

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider, Biskra
Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie
Département de Mathématiques



Mémoire présenté pour obtenir le diplôme de
Master en “**Mathématiques Appliquées**”

Option : **Statistique**

Par :

ADJAIMI CHOROUK

Titre :

**Sur le choix du noyau
de l'estimateur non paramétrique de la densité**

Membres du Comité d'Examen :

Mr. NECIR Abdelhakim	Prof.	U. Biskra	Président
Mr. YAHIA Djabrane	Prof.	U. Biskra	Rapporteur
Melle. ZOUAOUI Nour El-Houda	MCB	U. Biskra	Examinatrice

Soutenu Publiquement le 26/06/2022

Dédicace

Je dédie cet humble travail

À mes chers parents qui ont toujours été présent pour moi et qui m'ont donné la vie. Pour de millions des raisons, ils m'ont donné jour après jour autant d'amour et de confiance, ils ont veillé à m'encourager tout au long de ma vie, à me donner de l'aide et à me protéger. Que le grand Dieu me les protège.

À mes chers frères et sœurs, ma source de joie et de bonheur

À tous mes amis.

À tous mes proches.

À tous mes professeurs et tous ceux qui ont contribué à mon éducation.

À mes camarades de promotion 2021/2022.

À tous ceux qui m'aiment.

À tous ceux que j'aime.

Remerciements

Merci ALLAH de m'avoir donné la santé, la patience et le courage pendant et de longues années d'étude pour terminer mes études et arriver à cet humble travail.

Mes remerciements à mon cher père, ma mère et frères pour leurs sacrifices et leur patience avec moi pour réaliser mon succès et mes rêves, qu'Allah les protège tous.

Mes remerciements à mon Encadreur, Professeur Yahia Djabrane pour m'avoir aidé, conseillé et encadré cette année, je suis très heureuse d'avoir travaillé avec lui.

Mes remerciements aux membres de Jury : Professeur Necir A. (en qualité de Président) et Melle. Zouaoui (Examinatrice), pour avoir accepté d'étudier et d'évaluer ce travail.

Merci à tous mes professeurs et mes amis proches.

Merci à tous.

Résumé du mémoire

Dans ce mémoire, nous nous intéressons à l'estimation non paramétrique de la densité de probabilités par la méthode du noyau. En particulier, nous étudions le choix optimal du noyau pour l'obtention d'un meilleur estimateur possible de la fonction de densité.

Malgré qu'en estimation à noyau le choix de cette fonction est de faible effet. Nous mettons l'accent dans ce mémoire sur les types du noyau existents, leurs constructions et propriétés, le problème du choix optimal d'une telle fonction et l'influence de ce choix sur la qualité d'estimation, ainsi que le problème des effets de bord rencontré dans les cas d'estimation des densités à supports compacts.

Des exemples et simulations en utilisant le logiciel statistique R, nous permettent d'observer et d'illustrer ce phénomène.

Mots-clés : *Estimation non paramétrique, densité de probabilité, méthode du noyau, paramètre de lissage, choix optimal de noyau.*

Abréviations et Notations

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

Notation	Signification
(X_1, \dots, X_n)	échantillon de taille n .
F	fonction de répartition.
f	densité de probabilité.
F_n	Fonction de répartition
\hat{f}_n	L'estimateur à noyau de la densité de probabilité f .
$E(\cdot)$	Espérance mathématique.
$Var(\cdot)$	Variance.
\mathbb{R}	Ensemble des nombres réels.
\mathbb{N}	Ensemble des nombres entiers naturels.
K	Noyau, kernel.
h	Paramètre de lissage
K_{opt}^*	Fonction de noyau Epanochnikov.
h_{opt}^*	Paramètre de lissage optimale.

MSE	:	erreur quadratique moyenne.
$MIES$:	erreur quadratique moyenne intégré.
$MIES$:	Erreur quadratique moyenne asymptotique.
$AMISE$:	Erreur quadratique moyenne intégrée asymptotique.
\bar{X}	:	Moyenne empirique.
$b(.)$:	biais d'un estimateur
$[]$:	La partie entière
$v.a$:	Variable aléatoire
$i.i.d$:	indépendantes et identiquement distribuées
c à d	:	C'est-à-dire

Table des matières

Dédicace	i
Remerciements	ii
Résumé du mémoire	iii
Abréviations et Notations	iv
Table des matières	vi
Table des figures	viii
Liste des tableaux	ix
Introduction	1
1 Généralités sur l'estimation non paramétrique de la densité	4
1.1 Estimation de la densité par histogramme	5
1.2 Estimateur à noyau de la densité	7
1.2.1 Propriétés de l'estimateur à noyau	9
1.2.2 Majoration de risque quadratique	13

2	Choix du noyau dans l'estimation de la densité	16
2.1	Construction d'un noyau d'ordre ℓ	16
2.2	Erreurs quadratique moyenne et intégrée	18
2.2.1	Erreur quadratique moyenne (MSE)	18
2.2.2	Erreur quadratique moyenne intégrée	20
2.2.3	Choix du paramètre de lissage	21
2.2.4	Règle de référence à une distribution standard	22
2.3	Types du noyau	25
2.3.1	Noyaux usuels symétriques	25
2.3.2	Noyaux polynomiaux ou de Legendre	26
2.3.3	Noyaux de Gram-Charlier	27
2.3.4	Noyaux de Laguerre	28
2.3.5	Noyaux divers	28
2.4	Noyau optimal	29
2.5	Choix du noyau et problème des effets de bord	31
	Conclusion	34
	Bibliographie	35

Table des figures

1.1	Estimation par histogramme, $X \rightarrow N(0;1)$, $m = 20$ à gauche et $m = 50$ à droite	6
1.2	Biais, Variance et choix optimal de h : sur et sous lissage.	14
2.1	Estimation à noyau de la densité : K est gaussien, h varié, $n = 1000$.	24
2.2	Estimation à noyau de la densité : K est d'Epanechnikov, h varié, $n = 1000$	24
2.3	Courbes des noyaux usuels	27
2.4	Estimation NP de la densité par différents noyaux	31

Liste des tableaux

2.1	Efficacite des noyaux symetriques	30
-----	---	----

Introduction

À ce jour, la plupart des modèles statistiques utilisés pour décrire les différents phénomènes, sont des modèles (distributions) classiques tel que la distribution exponentielle, gaussienne et celle de Poisson,... etc. Dans le cadre classique de la statistique, nous utilisons souvent la statistique paramétrique dont le modèle est défini par un paramètre $\theta \in \mathbb{R}^k$ pour un certain entier k . Typiquement $M = \{P_\theta, \theta \in \mathbb{R}^k\}$ est le modèle qui décrit la distribution des variables aléatoires observées. Par opposition, en statistique non-paramétrique, le modèle n'est pas décrit par un nombre fini de paramètres. Mais il s'agit d'un cadre plus général et offre des avantages intéressants, car il ne suppose aucune hypothèse à priori sur les données.

La densité de probabilité est une fonction de très importantes outils et applications en statistique. Elle est rarement connue dans la pratique, son estimation à partir d'un ensemble fini d'observations est un problème fondamental en statistique. On trouve deux types d'approches d'estimation de la densité de probabilité : estimation paramétrique et non paramétrique. Dans le premier cas, le problème de l'estimation paramétrique (par maximum de vraisemblance ou la méthode des moments) on cherche à représenter la distribution des observations par une fonction $f(x, \theta)$ où le paramètre θ ($\theta \in \mathbb{R}^k$) est la seule inconnue. L'estimation non-

paramétrique consiste à estimer, à partir des observations, une fonction inconnue f . D'une manière équivalente estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble d'observations. Différentes méthodes existent pour l'estimation non paramétrique, on peut citer la méthode de l'histogramme et la méthode du noyau. Une bonne introduction à cette problématique peut être trouvée dans les livres de Silverman (1986) [12] et Tsybakov (2009) [13].

