

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

Université Abderahmane Mira de Béjaïa
Faculté des Sciences Exactes
Département de Recherche Opérationnelle



Mémoire de Fin d'Études
Pour l'obtention du diplôme de Master
Option : Modélisation Mathématique et Techniques de Décision

Thème

Estimation non paramétrique de la fonction densité par la méthode du noyau.

Présenté par :

M^r BOUTI Aniss
M^r RAMDANI Bachir

Soutenu publiquement le 06/07/2019 devant le jury :

M ^{me} DJERROUD Lamia	M.C.B	Univ. de Bejaia	Président
M ^{me} AMROUN Sonia	M.A.A	Univ. de Bejaia	Encadreur
M ^{me} ZIANE Yasmina	M.C.B	Univ. de Bejaia	Co-Encadreur
M ^{me} HARFOUCHE Lynda	Docteur	Univ. de Bejaia	Examinatrice
M ^{elle} MAKHLOUFI Cilya	Doctorante	Univ. de Bejaia	Examinatrice

2018-2019.

Remerciements

Au terme de ce travail qui marque la fin du cycle de notre formation au sein de l'université ABDERAHMANE MIRA, il nous est opportun d'exprimer notre gratitude à tous ceux qui, de loin ou de près, ont matériellement ou moralement contribué à la réalisation de notre modeste travail. Qu'ils trouvent ici l'expression de notre considération.

a notre Dieu, qui nous a donné la vie, l'intelligence et le courage de réaliser ce travail.

A nos chers parents par leur affection et amour de nous avoir donné la vie et l'éducation. Voilà aujourd'hui nous sommes comptés parmi les hommes intellectuels du monde. Qu'ils se réjouissent du fruit de leur progéniture.

Nous exprimons nos vifs remerciements, notre profonde gratitude et notre reconnaissance à notre encadreur **M^{me} AMROUN Sonia** et notre co-encadreur **M^{me} ZIANE Yasmina**, qui a dirigé ce travail.

Nous tenons également à remercier les membres du jury d'avoir accepté de juger notre travail.

Enfin, Nous remercions, de tout coeur, tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Dédicaces

Je dedie ce modeste travail tout d'abord à mes très chers parents qui nous ont soutenu tout le long de mon parcours d'études ;

À mes très chers parents ;

À mes sœurs ;Hasna et Oumaima

À mes ami(e)s : Bachir-mitcho-celia-walid-fraoussen-El
hech-akram-kamal-kader-abdesslam-aziz-saif-mouha-lamia... ;

À toute la famille RO ;

À toute la promotion 2018-2019 ;

M^r Bouti ANISS

Dédicaces

Je dedie ce modeste travail tout d'abord à mes très chers parents qui nous ont soutenu tout le long de mon parcours d'études;

À mes très cher père Hammo ;

À mes frères et sœurs ;

À toute la famille RAMDANI;

À mes amies ;

À toute la famille RO ;

À toute la promotion 2018-2019 ;

M^r Ramdani BACHIR

TABLE DES MATIÈRES

Liste des figures	8
Liste des tableaux	9
Introduction générale	10
1 Estimation non paramérique de la densité de probabilité	13
1.1 Introduction	13
1.2 Définitions	13
1.2.1 Espérance mathématique	13
1.2.2 Variance	14
1.2.3 Moment et moment centré d'ordre k	14
1.2.4 Biais	14
1.2.5 Intervalle de confiance	14
1.2.6 Estimateur sans biais	14
1.2.7 Erreurs quadratique moyenne (MSE)	15
1.2.8 Erreurs quadratique moyenne Intégrée	15
1.2.9 Erreurs quadratique Intégrée	15
1.2.10 Trace d'une matrice carrée	15
1.3 Estimateur par histogramme	15
1.3.1 proposition	16
1.3.2 Propriétés de l'estimateur	16
1.3.3 Choix du paramètre de lissage	17
1.4 Estimation par les fonctions orthogonales	17
1.4.1 Principe de la méthode	17
1.4.2 Propriétés statistiques de l'estimateur	18
1.4.3 Choix du paramètre de lissage	19
1.5 Estimation par la méthode du noyau	19
1.5.1 Définition	19
1.5.2 Propriétés de l'estimateur à noyau	20

1.6	Estimation par la méthode des fonctions Splines	20
1.6.1	Les fonctions splines	20
1.6.2	Estimation de la fonction de répartition avec les fonctions splines	21
1.6.3	Choix du paramètre de Lissage : Validation croisée généra- lisée (GCV)	23
1.7	Conclusion	23
2	Estimation non paramétrique de la fonction densité par la mé- thode du noyau	24
2.1	Introduction	24
2.2	Estimateur de Rosenblat(1956)	24
2.3	Estimateur de parzen (1962)	26
2.4	Estimateur à noyau	27
2.5	Noyaux symétriques	27
2.5.1	Quelque noyaux usuelles symétriques	27
2.6	Noyau asymétrique	31
2.6.1	Noyau Bêta	32
2.6.2	Noyaux Gamma	32
2.7	Propriétés de l'estimateur à noyau	32
2.7.1	Espérance mathématique	33
2.7.2	Variance de l'estimateur	34
2.7.3	Biais de l'estimateur	35
2.7.4	Erreur quadratique moyenne MSE	35
2.7.5	Erreur moyenne quadratique intégrée (MISE)	35
2.7.6	Comportement asymptotique	36
2.7.7	Intervalle de confiance	36
2.8	Critères de convergence	38
2.8.1	Convergence en moyenne quadratique	38
2.8.2	Convergence uniforme en probabilité	38
2.8.3	Convergence uniforme presque complète	39
2.8.4	Convergence en loi	39
2.9	Choix de noyau	40
2.10	Choix de la fenêtre	40
2.10.1	Méthodes plug-in	40
2.10.2	Méthodes de Validation Croisée (Cross Validation)	41
2.11	Conclusion	43
3	Application numérique	44
3.1	Introduction	44
3.2	Méthode plug-in	44
3.2.1	Algorithme de la méthode plug-in(Rull Of Thumb)	47
3.3	Validation croisée non biaisée	48
3.3.1	Algorithme de la méthode validation croisée non biaisée	50
3.4	Simulation	51

3.4.1	Plan de simulation	51
3.4.2	Résultats de simulation	53
3.5	Interprétation des résultats	58
3.5.1	Loi normale centrée réduite	58
3.5.2	Mélange de populations gaussiennes	58
3.6	Conclusion	59
	Conclusion générale	60
	BIBLIOGRAPHIE	61

TABLE DES FIGURES

2.1	Noyau Uniforme.	28
2.2	Noyau Triangulaire.	28
2.3	Noyau Gaussien.	29
2.4	Noyau Biweight.	30
2.5	Noyau Epanechnikov.	30
2.6	Noyau Sinus.	31
3.1	Loi normale $N(0, 1)$	52
3.2	Mélange de la normal $\frac{1}{2}N(1, \frac{16}{49}) + \frac{1}{2}N(-1, \frac{16}{49})$	52
3.3	Mélange de la normal $\frac{3}{4}N(0, 1) + \frac{1}{4}N(1.5, \frac{1}{9})$	52
3.4	Comparaison entre la densité théorique D1 et celle estimée avec h_{rot}	55
3.5	Comparaison entre la densité théorique D1 et celle estimée avec h_{ucv}	55
3.6	Comparaison entre la densité théorique D1 et celle estimée avec h_{rot}	55
3.7	Comparaison entre la densité théorique D1 et celle estimée avec h_{ucv}	55
3.8	Comparaison entre la densité théorique D2 et celle estimée avec h_{rot}	56
3.9	Comparaison entre la densité théorique D2 et celle estimée avec h_{ucv}	56
3.10	Comparaison entre la densité théorique D2 et celle estimée avec h_{rot}	56
3.11	Comparaison entre la densité théorique D2 et celle estimée avec h_{ucv}	56
3.12	Comparaison entre la densité théorique D3 et celle estimée avec h_{rot}	57
3.13	Comparaison entre la densité théorique D3 et celle estimée avec h_{ucv}	57

3.14	Comparaison entre la densité théorique D3 et celle estimée avec h_{rot}	57
3.15	Comparaison entre la densité théorique D3 et celle estimée avec h_{ucv}	57

LISTE DES TABLEAUX

3.1	Résultats des simulations effectuées sur la loi D1.	54
3.2	Résultats des simulations effectuées sur la loi D2.	54
3.3	Résultats des simulations effectuées sur la loi D3.	54

INTRODUCTION GÉNÉRALE

La théorie de l'estimation est une des préoccupations majeures des statisticiens. Cette théorie est habituellement divisée en deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. Le problème de l'estimation non paramétrique consiste, dans la majeure partie des cas, à estimer, à partir des observations, une fonction inconnue élément d'une certaine classe fonctionnelle. Plus particulièrement, on parle d'estimation non paramétrique lorsque celle-ci ne se ramène pas à l'estimation d'un nombre fini de paramètres réels associés à la loi de l'échantillon, c-à-d laisser les données parler d'elles même.

L'un des problèmes les plus fréquemment rencontrés en statistique paramétrique ou non paramétrique est celui de l'estimation de la fonction densité. Il s'agit d'un problème fondamental qui a connu durant ces quarante dernières années, des développements théoriques et pratiques à la fois rapides et nombreux. Le problème de l'estimation de la densité de probabilité est intéressant pour plusieurs raisons. D'une part, la fonction densité caractérise la répartition des données fournies par un échantillon (c'est une des premières quantités qu'un praticien peut étudier). D'autre part, on pourra facilement avoir, grâce à une estimation de la densité, des informations sur la symétrie ou la multimodalité de la loi de l'échantillon. De même, elle est une aide importante au choix d'un modèle approprié pour la loi de l'échantillon.

Il existe un certain nombre de méthodes d'estimation non paramétrique qui permettent d'estimer la fonction densité. La première méthode introduite à ce sujet, est l'estimation par histogramme. Son principe est de regrouper les données dans des classes, puis, représenter les fréquences empiriques par des aires rectangulaires.

D'autres méthodes plus robustes ont été proposées pour l'estimation de la fonction densité. Par exemple on peut citer la méthode de la fonction orthogonale proposée par (Kronmal tarter) [23] et la méthode des fonctions splines proposée par (Restle (2006)) [32]. La méthode la plus rencontrée, dans la littérature, est la méthode du noyau proposée par Rosenblatt(1956)[31], et Parzen(1962) [33]. C'est une méthode très utilisée vu sa souplesse d'utilisation et ses propriétés de

convergence. L'estimateurs à noyau est une fonction de deux paramètres K , appelé noyau, et h dit paramètre de lissage (largeur de fenêtre). Rosenblatt reprenait l'idée de Fix et Hodges en 1951, qui consistait à estimer la densité en un point, en comptant le nombre d'observations situées dans l'intervalle de longueur $2h$ et centré en ce point. Son expression théorique est extrêmement simple puisque l'estimateur s'écrit sous la somme de n variables aléatoires indépendantes et identiquement distribuées. L'estimateur à noyau converge vers la vraie densité en de nombreux sens, en particulier au sens L_1 . D'autres part, il est convergent dans tous les modes : en probabilité, en moyenne, presque sûr et presque complète. Cet estimateur est flexible, dans la mesure où il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau, mais encore dans le choix du paramètre de lissage. Les propriétés de convergence de l'estimateur à noyau ont été établies par Parzen [31], Silverman [44] et Nadaraya [28]. Les théorèmes relatifs à l'erreur moyenne quadratique et l'erreur moyenne quadratique intégrée ont été obtenus sous forme élémentaire par Parzen [31].

Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Actuellement, il n'existe pas de choix optimal pour ce paramètre de lissage. Le choix optimal qui minimise l'erreur relative globale (*MISE*) dépend de la dérivée seconde de la densité inconnue. Les auteurs se sont alors attachés à introduire des procédures de sélection automatiques et donc moins subjectives que le simple choix à l'oeil. L'étude de ce problème a nourri une littérature abondante, notamment vers le milieu des années quatre-vingt.

Parmi les objectifs de ce mémoire est de faire le point sur les procédures de sélection du paramètre de lissage h . Plusieurs méthodes ont été proposées dans la littérature. On rencontre une classe de méthode dite méthode plug-in (ré-injection) : la méthode plug-in itérée proposée par Scott, Tapia et Thompson [39], la méthode rule of thumb proposée par Silverman [44], la méthode surlissage (oversmoothing) proposée par Scott et Terrell [40], la méthode plug-in itéré moderne proposée par Park et Marron [30] et la méthode de Sheather et Jones [43]. Le principe de ces méthodes repose sur l'estimation d'une quantité qui dépend de la dérivée seconde de la densité de probabilité inconnue que l'on désire à estimer. Une autre classe de méthodes rencontrées est la classe des méthodes reposant sur la validation croisée. D'un point de vue pratique, un des principaux intérêts de ces méthodes est leurs caractérisation directe. Contrairement aux méthodes plug-in, le paramètre de lissage estimé dépend des observations. Habbema, Hermans et Vandenbroek [20] ont proposé une méthode fondée sur un critère non asymptotique du maximum de vraisemblance. Une autre méthode dite, validation croisée non biaisée, a été proposée par Rudemo [36] et Bowman [5]. Mais cette méthode possède de nombreux défauts (convergence très lente vers paramètre de lissage l'optimal,...). Pour remédier aux problèmes de cette méthode, deux autres approches ont été développées, la validation croisée biaisée proposée par Scott et Terrell [40] et la validation croisée lissée introduite par Hall, Marron et Park [27]. Parmi les objectifs de ce mémoire est de comparer deux méthodes de sélection du

paramètre de lissage, à savoir, la méthode plug-in et la méthode validation croisée sans biais (UCV). En se basant sur le critère d'erreur moyenne quadratique (MSE), nous comparons théoriquement les erreurs associées à chaque méthode et graphiquement les estimateurs donnés par chaque méthode sur plusieurs distributions connues.

Elles diffèrent au niveau du choix du critère à optimiser. Le mémoire est structuré de la manière suivante :

1. Le premier chapitre, nous présente les différentes méthodes non paramétriques d'estimation de la fonction densité de probabilité, à savoir, méthode des histogrammes, la méthode des fonctions orthogonales, la méthode des splines, la méthode du noyau.
2. Le deuxième chapitre, est consacré pour la méthode du noyau, le principe de la méthode, la forme de l'estimateur à noyau et ces différentes propriétés, ainsi que les différentes méthodes pour le choix de paramètre de lissage.
3. Le troisième chapitre, s'intéresse principalement aux deux méthodes de choix de paramètre de lissage (Rule Of Thumb et Validation croisée non biaisée), une étude comparative est réalisée sur des densités cibles.

Nous terminerons ce mémoire par une conclusion générale.

CHAPITRE 1

ESTIMATION NON PARAMÉRIQUE DE LA DENSITÉ DE PROBABILITÉ

1.1 Introduction

Dans de nombreuses applications, la densité f est inconnue et on dispose d'un n observations x_1, x_2, \dots, x_n , issues d'une variable aléatoire réelle X de densité de probabilité $f(x)$ inconnue. Comment obtenir un estimateur 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, à savoir :

1. L'estimateur par histogramme.
2. L'estimateur par la méthode du noyau.
3. L'estimateur par les fonctions orthogonales.
4. L'estimateur par les splines.

1.2 Définitions

1.2.1 Espérance mathématique

Soit une variable aléatoire X absolument continue de densité de probabilité $f(x)$ définie sur $[a, b]$. L'espérance mathématique de X est définie par :

$$E(X) = \int_a^b x f(x) dx. \quad (1.1)$$

On a la formule suivante pour une variable aléatoire discrète X de loi de probabilité (p_k) :

$$E(X) = \sum_{i=1}^{+\infty} p_i \times x_i. \quad (1.2)$$

1.2.2 Variance

La variance d'une variable aléatoire X absolument continue est définie par :

$$V(X) = E(X^2) - [E(X)]^2. \quad (1.3)$$

Si X une variable aléatoire discrète de loi (p_i) . La variance de X se calcule ainsi :

$$V(X) = \sum_{i=1}^{+\infty} p_i \times x_i^2 - E(X)^2. \quad (1.4)$$

1.2.3 Moment et moment centré d'ordre k

On appelle moment d'ordre k d'une variable aléatoire X , le nombre m_k défini par :

$$m_k = E(X^k) = \int x^k f(x) dx. \quad (1.5)$$

Et on appelle moment centré d'ordre k d'une variable aléatoire X , le nombre μ_k défini par :

$$\mu_k = E(X - E(X))^k = \int (x - E(X))^k f(x) dx. \quad (1.6)$$

1.2.4 Biais

Un biais est une démarche ou un procédé qui engendre des erreurs dans les résultats d'une étude. Si $\hat{\theta}$ est l'estimateur de θ , le biais est défini par :

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

1.2.5 Intervalle de confiance

Un intervalle de confiance de seuil (ou niveau de signification) $\alpha \in [0, 1]$ pour un paramètre θ , est un intervalle aléatoire I tel que $P(\theta \in I) = 1 - \alpha$.

