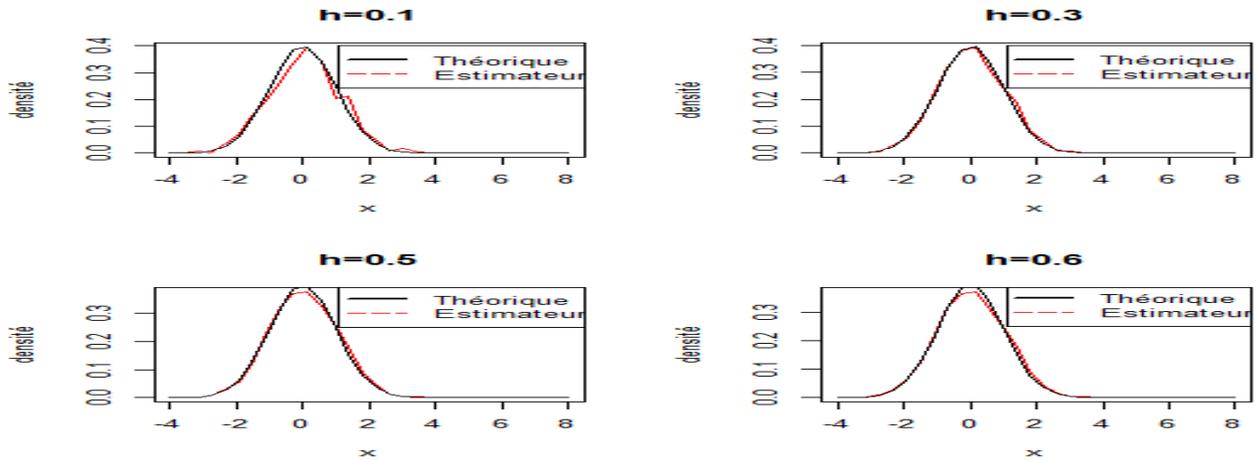


FIG. 3.6 – Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 1000$.

Remarque 3.2 Nous remarquons dans la figure précédente que le choix du paramètre de lissage a une grande influence sur la qualité de l'estimateur.

- En choisissant $h = h_{opt} = 2.34\sigma n^{-\frac{1}{5}}$ (avec $\sigma = \sqrt{Var(x)} = 1$ est l'écart type) par la règle simple de référence et on utiliserons le noyaux optimale k de Epanechnikov et la taille de l'échantillons $n = 100, n = 200, n = 1000$.

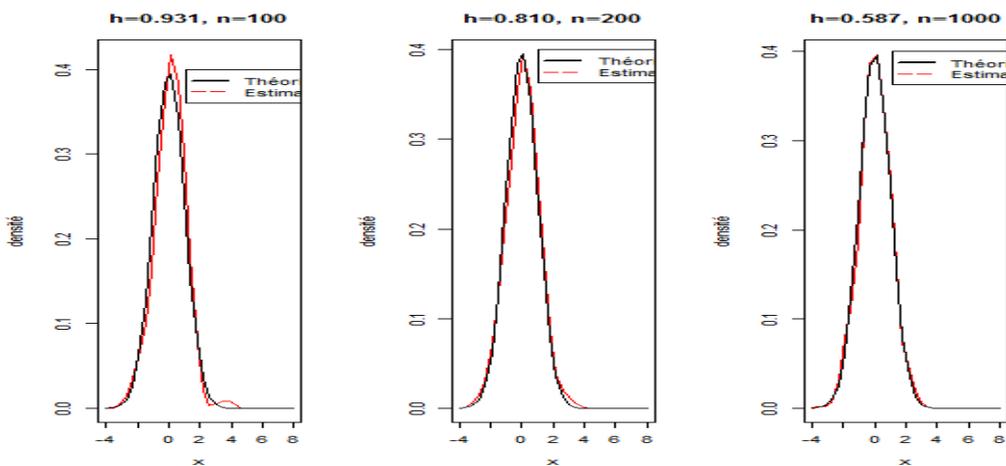


FIG. 3.7 – Comparaison entre la densité théorique $\mathcal{N}(0, 1)$ et celle estimée par un noyau d'Epanechnikov et un h "optimale" et pour n "varié"

Remarque 3.3 Nous remarquons dans la figure précédente que si on prend la fenêtre optimale pour AIMSE avec n est grand on obtient un bon estimateur de la densité.