La méthode d'estimation non paramétrique par noyaux, qui a été proposé par Rosenblatt en 1956 [10], puis amélioré par Parzen en 1962 [9] est basée sur deux quantités : une fonction appelée noyau (kernel en anglais) et une suite réel h_n que l'on appellera la fenêtre (bandwidth en anglais), qui dépend de la taille n de l'échantillon.

Souvent, les noyaux employés pour ce type d'estimation sont symétriques (dit aussi classiques). Comme exemple, le noyau triangulaire, parabolique (Epanechnikov en 1969 [5]), et gaussien,... Mais ces estimateurs connaissent quelques inconvénients comme le problème de bords ou des dérivées supérieures nulles. Ainsi, il existe des noyaux asymétriques tels que les noyaux bêta et gamma proposés par Chen en 1999 [1] et en 2000 [2] respectivement qui permettent de résoudre ces problèmes.

Notre objectif dans ce mémoire est le choix optimal du noyau pour l'obtention d'un meilleur estimateur possible de la fonction de densité.

Ce mémoire est composé de deux chapitres et d'une conclusion générale. Le premier chapitre est composé de deux parties bien distinctes. La première partie parle de la méthode d'estimation par histogramme, et dans la deuxième partie nous

donnons des détails sur la méthode d'estimation à noyau classique. En particulier, nous rappelons les différentes propriétés asymptotiques de cette méthode.

Le deuxième chapitre est composé de cinq parties. La première porte sur la construction d'un noyau. La deuxième partie est un aperçu sur les erreurs quadratique moyenne et intégrée et leurs propriétés. La troisième partie présente le problème du choix du paramètre de lissage. La dernière partie aborde le problème du choix du noyau. Des exemples et des simulations par le logiciel R, sont donnés pour une bonne explication du sujet.

Chapitre 1

Généralités sur l'estimation non paramétrique de la densité

Le but de ce chapitre est de présenter les différentes méthodes non paramétriques (NP) d'estimation de la fonction de densité de probabilités. En particulier, la méthode d'estimation par histogramme et celle du noyau. Nous donnons aussi leurs principales propriétés statistique.

Soit X une variable aléatoire (va) de densité de probabilité inconnue f :

$$P(X \in [a, b]) = \int_a^b f(x)dx, \quad \forall a, b \in \mathbb{R}$$

et de fonction de répartition F :

$$\forall x \in \mathbb{R}, \quad F : x \rightarrow F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

Supposons que nous avons n observations X_1, X_2, \dots, X_n i.e, un échantillon de taille n (suite indépendantes et identiquement distribuées notée *i.i.d*) provenant de X .

L'un des problèmes majeurs consiste à utiliser cet échantillon pour construire un estimateur qui soit le plus proche possible de la densité f .

1.1 Estimation de la densité par histogramme

L'estimateur le plus ancien et le plus utilisé de la densité est **l'histogramme**. Il s'agit d'approximation de f par une fonction étagée, qui permet d'étudier la répartition des données et parfois de justifier certains choix de modèles (*Comte*, [3]).

Supposons que la densité f est définie sur un intervalle borné $[a, b] \subset \mathbb{R}$. On dispose alors de m intervalles disjoints (classes forment $[a, b]$), dont l'union est $[a, b]$ de mêmes longueurs :

$$\forall k = 1, \dots, m; \quad C_{k,m} = \left[a + (k-1) \frac{b-a}{m}, a + k \frac{b-a}{m} \right[.$$

Pour chacun de ces intervalles, on compte le nombre d'observations parmi les X_1, X_2, \dots, X_n . Ce comptage est effectué en formulation mathématiques par une somme d'indicatrices que l'on écrit sous la forme :

$$\sum_{i=1}^n I_{X_i \in C_{k,m}}$$

où

$$I_{X_i \in C_{k,m}} = \begin{cases} 1 & \text{si } X_i \in C_{k,m}, \\ 0 & \text{si } X_i \notin C_{k,m}. \end{cases}$$

Ensuite, ce cardinal est normalisé par l'effectif n , et nous définissons l'estimateur

de f par histogramme à m classes comme suit :

$$\hat{f}_m(x) = \sum_{i=1}^m \left(\frac{1}{n} \sum_{i=1}^n I_{X_i \in C_{k,m}} \right) I_{C_{k,m}}(x), \quad \forall k = 1, \dots, m. \quad (1.1)$$

Remarque 1.1.1 *Un histogramme est dit régulier lorsque tous les intervalles $I_{k,D}$ de la partition ont la même longueur. De plus, l'estimateur donné par (1.1) est une fonction étagée et donc discontinue.*

Exemple 1.1.1 *Pour illustrer l'importance du nombre de classes utilisé dans l'histogramme, nous proposons dans la Figure (1.1) un histogramme de basé sur 5000 observations tirées aléatoirement d'une loi normale centrée réduite $N(0, 1)$. Ces observations sont réparties sur un intervalle de référence $\Omega = [-4, 4]$, avec le nombre de classes $m : 20, 50$. Il est clair, qu'une bonne estimation est assurée par un nombre de classes élevé.*

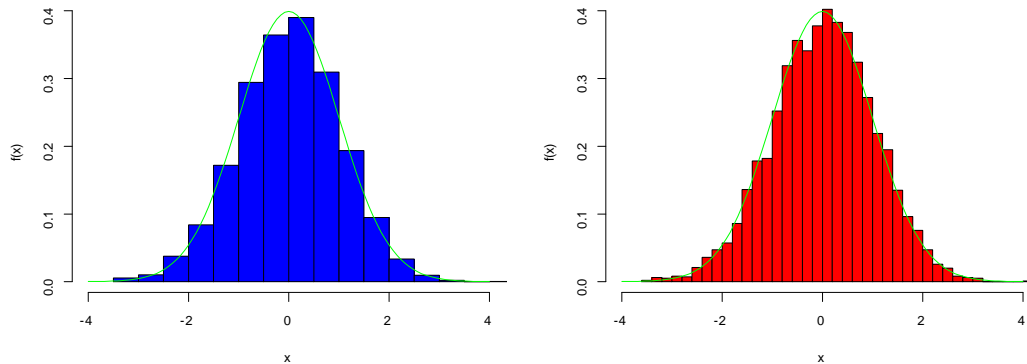


FIG. 1.1 – Estimation par histogramme, $X \rightarrow N(0; 1)$, $m = 20$ à gauche et $m = 50$ à droite

Propriété 1.1.1 *La vitesse de convergence de l'estimateur par histogramme de la densité est de l'ordre $n^{-2/3}$ (voir Comte, [3] pour plus de détails).*

1.2 Estimateur à noyau de la densité

Considérons la fonction de répartition empirique :