1.2.6 Estimateur sans biais

Un estimateur $\hat{\theta}$ est dit sans biais lorsque son espérance mathématique est égale au paramètre.

$$E(\hat{\theta}) = \theta. \quad (1.7)$$

Un estimateur $\hat{\theta}$ est dit asymptotiquement sans biais si le biais diminue avec l'augmentation de la taille de l'échantillon :

$$\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta. \quad (1.8)$$

1.2.7 Erreurs quadratique moyenne (MSE)

L'évaluation de la similarité entre l'estimateur \hat{f}_h et la vraie densité f à estimer, nécessite des critères d'erreur. On définit l'erreur moyenne quadratique (Mean Squared Error MSE) comme suite :

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

1.2.8 Erreurs quadratique moyenne Intégrée

Une mesure globale de l'efficacité de l'estimateur \hat{f}_h est obtenue en intégrant le MSE. Il s'agit de l'Erreur Quadratique Moyenne Intégrée en anglais Mean Integrated Squared Error (MISE), elle s'écrit :

$$\begin{aligned} MISE(\hat{f}_h) &= \int MSE(\hat{f}_h) dx \\ &= \int V(\hat{f}_h(x)) dx + \int Biais^2(\hat{f}_h(x)) dx \end{aligned}$$

1.2.9 Erreurs quadratique Intégrée

l'intégrale des erreurs quadratiques (ISE) est défini par :

$$ISE(\hat{f}_h) = \int [\hat{f}_h(x) - f(x)]^2 dx \quad (1.9)$$

1.2.10 Trace d'une matrice carrée

La trace d'une matrice $A = (a_{ij})_{1 \leq i, j \leq n}$ notée par $Tr(A)$ est donnée par :

$$Tr(A) = \sum_{i=1}^n a_{ii}. \quad (1.10)$$

1.3 Estimateur par histogramme

La méthode d'estimateur par histogramme consiste à estimer la densité de la variable aléatoire X en x par n_i le nombre d'occurrences de réalisations x_i appartenant à la i ème classe associée à la valeur x .

Etant données les observations (x_1, x_2, \dots, x_n) réalisations des variables aléatoires réelles indépendantes et identiquement distribuées X_1, X_2, \dots, X_n de densité f inconnue sur un intervalle fini $[a; b[$. On considère les sous intervalles (classe) de l'intervalle $[a, b[$: $(I_j)_{j=0,1,\dots,k-1}$ avec $I_j = [a_j, a_{j+1}[$.

Pour construire l'estimateur par histogramme nous devons choisir une origine

x_0 et une largeur d'intervalle h . la largeur contrôle principalement la qualité de lissage. Supposons que la droite est coupée en intervalles, l'estimateur de f sur $[a_j; a_{j+1}[$ est défini par :

$$\hat{f}_h(x) = \frac{1}{nh} \times \text{card}\{i, x_i \in [a_j, a_{j+1}[\} = \frac{n_j}{nh}, x \in [a, b[\quad (1.11)$$

Remarque :

la construction de $\hat{f}_h(x)$ peut être faite même si les données ne proviennent pas d'une loi continue.

1.3.1 proposition

D'après [18], on a

$$n_j \xrightarrow{\text{loi}} \beta(n, p_j). \quad (1.12)$$

où β est la loi binomiale de paramètre n et $p_j = P(a_j < x \leq a_{j+1})$ pour $x \in [a_j; a_{j+1}[$.

1.3.2 Propriétés de l'estimateur

En faisant le développement de Taylor à l'ordre 1, on obtient les propriétés suivantes (voir,[6]) :

1. L'espérance est donnée par :

$$E(\hat{f}_h(x)) = \frac{E(n_j)}{nh} = \frac{p_j}{h}. \quad (1.13)$$

2. La variance est donnée par :

$$\text{Var}[\hat{f}_h(x)] = \frac{f(x)}{nh} + O(n^{-1}) = \frac{p_j(1-p_j)}{nh^2}. \quad (1.14)$$

3. Le biais de l'estimateur est donné par :

$$\text{Biais}(\hat{f}_h(x)) = \frac{f'}{2}[h - 2(x - a_j)] + O(h^2). \quad (1.15)$$

4. L'erreur Quadratique Moyenne est donnée par :

$$\text{MSE}(\hat{f}_h(x)) = \frac{f(x)}{nh} + \frac{f'^2(x)}{4}[h - 2(x - a_j)]^2 + O(h^3) + O(n^{-1}). \quad (1.16)$$

5. L'erreur Quadratique Moyenne Intégrée est donnée par :

$$\text{MISE}(\hat{f}_h(x)) = \frac{1}{nh} \frac{h^2 \int f'^2(t) dt}{12} + O(h^3) + O(n^{-1}). \quad (1.17)$$

1.3.3 Choix du paramètre de lissage

Règle de sturges [37] : Prendre le nombre de k de classes égal à $1+3.322\log_{10}n$.
En pratique, cela revient à prendre (voir,[1]) :

$$h = \frac{x_{(k)} - x_{(1)}}{k}. \quad (1.18)$$

Où les $x_{(i)}$ sont des observations d'un échantillon ordonné par ordre croissant. la règle de Sturges a tendance à produire des histogrammes trop lisses.

Règle de Scott [38] : Le choix de h est pris de telle sorte qu'il minimise l'erreur moyenne quadratique intégrée, c-à-d :

$$h_{opt} = \operatorname{argmin} MISE = \frac{6}{\int f'^2(t)dt}. \quad (1.19)$$

Exemple :

Soit X une variable aléatoire $X \rightarrow N(\mu; \sigma^2)$ de densité f , d'après Scott [38], le paramètre de lissage optimale est donnée par :

$$h_{opt} = 3.491\sigma n^{-\frac{1}{3}}. \quad (1.20)$$

En estimant σ par l'écart type S de l'échantillon, on aura :

$$h_{opt} = 3.491S n^{-\frac{1}{3}}. \quad (1.21)$$

1.4 Estimation par les fonctions orthogonales

1.4.1 Principe de la méthode

Soit X_1, X_2, \dots, X_n une suite des variables aléatoires indépendantes et identiquement distribuées de densité de probabilité f sur \mathbb{R} .

On veut estimer f à partir des observations x_1, x_2, \dots, x_n . Pour cela on suppose que (voir,[11]) :

1. L'espace de Hilbert L^2 est de dimension infinie ;
2. $\{e_k, k \in \mathbb{N}\}$ est une base orthogonale dans L^2 ;
3. $f \in L^2$ tel que :

$$f(x) = \sum_{k=0}^{\infty} a_k e_k(x), x \in \mathbb{R}. \quad (1.22)$$

4. Le développement à l'ordre d_n de $f(x)$ est :

$$f_{d_n}(x) = \sum_{k=0}^{d_n} a_k e_k(x), x \in \mathbb{R}. \quad (1.23)$$

avec (d_n) est une suite d'entiers qui tend vers $+\infty$ lorsque $n \rightarrow \infty$ et $(a_k)_{k \in \mathbb{N}}$ sont les coefficients de Fourier associés à f donnés par :

$$a_k = \int_R e_k(x) f(x) dx = E[e_k(x)]. \quad (1.24)$$

Pour estimer $f(x)$ dans L^2 on se propose de construire un estimateur sans biais de $\hat{f}_{d_n}(x)$. Par la méthode des moments, on peut estimer les coefficients $\{a_k; k \in \mathbb{N}\}$ par :

$$\hat{a}_k = \frac{1}{n} \sum_{i=1}^n e_k(x_i). \quad (1.25)$$

Ainsi $f_{d_n}(x)$ peut être estimée par :

$$\hat{f}_{d_n}(x) = \sum_{k=0}^{d_n} \hat{a}_k e_k(x). \quad (1.26)$$

1.4.2 Propriétés statistiques de l'estimateur

1. Biais de l'estimateur :

$$\begin{aligned} \text{Biais}(\hat{f}_{d_n}(x)) &= E(\hat{f}_{d_n}(x)) - f(x), \\ &= \sum_{k=0}^{d_n} \hat{a}_k e_k(x) - \sum_{k=0}^{\infty} a_k e_k(x), \\ &= - \sum_{k=d_n+1}^{\infty} a_k e_k(x). \end{aligned}$$

2. Variance de l'estimateur :

$$\begin{aligned} \text{Var}(\hat{f}_{d_n}(x)) &= \\ &= \frac{d_n}{2} + \sum_{k=1}^{d_n} \cos(2\pi kx) \left(\frac{1}{n} + \frac{1}{\sqrt{2n}} a_{2k} - \frac{1}{n} a_k^2 \right) + \sum_{k=1}^{d_n} \sum_{j=1}^{d_n} \frac{1}{n} \left[\frac{1}{\sqrt{2}} a_{k+j} \frac{1}{\sqrt{2}} \right. \\ &\quad \left. a_{k-j} - a_k a_j \right] e_k(x) e_j(x). \end{aligned}$$

3. Erreur quadratique moyenne :

$$\begin{aligned} \text{MSE}(\hat{f}_{d_n}(x)) &= \\ &= \frac{d_n}{2} + \sum_{k=1}^{d_n} \cos(2\pi kx) \left(\frac{1}{n} + \frac{1}{\sqrt{2n}} a_{2k} - \frac{1}{n} a_k^2 \right) + \sum_{k=1}^{d_n} \sum_{j=1}^{d_n} \frac{1}{n} \left[\frac{1}{\sqrt{2}} a_{k+j} \frac{1}{\sqrt{2}} \right. \\ &\quad \left. a_{k-j} - a_k a_j \right] e_k(x) e_j(x) + \left[\sum_{k=d_n+1}^{\infty} a_k e_k(x) \right]^2. \end{aligned}$$

4. L'Erreur quadratique moyenne intégrée (voir, [46]) : si $\int f^2(x) < \infty$, alors :

$$\text{MISE}(\hat{f}_{d_n}(x)) = \int_0^1 f^2(x) dx - \sum_{k=0}^{d_n} a_k^2 + \sum_{k=0}^{d_n} \text{Var}(\hat{a}_k). \quad (1.27)$$

1.4.3 Choix du paramètre de lissage

de Kronmal-Tarter :

On ajoutera à la forme (1.26) le d_n terme si et seulement si :

$$\Delta_{d_n} = MISE(\hat{f}_{d_n}(x)) - (\hat{f}_{d_{n-1}}(x)) = \frac{n+1}{n} Var(e_{d_n}(X)) - E(e_{d_n}(X))^2 \leq 0.$$

On pose

$$\theta_i = e_{d_n}(x_i), i = 1, 2, \dots, n.$$

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \theta_i.$$

L'estimateur symétrique sans biais de Δ_{d_n} donné par :

$$\hat{\Delta}_{d_n} = \frac{1}{n} \left[\frac{n+1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta})^2 - \sum_{i=1}^n \theta_i^2 \right]. \quad (1.28)$$

On fixe un entier positif D , l'optimum d_n^* est alors de la forme :

$$d_n^* = \begin{cases} \inf\{d_n, 1 \leq d_n \leq D\} & \text{si } \hat{\Delta}_{d_n} \geq 0. \\ D & \text{sinon.} \end{cases}$$

Méthode de Bosq [3]

Bosq a proposé un nouveau estimateur de paramètre de lissage donné par :

$$\hat{d}_n = \max\{j : 0 \leq j \leq d_n, |\hat{a}_j| \geq \gamma_n\},$$

Avec

$$\gamma_n = c \sqrt{\frac{\log n}{n}}, c > 0. \quad (1.29)$$

1.5 Estimation par la méthode du noyau

1.5.1 Définition

Soit X_1, X_2, \dots, X_n un n-échantillon de variables aléatoires indépendantes et identiquement distribuées de fonction de répartition F et de densité f .

L'estimateur à noyau de f , notée \hat{f} est définie par (voir,[31],[33])

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \quad (1.30)$$

Où K est une fonction noyau, et h est un paramètre de lissage ou fenêtre.

Remarque :

1. $\int_{\mathbb{R}} K(u) du = 1$ et $K(u) \geq 0$.(voir,[35])
2. L'estimateur à noyau \hat{f}_h est une fonction de densité (voir,[35]).
3. L'estimateur \hat{f}_h a les mêmes propriétés de continuité et de différentiabilité que K (voir,[35],[14]) :
 - (a) Si K est continue, \hat{f} alors une fonction continue.
 - (b) Si K est différentiable, \hat{f} alors une fonction différentiable.

1.5.2 Propriétés de l'estimateur à noyau

Nous avons établi les propriétés suivantes (voir, [1]) :

1. Le biais est donné par :

$$\text{Biais}(\hat{f}_h(x)) = \frac{h^2}{2} f''(x) \sigma_k^2 + O(h^2). \quad (1.31)$$

2. La variance est donnée par :

$$\text{Var}(\hat{f}_h(x)) = \frac{f(x)}{n} R(k) + O\left(\frac{1}{nh}\right). \quad (1.32)$$

3. L'erreur quadratique moyenne est donnée par :

$$\text{MSE}(\hat{f}_h(x)) = \frac{h^4}{4} (f''(x))^2 \sigma_k^4 + \frac{f(x)}{nh} R(k) + O\left(\frac{1}{nh}\right) + O(h^4) \quad (1.33)$$

4. L'erreur moyenne quadratique intégrée est donnée par :

$$\text{MISE}(\hat{f}_h(x)) = \frac{h^4}{4} R(f'') \sigma_K^4 + \frac{1}{nh} R(k) + O(h^4 + \frac{1}{nh}) \quad (1.34)$$

5. L'approximation asymptotique du MSE par :

$$\text{AMSE}(\hat{f}_h(x)) = \frac{h^4}{4} (f''(x))^2 \sigma_k^4 + \frac{f(x)}{nh} R(k) \quad (1.35)$$

6. L'approximation asymptotique du MISE :

$$\text{AMISE}(\hat{f}_h(x)) = \frac{h^4}{4} \sigma_k^4(K) R(f'') + \frac{1}{nh} R(k) \quad (1.36)$$

où $\sigma_K^2 = \int u^2 K(u) du$ et $R(K) = \int k^2(u) du$ et $R(f'') = \int (f''(x))^2 dx$.

1.6 Estimation par la méthode des fonctions Splines

1.6.1 Les fonctions splines

Soient $I = [a, b]$ un intervalle de \mathbb{R} , $a < x_1 < \dots < x_n < b$ une partition de l'intervalle I et soit $f : I \rightarrow \mathbb{R}$ une fonction. Nous définissons les ensembles suivants :

1. $\mathcal{C}^r(I) = \{f : I \rightarrow \mathbb{R}, f \text{ est continûment dérivable jusqu'à l'ordre } r\}$.
2. $\mathcal{P}_r = \{p : \mathbb{R} \rightarrow \mathbb{R}, P \text{ polynôme de degré } \leq r, r \in \mathbb{N}\}$.
3. $W_2^r = \{f : I \rightarrow \mathbb{R}; f^{(k)} \text{ absolument continue pour } k = 0, \dots, r - 1, \text{ et } \int_I (f^{(r)}(x))^2 < \infty\}$, f est dite absolument continue, si il existe une fonction g tels que : $f(x) = \int_a^x g(t) dt$.

Définition : La fonction f est appelée spline polynomiale de degré $(r-1)$ (ou d'ordre r) ayant pour noeuds x_1, \dots, x_n si elle vérifie les propriétés suivantes :

- f est continûment dérivable jusqu'à l'ordre $(r-2)$ (c-a-d $f \in \mathcal{C}^{r-2}(I)$),
- La restriction de f aux intervalles inter-noeuds $[a, x_1], \dots, [x_i, x_{i+1}], \dots, [x_n, b]$ est un polynôme de degré inférieur ou égal à $(r-1)$ (c-a-d $f|_{[x_i, x_{i+1}]} \in \mathcal{P}_{(r-1)}$ pour tout $i = 1, \dots, n-1$).

L'ensemble des splines sera noté $\mathcal{S}_r(x_1, \dots, x_n)$. Il contient l'ensemble des polynômes de degré inférieur ou égal à $(r-1)$.

Remarque :

1. $r = 2$ est une spline de degré 1 (d'ordre 2) est donc une fonction continue et linéaire par morceaux.
2. $r = 4$ est une de degré 3 (d'ordre 4), ce sont les splines les plus fréquemment utilisées sont les splines dites splines cubiques.
3. Tout polynôme de degré k sur $[a, b]$ est une spline, mais en pratique, on considère les splines constituées de n polynômes différents sur chaque sous-intervalle.

1.6.2 Estimation de la fonction de répartition avec les fonctions splines

Soit T_1, \dots, T_n un échantillon issu d'une variable aléatoire X de fonction de répartition F . Soit (a, b) , tel que $-\infty < a < b < +\infty$ le support de F , $T_{(i)}$ la $i^{\text{ème}}$ statistique d'ordre de l'échantillon et F_n la fonction de répartition empirique de l'échantillon, définie par :

$$F_n(x) = \frac{1}{n+1} \sum_{i=1}^{n+1} 1_{\{T_{(i)} \leq x\}}. \quad (1.37)$$

On divise par $n+1$ et non par n pour éviter d'atteindre la valeur 1 à la dernière observation, (voir [32]). $\frac{1}{n+1}$ est l'espérance des écarts $h_i = T_{(i+1)} - T_{(i)}$ dans le cas de la distribution uniforme.

Le lissage des points $(T_{(1)}, y_1 = F_n(T_{(1)})), \dots, (T_{(n)}, y_n = F_n(T_{(n)}))$ est donné en minimisant la somme des carrés pondérée pénalisée donnée par :

$$C(G) = \sum_{i=1}^n w_i (F_n(T_{(i)}) - G(T_{(i)}))^2 + \alpha \int_a^b G''(x)^2 dx, \quad (1.38)$$

sur l'espace W_2^2 de Sobolev défini par :

$$W_2^2 = \{G : (a, b) \rightarrow \mathbb{R} / G, G' \text{ sont absolument continues, } G'' \in L_2(a, b)\},$$

où $w_i = \frac{1}{y_i(1-y_i)} = \text{Var}(U_i)/(n+2)$ pour la $i^{\text{ème}}$ statistique d'ordre U_i d'un échantillon uniforme de taille n , α varie entre 0 et $+\infty$ est le paramètre de lissage. La solution \hat{F}_n du problème de minimisation précédent est une spline cubique

naturelle sur (a, b) ayant pour noeuds $T_{(1)}, \dots, T_{(n)}$.