2. Etude du deuxième type de loi (La loi exponentielle $\xi(1)$)

Choix de noyau

En choisissant $h = h_{opt} = 1.06\sigma(n^{-\frac{1}{5}})$ (avec $\sigma = \sqrt{\frac{1}{\lambda^2}} = 1$ est l'écart type) fixé et on utilisera quelques noyaux k (Triangulaire, Gaussien, Epanechnikov, Quartique.) et la taille de l'échantillons $n = 100, n = 200, n = 1000$.

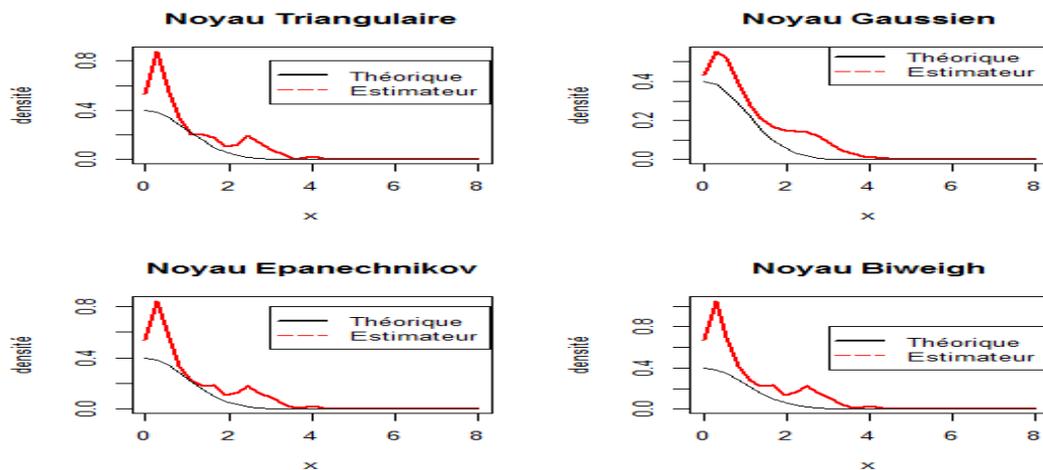


FIG. 3.8 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 100$.

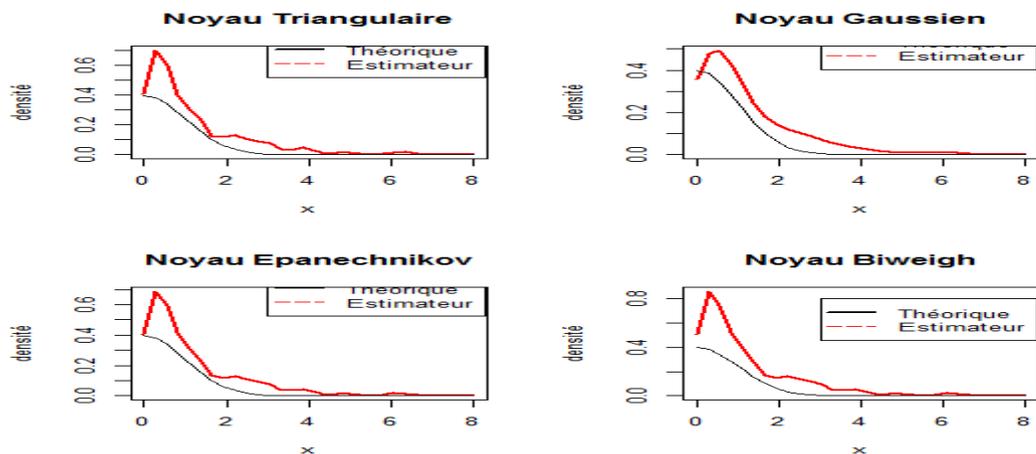


FIG. 3.9 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 200$.