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

$$= \begin{cases} 0 & \text{si } X_{(1)} > x \\ \frac{i}{n} & \text{si } X_{(i)} \leq x < X_{(i+1)} \\ 0 & \text{si } X_{(n)} < x \end{cases} .$$

où $I(\cdot)$ désigne la fonction indicatrice. Il s'agit d'un estimateur sans biais de F et de variance

$$Var(F_n(x)) = \frac{1}{n} F(x)(1 - F(x)).$$

Donc F_n est un estimateur convergent (consistant) de F . D'après la loi forte des grands nombres, presque sûrement, quand $n \rightarrow \infty$

$$F_n(x) \rightarrow F(x), \quad \forall x \in \mathbb{R}.$$

Pour $h > 0$ assez petit on a :

$$f(x) = \frac{dF}{dx}(x) \approx \frac{F(x+h) - F(x-h)}{2h}.$$

En remplaçant ici F par F_n , on obtient

$$\hat{f}_n^R(x) = \frac{F_n(x+h) - F_n(x-h)}{2h}.$$

La fonction \hat{f}_h^R est un estimateur de f appelé estimateur de *Rosenblatt* [10]. On peut aussi l'écrire sous la forme :

$$\begin{aligned}\hat{f}_n^R(x) &= \frac{1}{2h} \left[\frac{1}{n} \sum_{i=1}^n I(X_i \leq x+h) - \frac{1}{n} \sum_{i=1}^n I(X_i \leq x-h) \right], \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} I(x-h < X_i \leq x+h), \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} I\left(-1 < \frac{X_i - x}{h} \leq 1\right), \\ &= \frac{1}{nh} \sum_{i=1}^n K_0\left(\frac{X_i - x}{h}\right).\end{aligned}$$

où

$$K_0(u) = \frac{1}{2} I(-1 < u \leq 1), \quad \forall u \in \mathbb{R}.$$

Ce cas particulier nous amène à la généralisation suivante, qui revient à *Parzen* (1962) et qui consiste à prendre comme estimateur de la densité :

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right). \quad (1.2)$$

où $K : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction intégrable, et telle que $\int_{-\infty}^{\infty} K(u)du = 1$. C'est l'estimateur à noyau de la densité ou estimateur de *Parzen - Rosenblatt*. La fonction K est dite noyau (Kernel en anglais) et h est appelée paramètre de lissage ou fenêtre (*bandwidth* en anglais).

Pour $n \in \mathbb{N}^*$, on supposera que la fenêtre, h dépend de n et on la notera h_n , la suite $(h_n)_{n \geq 1}$ tendant vers 0 lorsque $n \rightarrow \infty$.

Remarque 1.2.1 (*Tsybakov* (2003), p 3) *Si K est un noyau positif et si X_1, X_2, \dots, X_n sont fixés. Alors, la fonction $x \rightarrow \hat{f}_n(x)$ est une densité de probabilité.*

1.2.1 Propriétés de l'estimateur à noyau

Soit $x_0 \in \mathbb{R}$ fixé. On définit le risque quadratique moyenne (Mean Squared Error MSE) de \hat{f}_n au point x_0 par :

$$MSE(x_0) = E \left[\left(\hat{f}_n(x_0) - f(x_0) \right)^2 \right]$$

où E désigne l'espérance par rapport à loi de (X_1, X_2, \dots, X_n) . On peut décomposer cette quantité en 2 termes, respectivement le biais $b(\cdot)$ et la variance $\sigma^2(\cdot)$ de l'estimateur au point x_0 :

$$MSE(x_0) = \text{Biais}^2(\hat{f}_n(x_0)) + \text{Var}(\hat{f}_n(x_0)) := b^2(x_0) + \sigma^2(x_0)$$

En notant

$$b(x_0) = E \left[\hat{f}_n(x_0) \right] - f(x_0) \text{ et } \sigma^2(x_0) = E \left[\left(\hat{f}_n(x_0) - E \left[\hat{f}_n(x_0) \right] \right)^2 \right].$$

Proposition 1.2.1 *Si f est bornée, et si K est de carré intégrable (i.e. $\int_{\mathbb{R}} K^2(u) du < \infty$).*

Alors, $\forall x_0 \in \mathbb{R}$, $h > 0$ et $n \geq 1$,

$$\text{Var}(\hat{f}_n(x_0)) \leq \frac{C_1}{nh}.$$

avec

$$C_1 := \|f\|_{\infty} \int_{\mathbb{R}} K^2(u) du.$$

Preuve. (voir Comte, p.48). En effet,

$$\begin{aligned}
 \text{Var}(\hat{f}_n(x_0)) &= \text{Var}\left(\frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x_0}{h}\right)\right) = \frac{1}{(nh)^2} \text{Var}\left(\sum_{i=1}^n K\left(\frac{X_i - x_0}{h}\right)\right) \\
 &= \frac{1}{(nh)^2} \sum_{i=1}^n \text{Var}\left(K\left(\frac{X_i - x_0}{h}\right)\right), \quad (\text{car les } X_i \text{ sont indépendantes}), \\
 &\leq \frac{1}{nh^2} E\left(K^2\left(\frac{X_i - x_0}{h}\right)\right) \leq \frac{1}{nh^2} \int_{\mathbb{R}} K^2\left(\frac{t - x_0}{h}\right) f(t) dt, \\
 &\leq \frac{1}{nh^2} \|f\|_{\infty} \int_{\mathbb{R}} K^2\left(\frac{t - x_0}{h}\right) dt, \quad (\text{car } f \text{ est bornée et } K \text{ intégrable}), \\
 &\leq \frac{1}{nh^2} \|f\|_{\infty} \int_{\mathbb{R}} h K^2(u) du, \quad \left(\text{en posant } u = \left(\frac{t - x_0}{h}\right)\right), \\
 &\leq \frac{1}{nh} \|f\|_{\infty} \int_{\mathbb{R}} K^2(u) du \leq \frac{C_1}{nh}.
 \end{aligned}$$

■

Remarque 1.2.2 Si $h = h_n$, et si $hn \rightarrow \infty$, alors $\sigma^2(x_0) \rightarrow 0$.

Définition 1.2.1 (voir Tsybakov (2003), p. 5) soit $\ell \in \mathbb{N}^*$, on dite que $K : \mathbb{R} \rightarrow \mathbb{R}$ est une noyau d'ordre ℓ si les fonctions $u \rightarrow u^j K(u)$, $j = 1, \dots, \ell$, sont intégrables et vérifient :

$$\int_{\mathbb{R}} u^j K(u) du = 0, \quad \int_{\mathbb{R}} K(u) du = 1, \quad \forall j = 1, \dots, \ell.$$

Pour la valeur de biais, on a

$$b(x_0) = E\left[\hat{f}_n(x_0)\right] - f(x_0) = \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{t - x_0}{h}\right) f(t) dt - f(x_0).$$

On va étudier le comportement de $b(x_0)$ en fonction de h , en supposant que la densité f et le noyau K sont suffisamment régulier. Dans la suite $[\beta]$ désigne le plus grand entier strictement inférieur à β ($\beta \in \mathbb{R}$).

Définition 1.2.2 Soit T un intervalle de \mathbb{R} et soient $\beta > 0$, $L > 0$. La classe de Höldre $\sum(\beta, L)$ sur T est définie comme l'ensemble de toutes les fonctions $f : T \rightarrow \mathbb{R}$ telles que la dérivée $f^{(\ell)}$ existe, $\ell = \lfloor \beta \rfloor$ et vérifie

$$|f^{(\ell)}(x_1) - f^{(\ell)}(x_2)| \leq L |x_1 - x_2|^{\beta - \ell}, \quad \forall x_1, x_2 \in T.$$

Proposition 1.2.2 (Comte, [3, p. 50]) Si la densité f est dans la classe de Höldre $\sum(\beta, L)$ sur $T = \mathbb{R}$ et si K est un noyau d'ordre $\ell = \lfloor \beta \rfloor$ tel que

$$\int_{\mathbb{R}} |u|^\beta |K(u)| du < \infty.$$

Alors, $\forall x_0 \in \mathbb{R}$, $h > 0$

$$|b(x_0)| \leq C_2 h^\beta$$

avec

$$C_2 := \frac{L}{\ell!} \int_{\mathbb{R}} |u|^\beta |K(u)| du.$$

Preuve. On a,

$$\begin{aligned} b(x_0) &= E \left[\hat{f}_n(x_0) \right] - f(x_0), \\ &= \frac{1}{nh} E \left(\sum_{i=1}^n K \left(\frac{X_i - x_0}{h} \right) \right) - f(x_0), \\ &= \frac{1}{h} E \left(K \left(\frac{X_1 - x_0}{h} \right) \right) - f(x_0), \\ &= \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{t - x_0}{h} \right) f(t) dt - f(x_0), \\ &= \int_{\mathbb{R}} K(u) [f(x_0 + uh) - f(x_0)] du, \quad \left(\text{en posant } u = \left(\frac{t - x_0}{h} \right) \right), \end{aligned}$$