L'estimateur de la fonction de répartition F est \hat{F}_n défini par :

$$\hat{F}_n = (I + \alpha W^{-1}K)^{-1}F_n = A_\alpha F_n. \quad (1.39)$$

où

1. I est la matrice identité.
2. W est une matrice diagonale avec les éléments diagonaux sont égaux à w_i .
3. K est une matrice qui dépend uniquement des écarts $H_i = T_{(i+1)} - T_{(i)}$ et s'écrit sous la forme $K = QR^{-1}Q^t$, avec Q et R des matrices bandes de terme général respectifs Q_{ij} et R_{ij} tel que :

— La matrice Q est de taille $n \times (n - 2)$ tel que :

$$\begin{aligned} Q_{j-1,j} &= \frac{1}{h_{j-1}}, \\ Q_{jj} &= -\left(\frac{1}{h_{j-1}} + \frac{1}{h_j}\right), \\ Q_{j+1,j} &= \frac{1}{h_j} \\ Q_{ij} &= 0, \text{ si } |i - j| \geq 2. \end{aligned}$$

— La matrice R est symétrique de taille $(n - 2) \times (n - 2)$ de composantes

$$\begin{aligned} r_{ij} : \\ R_{ii} &= \frac{1}{3}(h_{i-1} + h_i), \quad i = 2, \dots, n - 1, \\ R_{i,i+1} &= R_{i+1,i} = \frac{1}{6}h_i, \quad i = 2, \dots, n - 2, \\ R_{ij} &= 0, \text{ si } |i - j| \geq 2. \end{aligned}$$

R est une matrice à diagonale strictement dominante, alors R est définie positive, l'inverse R^{-1} existe et donc \hat{F}_n est bien défini.

On peut obtenir l'estimateur de la fonction densité de probabilité, en dérivant l'estimateur de la fonction de répartition. La fonction spline cubique \hat{F} peut s'écrire sous la forme polynomiale suivante :

$$\hat{F}(x) = \sum_{i=1}^n P_i(x) 1_{[T_{(i)}, T_{(i+1)}[}(x),$$

avec

$$P_i(x) = a_i(x - T_{(i)})^3 + b_i(x - T_{(i)})^2 + c_i(x - T_{(i)}) + d_i, \quad x \in [T_{(i)}, T_{(i+1)}[.$$

La dérivée première de \hat{F} est :

$$\hat{f}(x) = \hat{F}'(x) = \sum_{i=1}^n P'_i(x) 1_{[T_{(i)}, T_{(i+1)}[}(x) = \sum_{i=1}^n 3a_i(x - T_{(i)})^2 + 2b_i(x - T_{(i)}) + c_i, \quad T_{(i)} \leq x < T_{(i+1)},$$

les coefficients a_i, b_i, c_i et d_i sont déterminés à partir des conditions imposées sur la spline et ses dérivées (les conditions d'interpolation et les conditions de continuité pour la première et la deuxième dérivées à chaque tous noeuds et sur chaque sous intervalle).

Notons par $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$, $c = (c_1, \dots, c_n)$, et $d = (d_1, \dots, d_n)$ les vecteurs coefficients.

On définit les vecteurs suivants :

1. $U = [6(1-s)QQ^t + sR]^{-1}QF_n, \quad (n-2 \times 1).$
2. $V = F_n - 6(1-s)Q^tU, \quad (n \times 1).$
3. $W = [0, sU^t, 0]^t, \quad (n \times 1),$ avec $s = (1 + \alpha)^{-1}.$

Les coefficients sont donnés par :

1. $a = (\frac{\Delta w}{h_i})^t, \quad i = 1, \dots, n-1$ et $\Delta w = w_{i+1} - w_i.$
2. $b = [0, 3sU^t]^t.$
3. $c = [\frac{\Delta v}{h_i} - h_i[(0, 2sU^t)^t + (sU^t, 0)^t]]^t, \quad \Delta v = v_{i+1} - v_i.$
4. $d = (v_i), \quad i = 1, \dots, n-1.$

1.6.3 Choix du paramètre de Lissage : Validation croisée généralisée (GCV)

Le score de la validation croisée généralisée est définie par :

$$GCV(\alpha) = \frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{F}_n(T_{(i)}))^2}{(1 - \frac{1}{n} Tr(A_\alpha))^2}. \quad (1.40)$$

α est choisi de telle manière qu'il minimise $GCV(\alpha).$

$$\alpha_{GCV} = \arg \min GCV(\alpha). \quad (1.41)$$

1.7 Conclusion

Dans ce chapitre, nous avons les différentes méthodes d'estimation de la fonction densité, à savoir, la méthode de l'histogramme, la méthode du noyau, la méthode des fonctions orthogonales, et la méthode des splines. Nous avons donné les propriétés statistiques et asymptotiques pour chaque méthode sauf la méthode des splines. Dans le chapitre suivant, nous allons présenter en détail l'estimateur à noyau ainsi que ses propriétés. Nous parlons aussi de son intérêt particulier sur le choix du paramètre de lissage.

CHAPITRE 2

ESTIMATION NON PARAMÉTRIQUE DE LA FONCTION DENSITÉ PAR LA MÉTHODE DU NOYAU

2.1 Introduction

Plusieurs méthodes non paramétriques ont été proposées pour estimer la fonction densité. La méthode la plus rencontrée dans la littérature est la méthode du noyau. cette méthode à été proposée par Rosenblatt en 1956 (voir,[33]) puis développée par Parzen en 1962 (voir,[31]).

Dans ce chapitre , nous présentons l'estimateur de la fonction densité par la méthode du noyau ainsi que ses différentes propriétés.

2.2 Estimateur de Rosenblatt(1956)

L'idée est d'utiliser la définition de la dérivée pour estimer $f(x), \forall x \in R$ (voir,[33]).

$$f(x) \stackrel{\text{def}}{=} F'(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h}, \quad (2.1)$$

h est très petit donc on peut écrire la formule (2.3) comme suit :

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

Donc l'estimateur de f est défini par :

$$\hat{f}(x) = \frac{1}{2h} [F_n(x+h) - F_n(x-h)]. \quad (2.2)$$

avec F étant la fonction de répartition. On peut la remplacer par son estimateur empirique F_n telle que :

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

$$= \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, x]}(X_i)$$

avec :

$$1_{]-\infty, x]}(X_i) = \begin{cases} 1 & \text{si } X_i \leq x, \\ 0 & \text{sinon.} \end{cases}$$

Donc :

$$F_n(x - h) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, x-h]}(X_i),$$

$$F_n(x + h) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, x+h]}(X_i),$$

On obtient :

$$\hat{f}(x) = \frac{1}{2nh} \sum_{i=1}^n 1_{]x-h, x+h]}(X_i). \quad (2.3)$$

cette estimateur est meilleur que celui de l'histogramme car on construit une classe de largeur $2h$ en chaque point x où l'on veut estimer f .

Son inconvénient est que \hat{f} est une fonction en escalier, par conséquent \hat{f} est discrète on a alors :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} 1_{]x-h, x+h]}(X_i) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} 1_{]-1, 1]} \left(\frac{x - X_i}{h} \right). \quad (2.4)$$

On a :

$$\begin{aligned} 1_{]x-h, x+h]}(X_i) &= \begin{cases} 1 & \text{si } x - h < X_i \leq x + h \\ 0 & \text{sinon} \end{cases}, \\ &= \begin{cases} 1 & \text{si } -h - x < -X_i \leq h - x \\ 0 & \text{sinon} \end{cases}, \\ &= \begin{cases} 1 & \text{si } -h < x - X_i \leq h \\ 0 & \text{sinon} \end{cases}, \\ 1_{]-1, 1]} \left(\frac{x - X_i}{h} \right) &= \begin{cases} 1 & \text{si } -1 < \frac{x - X_i}{h} \leq 1 \\ 0 & \text{sinon} \end{cases}. \end{aligned}$$

on pose $K(x) = \frac{1}{2} 1_{]-1, 1]}(x)$. $K(x)$ est une densité de probabilité ;

en effet :

$$\int K(x) dx = \frac{1}{2} \int 1_{]-1, 1]}(x) dx = \frac{1}{2} \int_{-1}^1 dx = \frac{1}{2} [x]_{-1}^1 = 1,$$

alors :

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

2.3 Estimateur de parzen (1962)

Cet estimateur consiste à remplacer la fonction densité K par une fonction continue pour que $\hat{f}(x)$ soit continue. Cette méthode s'appelle méthode de noyau (voir,[31]).

définition :

Un estimateur à noyau de la fonction densité f au point x est défini par :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right). \quad (2.5)$$

où h est un paramètre réel positif qui tend vers 0, il détermine le voisinage de x , et K est une fonction densité de probabilité appelé noyau, elle détermine la forme du voisinage autour du point x .

Remarque :

- $\hat{f}(x)$ a les mêmes propriétés de continuité et de dérivabilité que K .
 - $\hat{f}(x)$ est une densité de probabilité .
- en effet :

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

posons :

$$\mu = \frac{x - X_i}{h} \Rightarrow h\mu = x - X_i,$$

alors

$$x = h\mu + X_i \Rightarrow dx = h d\mu,$$

$$\begin{aligned} \int_{\mathbb{R}} \hat{f}(x) dx &= \frac{1}{h} \int_{\mathbb{R}} K(\mu) h d\mu, \\ \int_{\mathbb{R}} \hat{f}(x) dx &= \int_{\mathbb{R}} K(\mu) d\mu = 1. \end{aligned}$$

Donc $\hat{f}(x)$ est bien une densité de probabilité

2.4 Estimateur à noyau

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

Soit $K : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction intégrable tel que $\int_{\mathbb{R}} K(\mu) d\mu = 1$. K est dit noyau pour $n \in \mathbb{N}^*$, on appelle $h_n > 0$ la fenêtre ou paramétré de lissage. et $\hat{f}(x)$ l'estimateur à noyau de la densité de probabilité f (estimateur de Parzen-Rozenblatt)[31][33] définit pour tout $x \in \mathbb{R}$ par

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

On note $h_n = h$. Dans cette partie, nous présentons l'estimateur de la densité de probabilité à noyau, puis nous donnons les différentes propriétés de cet estimateur. Il existe deux types de noyau (symétrique et asymétrique).

2.5 Noyaux symétriques

Un noyau continu classique est dit symétrique si, pour tout μ dans son domaine de définition $K(\mu) = K(-\mu)$. ce qui implique l'égalité suivante :

$$\int_{\mathbb{R}} \mu K(\mu) d\mu = 0.$$

De plus, elle est de carré intégrable, c-à-d :

$$\int K^2(\mu) d\mu < +\infty.$$

2.5.1 Quelques noyaux usuelles symétriques

Noyau Uniforme (Rosenblatt)

Ce noyau a été proposé par Rosenblatt en 1956 [33], l'avantage de ce noyau est la simplicité de sa forme. Il s'écrit sous la forme (voir, [13]) :

$$K(x) = \begin{cases} \frac{1}{2} & \text{si } |x| \leq 1; \\ 0 & \text{sinon.} \end{cases}$$

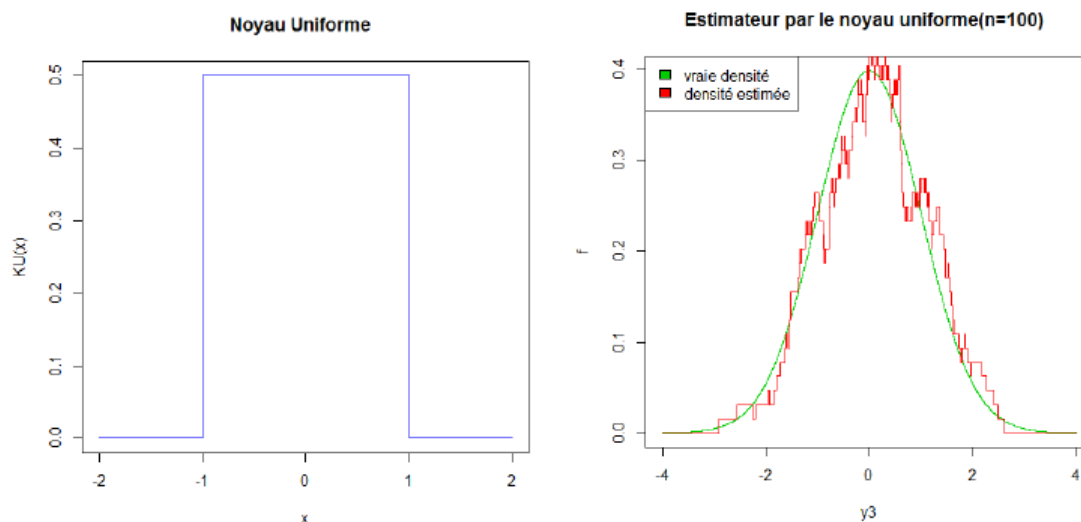


FIGURE 2.1 – Noyau Uniforme.

Noyau triangulaire

L'avantage de ce noyau par rapport au précédent est sa continuité partout, ce qui conduit à un estimateur continu. Ce noyau s'écrit sous la forme (voir,[48]) :

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

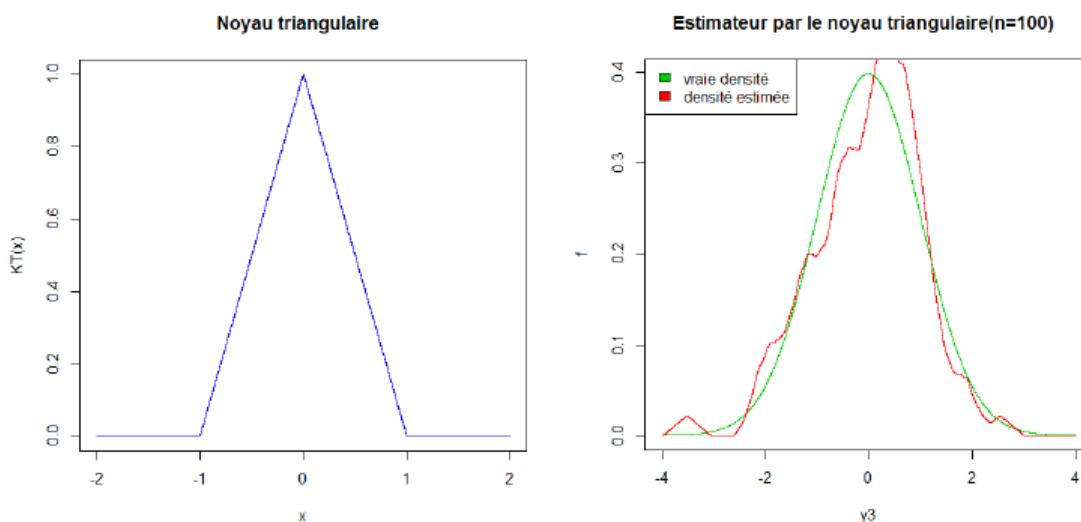


FIGURE 2.2 – Noyau Triangulaire.

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 :

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

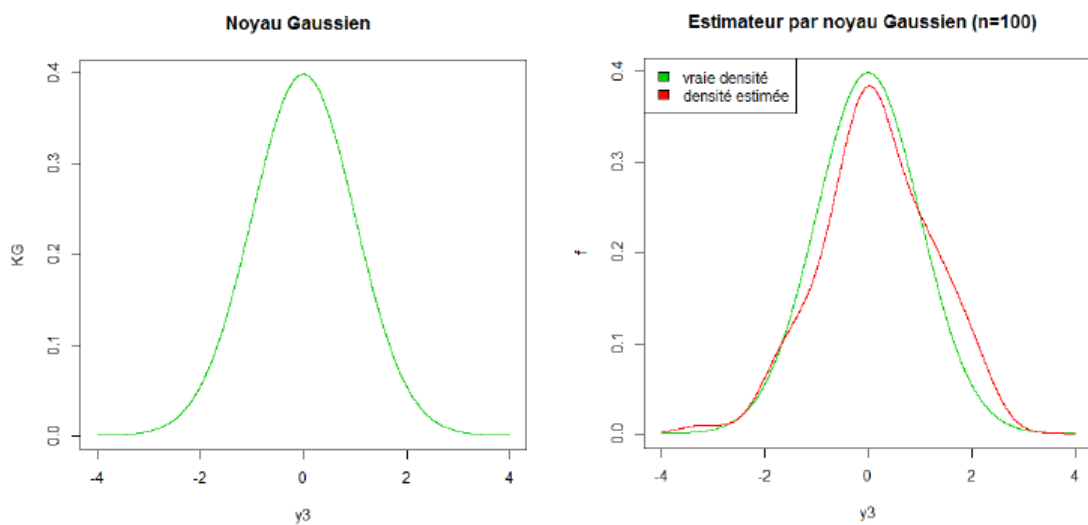


FIGURE 2.3 – Noyau Gaussien.