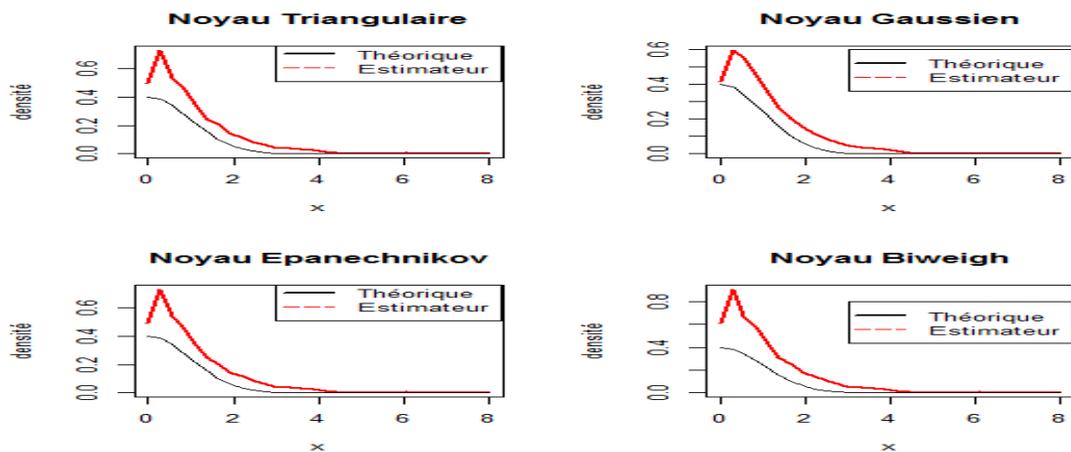


FIG. 3.10 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par des quelques noyau avec $h_{opt} = 1.06\hat{\sigma}n^{-\frac{1}{5}}$ et pour $n = 1000$.

Choix de paramètre de lissage

1. En choisissant $h = 0.1, h = 0.3, h = 0.5, h = 0.6$ ("h" varié) et on utilisera le noyau optimal k de Epanechnikov et la taille de l'échantillon $n = 100, n = 200, n = 1000$.

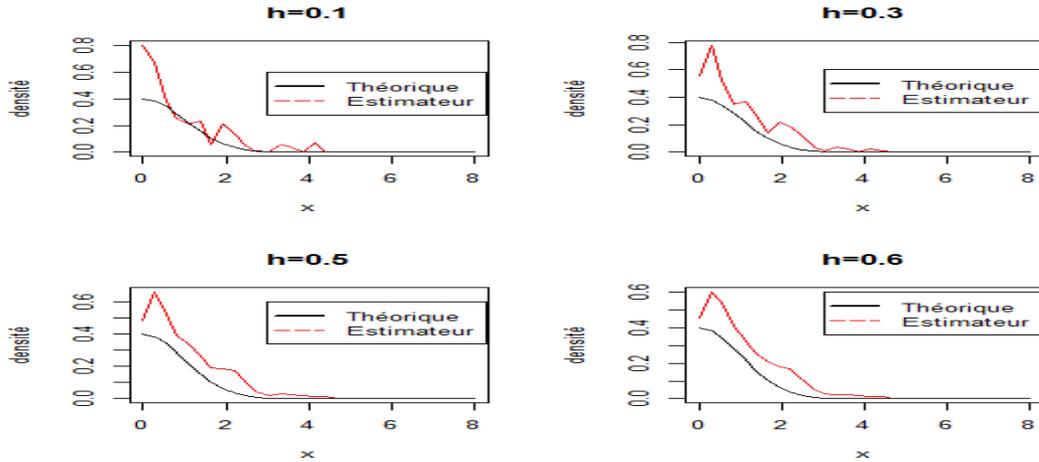


FIG. 3.11 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 100$.