En effectuons un développement de Taylor au voisinage de x_0 d'ordre ℓ , à $f(x_0 + uh)$,

$$f(x_0 + uh) = f(x_0) + \frac{(uh)^1}{1!} f^{(1)}(x_0) + \dots + \frac{(uh)^\ell}{\ell!} f^{(\ell)}(x_0 + \theta uh)$$

où $\theta \in [0, 1]$, comme K est d'ordre $l = \lfloor \beta \rfloor$, on a donc

$$\begin{aligned} b(x_0) &= \int_{\mathbb{R}} K(u) \left[f(x_0) + \frac{(uh)^1}{1!} f^{(1)}(x_0) + \dots + \frac{(uh)^\ell}{\ell!} f^{(\ell)}(x_0 + \theta uh) - f(x_0) \right] du \\ &= \int_{\mathbb{R}} K(u) \frac{(uh)^\ell}{\ell!} f^{(\ell)}(x_0 + \theta uh) du \\ &= \int_{\mathbb{R}} K(u) \frac{(uh)^\ell}{\ell!} (f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0)) du. \end{aligned}$$

En utilisant le fait que f est dans une classe de Höldre $\sum(\beta, L)$ sur \mathbb{R} , on a :

$$|f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0)| \leq L |(x_0 + \theta uh) - x_0|^{\beta - \ell} = L |\theta uh|^{\beta - \ell}$$

et puisque $\theta \in [0, 1]$,

$$|f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0)| \leq L |uh|^{\beta - \ell}, \quad \forall u \in \mathbb{R}.$$

Ainsi, $\forall h > 0$,

$$\begin{aligned} |b(x_0)| &= \left| \int_{\mathbb{R}} K(u) \frac{(uh)^\ell}{\ell!} (f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0)) du \right| \\ &\leq \int_{\mathbb{R}} |K(u)| \frac{|uh|^\ell}{\ell!} |f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0)| du \\ &\leq \frac{|h|^\ell}{\ell!} \int_{\mathbb{R}} |K(u)| |u|^\ell |(f^{(\ell)}(x_0 + \theta uh) - f^{(\ell)}(x_0))| du \end{aligned}$$

Donc

$$\begin{aligned}
 |b(x_0)| &\leq \frac{|h|^\ell}{\ell!} \int_{\mathbb{R}} |K(u)| |u|^\ell L |uh|^{\beta-\ell} du \\
 &\leq L \frac{|h|^\beta}{\ell!} \int_{\mathbb{R}} |K(u)| |u|^\beta du \\
 &\leq L \frac{h^\beta}{\ell!} \int_{\mathbb{R}} |K(u)| |u|^\beta du \\
 &\leq C_2 h^\beta,
 \end{aligned}$$

avec $C_2 := \frac{L}{\ell!} \int_{\mathbb{R}} |K(u)| |u|^\beta du$. ■

1.2.2 Majoration de risque quadratique

Si f et K vérifient les hypothèses des Propositions (1.1.1) et (1.1.2). On obtient

$$MSE(x_0) = \text{Biais}^2(\hat{f}_n(x_0)) + \text{Var}(\hat{f}_n(x_0)) \leq C_2^2 h^{2\beta} + \frac{C_1}{nh} = \psi(h).$$

La valeur de la fenêtre h qui minimise ce majorant du MSE est :

$$h_{opt} = \text{Arg min}_h MSE(\hat{f}_n(x_0)), \text{ tel que } \frac{d}{dh} \psi(h_{opt}) = 0.$$

En dérivant ψ par rapport à h , on obtient donc,

$$\psi'(h_{opt}) = 2\beta h^{2\beta-1} - \frac{C_1}{nh^2} = 0.$$

Par conséquent,

$$h_{opt} = \left(\frac{C_1}{2\beta C_2^2} \right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}}. \quad (1.3)$$

Remarque 1.2.3 Le choix h_{opt} est inaccessible en pratique car il dépende de β

et d'autres constantes qui sont inconnues. En particulier, si nous substituons la valeur de β avec $\beta = 2$ dans (1.3), nous trouvons $h_{\text{opt}} = \left(\frac{C_1}{4C_2^2}\right)^{\frac{1}{5}} n^{-4/5}$. Cela montre que la vitesse de convergence de l'estimateur à noyau est d'ordre $n^{-4/5}$. Elle est donc meilleure que la vitesse $n^{-2/3}$ obtenue pour les histogrammes, (voir Comte, [3] pour plus de détails).

Remarque 1.2.4 Lorsque la fenêtre h est très petite ($h \rightarrow 0$), le biais de l'estimateur à noyau est très petit face à sa variance. Dans ce cas, on parle de sous-lissage. Lorsque h grandit, la variance devient petite et c'est le biais qui devient dominant. On parle alors d'un effet de sur-lissage, voir la figure 1.2 suivante.

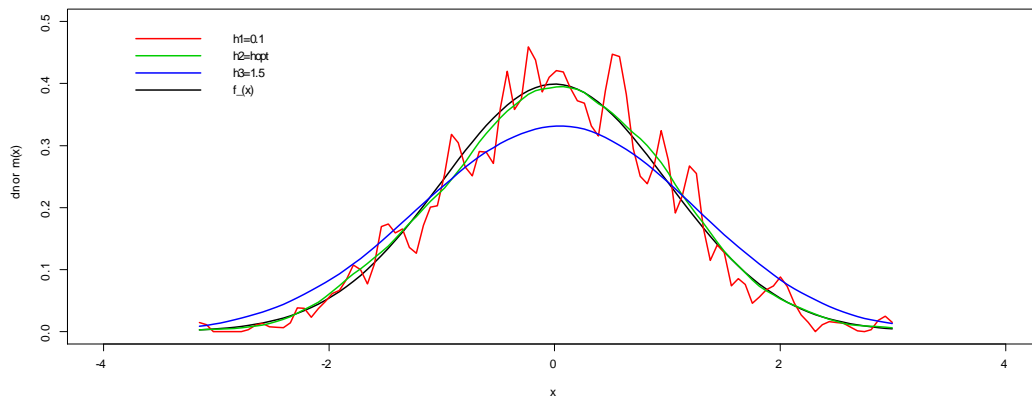


FIG. 1.2 – Biais, Variance et choix optimal de h : sur et sous lissage.

Finalement, la fenêtre h a un rôle important dans l'estimation à noyau. Elle détermine la qualité d'estimation et contrôle le lissage (sur lissage, sous lissage) d'une façon très sensible.

En effet, le choix de h tourne autour de la condition

$$h \rightarrow 0 \quad \text{et} \quad nh \rightarrow \infty. \quad (1.4)$$

Des faibles valeurs de h implique un sous-lissage ($Biais \rightarrow 0$) mais la variance augmente (la courbe en rouge dans la figure 1.2). Lorsque h grandit, la variance devient petite et c'est le biais qui augmente (la courbe en bleu dans la figure 1.2), donc sur-lissage de l'estimateur. En pratique, il est primordial donc de trouver la bonne valeur de h qui permet d'éviter le sous-lissage et le sur-lissage. Ce problème est étudié par plusieurs auteurs (citons par exemple les références [12, 13, 14]).