Noyau Biweight (Tukey)

Le noyau de biweight est très intéressant car il donne un estimateur dérivable partout. En fait, il s'agit du noyau le plus simple parmi les noyaux de forme polynômial dérivable partout. Ainsi, il assure le lissage locale de la fonction \hat{f} . Ce noyau est d'une forme très proche du noyau gaussien, il est donc préférable de l'utiliser. il s'écrit sous la forme :

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & \text{si } |x| \leq 1; \\ 0 & \text{sinon.} \end{cases}$$

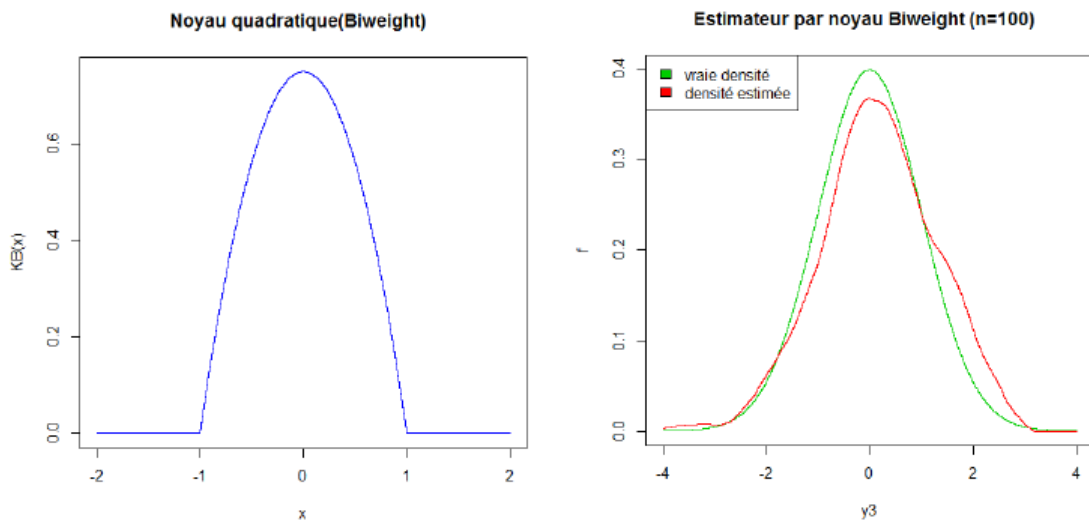


FIGURE 2.4 – Noyau Biweight.

Noyau Epanechnikov

En (1969), Epanechnikov [48], a donné la forme du noyau K_E défini par :

$$K_E(x) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - \frac{x^2}{5}) & \text{si } |x| \leq \sqrt{5}; \\ 0 & \text{sinon.} \end{cases}$$

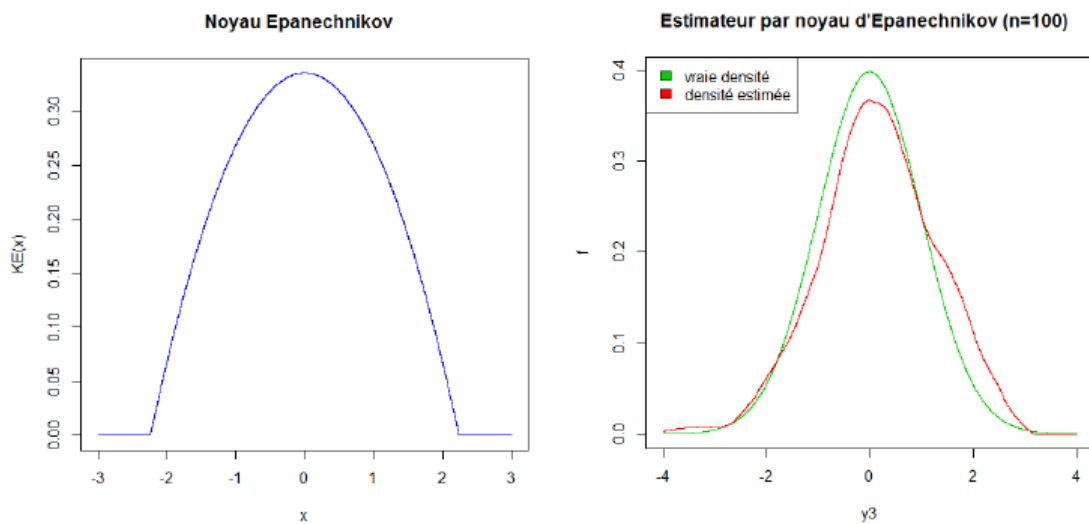


FIGURE 2.5 – Noyau Epanechnikov.

Noyau Sinus

$$K(x) = \frac{1}{2\pi} \left(\frac{\sin(\frac{x}{2})}{\frac{x}{2}} \right)^2 \quad \text{si } x \neq 0.$$

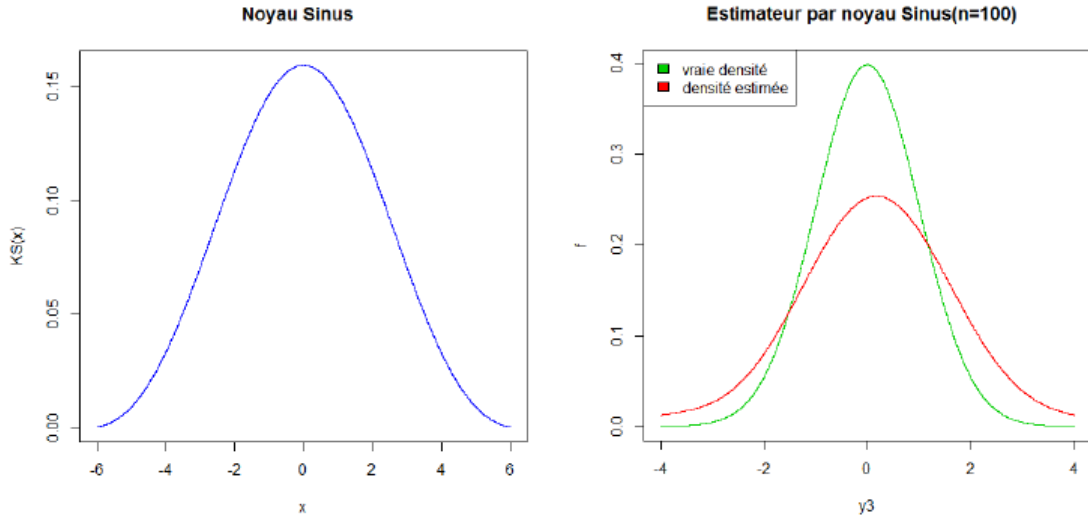


FIGURE 2.6 – Noyau Sinus.

2.6 Noyau asymétrique

Quoique les méthodes précédentes diminuent le biais, aux bornes, elles restent peu efficaces car le biais reste considérable si on le compare aux biais de l'intérieur du support. Pour obtenir un biais aux bornes de même ordre que celui de l'intérieur, Devroy et Györfi (1985)[15] et Marron et Ruppert (1994)[26], ont proposé d'appliquer une transformation sur les données originales de telle façon que la dérivée d'ordre 1 de la densité des variables transformées soit égale à zéro et ensuite utiliser la méthode de réflexion pour estimer la densité des données transformées. L'objectif étant de trouver cette fois un biais du même ordre mais sans transformation des données. Plusieurs autres auteurs ont proposé d'utiliser les noyaux adaptés dans la région des bornes et le noyau standard à l'intérieur du support (voir Jones (1993) [21]). Pour l'estimation à noyau aux bornes, Müller (1991) [25] pour l'estimateur à noyau optimal aux bornes et Lejeune et Sarda (1992) [24] pour l'estimation linéaire locale.

L'inconvénient de ces estimateurs est qu'ils attribuent des poids négatifs aux valeurs du voisinage des bornes.

La solution la plus récente est d'utiliser des noyaux asymétriques et adaptés qui n'assignent aucun poids à l'extérieur du support. Chen (1999)[10] et Chen (2000)[9] propose respectivement le noyau Bêta pour les densités à support compact et le noyau gamma pour les densités à variables à support positif (c'est-à-dire

sur $[0; +\infty[$.

2.6.1 Noyau Bêta

le noyau bêta proposé par Brown (1999)[45], et Chen (2000)[9] pour l'estimation non paramétrique de la densité unidimensionnelle défini sur un support compact. L'idée de Harrell et Davis (1982)[19], Chen (1999,2000)[[10], [9]] est d'utiliser le noyau bêta pour estimer la densité de probabilité à support compact $[0,1]$. L'estimateur de la densité sera alors de la forme :

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K(X_i, \frac{x}{h} + 1, \frac{1-x}{h} + 1). \quad (2.6)$$

où $K(x, \alpha, \beta)$ représente la densité de la distribution Bêta de paramètres α et β ,

$$K(x, \alpha, \beta) = \frac{x^\alpha(1-x)^\beta}{B(\alpha, \beta)} \quad x \in [0, 1].$$

avec :

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx.$$

2.6.2 Noyaux Gamma

L'objectif est d'estimer la fonction $f(x)$ définie sur un support $[0, +\infty[$. Il existe deux versions du Noyau Gamma (voir, [27]).

Version de Bouezmarni et Scaillet :

$$K_{(\frac{x}{h}+1, h)}(t) = \frac{t^{(x/h)} e^{-(t/h)}}{h^{(x/h)+1} \Gamma((x/h) + 1)}.$$

Version de Chen :

$$K_{(\rho_h(x), h)}(t) = \frac{t^{\rho_h(x)-1} e^{-\frac{t}{h}}}{h^{\rho_h(x)} \Gamma(\rho_h(x))}$$

et

$$\rho_h(x) = \begin{cases} \frac{x}{h} & \text{si } x \geq 2h. \\ \frac{1}{4}(\frac{x}{h})^2 + 1 & \text{si } 0 \leq x < 2h. \end{cases}$$

2.7 Propriétés de l'estimateur à noyau

Soit un échantillon X_1, X_2, \dots, X_n , issu de X de fonction de répartition F et de densité f inconnues de avec le support \mathbb{T} .

2.7.1 Espérance mathématique

L'estimateur à noyau continu classique de la fonction densité est défini par :

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

On a

$$h \rightarrow 0 \quad \text{et} \quad n \rightarrow \infty \quad \text{et} \quad \int_{\mathbb{R}} K(\mu) d\mu = 1$$

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

on pose

$$\begin{aligned} \Rightarrow h\mu &= x - t, \\ \Rightarrow t &= x - h\mu, \\ \Rightarrow dt &= -h d\mu, \end{aligned}$$

$$\begin{aligned} \Rightarrow E[\hat{f}(x)] &= \frac{1}{h} \int_{\mathbb{T}} K(\mu) f(x - h\mu) h d\mu \\ E[\hat{f}(x)] &= \int_{\mathbb{T}} K(\mu) f(x - h\mu) d\mu \\ E[\hat{f}(x)] &= \int_{\mathbb{T}} K(\mu) f(x - h\mu) d\mu \end{aligned}$$

En faisant le développement de Taylor de $f(x - h\mu)$ au point $x = 0$ jusqu'à l'ordre 2.

On obtient :

$$f(x - h\mu) = f(x) - h\mu f'(x) + \frac{h^2 \mu^2}{2} f''(x) + o(h^2);$$

On remplace dans la formule de l'espérance on trouve :

$$\begin{aligned} E[\hat{f}(x)] &= \int_{\mathbb{T}} K(\mu) [f(x) - h\mu f'(x) + h^2 \mu^2 f''(x)] d\mu + o(h^2), \\ E[\hat{f}(x)] &= f(x) \int_{\mathbb{T}} K(\mu) d\mu - h f'(x) \int_{\mathbb{T}} \mu K(\mu) d\mu + h^2 f''(x) \int_{\mathbb{T}} \mu^2 K(\mu) d\mu + o(h^2), \\ E[\hat{f}(x)] &= f(x) + h^2 f''(x) \int_{\mathbb{T}} \mu^2 K(\mu) d\mu + o(h^2). \end{aligned}$$

On pose $\int_{\mathbb{R}} \mu^2 K(\mu) d\mu = \varphi_2(K)$

$$\begin{aligned} v(K) &= E(K^2) - (E(K))^2, \\ v(K) &= \int_{\mathbb{T}} \mu^2 K(\mu) d\mu, \end{aligned}$$

alors :

$$E[\hat{f}(x)] = f(x) + h^2 f''(x) \varphi_2(K) + o(h^2).$$

Donc $\hat{f}(x)$ est un estimateur avec biais.

2.7.2 Variance de l'estimateur

La variance de $\hat{f}(x)$ est donnée par (voir, [42]) :

$$\begin{aligned} V[\hat{f}(x)] &= V\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right), \\ &= \frac{1}{n^2 h^2} \cdot n \cdot V\left[K\left(\frac{x - X}{h}\right)\right], \\ &= \frac{1}{nh^2} [E[K^2\left(\frac{x - X}{h}\right)] - E\left[K\left(\frac{x - X}{h}\right)\right]^2], \\ &= \frac{1}{nh^2} \left[\int_{\mathbb{T}} K^2\left(\frac{x-t}{h}\right) f(t) dt - \left(\int_{\mathbb{T}} K\left(\frac{x-t}{h}\right) f(t) dt \right)^2 \right], \\ &= \frac{1}{nh^2} \int_{\mathbb{T}} K^2\left(\frac{x-t}{h}\right) f(t) dt - \frac{1}{n} \left(\frac{1}{h} \int_{\mathbb{T}} K\left(\frac{x-t}{h}\right) f(t) dt \right)^2, \end{aligned}$$

$$\text{On pose } \mu = \frac{x-t}{h} \Rightarrow h\mu = x-t \Rightarrow t = x-h\mu \Rightarrow dt = -h d\mu,$$

On aura :

$$\begin{aligned} \frac{1}{h} \int_{\mathbb{T}} K\left(\frac{x-t}{h}\right) f(t) dt &= \frac{1}{h} \int_{\mathbb{T}} K(\mu) f(x-h\mu) h d\mu, \\ &= \int_{\mathbb{T}} K(\mu) f(x-h\mu) d\mu, \\ &= E[\hat{f}(x)]. \end{aligned}$$

Alors

$$\begin{aligned} V(\hat{f}(x)) &= \frac{1}{nh^2} \int_{\mathbb{T}} K^2(\mu) f(x-h\mu) (-h d\mu) + (E[\hat{f}(x)])^2, \\ &= \frac{1}{nh} \int_{+\infty}^{-\infty} K^2(\mu) f(x-h\mu) d\mu + (E[\hat{f}(x)])^2. \end{aligned}$$

Le développement de Taylor de $f(x-h\mu)$ au $v(\mu=0)$ jusqu'à l'ordre 1, on obtient :

$$f(x-h\mu) = f(x) - h\mu f'(x) + o(h).$$

On remplace dans l'expression de la variance, donne :

$$\begin{aligned} V(\hat{f}(x)) &= \frac{1}{nh} \left[\int_{+\infty}^{-\infty} K^2(\mu) f(x) - \int_{+\infty}^{-\infty} K^2(\mu) \mu h f'(x) d\mu + o(h) \right] + (E[\hat{f}(x)])^2, \\ &= \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(\mu) d\mu - \frac{f'(x)}{n} \int_{-\infty}^{+\infty} \mu K^2(\mu) d\mu + o(h) + (E[\hat{f}(x)])^2. \end{aligned}$$

On suppose que $n \rightarrow \infty$ et $h \rightarrow \infty$, la variance devient :

$$V(\hat{f}(x)) = \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(\mu) d\mu + o\left(\frac{1}{nh}\right).$$

2.7.3 Biais de l'estimateur

Le biais est donné par, (voir,[42]) :

$$\begin{aligned} \text{Biais}(\hat{f}) &= E[\hat{f}(x)] - f(x), \\ &= f(x) + h^2 f''(x) \varphi_2(K) + o(h^2) - f(x), \\ &= h^2 f''(x) \varphi_2(K) + o(h^2). \end{aligned}$$

2.7.4 Erreur quadratique moyenne MSE

L'analyse de la performance de l'estimateur à noyau exige la spécification d'un critère d'erreur approprié, afin de mesurer l'erreur d'estimation à un simple point aussi bien que sur l'ensemble des points. Car L'estimateur \hat{f} dépend du noyau K et du paramètre de lissage h , cette dépendance n'est généralement pas exprimée explicitement. Lorsque nous considérons l'estimation à un point, une mesure naturelle de la dispersion est l'erreur quadratique moyenne.

$$MSE[\hat{f}(x)] = \text{biais}^2(\hat{f}(x)) + V(\hat{f}(x)).$$

Alors MSE est donnée par :

$$MSE(\hat{f}(x)) = \frac{h^4}{4} \int_{\mathbb{R}} (f''(x))^2 dx \varphi_2^2(K) + \frac{f(x)}{nh} \int_{-\infty}^{+\infty} K^2(\mu) d\mu + o\left(h^4 + \frac{1}{nh}\right).$$

2.7.5 Erreur moyenne quadratique intégrée (MISE)

L'Erreur moyenne quadratique intégrée (MISE) est donnée par (voir,[8]) :

$$\begin{aligned} MISE(\hat{f}(x)) &= \int_{\mathbb{T}} MSE(\hat{f}(x)) dx, \\ &= \varphi_2^2(x) \frac{h^4}{4} \int_{\mathbb{T}} (f''(x))^2 dx + \frac{1}{nh} \int_{-\infty}^{+\infty} K^2(\mu) d\mu \int_{\mathbb{T}} f(x) dx + o\left(h^4 + \frac{1}{nh}\right), \end{aligned}$$

$$MISE(\hat{f}(x)) = \frac{h^4}{4} \varphi_2^2(x) R(f'') + \frac{1}{nh} R(K) + o\left(h^4 + \frac{1}{nh}\right). \text{ avec } R(t) = \int_{\mathbb{T}} t^2(x) dx,$$

et t une fonction continue.