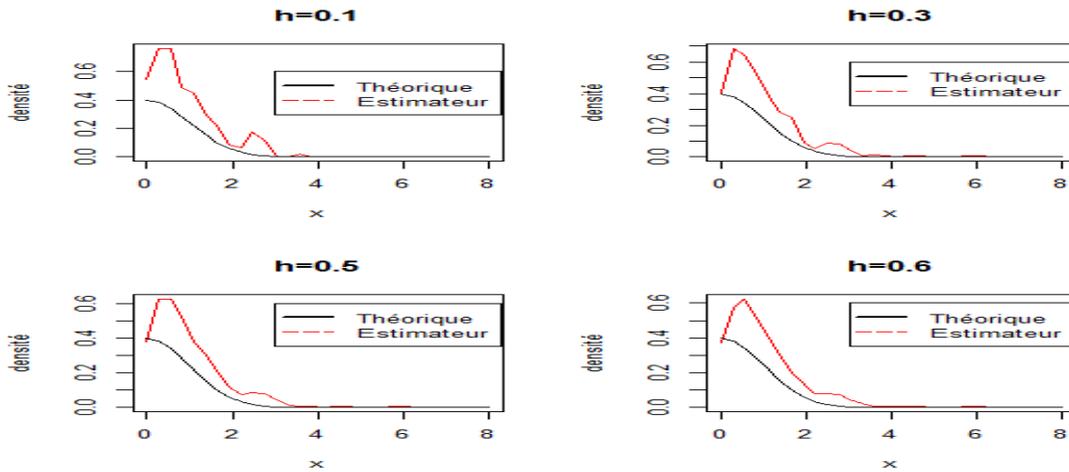


FIG. 3.12 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 200$.

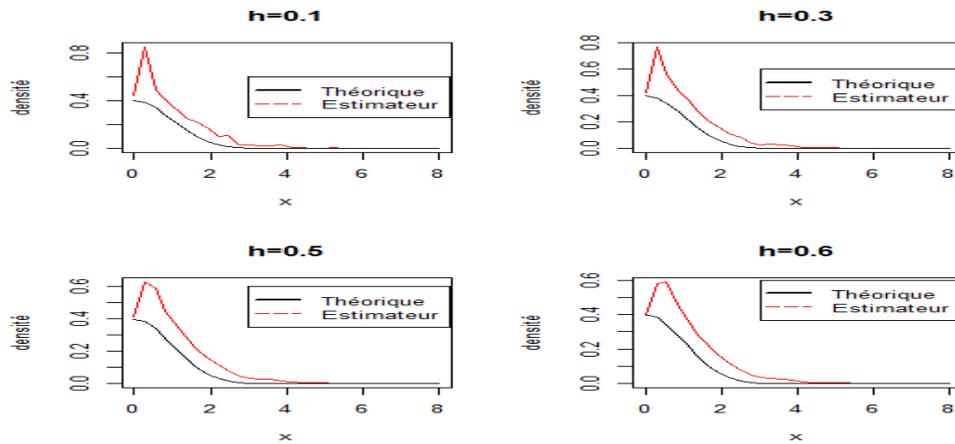


FIG. 3.13 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov avec "h" varié et pour $n = 1000$.

- En choisissant $h = h_{opt} = 2.34\sigma n^{-\frac{1}{5}}$ (avec $\sigma = 1$ est l'écart type) par la règle simple de référence et on utilisera le noyau optimal k d' Epanechnikov et la taille de l'échantillon $n = 100, n = 200, n = 1000$.

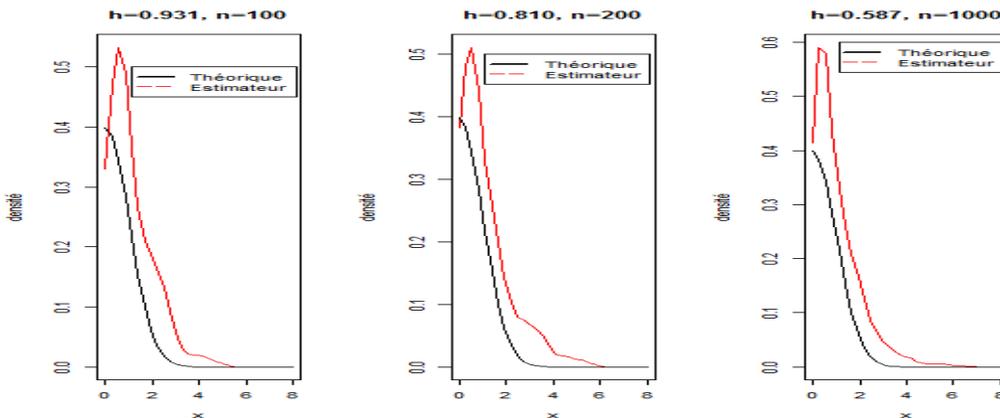


FIG. 3.14 – Comparaison entre la densité théorique $\xi(1)$ et celle estimée par un noyau d'Epanechnikov et un h "optimale" et pour n "varié"