Chapitre 2

Choix du noyau dans l'estimation de la densité

Mais malgré qu'en estimation à noyau c'est le paramètre h qui contrôle la qualité d'estimation et que le noyau K est de faible effet, il est important d'étudier le problème du choix du noyau d'une façon détaillée. Cela est le sujet principal du mémoire. Dans ce deuxième chapitre, nous mettons l'accent sur les types du noyau existents, leurs constructions et propriétés, le problème du choix optimal d'une telle fonction K , et l'influence de ce choix sur la qualité d'estimation.

2.1 Construction d'un noyau d'ordre ℓ

Nous suivons la même démarche décrite dans *Tsybakove* (2003), p. 8, pour construire un noyau d'ordre ℓ . Soit $\{\varphi_m(\cdot)\}_{m=0}^{\infty}$ la base orthonormée des polynômes de Legendre dans $L_2([-1, 1], dx)$ définie par :

$$\varphi_0(x) = \frac{1}{\sqrt{2}}, \quad \varphi_m(x) = \sqrt{\frac{2m+1}{2}} \frac{1}{2^m m!} \frac{d^m}{dx^m} \left[(x^2 - 1)^m \right], \quad \forall m = 1, 2, \dots$$

Pour $x \in [-1, 1]$,

$$\int_{-1}^1 \varphi_m(u) \varphi_k(u) du = \delta_{mk}. \quad (2.1)$$

avec

$$\delta_{mk} = \begin{cases} 1 & \text{si } m = k \\ 0 & \text{si } m \neq k. \end{cases}$$

Proposition 2.1.1 *La fonction K définie par*

$$K(u) = \sum_{m=0}^{\ell} \varphi_m(0) \varphi_m(u) I_{(|u| \leq 1)}. \quad (2.2)$$

est un noyau d'ordre ℓ .

Preuve. Décomposons le polynôme pour tout $j = 0, 1, \dots, \ell$, dans la base orthonormée des polynômes de Legendre. Il existe des réels b_k tels que

$$u^j = \sum_{k=0}^j b_k \varphi_k(u), \quad \forall u \in [-1, 1]. \quad (2.3)$$

Soit K le noyau (2.2). Alors d'après (2.1) et (2.3) :

$$\begin{aligned} \int u^j K(u) du &= \int_{-1}^1 \left(\sum_{k=0}^j b_k \varphi_k(u) \right) \left(\sum_{m=0}^{\ell} \varphi_m(0) \varphi_m(u) \right) du, \\ &= \sum_{k=0}^j \sum_{m=0}^{\ell} \varphi_m(0) b_k \int_{-1}^1 \varphi_k(u) \varphi_m(u) du, \\ &= \sum_{k=0}^j \sum_{m=0}^{\ell} \varphi_m(0) b_k \delta_{mk}, \\ &= \sum_{k=0}^j \varphi_k(0) b_k, \\ &= 0^j = \begin{cases} 1 & \text{si } j = 0 \\ 0 & \text{si } j = 1, \dots, \ell \end{cases}. \end{aligned}$$

■

Exemple 2.1.1 Les deux premiers polynômes de Legendre de degré pair sont :

$$\varphi_0(x) = \frac{1}{\sqrt{2}}, \quad \varphi_2(x) = \sqrt{\frac{5}{2}} \frac{3x^2 - 1}{2}.$$

Donc, un noyau d'ordre 2 est donné par

$$K(u) = \left(\frac{9}{8} - \frac{15}{8}u^2 \right) I_{(|u| \leq 1)}.$$

2.2 Erreurs quadratique moyenne et intégrée

2.2.1 Erreur quadratique moyenne (MSE)

Dans cette section nous supposons les hypothèses suivantes :

- i) La densité f est telle que sa dérivée seconde f'' est continue ($f \in C^2$), de carré intégrable et monotone (sur $(-\infty, M)$ et (∞, M) pour certains $M > 0$).
- ii) La suite $h_n := h$ vérifiée

$$h \rightarrow 0 \quad \text{et} \quad nh \rightarrow \infty, \quad \text{quand } n \rightarrow \infty.$$

- iii) La fonction noyau K est supposée (densité, bornée, symétrique, de moment d'ordre 4 fini).

Proposition 2.2.1 Supposons que la densité f à estimer est bornée sur \mathbb{R} , continue dans un voisinage de x et $f(x) > 0$, et que K est un noyau tel que

$$\int_{\mathbb{R}} K^2(u) du < \infty \quad (\text{i.e., de carré intégrable}), \quad \int_{\mathbb{R}} K(u) du = 1 \quad (\text{densité}).$$

Alors, la variance de l'estimateur \hat{f}_n est donnée par :

$$\text{Var}(\hat{f}_n(x)) = \frac{1}{nh} f(x) R(K) + o\left(\frac{1}{nh}\right). \quad (2.4)$$

avec $R(g) = \int_{\mathbb{R}} g^2(u) du$.

Preuve. Voir Wand et Jones (1995), p. 21. ■

Proposition 2.2.2 *Supposons que la densité f est de classe $C^2(\mathbb{R})$, la dérivée f'' est bornée sur \mathbb{R} , $f''(x) \neq 0$, et que K est un noyau d'ordre 1 tel que*

$$\int_{\mathbb{R}} u^2 |K(u)| du < \infty \text{ (moment d'ordre 2 fini)}, \quad \int_{\mathbb{R}} u^2 K(u) du \neq 0.$$

Alors, le biais de l'estimateur \hat{f}_n en x vérifie, pour tout $n \geq 1$,

$$\text{Biais}(\hat{f}_n(x)) = \frac{h}{2} \mu_2(K) f''(x) + o(h^2). \quad (2.5)$$

avec $\mu_2 = \int_{\mathbb{R}} u^2 K(u) du$.

Preuve. Voir Wand et Jones (1995), p.20. ■

Théorème 2.2.1 (Tsybakov (2003), p. 14) *Supposons que*

i) La fonction K d'ordre 1 vérifiant les conditions

$$\int_{\mathbb{R}} K^2(u) du < \infty, \quad \int_{\mathbb{R}} u^2 |K(u)| du < \infty, \quad \mu_2 = \int_{\mathbb{R}} u^2 K(u) du \neq 0.$$

ii) La densité f est de classe $C^2(\mathbb{R})$ avec $f(x) > 0$, $f''(x) \neq 0$, et la dérivée f'' est bornée sur \mathbb{R} . Alors, l'erreur quadratique moyenne de l'estimateur \hat{f}_n est

donnée par :

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

Preuve. Voir *Tsybakov* (2003), p. 14. En effet, il suffit de rappeler que

$$MSE(\hat{f}_n(x)) = \text{Biais}^2(\hat{f}_n(x)) + \text{Var}(\hat{f}_n(x)),$$

et d'appliquer les Propositions (2.2.1) et (2.2.2). ■

Corollaire 2.2.1 *Du théorème précédent nous déduisons la MSE asymptotique :*

$$AMSE(\hat{f}_n(x)) = \frac{h^4}{4} \mu_2^2(K) f''(x)^2 + \frac{1}{nh} f(x) R(K). \quad (2.6)$$

2.2.2 Erreur quadratique moyenne intégrée

L'erreur quadratique moyenne intégrée (*MISE*) est définie en ce basant sur la formule (2.6) comme suit :

$$\begin{aligned} MISE(\hat{f}_n(x)) &= \int E \left\{ \hat{f}_n(x) - f(x) \right\}^2 dx = \int MSE(\hat{f}_n(x)) dx \\ &= \int_{\mathbb{R}} \text{Biais}^2(\hat{f}_n(x)) dx + \int_{\mathbb{R}} \text{Var}(\hat{f}_n(x)) dx \\ &= \int_{\mathbb{R}} \left[\frac{h^4}{4} \mu_2^2(K) f''(x)^2 + \frac{1}{nh} f(x) R(K) + o\left(\frac{1}{nh}\right) + o(h^2) \right] dx \\ &= \frac{h^4}{4} \mu_2^2(K) R(f'') + \frac{1}{nh} R(K) + o\left(\frac{1}{nh}\right) + o(h^4). \end{aligned}$$