2.7.6 Comportement asymptotique

Une approximation asymptotique de l'espérance de l'estimateur \hat{f} est donnée sous les conditions suivantes sur f , h et K [47] :

1. La fonction densité f admet la dérivée seconde f'' absolument continue, de carré intégrable et monotone sur $] -\infty; -M]$ et $[M; +\infty[$ pour $M > 0$.
2. Le paramètre de lissage h est positif et satisfait $\lim_{n \rightarrow +\infty} h = 0$ et $\lim_{n \rightarrow +\infty} nh = +\infty$
3. Pour que $\hat{f}(x)$ soit une densité, on suppose que $K(\mu) \geq 0$ et $\int K(\mu) d\mu = 1$, La fonction noyau est supposée être symétrique autour de zéro, c-à-d $\int \mu K(\mu) d\mu = 0$, et possède un moment d'ordre 2 fini, c-à-d $\int \mu^2 K(\mu) d\mu < \infty$.

Le biais et la variance sont définis par :

$$\begin{aligned} \text{Biais} \hat{f}(x) &= \frac{h^2}{2} \varphi_2(K) f''(x) + o(h^2). \\ V(\hat{f}(x)) &= \frac{R(K)}{nh} f(x) + o\left(\frac{1}{nh}\right). \end{aligned}$$

où $\varphi_2(K) = \int \mu^2 K(\mu) d\mu$ et $R(g) = \int g^2(\mu) d\mu$ pour une fonction g de carré intégrable

Les expressions asymptotiques du biais et de la variance nous permettent de trouver des expressions asymptotiques de MSE et MISE. Ces expressions ont été obtenues sous la condition (3) sur K et en supposant que la densité de probabilité f a des dérivées continues. On peut obtenir les approximations asymptotiques suivantes pour les erreurs MSE et MISE sous des conditions appropriées d'intégrabilité de f et ses dérivées. On note l'approximation asymptotique du MSE par :

$$AMSE(\hat{f}(x)) = \frac{R(K)}{nh} f(x) + \frac{h^2}{4} \varphi_2^2(K) f''^2(x), \quad (2.7)$$

et l'approximation asymptotique du MISE par :

$$AMISE(\hat{f}(x)) = \frac{R(K)}{nh} + \frac{h^4}{4} \varphi_2^2(K) R(f''(x)). \quad (2.8)$$

2.7.7 Intervalle de confiance

On suppose que le nombre d'observation est très grand, construisons l'intervalle de confiance pour $f(x)$ au niveau $1 - \alpha$ (voir, [1]).

On a

$$I_c(f(x)) = [\hat{f} - a_\alpha, \hat{f} + a_\alpha],$$

alors

$$P(f(x) \in I_c) = 1 - \alpha.$$

Avant tout, on doit étudier la convergence en probabilité de $\hat{f}(x)$.

$$\begin{aligned} E[\hat{f}(x)] &= f(x) + \frac{h^2}{2} f''(x) \varphi_2 + o(h^2), \\ &= f(x) + \text{Biais}(\hat{f}(x)), \\ E[\hat{f}(x)] &\neq f(x). \end{aligned}$$

Cet estimateur est avec biais, il est asymptotiquement sans biais si et seulement si $h = h(n)$ est très petit ($h = h(n) \xrightarrow{\text{c.v.}} 0$) quand $n \rightarrow +\infty$

$$\lim_{n \rightarrow +\infty} \text{Biais}(f(\hat{x})) = \lim_{h=h(n) \rightarrow 0} \left[\frac{h^2}{2} f''(x) \varphi_2 + o(h^2) \right] = 0.$$

Alors :

$$E[\hat{f}(x)] \xrightarrow{\text{c.v.}} f(x).$$

D'après le théorème central limite si n est très grand et $h = h(n) \rightarrow 0$ est très petit alors on a :

$$\begin{aligned} &— E[\hat{f}(x)] \xrightarrow{\text{c.v.}} f(x). \\ &— V(\hat{f}(x)) = \frac{1}{nh} R(K) f(x) + o(h^2). \\ &\quad \text{avec } R(K) = \int_{\mathbb{R}} K^2(\mu) d\mu. \end{aligned}$$

Donc

$$\hat{f}(x) \xrightarrow{\text{c.v.}} \mathcal{N}\left(f(x), \frac{1}{nh} R(K) f(x)\right).$$

$$\left(\frac{\hat{f}(x) - f(x)}{\sqrt{\frac{f(x) \cdot R(K)}{nh}}} \right) \xrightarrow{\text{c.v.}} \mathcal{N}(0, 1).$$

Construction de l'intervalle de confiance de f au point x -fixé, au niveau $1 - \alpha$

$$I_c(f(x)) = [\hat{f}(x) - a_\alpha, \hat{f}(x) + a_\alpha].$$

$$P(f(x) \in I_c) = P(\hat{f}(x) - a_\alpha \leq f(x) \leq \hat{f}(x) + a_\alpha) = 1 - \alpha,$$

$$\Leftrightarrow P(f(x) - a_\alpha \leq \hat{f}(x) \leq f(x) + a_\alpha) = 1 - \alpha,$$

$$\Leftrightarrow P(f(x) - a_\alpha - f(x) \leq \hat{f}(x) - f(x) \leq f(x) + a_\alpha - f(x)) = 1 - \alpha,$$

$$P(-a_\alpha \leq \hat{f}(x) - f(x) \leq a_\alpha) = 1 - \alpha,$$

$$P(|\hat{f}(x) - f(x)| \leq a_\alpha) = 1 - \alpha,$$

$$P\left(\frac{\sqrt{nh} |\hat{f}(x) - f(x)|}{\sqrt{f(x) - R(K)}} \leq \frac{a_\alpha \sqrt{nh}}{\sqrt{f(x) - R(K)}}\right) = 1 - \alpha,$$

$$\Phi\left(\frac{a_\alpha \sqrt{nh}}{\sqrt{f(x) - R(K)}}\right) = 1 - \alpha,$$

$$\Leftrightarrow \frac{a_\alpha \sqrt{nh}}{\sqrt{f(x) - R(K)}} = \phi^{-1}(1 - \alpha),$$

alors

$$a_\alpha = \frac{1}{\sqrt{nh}} \sqrt{f(x) - R(K)} \phi^{-1}(1 - \alpha)$$

On à a_α et $f(x)$ ne sont pas connues, il suffit de remplacer $f(x)$ par $\hat{f}(x)$.

donc :

$$a_\alpha = \frac{1}{\sqrt{nh}} \sqrt{\hat{f}(x) - R(K)} \phi^{-1}(1 - \alpha),$$

avec :

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

$$\Rightarrow I_c(f(x)) = [\hat{f}(x) - a_\alpha, \hat{f}(x) + a_\alpha].$$

2.8 Critères de convergence

Parmi toutes les qualités que peut avoir un estimateur, on s'intéresse souvent à sa consistance, c'est à dire, au fait qu'un estimateur \hat{f} converge ou non vers f . La convergence d'un estimateur peut être faible (en probabilité) ou forte (presque sûrement ou en moyenne quadratique). On donne quelques résultats de convergence des estimateurs à noyaux .

2.8.1 Convergence en moyenne quadratique

Théorème[Parzen][33].

Soit f une densité continue et \hat{f} son estimateur. Si le noyau K vérifie :

- $\int K(\mu) d\mu = 1$ et $\int |K(\mu)| d\mu < \infty$;
- $\sup |K(\mu)| < \infty$ et $\lim_{\mu \rightarrow \infty} |\mu K(\mu)| = 0$.

Si le paramètre h satisfait : $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh = \infty$

Alors \hat{f} est un estimateur convergent en moyenne quadratique c'est à dire :

$$MSE(\hat{f}(x)) \xrightarrow[n \rightarrow \infty]{} 0.$$

2.8.2 Convergence uniforme en probabilité

Théorème[Parzen] [33]. Soit f la densité à estimer et \hat{f} son estimateur, si les conditions suivantes sont vérifiées :

- $\int K(\mu) d\mu = 1$ et $\int |K(\mu)| d\mu < \infty$;

- $\sup |K(\mu)| < \infty$ et $\lim_{\mu \rightarrow \infty} |\mu K(\mu)| = 0$;
- $\lim_{n \rightarrow \infty} nh^2 = \infty$;
- la transformée de Fourier $\mathcal{TF}(z) = \int \exp(-izu)K(u)du$. est absolument intégrable ;

alors \hat{f} est convergent uniformément en probabilité c'est à dire :

$$\forall \varepsilon > 0 \quad P(\sup_{x \in \mathbb{R}} |\hat{f}(x) - f(x)| < \varepsilon) = 1.$$

2.8.3 Convergence uniforme presque complète

Théorème[Nadaraya][11].

Soit f une densité uniformément continue et son estimateur à noyau K positif et à variations bornées. Si $\lim_{n \rightarrow \infty} h = 0$ et $\sum_{i=1}^{\infty} \exp(-\varepsilon nh^2) < \infty, \forall \varepsilon > 0$

alors \hat{f} est convergent uniformément avec probabilité 1 c'est à dire :

$$\sup |\hat{f}(x) - f(x)| \rightarrow 0.$$

Théorème[Silverman][44] Soit f une densité uniformément continue et \hat{f} son estimateur à noyau K positif et à variations bornées.

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

Alors

$$\sup |\hat{f}(x) - f(x)| \xrightarrow{\text{p.s.}} 0.$$

où $\xrightarrow{\text{p.s.}}$ est la convergence presque-sûre.

2.8.4 Convergence en loi

Théorème[parzen][33]

Soit f une densité continue et \hat{f} son estimateur. Si le noyau K vérifie :

- $\int K(\mu)d\mu = 1$ et $\int |K(\mu)|d\mu < \infty$.
- $\sup |K(\mu)| < \infty$ et $\lim_{\mu \rightarrow \infty} |\mu K(\mu)| = 0$.

si $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh = \infty$.

alors :

$$\frac{\hat{f}(x) - E[\hat{f}(x)]}{\sqrt{V[\hat{f}(x)]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

où $\xrightarrow{\mathcal{L}}$ représente la convergence en loi.

2.9 Choix de noyau

- On choisit le noyau K selon le support des données, c-à-d
- Si le support des données est $] - \infty, +\infty[$ on peut prendre :
 1. K noyau Gaussian $\mathcal{N}(\lambda, \sigma^2)$

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

2. K noyau EPANECHNIKOV

$$K(\mu) = \frac{3}{4}(1 - \mu^2) \quad -1 \leq \mu \leq 1.$$

- Si le support de données est $[0, +\infty[$ on peut prendre le noyau Gamma.
- Si le support est $[0,1]$ on peut prendre le noyau beta.

2.10 Choix de la fenêtre

L'estimation de la fonction de probabilité par la méthode du noyau est principalement conditionnée par le paramètre de lissage h , ce paramètre est un facteur important dans l'estimation par la méthode du noyau. Il représente en quelque sorte une fenêtre qui permet de déterminer le degré de lissage de l'estimation d'une fonction densité. Un faible paramètre de lissage implique un faible degré de lissage et résulte en une fonction densité irrégulière. À l'opposé, une large valeur de h conduit à une estimation sur lissée.

Plusieurs méthodes existent pour le choix de h , elles peuvent être classées en deux catégories. La première, est l'approche fréquentistes dont les méthodes plug-in et les méthodes validation croisée. La seconde, est l'approche bayésienne qui regroupe, l'approche bayésienne globale qui fait appel aux méthodes d'approximation MCMC et l'approche bayésienne variable (locale et adaptative).

2.10.1 Méthdes plug-in

Estimateur optimal(remplacement)

L'idée de cette méthode est de chercher le h optimale qui le minimiser $MISE$; alors le paramètre de lissage optimale prend la forme suivant :

$$h^* = \left[\frac{R(K)}{\varphi_2^2(x)R(f'')} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (2.9)$$

Rule Of Thumb

Valeur h^* notée dans ce cas par h_{rot} est donne par la formule suivant :

$$h_{rot} = 1.06\sigma^2 n^{-\frac{1}{5}}.$$

Surlissage (Oversmoothing)

L'estimateur h_{os} , dit paramètre de surlissage (oversmoothing parameter en anglais), est défini par :

$$h_{os} = 3 \left[\frac{R(K)}{35\varphi_2^2(K)} \right]^{\frac{1}{5}} \sigma n^{-\frac{1}{5}}.$$

pour le noyau gaussien on a :

$$h_{os} = 1.144\sigma n^{-\frac{1}{5}}.$$

où σ représente l'écart-type empirique de l'échantillon. Terrell et Scott [40] ont montré sous les conditions suivantes :

f admet une dérivée seconde absolument continue, $f'' \in L^2$, le noyau $K \in L^2$ est une densité de probabilité continue, symétrique de variance $\varphi_2(K)$ et sous conditions $nh \rightarrow 0$ et $nh \rightarrow \infty$, que l'ensemble des largeurs de fenêtre optimales h^* admettent une borne supérieure. h_{os} est alors construit comme un estimateur de cette borne supérieure. La largeur de fenêtre h_{os} devrait donner de bons résultats lorsque la densité à estimer présente un faible niveau de complexité structurelle (unimodalité, par exemple). En revanche, sa qualité devrait se dégrader pour des densités cibles plus complexes, la fenêtre étant alors trop large pour permettre à l'estimateur à noyau de rendre compte des variations de la densité (phénomène de surlissage).

Plug-in itéré

Scott, Tapia et Thompson [39] .le principe de cette méthode est que l'on remplace f par son estimateur \hat{f} et on propose une longueur h_0 initiale tel que à chaque fois on calcule le nouveau h_{i+1} a partir h_i et \hat{f} tel que on remplace cette dernière dans h^* et on continue ce processus jusqu'à ce que la condition suivante soit vérifiée :

$$\left| \frac{h_{i+1} - h_i}{h_i} \right| \leq \epsilon.$$

2.10.2 Méthodes de Validation Croisée (Cross Validation)

Parallèlement aux méthodes de sélection de type plug-in, d'autres méthodes se sont également révélées performantes. L'idée de base des méthodes de validation croisée consiste à trouver une fonction de score $CV(h)$ ayant la même structure que le $MISE$ et dont le calcul soit plus simple. On sélectionne alors la fenêtre h minimisant ce critère dont on attend le même comportement asymptotique que h^* . La fenêtre h n'est alors plus déterministe, elle dépend des observations.

Validation croisée de la vraisemblance (Cross validation Likelihood)

Pour un estimateur à noyau \hat{f} de f défini par (1.13) et de largeur h , la sélection par validation croisée Likelihood est une approche classique. C'est Habbema, Hermans et Vandenbroek [20] en 1974 qui ont proposé cette méthode fondée sur un critère non asymptotique du maximum de vraisemblance, mais l'interpolation entre les données est purement heuristique. Il s'agit de maximiser par rapport à h la vraisemblance pour l'échantillon $(X_i)_{1 \leq i \leq n}$ définie par :

$$LCV(h) = \prod_{i=1}^n \hat{f}_i(x_i),$$

$$\hat{f}_i(x_i) = \frac{1}{(n-1)h} \sum_{\substack{1 \leq i \leq n \\ j \neq i}} K\left(\frac{x_i - x_j}{h}\right),$$

est l'estimateur à noyau basé sur les $(n-1)$ observations différentes de x_i . La vraisemblance est alors :

$$LCV(h) = \prod_{i=1}^n \frac{1}{(n-1)h} \sum_{\substack{1 \leq i \leq n \\ j \neq i}} K\left(\frac{x_i - x_j}{h}\right). \quad (2.10)$$

Validation croisée non biaisée

Une méthode appelée validation croisée non biaisée a été proposée par Rudemo [27] et Bowman [5] en 1984.