Remarque 3.4 *Nous pouvons faire les mêmes remarques que dans l'étude de l'estimateur de la première fonction de densité.*

Conclusion de simulation

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

Conclusion

L'estimation de la fonction de densité joue un rôle central dans l'estimation fonctionnelle.

Dans ce mémoire nous avons étudié l'estimation de la densité de probabilité dans le cadre de données indépendantes et identiquement distribuées (i.i.d).

On a vu dans ce mémoire deux méthodes pour estimer la fonction de densité tel que l'estimation paramétrique et l'estimation non-paramétrique, dans cette dernière on parle sur la méthode de l'histogramme et la méthode du noyau qui est proposée par Rosenblatt en 1956 puis améliorée par Parzen en 1962 , cette méthode de noyau est la plus utilisée et basée sur une fonction k appelée noyau et une fenêtre h (paramètre de lissage) qui joue un rôle important dans la qualité de l'estimation. On a vu que l'estimateur à noyau de la densité dépend de la taille de l'échantillon n mais aussi de la fenêtre h et du noyau k .

En conclusion, nous disons que le choix du noyau n'a pas d'influence majeure sur la qualité de l'estimateur par contre le choix de la fenêtre h est crucial, ceci est illustré par les résultats de simulations.

Bibliographie

- [1] Achour, S., 2015. Le choix du noyau et de la fenêtre dans l'estimation de la densité. Université de Biskra.
- [2] Belahcene, I., 2017. Mémoire Master Estimation non paramétrique de La fonction densité de probabilité avec un noyau. Université de Ouargla
- [3] Bosq, D. and Lecoutre, J., 1987. Théorie de l'estimation fonctionnelle. Paris Economica. [13]
- [4] Dusart, P., 2015. Cours de Statistiques inférentielles. Licence 2-S4 SI-Mass.
- [5] Derrasse Chouk, B., 2015. Equivalence du choix de la fenêtre pour l'estimation des fonctions de densité et d'intensité.
- [6] Devroye, L., Györfi, L., 1985. Non parametric Density Estimation : The view. Wiley, New York.
- [7] Epanechnikov, V. A., 1969. Nonparametric Estimation of a Multivariate Probability Density. Theory of Probability and its Applications. 14 (1) : (153 – 158).
- [8] Nadaraya, E. A., 1964. On Estimating Regression. Theory of Probability and its Applications, 1964; 9 : 1 ; 141 – 142.
- [9] Nadaraya, E. A., 1976. On the nonparametric estimator of Bayesian risk in the classification problem. Proc. AN. Georg SSR, 82(2); 277280 (in Russian).
- [10] Boukhames, O., Zaidi, M., 2016. Mémoire Master Estimation Non Paramétrique De La densité De Probabilité Et De La Fonction De Régression. Université de Boumerdes.

- [11] Parzen, E., 1962. On the estimation of a probability density and mode, *Ann. Math. Statist.* 33, 1065 – 1076.
- [12] Ronsenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math.Statist.* 27, 832 – 837.
- [13] Scott, D.W., Tapia, R.A, and Thomson, J.R. Kernel density estimation revisited. *Nonlinear Analysis, Theory , Methods and Applications*, (1) : 339-372, 1977.
- [14] Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman Hall, London.
- [15] Simonoff, J.S., 1996. *Smoothing Methods in Statistics*. Springer-Verlag.
- [16] STAT., 2413, 2002 – 2003. Chapitre 3 Estimation non paramétrique d’une fonction de répartition et d’une densité.
- [17] Veysseyre, R., 2006. *Aide-mémoire Statistique et probabilités pour l’ingénieur*. Dunod.