Asymptotiquement, on a

$$AMISE\left(\hat{f}_n(x)\right) = \frac{h^4}{4}\mu_2^2(K)R(f'') + \frac{1}{nh}R(K). \quad (2.7)$$

2.2.3 Choix du paramètre de lissage

Le paramètre de lissage h est un réel positif dont le choix est dominant par rapport au choix du noyau K . On voit que plus le paramètre h est faible plus le biais diminue mais plus la variance augmente, donc sous-lissage de l'estimateur, et lorsque h grandit, la variance devient petite et c'est le biais qui augmente, donc sur-lissage de l'estimateur. Il existe un optimum (**global**) qui est la valeur de h qui minimise l' $AMISE$:

$$h_{opt}^* = \mathbf{Arg\,min}_h \mathbf{AMISE}\left(\hat{f}_n(x)\right).$$

Supposons d'abord que le noyau K est fixé en dérivant (2.7) par rapport à h , on obtient :

$$h_{opt}^* = \left(\frac{R(K)}{\mu_2^2(K)R(f'')}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \quad (2.8)$$

Par minimisation en h de l' $AMSE$, on trouve la fenêtre optimale (**locale**), notée h_{opt} :

$$h_{opt} = \mathbf{Arg\,min}_h \mathbf{AMSE}\left(\hat{f}_n(x)\right),$$

ce qui est équivalent à

$$h_{opt} = \left(\frac{f(x)R(K)}{\mu_2^2(K)f''(x)^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \quad (2.9)$$

Remarque 2.2.1 Ces résultats (2.8) et (2.9) qui donnent le h_{opt} et h_{opt}^* sont ba-

sés sur un choix optimal théorique qui n'est pas utilisable en pratique, car il dépend de la quantité inconnue f'' . Plusieurs méthodes ont été développées pour résoudre ce problème, nous citons :

- 1) Validation croisée des moindres carrés (Silverman (1986), p. 48 – 52).
- 2) Validation croisée biaisée voir (Wand et Jones (1995), p. 65 et 66).
- 3) Validation croisée des probabilités (Silverman (1986), p. 52 – 55).
- 4) Validation croisée lissée (Wand et Jones (1995), p.75 et 79).
- 5) La méthode Plug-in (Wand et Jones (1995), p. 71-75).
- 6) Référence à une distribution standard (Silverman (1986), p. 45 – 48).

2.2.4 Règle de référence à une distribution standard

Un approche très simple et naturelle pour le choix de h consiste à utiliser une famille standard de distribution pour attribuer une valeur au terme $\int_{\mathbb{R}} f''(x)^2 dx$. Silverman (1986) [12] à proposer de se référer à une loi normale pour le calcul de h_{opt}^* . Cette règle de Silverman est basée sur le résultat classique du théorème central limite, qui assure l'approximation de la moyenne empirique par la loi normale, quelque soit la loi de la variable initiale considérée (sous des conditions sur la moyenne et la variance).

Soit X_1, X_2, \dots, X_n une suite de va's de densité de probabilité f , supposons que f appartient à une famille de distributions normales $N(\mu, \sigma^2)$. La quantité inconnue $R(f'')$, s'écrit alors :

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

Donc, l'expression du paramètre de lissage optimal devient :

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

où $\hat{\sigma}^2$ est la variance empirique de la variance σ^2 de X :

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

Remarque 2.2.2 Comme cas particulier de la règle de Silverman, nous avons

a) Si K est un noyau d'Epanechnikov (2.13), alors

$$h_{opt}^* = 2.43\hat{\sigma}n^{-\frac{1}{5}}.$$

b) Si K est un noyau gaussien (2.11), alors

$$h_{opt}^* = 1.06\hat{\sigma}n^{-\frac{1}{5}}.$$

c) Si K est un noyau Quartique (2.12), alors

$$h_{opt}^* = 2.78\hat{\sigma}n^{-\frac{1}{5}}.$$

Exemple 2.2.1 On suppose que l'on a observé un échantillon de taille $n = 1000$, issu d'une va gaussienne et $\hat{f}_n(x)$ l'estimateur à noyau de la densité donné par :

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

Nous prenons différentes valeurs de h : $h = 0.1, 0.4, 0.6$ et 0.9 . Pour le noyau K , nous considérons deux cas :

cas 1 : dans la figure 2.1, nous choisissons un noyau Gaussien à support non compact.

cas 2 : dans la figure 2.2, nous choisissons un noyau d'Epanechnikov à support compact.

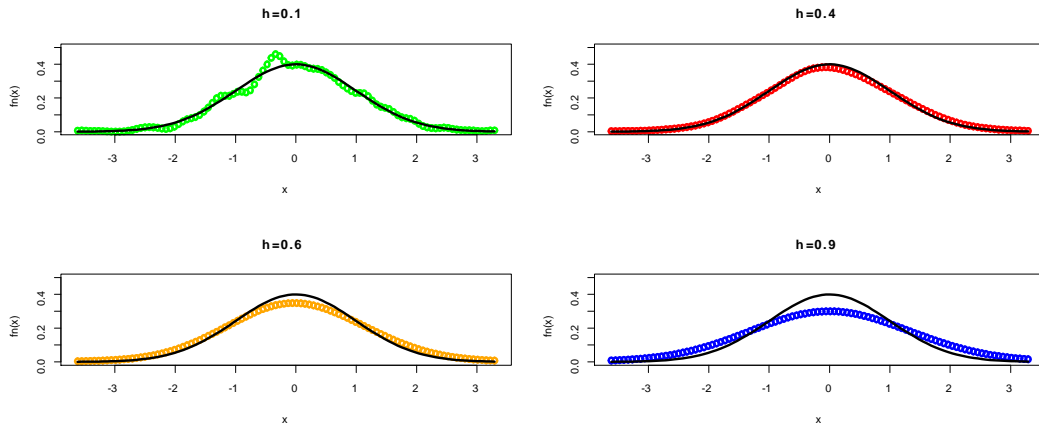


FIG. 2.1 – Estimation à noyau de la densité : K est gaussien, h varié, $n = 1000$.

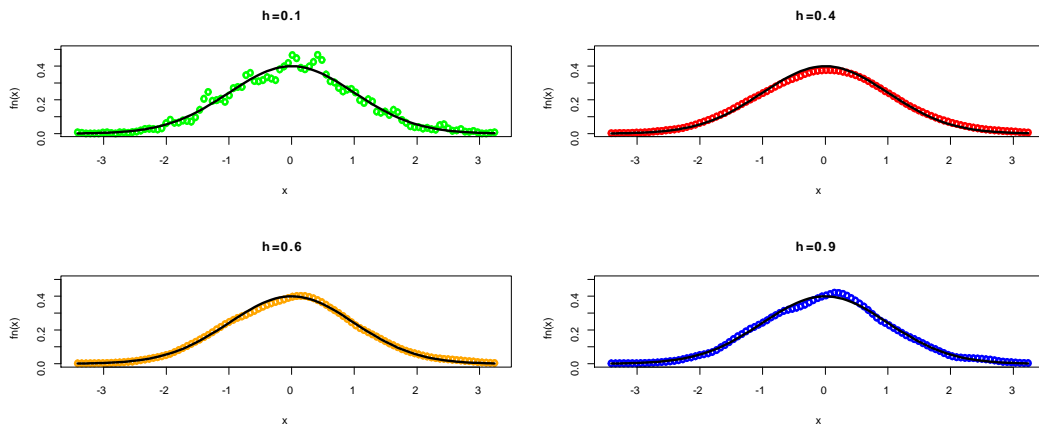


FIG. 2.2 – Estimation à noyau de la densité : K est d'Epanechnikov, h varié, $n = 1000$.