L'idée est de trouver le h optimale qui minimise le ISE tel que

$$ISE(h) = \int_{\mathbb{T}} [\hat{f}(x) - f(x)]^2 dx = \int_{\mathbb{T}} \hat{f}^2(x) dx - 2 \int_{\mathbb{T}} \hat{f}(x) f(x) dx + \int_{\mathbb{T}} f^2(x) dx.$$

- le dernier terme $\int_{\mathbb{T}} f^2(x) dx$ ne dépend pas du paramètre de lissage h , On peut le supprimer.
- le second terme $\int_{\mathbb{T}} \hat{f}(x) f(x) dx = E[\hat{f}(x)]$.
on peut estimer par :

$$\frac{1}{n} \sum_{j=1}^n f(x_j),$$

avec

$$\hat{f}(x_j) = \frac{1}{(n-1)h} \sum_{i=1, i \neq j}^n K\left(\frac{x_j - x_i}{h}\right).$$

La fonction $UCV(h)$:

$$\frac{1}{n(n-1)h} \sum_{j=1}^n \sum_{i=1, i \neq j}^n K\left(\frac{x_j - x_i}{h}\right).$$

l'estimateur de $ISE(h)$ est alors :

$$UCV(h) = \int_{\mathbb{T}} \hat{f}^2(x) dx - \frac{2}{n(n-1)h} \sum_{j=1}^n \sum_{i=1, i \neq j}^n K\left(\frac{x_j - x_i}{h}\right).$$

et

$$h_{UCV} = \arg \min_h UCV(h).$$

Validation croisée biaisée

La notion de validation croisée biaisée, a été introduite par Scott et Terrell en 1987 [40] , l'idée de cette méthode est de trouver la valeur de h qui minimise un estimateur du $AMISE$. définit par :

$$AMIS(h) = \frac{h^4}{4} \varphi_2^2(x) \int_R [f''(x)]^2 dx + \frac{1}{nh} \int_R [K(\mu)]^2 d\mu.$$

Le paramètre de lissage basé sur la méthode de la validation croisée biaisée est la valeur de h qui minimise un estimateur du $AMISE$. On peut estimer le $AMISE$ si l'on connaît $\int_R [f''(x)]^2 dx$, pour cette raison un estimateur de ce terme est donné par $\int_R [\hat{f}''(x)]^2 dx$ Scott et Terrell (1987) [40] ont montré que

$$\mathbb{E} \left[\int_R [\hat{f}''(x)]^2 dx \right] = \int_R [f''(x)]^2 dx + \frac{1}{nh^5} \int_R [K''(\mu)]^2 d\mu.$$

Ils proposent alors d'estimer $\int_R [\hat{f}''(x)]^2 dx$ par $\int_R [\hat{f}''(x)]^2 dx - \frac{1}{nh^5} \int_R [K''(\mu)]^2 d\mu$. On peut donc estimer le $AMISE$ par :

$$BCV = \frac{h^4}{4} \varphi_2^2(x) \left[\int_R [\hat{f}''(x)]^2 dx - \frac{1}{nh^5} \int_R [K''(\mu)]^2 d\mu \right] + \frac{1}{nh} \int_R [K(\mu)]^2 d\mu.$$

Le paramètre de lissage h choisi par cette méthode est la valeur de h qui minimise $BCV(h)$

$$h_{BCV} = \arg \min_h BCV(h).$$

2.11 Conclusion

Dans ce chapitre, nous avons présenté l'estimateur à noyau pour la fonction densité de probabilité, nous avons étudié ses propriétés stochastiques, et asymptotique et nous avons présenté quelques méthodes pour la sélection du paramètre de lissage.

Dans le chapitre suivant, nous allons présenter en détail les méthodes Plug-in et UCV pour le choix du paramètre de lissage et réaliser une étude comparative de ces deux méthodes .

3.1 Introduction

L'estimation de la fonction de probabilité par la méthode du noyau est principalement conditionnée par le paramètre de lissage h , ce paramètre est un facteur important dans l'estimation par la méthode du noyau. Il représente en quelque sorte une fenêtre qui permet de déterminer le degré de lissage de l'estimation d'une fonction de densité.

Un faible paramètre de lissage implique un faible degré de lissage et résulte en une fonction de densité irrégulière. À l'opposé, une large valeur de h conduit à une estimation trop lissée.

Dans ce chapitre nous allons comparer et parler de la méthode Rule Of Thumb et la méthode validation croisée non biaisée.

3.2 Méthode plug-in

La décision d'un choix optimal pour le paramètre de lissage suppose la spécification d'un critère d'erreur qui puisse être optimisé. Bien sûr, l'optimalité n'est pas un concept absolu : elle est intimement liée aux choix du critère, qui peut faire intervenir à la fois la densité inconnue f et l'estimateur f_h (voir,[50]).

Dans ce cas, on cherche à minimiser l'Erreur Quadratique Moyenne Intégrée (MISE).

$$MISE = \int E[\hat{f}_h(x) - f(x)]^2 dx. \quad (3.1)$$

En utilisant la formule (1.34), le MISE est défini par :

$$MISE(\hat{f}_h(x)) = \frac{h^4}{4} R(f'') \sigma_K^4 + \frac{1}{nh} R(k) + O(h^4 + \frac{1}{nh}). \quad (3.2)$$

et l'Erreur Quadratique Moyenne Intégrée Asymptotique $AMISE = MISE - O(h^4 + \frac{1}{nh})$ est :

$$AMISE(\hat{f}_h(x)) = \frac{h^4}{4} R(f'') \sigma_K^4 + \frac{1}{nh} R(k). \quad (3.3)$$

De l'expression (3.3), on peut déterminer le paramètre de lissage h^* qui minimise l'Erreur Quadratique Moyenne Intégrée Asymptotique :

$$\frac{dAMISE}{dh} = h^3 \sigma_K^4 R(f'') - \frac{1}{nh^2} R(K) = 0.$$

Si on multiplie le tous fois (nh^2) on obtient

$$nh^5 \sigma_K^4 R(f'') - R(K) = 0.$$

Alors

$$h^5 = \frac{R(K)}{n \sigma_K^4 R(f'')}. \quad (3.4)$$

$$h^* = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

$$\frac{d^2 AMISE}{dh^2} = 3h^2 \sigma_K^4 R(f'') + \frac{2}{nh^3} R(K) > 0 \implies h^* \text{ minimise AMISE.}$$

Cette largeur de fenêtre n'est donc pas directement calculable. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R(f'')$ par un estimateur approprié. (voir, [50]) Si on choisit f comme étant la distribution normale de moyenne 0 et de variance σ^2 on a alors :

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

ainsi

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

d'où :

$$f' = - \left(\frac{x-\mu}{\sigma} \right) f(x);$$

ainsi

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

d'où :

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

La moyenne n'est qu'un paramètre de position, tandis que l'écart-type joue un rôle dans la forme de la distribution. On désire donc que h soit relié à l'écart-type. Ainsi, on pose la moyenne égale à zéro. Par conséquent, en considérant une distribution de

variance σ^2 et de moyenne $\mu = 0$ on a alors :

$$\begin{aligned} f''(x) &= \frac{1}{\sigma^2} \left(\left(\frac{x}{\sigma} - 1 \right)^2 \right) f(x); \\ &= \frac{1}{\sigma^3 \sqrt{2\pi}} \left(\left(\frac{x}{\sigma} \right)^2 - 1 \right) e^{-\frac{1}{2} \left(\frac{x}{\sigma} \right)^2}. \end{aligned}$$

En effectuant le changement de variable $z = x/\sigma$ on obtient :

$$\begin{aligned} R(f'') &= \left[\frac{1}{\sigma^3 \sqrt{2\pi}} \left(\left(\frac{x}{\sigma} \right)^2 - 1 \right) e^{-\frac{1}{2} \left(\frac{x}{\sigma} \right)^2} \right]^2; \\ &= \frac{1}{\sigma^5 2\pi} \int (z^2 - 1)^2 e^{-z^2} dz; \\ &= \frac{1}{\sigma^5 2\pi} \int (z^4 - 2z^2 + 1) e^{-z^2} dz; \\ &= \frac{1}{\sigma^5 2\pi} \left[\int z^4 e^{-z^2} dz - 2 \int z^2 e^{-z^2} dz + \int e^{-z^2} dz \right]; \end{aligned}$$

En effectuant le changement de variable $t = z\sqrt{2}$, on obtient :

$$\begin{aligned} R(f'') &= \frac{1}{\sigma^5 2\pi} \left[\int \left(\frac{t}{\sqrt{2}} \right)^4 e^{-\left(\frac{t}{\sqrt{2}}\right)^2} \frac{dt}{\sqrt{2}} - 2 \int \left(\frac{t}{\sqrt{2}} \right)^2 e^{-\left(\frac{t}{\sqrt{2}}\right)^2} \frac{dt}{\sqrt{2}} + \int e^{-\left(\frac{t}{\sqrt{2}}\right)^2} \frac{dt}{\sqrt{2}} \right]; \\ &= \frac{1}{\sigma^5 2\pi} \left[\int \frac{t^4}{4\sqrt{2}} e^{-\frac{t^2}{2}} dt - \frac{t^2}{\sqrt{2}} e^{-\frac{t^2}{2}} dt + \int \frac{e^{-\frac{t^2}{2}}}{\sqrt{2}} dt \right]; \\ &= \frac{1}{\sigma^5 2\pi} \left[\frac{\sqrt{\pi}}{4} \int \frac{t^4}{\sqrt{\pi}} e^{-\frac{t^2}{2}} dt - \sqrt{\pi} \int \frac{t^2}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \sqrt{\pi} \int \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt \right]; \\ &= \frac{1}{\sigma^5 2\pi} \left[\frac{\sqrt{\pi}}{4} E(T^4) - \sqrt{\pi} E(T^2) + \sqrt{\pi} \int \phi(t) dt \right]. \end{aligned}$$

En considérant que $E(T^2) = V(T) = \int t^2 \phi(t) dt = 1$ pour une loi normale centrée réduite, et en considérant que l'intégrale de la densité ϕ sur tout son domaine est égale à 1, et $E(T^4) = V(T) = \int t^4 \phi(t) dt = 3$ pour la loi normale centrée réduite, on a finalement :

$$\begin{aligned} R(f'') &= \frac{1}{\sigma^5 2\pi} \left[\frac{\sqrt{\pi}}{4} 3 - \sqrt{\pi} + \sqrt{\pi} \right]; \\ &= \frac{3}{8\sigma^5 \sqrt{\pi}}. \end{aligned}$$

De plus si on utilise un noyau gaussien, on aura :

$$R(K) = \int K^2(u) du = \int \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \right)^2 du = \frac{1}{2\pi} \int e^{-u^2} du = \frac{1}{2\sqrt{\pi}}. \quad (3.5)$$

Alors la valeur pour h^* notée dans ce cas h_{rot} est obtenue en substituant les valeurs obtenue dans la formule (3.4) on obtien :

$$h_{rot} = 1,06\sigma n^{-\frac{1}{5}}. \quad (3.6)$$

Il suffit donc d'estimer σ à partir des données et de substituer cet estimateur dans la formule ci-dessus. D'après Silverman en (1986), cette formule donnera de bons résultats si la population est réellement normalement distribuée. Celle ci peut aussi donner une distribution trop lissée si la population est plutôt multimodale. Dans ce cas, de meilleurs résultats peuvent être obtenus si l'on utilise une mesure plus robuste de l'étendue. Au lieu d'utiliser σ on utilise plutôt l'interquartile IQ défini par (voir,[50]) :

$$IQ = \frac{X_{3n/4} - X_{n/4}}{1.34}. \quad (3.7)$$

où $X_{(n/4)}$ et $X_{(3n/4)}$ sont respectivement le premier et le troisième quartile. Dans le cas où X suit une loi normale l'écart interquartile est $IQ = 1.394$. h_{rot} de l'équation (3.6) devient alors (voir,[50]) :

$$h_{rot} = 1,06IQ n^{-\frac{1}{5}}. \quad (3.8)$$

Cette formule peut aussi donner une distribution trop lissée si la vraie densité est multimodale. La meilleure des deux méthodes peut être obtenue en utilisant un estimateur adaptatif de l'étendue. C'est à dire, en utilisant A au lieu de σ dans la formule (3.6) où A est défini par $A = \min(\sigma, IQ)$. La formule pour h_{rot} devient alors (voir,[50]) :

$$h_{rot} = 1,06A n^{-\frac{1}{5}}. \quad (3.9)$$

3.2.1 Algorithme de la méthode plug-in(Rull Of Thumb)

Les étapes de l'algorithme sont :

Algorithme(méthode plug-in(Rull Of Thumb)) :

Début (Entrer $X_{1 \leq i \leq n}$, un échantillon de taille n)

Si L'échantillon $X_{1 \leq i \leq n}$ issu d'une variable aléatoire de la loi $N(\mu, \sigma^2)$ alors :

$A = \min(1.394, \sigma)$,

sinon

Ordonnet l'échantillon par ordre croissant,

$$IQ = (X_{(3n/4)} - X_{(n/4)})/1.34,$$

$$A = \min(IQ, \sigma),$$

Fin Si

$$h_{rot} = 1.06An^{-1/5},$$

Fin

3.3 Validation croisée non biaisée

Description de la méthode "validation croisée non biaisée"

Une méthode appelée Validation croisée non biaisée a été proposée par Rudemo et Bowman en 1984. Le critère consiste à choisir le paramètre de lissage qui minimise un estimateur convenable (voir, [29]) :

$$ISE(h) = \int_{\mathbb{T}} [\hat{f}_h(x) - f(x)]^2 dx = \int_{\mathbb{T}} \hat{f}_h^2(x) dx - 2 \int_{\mathbb{T}} \hat{f}_h(x) f(x) dx + \int_{\mathbb{T}} f^2(x) dx. \quad (3.10)$$

Puisque $\int_{\mathbb{T}} f^2(x) dx$ 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 :

$$UCV(h) = ISE(h) - \int_{\mathbb{T}} f^2(x) dx = \int_{\mathbb{T}} \hat{f}_h^2(x) dx - 2 \int_{\mathbb{T}} \hat{f}_h(x) f(x) dx. \quad (3.11)$$

On veut premièrement trouver un estimateur de $\int_{\mathbb{T}} \hat{f}_h(x) f(x) dx$. Nous remarquons que

$$\int_{\mathbb{T}} \hat{f}_h(x) f(x) dx = E(f_h(x)) \quad (3.12)$$

L'estimateur empirique de $\int_{\mathbb{T}} \hat{f}_h(x) f(x) dx$, est alors $\frac{1}{n} \sum_{i=1}^n \hat{f}_{h,i}(x_i)$.

Le critère à optimiser est alors :

$$UCV(h) = \int_{\mathbb{T}} \hat{f}_h^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{h,i}(x_i). \quad (3.13)$$

Où

$$\hat{f}_{h,i}(x_i) = \frac{1}{(n-1)h} \sum_{1 \leq i \leq n, i \neq j} K\left(\frac{x - x_j}{h}\right)$$

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 :

$$MISE(h) - R(f) = E\left[\int (f(x) - f_h(x))^2 dx\right] - R(f) = E\left[\int f_h^2(x) dx - 2 \int f_h(x) f(x) dx\right]. \quad (3.14)$$

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

On a d'une part,

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

D'autre part,

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

Ce qui implique que

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

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

En utilisant l'équation (3.13), le critère UCV(h) devient :

$$UCV(h) = \frac{R(K)}{nh} + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \left[\int \frac{1}{n^2 h^2} K\left(\frac{x-x_i}{h}\right) K\left(\frac{x-x_j}{h}\right) dx - \frac{2}{n(n-1)h} K\left(\frac{x_i-x_j}{h}\right) \right]. \quad (3.15)$$

$R(k) = \int k^2(u) du$. Nous noterons h_{UCV} l'estimateur de h qui minimise $UCV(h)$

$$h_{UCV} = \underset{h}{\operatorname{argmin}} UCV(h).$$

De plus, si K est un noyau gaussien alors le critère $UCV(h)$ est donné par la proposition suivante :

Proposition. 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 utilisant le noyau gaussien on obtient :

$$UCV(h) = \frac{2}{n(n-1)h\sqrt{2\pi}} \left(n - \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{h}\right)^2\right) \right) + \frac{1}{2\sqrt{\pi}n(n-1)h} \sum_{i=1}^n \sum_{j=1, j \neq i}^n \exp\left(-\frac{1}{4}\frac{(x_i - x_j)^2}{h^2}\right). \quad (3.16)$$

Remarque Cette méthode présente deux points faibles(voir,[17]) :

1. le résultat de simulation peut se révéler extrêmement variable d'un échantillon à l'autre.
2. la fonctionnelle à minimiser a souvent tendance à présenter plusieurs minimums locaux.

3.3.1 Algorithme de la méthode validation croisée non biaisée

Pour trouver le paramètre de lissage optimale, noté h_{ucv} par validation croisée non biaisée, on minimise numériquement $UCV(h)$. Par exemple on peut faire une recherche exhaustive du minimum $UCV(h)$ sur un grand nombre de valeurs possibles du paramètre de lissage ou il suffit de donner un algorithme afin de calculer le paramètre de lissage optimal. En utilisant le noyau gaussien. Les étapes de l'algorithme sont(voir,[50]) :

Algorithme(validation croisée non biaisée UCV) :

Début (Génération d'un échantillon $x_{1 \leq i \leq n}$)

Somme1 = 0, Somme2 = 0 ;

Pour i = 1 à n faire

Pour j = 1 à n faire

Si $i \neq j$

som1 = $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_i - x_j}{h}\right)^2\right)$;

Somme1 = Somme1 + Som1 ;

Som2 = $\frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}\frac{(x_i - x_j)^2}{h^2}\right)$

Somme2 = Somme2 + Som2 ;

Fin pour

Fin pour

$UCV(h) = \frac{2}{(n-1)h\sqrt{2\pi}} - \frac{2}{n(n-1)h} \text{somme1} + \frac{1}{n(n-1)h} \text{somme2}$;

Fin

$h_{ucv} = \min_h UCV(h)$.

Nous souhaitons tester les deux méthodes de sélection considérées dans notre

étude. Notre objectif est de :
 Comparer ces deux méthodes de sélection du paramètre de lissage.

3.4 Simulation

3.4.1 Plan de simulation

Nous nous contenterons de faire des simulations et d'observer le comportement asymptotique de l'estimateur à noyau calculé à partir d'un échantillon simulé à partir d'une loi connue f . Ceci, nous permettra de savoir si empiriquement l'estimateur f_h converge vers f .

Nous utilisons pour les simulations des échantillons de lois connues de taille de plus en plus grande (10, 50, 500 et 1000). Pour chaque type de densité et pour chaque taille d'échantillon 50 répétitions d'expérience ont été conduites.