Annexe : Abréviations et Notations

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

Notation	Signification
Ω	un ensemble fondamental.
\mathcal{F}	ensemble des événements.
P	loi de probabilité.
(Ω, \mathcal{F}, P)	espace probabilisé.
v.a.r	variable aléatoire réelle.
i.i.d	indépendant et identiquement distribués.
f	densité de probabilité.
F	fonction de répartition.
$E(X)$ ou μ	espérance mathématique ou moyenne de X .
$Var(X)$ ou σ^2	variance de X .
\mathbb{R}	ensemble des nombres réels.
\mathbb{N}	ensemble des nombres naturel.
$\mathcal{N}(\mu, \sigma^2)$	loi normale (ou de Gauss) à deux paramètre $\mu \in \mathbb{R}$ et $\sigma^2 > 0$.
$\mathcal{N}(0, 1)$	loi normale standard (centrée réduite).
$\xi(\lambda)$	loi exponentielle de paramètre λ .
\xrightarrow{P}	convergence en probabilité.
\xrightarrow{L}	Convergence en loi.

Notation	Signification
$\xrightarrow{m.q}$	Convergence en moyenne quadratique.
$\xrightarrow{p.s}$	convergence presque sûre.
θ	paramètre inconnu.
Θ	ensemble des valeurs possibles du paramètre θ .
T	un estimateur.
\bar{X}	estimateur de la moyenne μ .
S^2	estimateur sans biais de la variance σ^2 .
\tilde{S}^2	estimateur asymptotiquement sans biais de la variance σ^2 .
e_{\min} (resp. e_{\max})	le plus petit (resp. plus grand) élément de Ω .
C_k	classe.
p	le nombre de classe.
ν_k	Le nombre d'observations appartenant à chaque classe C_k .
h	fenêtre.
n	Le nombre d'observations.
\tilde{f}_h	estimateur de la densité f par l'histogramme.
$O(\cdot)$	terme résiduel.
f' (resp. f'')	la dérivée de f (resp. f').
MSE	erreur quadratique moyenne.
$MISE$	erreur quadratique moyenne intégrée.
$AMISE$	erreur quadratique moyenne intégrée asymptotique.
k	noyau.
h_{opt}	paramètre de lissage (h) optimale.
eff	efficacité.
\hat{f}_n	estimateur de la densité f par noyau.
\bar{f}	estimateur paramétrique de la densité f .

Résumé

Dans ce mémoire, on s'intéresse à l'estimation de la fonction de la densité de probabilité. Dans un premier temps, on fait un rappel sur l'estimation paramétrique excitante. Puis on mentionne quelques méthodes de l'estimation non paramétrique de la fonction de densité et en particulier la méthode du noyau qui est basée sur le choix d'un noyau k et d'un paramètre de lissage h (fenêtre).

Finalement, on a utilisé le logiciel R, pour réaliser des simulations, qui nous permettent d'observer l'influence cruciale du paramètre de lissage h dans l'estimation à noyau par contre le choix du noyau a un petit rôle dans la qualité de l'estimation.

ملخص

في هذه المذكرة، نحن مهتمون بتقدير دالة كثافة الاحتمال. في البداية نقوم بالتذكير بطريقة التقدير الوسيطي. من ثم التقدير الغير الوسيطي بحيث نتطرق لبعض الطرق و من بين هذه الطرق الأكثر فعالية هي طريقة النواة التي تتميز باختيار النواة k و وسيط التنعيم h .

أخيرا. استخدمنا البرنامج R، لإجراء عمليات المحاكاة، والتي تسمح لنا بمراقبة التأثير الفعال لمعامل التنعيم h في التقدير الغير الوسيطي بطريقة النواة، ومن ناحية أخرى، فإن اختيار النواة له دور صغير في جودة التقدير.

Abstract

In this dissertation, we are interested in estimating the function of the probability density. First, we do a reminder on the exciting parametric estimation. Then we mention a few methods of nonparametric estimation of the density function and in particular the kernel method which is based on the choice of a kernel k and a smoothing parameter h (window).

Finally, we used the software R, to carry out simulations, which allow us to observe the crucial influence of the smoothing parameter h in the kernel estimation on the other hand the choice of the kernel has a small role in the quality of the estimate.