2.3 Types du noyau

L'estimation non paramétrique d'une densité de probabilité par la méthode du noyau nécessite un choix du noyau K . Les noyaux usuels se limitent à la liste suivante :

2.3.1 Noyaux usuels symétriques

Noyau Uniforme (Rosenblatt) : Ce noyau a été proposé par *Rosenblatt* en 1956 [10], l'avantage de ce noyau est la simplicité de sa forme. Il s'écrit sous la forme :

$$K(u) = \begin{cases} \frac{1}{2} & \text{si } |u| \leq 1 \\ 0 & \text{sinon} \end{cases} . \quad (2.10)$$

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 :

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

Noyau Gaussien : Le noyau Gaussien s'écrit sous la forme :

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2), \quad u \in \mathbb{R}. \quad (2.11)$$

Noyau Quartique (biweight) : Ce noyau 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. Ce noyau est d'une forme très

proche du noyau gaussien. Il s'écrit sous la forme :

$$K(u) = \begin{cases} \frac{15}{16} (1 - u^2)^2 & \text{si } |u| \leq 1 \\ 0 & \text{sinon} \end{cases} . \quad (2.12)$$

Noyau Triweight : Il s'écrit sous la forme :

$$K(u) = \begin{cases} \frac{35}{32} (1 - u^2)^3 & \text{si } |u| \leq 1 \\ 0 & \text{sinon} \end{cases} .$$

Noyau Epanechnikov : En 1969, Epanechnikov [5], a donné la forme du noyau K défini par :

$$K(u) = \begin{cases} \frac{3}{4} (1 - u^2) & \text{si } |u| \leq 1 \\ 0 & \text{sinon} \end{cases} . \quad (2.13)$$

Noyau Silverman : Le noyau dû à Silverman s'écrit sous la forme :

$$K(u) = \frac{1}{2} \exp\left(\frac{-|u|}{\sqrt{2}}\right) \sin\left(\frac{|u|}{\sqrt{2}} + \frac{\pi}{4}\right), \quad u \in \mathbb{R}.$$

La représentation graphique des noyaux définis ci-dessus est donnée par la figure 2.3 suivante

2.3.2 Noyaux polynomiaux ou de Legendre

Deheuvels [4], p 10 – 11, décrit une classe de noyaux polynomiaux ou de Legendre. En général

$$K(u) = P(u), \quad |u| < C, \quad (2.14)$$

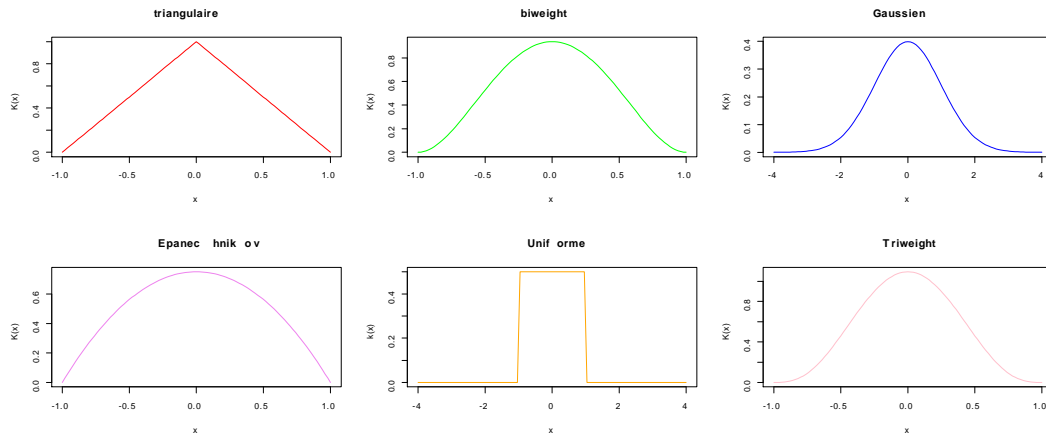


FIG. 2.3 – Courbes des noyaux usuels

où P est un polynôme, $K(u) = 0$ pour $|u| \geq C$. Nous donnons ici des exemples importants de $K(u)$:

- (1) Noyau unité : $K(u) = 1$, si $|u| < \frac{1}{2}$.
- (2) Noyau de Legendre d'ordre 1, $K(u) = \frac{3}{8}(3 - 5u^2)$, si $|u| < 1$.
- (3) Noyau de Legendre d'ordre 2, $K(u) = \frac{15}{128}(15 - 70u^2 + 63u^4)$, si $|u| < 1$.

2.3.3 Noyaux de Gram-Charlier

De même, les noyaux de Gram-Charlier sont construits comme suit :

$$K(u) = P(u) \exp\left(-\frac{1}{2}u^2\right), \quad (2.15)$$

P est un polynôme. Les exemples importants sont :

- (4) Noyau de Gram-Charlier d'ordre 1, $K(u) = \frac{1}{2}(3 - u^2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$.
- (5) Noyau de Gram-Charlier d'ordre 2, $K(u) = \frac{1}{8}(u^4 - 10u^2 + 15) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$.

Remarque 2.3.1 *En particulier, le noyau Gaussien est un noyau de Gram-Charlier.*

2.3.4 Noyaux de Laguerre

La troisième classe de noyaux est celle de Legendre, définie par

$$K(u) = P(|u|) \exp(-|u|), \quad (2.16)$$

P est un polynôme. Elle contient les exemples suivants :

(6) Noyau de Picard : $K(u) = \frac{1}{2} \exp(-|u|)$.

(7) Noyau de Laguerre d'ordre 1, $K(u) = \frac{1}{2}(\frac{1}{2}u^2 - 3|u| + 3) \exp(-|u|)$.

2.3.5 Noyaux divers

Il existe d'autre type de noyaux (citons Deheuvels [4], pour plus de détails), tels que :

(8) Noyau de Cauchy d'ordre r : $K(u) = \frac{2^{2r-2} \Gamma(r)^2}{\pi(2r-1)(1+u^2)^r}$.

(9) $K(u) = \frac{(1-|u|^a)^{(a+1)}}{2a}$, $|u| \leq 1$, dont le cas particulier pour $a = 1$, est le noyau triangulaire : $K(u) = 1 - |u|$.

(10) $K(u) = \frac{\exp(-|u|^a)}{2\Gamma(\frac{1}{a})}$.

(11) Noyau de Fejer, de la Vallée, Poussin : $K(u) = \frac{1}{\pi} \left(\frac{1}{u} \sin(u)\right)^2$.

(12) Noyau de Jackson, de la Vallée, Poussin : $K(u) = \frac{3}{\pi} \left(\frac{1}{u} \sin(u)\right)^4$.

(13) Noyau de Fourier : $K(u) = \frac{1}{\pi} \left(\frac{1}{u} \sin(u)\right)$.

(14) Noyau cosinus : $K(u) = \cos(u)$, pour $|u| \leq \frac{1}{2}\pi$.