Nous utilisons 3 densités tests. Nous avons choisi des densités présentant différents aspects :

D1 : La loi normale $N(0,1)$; $x \rightarrow \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$.

D2 : La combinaison linéaire de deux densités de loi normale $N(1, \frac{16}{49})$ et $N(-1, \frac{16}{49})$ affectées par le même coefficient $\frac{1}{2}$; $x \rightarrow \frac{1}{2}N(1, \frac{16}{49}) + \frac{1}{2}N(-1, \frac{16}{49})$.

D3 : La combinaison linéaire de deux densités de loi normale $N(0, 1)$ et $N(1.5, \frac{1}{3})$ affectées par les coefficients $\frac{3}{4}$ et $\frac{1}{4}$; $x \rightarrow \frac{3}{4}N(0, 1) + \frac{1}{4}N(1.5, \frac{1}{3})$.

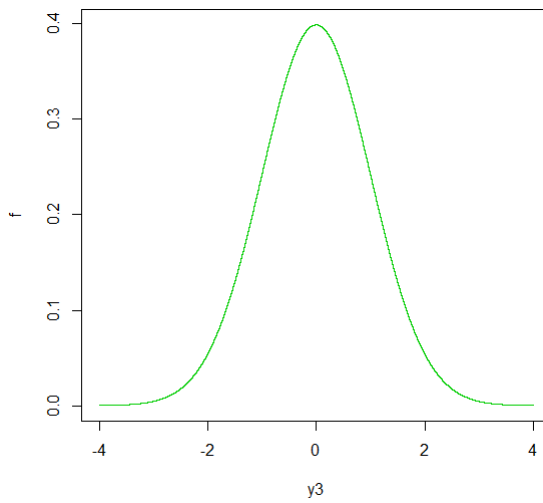


FIGURE 3.1 – Loi normale $N(0, 1)$.

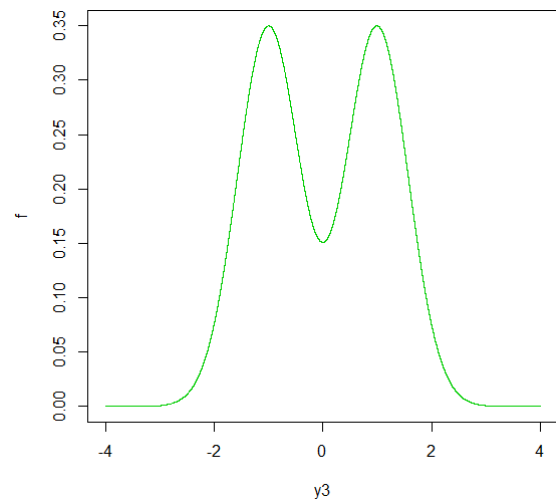


FIGURE 3.2 – Mélange de la normal $\frac{1}{2}N(1, \frac{16}{49}) + \frac{1}{2}N(-1, \frac{16}{49})$.

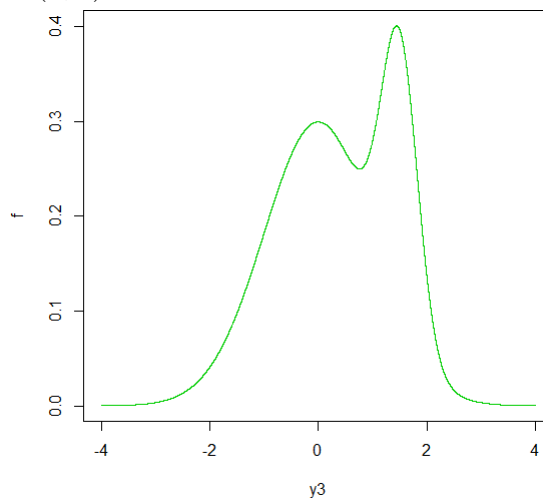


FIGURE 3.3 – Mélange de la normal $\frac{3}{4}N(0, 1) + \frac{1}{4}N(1.5, \frac{1}{9})$.

Algorithme

L'algorithme de simulation que nous avons utilisé comporte quatre phases :

1. Simuler un échantillon de taille n pour une densité qui appartient à une famille paramétrique connue : **D1**, **D2** ou **D3** ;
2. Calculer le paramètre de lissage optimal h_{opt} en utilisant la méthode Rull Of Thumb et la méthode validation croisée non biaisée ;
3. Construire l'estimateur par la méthode du noyau (en utilisons le noyau gaussien) à partir des observations ;
4. Tracer les deux courbes : la densité test et la densité estimée.

Notons, que dans la programmation des méthodes de sélection du paramètre de lissage h , nous avons choisi comme noyau, le noyau gaussien.

Les simulations et les graphiques ont été réalisés à l'aide du logiciel R. Nous avons utilisé la version 3.5.3 pour la programmation.

R est un système d'analyse statistique et graphique créé par Ross Ihaka et Robert Gentleman (voir, [29]). Il est à la fois un langage et un logiciel qui comporte de nombreuses fonctions pour les analyses statistiques et les graphiques.

Le langage R permet, par exemple, de programmer des boucles qui vont analyser successivement différents jeux de données. Il est aussi possible de combiner dans le même programme différentes fonctions statistiques pour réaliser des analyses plus complexes (voir, [29]).

3.4.2 Résultats de simulation

Les résultats de simulation sont donnés sous forme de tableaux et de graphiques. Les tableaux 3.1, 3.2 et 3.3 contiennent les résultats suivants :

1. N : la taille de l'échantillon .
2. h_{rot} : Le paramètre de lissage asymptotique optimal pour le modèle de densité testé, pour la méthode Rull Of Thumb.
3. h_{ucv} : Le paramètre de lissage asymptotique optimal pour le modèle de densité testé, pour la méthode validation croisée non biaisée.
4. MSE : L'erreur quadratique moyenne pour chaque méthode de sélection.
5. \bar{h}_{rot} : La moyenne des paramètres de lissage pour 50 répétitions, pour la méthode Rull Of Thumb.
6. \bar{h}_{ucv} : La moyenne des paramètres de lissage pour 50 répétitions, pour la méthode validation croisée non biaisée.
7. \overline{MSE} : La moyenne des erreurs moyenne quadratique pour chaque méthode de sélection, avec pour 50 répétitions.

CHAPITRE 3. APPLICATION NUMÉRIQUE

Les tableaux 3.1, 3.3 et 3.2 nous permettent de comparer les résultats obtenus par les deux méthodes, pour les différentes densités tests : D1, D2 et D3.

N	10	50	100	500	1000
h_{rot}	0.625965400	0.55571260000	0.4388309000	0.3166028000	0.2362395000
h_{ucv}	0.999948200	0.63907480000	0.4244909000	0.3469260000	0.2544716000
$MSE(h_{rot})$	0.001532281	0.00054070000	0.0002906710	5.471696e-05	3.234213e-05
$MSE(h_{ucv})$	0.001533013	0.00045130000	0.0002914687	5.181784e-05	3.170548e-05
\bar{h}_{rot}	0.650721700	0.48346020000	0.4193614000	0.3055098000	0.2317135000
\bar{h}_{ucv}	0.933311400	0.71178950000	0.5628688000	0.3996117000	0.2907629000
$\overline{MSE}(h_{rot})$	0.001631175	0.00078074222	0.0002944770	5.751253e-05	3.282965e-05
$\overline{MSE}(h_{ucv})$	0.002504886	0.00075465590	0.0004474740	6.846421e-05	3.894635e-05

TABLE 3.1 – Résultats des simulations effectuées sur la loi D1.

N	10	50	100	500	1000
h_{rot}	0.572737800	0.500243100	0.4551731000	0.4508890000	0.4500414800
h_{ucv}	0.693997300	0.485436800	0.4913712000	0.4577113000	0.4474148000
$MSE(h_{rot})$	0.002262460	0.001117000	0.0005026443	1.135079e-05	4.426034e-06
$MSE(h_{ucv})$	0.002736698	0.001128328	0.0004193273	1.026782e-05	4.422257e-06
\bar{h}_{rot}	0.653253800	0.486267200	0.4209117000	0.3065183000	0.3000183000
\bar{h}_{ucv}	0.959549000	0.739914900	0.5913158000	0.4087617000	0.3135808000
$\overline{MSE}(h_{rot})$	0.002562002	0.001153219	0.0007599637	0.0001363710	2.957375e-06
$\overline{MSE}(h_{ucv})$	0.004774535	0.002622120	0.0009396884	0.0001716434	2.957375e-05

TABLE 3.2 – Résultats des simulations effectuées sur la loi D2.

N	10	50	100	500	1000
h_{rot}	0.8407661000	0.42425960	0.4010610000	0.3195786000	0.2724488000
h_{ucv}	0.9233585000	0.48466840	0.3785311000	0.3551464000	0.3142217000
$MSE(h_{rot})$	0.0054559340	0.00226462	0.0004057155	0.0003583674	0.0001745924
$MSE(h_{ucv})$	0.005428990	0.002328785	0.0004093226	0.0003172068	0.0001057796
\bar{h}_{rot}	0.658198900	0.482003100	0.4220939000	0.3061514000	0.2662382000
\bar{h}_{ucv}	0.960633500	0.723799300	0.5818788000	0.3897641000	0.3388245000
$\overline{MSE}(h_{rot})$	0.006338295	0.002325210	0.0004357713	0.0004045860	0.0002004889
$\overline{MSE}(h_{ucv})$	0.005543448	0.002570753	0.0011182060	0.0004206431	0.0001456385

TABLE 3.3 – Résultats des simulations effectuées sur la loi D3.

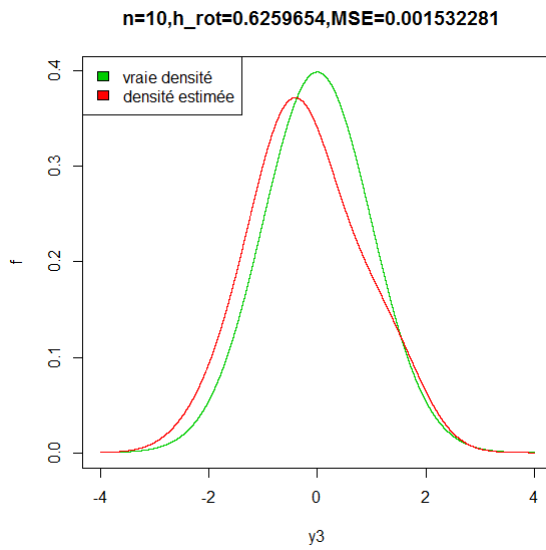


FIGURE 3.4 – Comparaison entre la densité théorique **D1** et celle estimée avec h_{rot} .

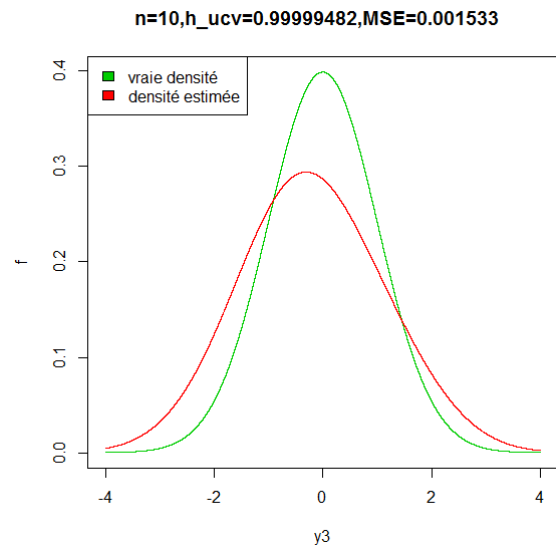


FIGURE 3.5 – Comparaison entre la densité théorique **D1** et celle estimée avec h_{ucv} .

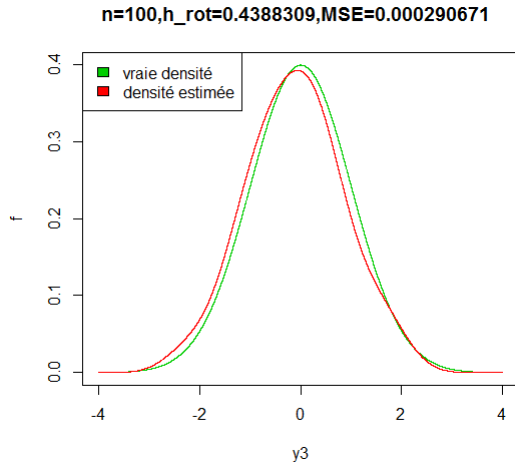


FIGURE 3.6 – Comparaison entre la densité théorique **D1** et celle estimée avec h_{rot} .

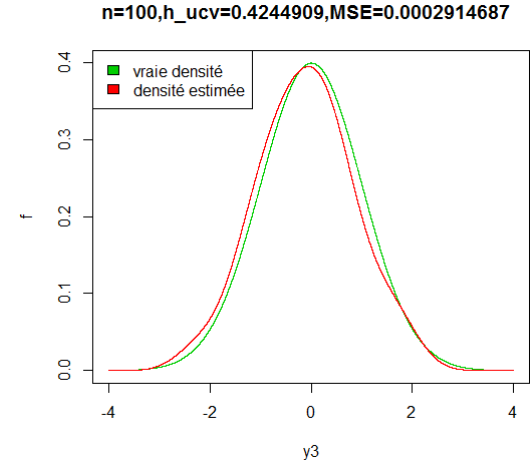


FIGURE 3.7 – Comparaison entre la densité théorique **D1** et celle estimée avec h_{ucv} .

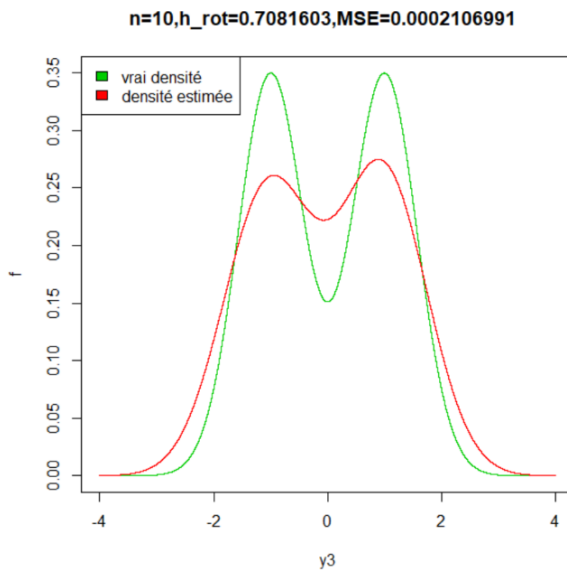


FIGURE 3.8 – Comparaison entre la densité théorique **D2** et celle estimée avec h_{rot} .

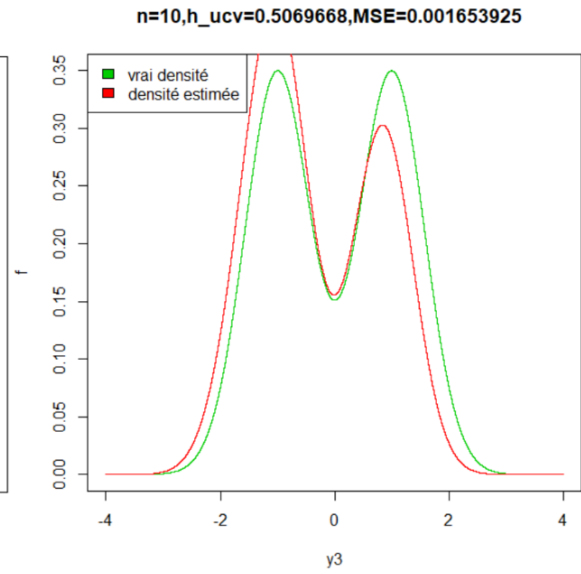


FIGURE 3.9 – Comparaison entre la densité théorique **D2** et celle estimée avec h_{ucv} .

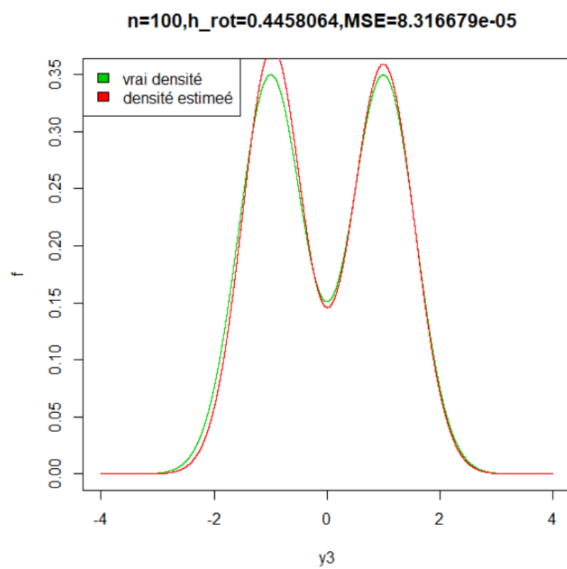


FIGURE 3.10 – Comparaison entre la densité théorique **D2** et celle estimée avec h_{rot} .

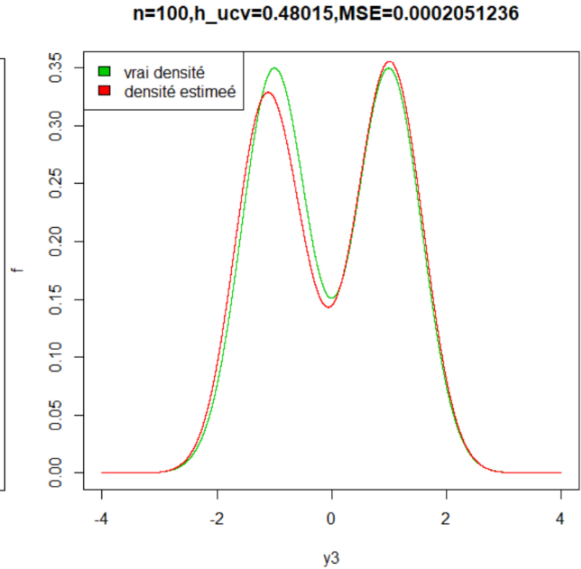


FIGURE 3.11 – Comparaison entre la densité théorique **D2** et celle estimée avec h_{ucv} .

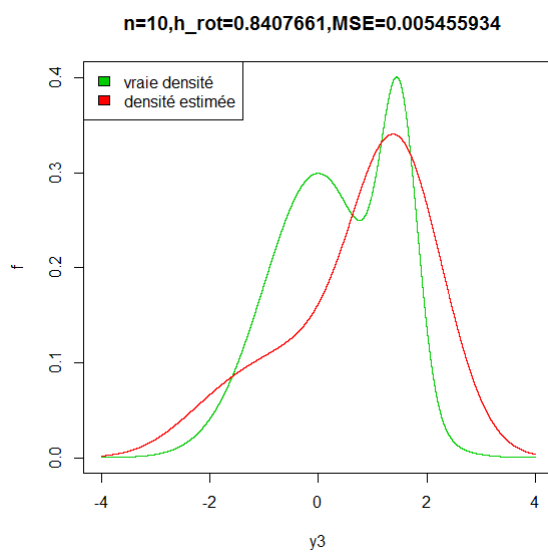


FIGURE 3.12 – Comparaison entre la densité théorique **D3** et celle estimée avec h_{rot} .

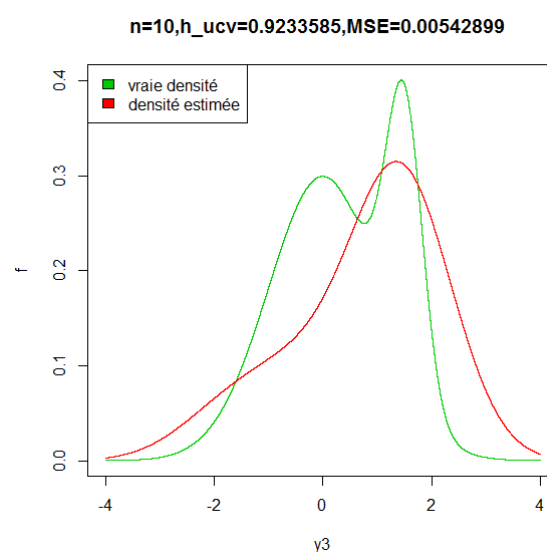


FIGURE 3.13 – Comparaison entre la densité théorique **D3** et celle estimée avec h_{ucv} .

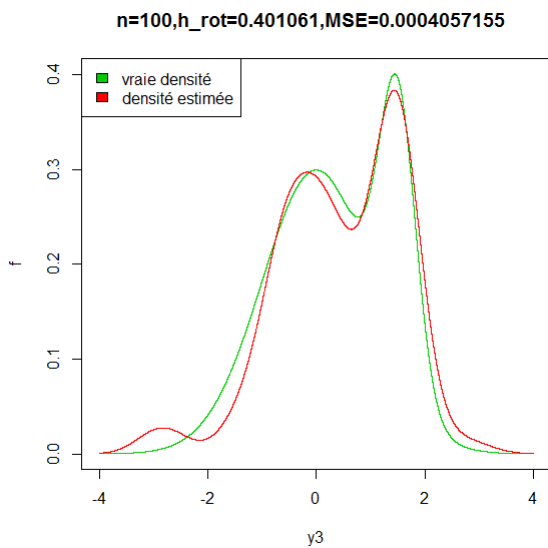


FIGURE 3.14 – Comparaison entre la densité théorique **D3** et celle estimée avec h_{rot} .

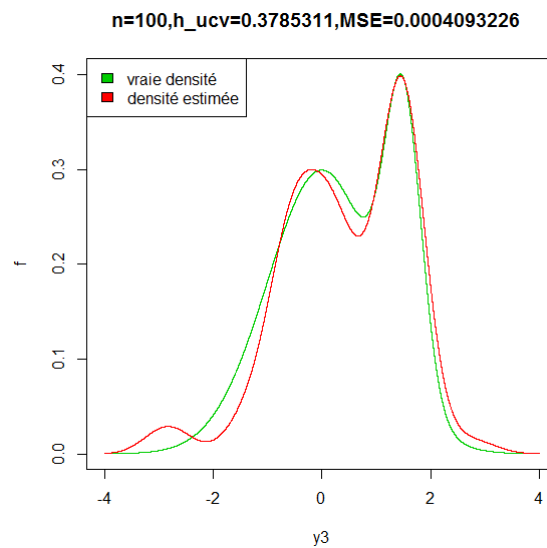


FIGURE 3.15 – Comparaison entre la densité théorique **D3** et celle estimée avec h_{ucv} .

3.5 Interprétation des résultats

3.5.1 Loi normale centrée réduite

Apartir du tableau 3.1 on peut déduire :

1. Le h_{rot} et le h_{ucv} et leur MSE diminuent avec l'augmentation de la taille de l'échantillant la taille de l'échantillant, ça veut dire que Les performances moyennes des estimateurs s'améliorent avec l'augmentation de la taille de l'échantillant. Cette observation est confirmée spar les graphes 3.5, 3.7, 3.4 et 3.6, où la courbe de l'estimateur se rapproche de la densité quand la taille de l'échantillon augmente.
2. Le \overline{MSE}_{rot} est meilleur que \overline{MSE}_{ucv} pratiquement pour toutes les tailles de l'échantillant sauf pour N=50, Ce qui signifie que dans ce cas, la méthode Rulle Of Thumb meilleur que la méthode Validation croisée non biaisée.

3.5.2 Mélange de populations gaussiennes

1. **La loi mixte** $f(x) = \frac{3}{4}N(0, 1) + \frac{1}{4}N(1.5, \frac{1}{9})$

Apartir du tableau 3.2 on peut déduire :

- (a) Le h_{rot} et le h_{ucv} et leur MSE diminuent avec l'augmentation de la taille de l'échantillant la taille de l'échantillant, ça veut dire que Les performances moyennes des estimateurs s'améliorent avec l'augmentation de la taille de l'échantillant. Cette observation est confirmée par les graphes 3.13, 3.15, 3.12 et 3.14, où la courbe de l'estimateur se rapproche de la densité quand la taille de l'échantillon augmente.
 - (b) Le \overline{MSE}_{rot} est moins que le \overline{MSE}_{ucv} pour N=50, N=100 et N=500 mais pour N=10 et N=1000 on voit le contraire, tell que le MSE moyenne de h_{ucv} est moins que le MSE moyenne de h_{rot} , ça veut dire que la méthode Rulle Of Thumb est meilleur que la méthode Validation croisée non biaisée. Ce qui signifie que dans ce cas, on ne peut pas dire qu'une Méthode de sélection est meilleure qu'une autre méthode.
2. **La loi mixte** $f(x) = \frac{1}{2}N(-1, \frac{16}{49}) + \frac{1}{2}N(1, \frac{16}{49})$

Apartir du tableau 3.3 on peut déduire :

- (a) Le h_{rot} et le h_{ucv} et leur MSE diminuent plus on augmente la taille de l'échantillant, ça veut dire que Les performances moyennes des estimateurs s'améliorent lorsque la taille de l'échantillon augmente. Cette observation est confirmée sur le graphe 3.8, 3.10, 3.9 et 3.11, où la courbe de l'estimateur et adapte à la densité quand la taille de l'échantillon augmente.

- (b) Le \overline{MSE}_{rot} est moins que le \overline{MSE}_{ucv} Quelles que soient la taille de l'échantillant, ça veut dire que la méthode Rulle Of Thumb, Ce qui signifie que dans ce cas, la Méthode Rulle Of Thumbe est meilleure que la méthode Validation croisée non biaisée.

3.6 Conclusion

Dans ce chapitre, nous avons comparé entre les deux méthodes : Rull Of Thumb et validation croisée non biaisé, et nous avons conclu que la méthode Rull Of Thumb est meilleure que la méthode Validation croisée non biaisée dans le cas ou l'échntillan simulér corespondant a une population gaussian. **Remarque :** Nous avons effectué 50 expériences, si le nombre d'essais était de 100 ou 1000, les résultats seraient meilleurs.

CONCLUSION GÉNÉRALE

Dans ce travail, nous nous sommes intéressés à l'estimation non paramétrique de la fonction densité de probabilité.

Dans un premier temps, nous avons présenté les différentes méthodes d'estimation, à savoir, la méthode des histogrammes, la méthode des fonctions orthogonales l'estimateur par les fonctions splines. D'autre part, nous nous sommes intéressés à l'estimateur par la méthode du noyau. Nous avons présenté l'estimateur en détails, Nous avons donné ses propriétés statistique et asymptotique.

Un intérêt bien particulier est porté sur le choix du paramètre de lissage. Nous avons présenté les différentes méthodes de sélection de cette fenêtre de lissage, ainsi que la fenêtre obtenue de chaque méthode. Une étude comparative est réalisée pour comparer deux méthodes de sélection de la fenêtre de lissage, à savoir, la méthode plug-in et la méthode validation croisée non biaisée. Afin d'analyser et de comparer les deux méthodes, une étude de simulation est effectuée sur des densités cibles et pour différentes tailles d'échantillon. Les résultats obtenus montrent que presque pour toutes les tailles de l'échantillon, l'erreur MSE associée à l'approche Rull Of Thumb est plus petite que celle associée à validation croisée non biaisée. Cette constatation est confirmée graphiquement .

BIBLIOGRAPHIE

- [1] S.Adjabi, Cours stat : Statistique non paramétrique, 2019 - 2020.
- [2] S.Adjabi, N. Saadi. On the estimation of the probability density by trigonometric series. Communications in Statistics-Theory and Methods, 38(3583-3595), 2009.
- [3] H.Benchoulak. Bandes de confiance pour les fonctions de densité et de régression. memoire de magistère en mathématiques .université mentouri-constantine .2012.
- [4] A.Berlinet, L. A.Devroye, comparison of kernel density estimates. Pub.Inst. Stat. Univ. Paris XXXVIII. (1994).
- [5] A. W.Bowman, An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 1984.
- [6] D.Bosq, J.Lecoutre. Théorie de l'estimation fonctionnelle. Economicaedition, 1967.
- [7] A.W. Bowman. An alternative method of cross-validation for the smoothing density estimates. Biometrika, (71) :353-360, 1984.
- [8] B.M. Brown and S. Chen. Beta-bernstein smoothing for regression curves with compact supports. Scandinavian Journal of Statistics, 1999.
- [9] S. Chen. Beta kernel for regression curve. Statistica Sinica, (10) :73-92, 2000.
- [10] S. Chen. A beta kernel estimation for density functions. Computational Statistics and Data Analysis, 1999.
- [11] N. Cencov. Evaluation of unknow distribution density from observation. SovMaths, (3) :1559-1562, 1962.
- [12] R. Cuevas, A.Cao, and W. A.Gonzdez-Manteiga, comparative study of several smoothing methods in density estimation. Computational Statistics and Data Analysis, 1994.
- [13] R. Coudret. G.Durrieu, and J.Saracco, Estimateurs a noyau bimodaux densité bimodale et comparaison avec estimateurs non paramériques, 2012.

- [14] P.Deheuvels and Hominal, Estimation non paramétrique de la densité compte tenu d'informations sur le support. *Revue de Statistique Appliquée*, 27, 47,68 (1979).
- [15] L. Devroye. *Nonparametric density estimation : The 11, view, new york.* John Wiley, 1985.
- [16] V. A. Epanechnikov, Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl*, 1969.
- [17] P. Hall and J. S. Marron, Local minima in cross-validation function. *Journal of the royal statistical society*, 1991.
- [18] P.Hall,R.Wolff and Q.Yao, Methods for estimating a conditional distrib function.*Journal of american statistical Association* 94,p :154-163.(1999).
- [19] F.E. Harrell and C.E. Davis. A new distribution-free quantile estimator. *Biometrika*, 1982.
- [20] J. Hermans and K. Vandebroek. A stepwise discriminant analysis program using density estimation. in *Compstat*, 1974.
- [21] M.C. Jones. Simple boundary correction for kernel density estimation. *Statistical Computing*, 1993.
- [22] S.Julian, An Assessment of Hermite Function Based Approximation of Mutuel Infor- mation Applied to Independant Component Analysis.10.745-756.(2008).
- [23] R. Kronmal, and M. Tarter, SThe estimation of probabilitydensities and cumulatives by fourierseries methods, *J. Amer. Statist. Assoc*,925-952, 1968.
- [24] M. Lejeune and P. Sarda. Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis*, 1992.
- [25] H.G.Müller, Smooth optimum kernel estimators near endpoints. *Biometrika*, (78) :521-530, 1991.
- [26] J . S. Marron and D. Ruppert, Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society, Series B*, 1994.
- [27] P.Marron, Smoothed cross-validation. *Probability Theory and Related Fields*,1992.
- [28] E. Nadaraya , On non parametric estimation density function and regression. *Theory Probab P.P.L*,1965.
- [29] B.Nehme, Techniques non-additives d'estimation de la densité de probabilité.Thèse docteur.Universié montpellier 2.2010.
- [30] B.U park and S.J.Marron, Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 1990.
- [31] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist*, (33) :1065-1076, 1962.

-
- [32] M. Restle, Estimating Cumulative Distributions Par Spline Smoothing. Thèse doctorat. école polytechnique fédérale de lausanne :{23, 41} ,2001.
- [33] M. Rosenblatt, Remarks in some nonparametric estimates of a density function. *Ann. Math. Statist*, (27) :832-837, 1956.
- [34] M. Rudemo, Empirical choice of histogram and kernel density estimators, *Scandinavian Journal of Statistics*, (9) :65-78, 1982.
- [35] G. Roussas, Asymptotic normality of the kernel estimate of a probability density function under association. *Statist. Probab. Lett. Statistics. Probability Letters*, (2000), 1–12.
- [36] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 1982.
- [37] G. Saporta. *Probabilités, Analyse des données et Statistiques*. Technip, Paris, 1990.
- [38] D. Scott, R. A. Tapia and J.R. Thompson. Kernel density estimation revisited. *Nonlinear Analysis, Theory, Method and Applications*, 1 :339-372, 1977.
- [39] D. W. Scott, R. A. Tapia, and J. R. Thompson, Non parametric probability density estimation by discrete maximum penalized-likelihood criteria. *The annals of statistics*, 1980.
- [40] D.W scott and G.R.terrell, Biased and unbiased cross-validation in density estimation, *Journal of the American Statistical Association*,1987.
- [41] D.W. Scott and G.R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, (82) :{1131 – 1146}, 1987 .
- [42] D. Scott , W. Multivariate Density Estimation :Theory, Practice, and Visualization. Wiley Interscience, New York, 1992.
- [43] S.J. Sheather, M. A. Jones, reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, 1991.
- [44] B. W. Silverman, *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.
- [45] J. Simonoff, *S. Smoothing methods in statistics*. Springer Science et Business Media, 2012.
- [46] M. Tarter, M. Lock. *Model-free curve estimation*. New York : chapman and Hall, 1993.
- [47] P. M. Wand, and M. C. Jones, *Kernel smoothing*. Crc Press, 1994.
- [48] W. E. Wansouw, Kokonendji, and Kolyang, D. T. *Nonparametric estimation for probability mass function with Disake*, 1995.
- [49] Y. Ziane .L'estimation non paramétrique de l'indice de variabilité et la distribution des densités heavy tailed. thèse doctorat. université A. Mira-Bejaia :{51, 60} ,2016.
-

- [50] N.Zougab .Etude comparative des méthodes de sélection du paraètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau.mémoire de magister en mathématiques appliquées.université de Béjaia, Mai 2007.

Résumé : Dans ce travail, nous avons présenté L'estimateur de la fonction densité de probabilité par la méthode du noyau. nous présentons les estimateurs de la densité et leurs propriétés statistiques (biais, variance, erreur moyenne quadratique), ainsi que les propriétés asymptotiques, Nous nous intéressons à l'estimateur à noyau. Le paramètre de lissage qui intervient dans la forme de l'estimateur est sélectionné par la méthode classique validation croisée non biaisée (UCV) et la méthode plug-in (Rull Of Thumb). Des études de simulation sur des modèles de densités connues sont effectuées pour comparer la méthode validation croisée non biaisée (UCV) avec la méthode plug-in (Rull Of Thumb).

Mots clés : Densité de probabilité, méthode du noyau, paramètre de lissage, UCV, Rull Of Thumb.

Abstract : In this work, we presented the estimator of the probability density by the kernel method. We present density estimator and their statistical properties (bias, variance, mean square error) as well as asymptotic properties. We are intersted in the kernel estimator. The smoothing parameter that occurs in the form of the estimator is selected by the classical unbiased cross-validation method (UCV) and the Rull Of Thumb method. Simulation studies on known density models are performed to compare the cross-validation method whith the Rull Of Thumb method.

Key words : probability density, kernel method, smoothing parameter, UCV, plug-in.