Remarque 2.3.2 *Il existe aussi dans la littérature des noyaux (de support) asymétriques, tels que les noyaux bêta et gamma, qui sont proposés par Chen en 1999 en 2000 [1, 2], respectivement.*

2.4 Noyau optimal

Le problème du choix optimal de K , consiste à chercher un noyau optimal sous la contrainte de positivité, ($K \geq 0$). Supposons que l'on a choisi le paramètre de lissage h de telle sorte que l' $AMISE$ soit minimum. On fait le rappel de l'expression asymptotique de cette erreur :

$$AMISE\left(\hat{f}_n(x)\right) = \frac{h^4}{4} \mu_2^2(K) R(f'') + \frac{1}{nh} R(K).$$

avec

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

On remarque que le h_{opt}^* dépend de la densité f qui est inconnue. On remarque aussi que h_{opt}^* convergera vers zéro mais de façon très lente quand n augmente. En substituant h_{opt}^* dans la formule, on montre que si h est choisi de manière optimale $AMISE$ sera :

$$\begin{aligned} AMISE_{h_{opt}^*}\left(\hat{f}_n(x)\right) &= \frac{5}{4} \left\{ \mu_2^2(K) R(f'') R^4(K) \right\}^{\frac{1}{5}} n^{-\frac{4}{5}}. \\ &= \frac{5}{4} C(K) R(f'')^{\frac{1}{5}} n^{-\frac{4}{5}}. \end{aligned}$$

où

$$C(K) = \left\{ \mu_2^2(K) R^4(K) \right\}^{\frac{1}{5}}. \quad (2.17)$$

Cette dernière expression ne dépend plus de h ou de x , elle dépend seulement du noyau K . Pour désigner un noyau optimal dans l'estimation NP de la densité, il faut choisir le noyau K qui minimise la valeur $C(K)$ dans (2.17). Hodges et Lehmann (1956) [6] ont montré que ce problème de minimisation est résolu en

définissant $K_{opt}^*(u)$ comme suit :

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

La notation $K_{opt}^*(u)$ est appelé le noyau d'Epanechnikov (Epanechnikov, 1969 [5]).

Nous pouvons maintenant considérer l'efficacité d'un noyau quelconque K , notée $eff(K)$. En le comparant avec le noyau optimal d'Epanechnikov K_{opt}^* .

Définition 2.4.1 *L'efficacité relative d'un noyau K par rapport à K_{opt}^* est donnée par :*

$$eff(K) = \frac{AMISE(K_{opt}^*)}{AMISE(K)} = \left(\frac{\mu_2^2(K_{opt}^*) R^4(K_{opt}^*)}{\mu_2^2(K) R^4(K)} \right)^{\frac{1}{5}} \leq 1.$$

Nous résumons dans le Tableau 2.1, quelques noyaux et leurs efficacités respectives. Le lecteur peut éventuellement se référer à Silverman [11] et Deheuvels [4] pour plus de détails.

Noyau $K(u)$	$eff(K)$
Noyau Epanechnikov	1
Noyau Quartique (biweight)	$\left(\frac{3087}{3125}\right)^{\frac{1}{2}} = 0.9939$
Noyau Gaussien	$\left(\frac{36\pi}{125}\right)^{\frac{1}{2}} = 0.9512$
Noyau triangulaire	$\left(\frac{243}{250}\right)^{\frac{1}{2}} = 0.9859$
Noyau Uniforme (Rosenblatt)	$\left(\frac{108}{125}\right)^{\frac{1}{2}} = 0.9295$
Noyau Noyau unité	0.9424
Noyau Noyau de Picard	0.8051
Noyau de Cauchy d'ordre 1	0
Noyau de Cauchy d'ordre 2	0.7380

TAB. 2.1 – Efficacite des noyaux symetriques

Exemple 2.4.1 *L'objectif de cet exemple est la comparaison entre différents noyaux.*

Soit un échantillon de taille $n = 1000$ d'une loi normale centrée et réduite $X_i \sim N(0, 1)$, le paramètre de lissage ou la fenêtre h est fixé $h = 0.4$. Ensuite, calculons l'estimateur à noyau $\hat{f}_n(x)$ basé sur les données X_i par les différents choix de K , la sortie graphique est donnée par la figure 2.4 suivante :

- 1) *Noyau Uniforme (2.10) à support compact (couleur vert).*
- 2) *Noyau Biweight (2.12) à support compact (couleur orange).*
- 3) *Noyau d'Epanechnikov (2.11) à support compact (couleur rouge).*
- 4) *Noyau Gaussien (2.13) à support non compact (couleur bleu).*

On remarque que le meilleur noyau est celui d'Epanechnikov.

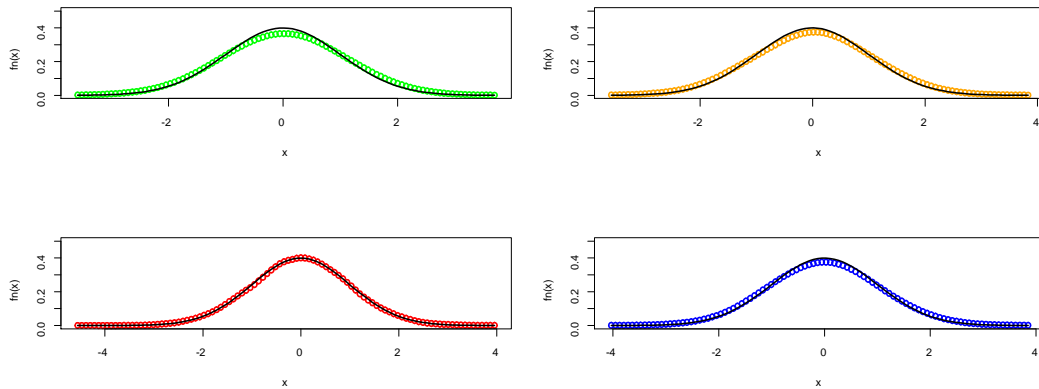


FIG. 2.4 – Estimation NP de la densité par différents noyaux

2.5 Choix du noyau et problème des effets de bord

Nous terminons ce chapitre par l'étude du problème des effets de bord et sa relation avec le choix du noyau. L'un des objectifs en estimation à noyau est de

déterminer quelle forme de densité est la plus facile afin d'avoir de petites valeurs de la quantité

$$R(f'') = \int_{\mathbb{R}} f''(x)^2 dx.$$

Dans certains cas (loi uniforme ou les lois polynômiales par exemple), la densité f est telle que la première et la second dérivée sont nulles, i.e., $f'(x) = 0$, $f''(x) = 0$. Donc, $R(f'') = 0$, cela impliquerait

$$h_{opt}^* = \left(\frac{R(K)}{\mu_2^2(K) R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} = \left(\frac{R(K)}{0} \right)^{\frac{1}{5}} n^{-\frac{1}{5}} = \infty.$$

Ce résultat nous pose un problème dans le choix du noyau dans l'estimation NP de f .

Lorsqu'on veut estimer des densités à support borné au moins d'un côté, l'estimateur à noyau classique devient non consistant, à cause des effets du bord (le problème des effets de bord). Bien que les noyaux précédents diminuent le biais aux bornes, ils ne sont pas très efficaces car le biais reste grand si on le compare aux biais de l'intérieur du support, par exemple : le noyau Gamma.

Plusieurs solutions ont été proposées dans la littérature pour résoudre ces problèmes. *Schuster* (1985) [11] a proposé une méthode dite des données reflétées (reflection data method, en anglais), cette méthode consiste à remplacer la partie $K\left(\frac{X_i-x}{h}\right)$ par $K\left(\frac{X_i-x}{h}\right) + K\left(\frac{X_i+x}{h}\right)$. L'estimateur de la densité est donné dans ce cas par

$$\hat{f}_n(x) = \frac{1}{2nh} \sum_{i=1}^n \left\{ K\left(\frac{X_i-x}{h}\right) + K\left(\frac{X_i+x}{h}\right) \right\}. \quad (2.18)$$

Marron et Ruppert (1994) [7] ont proposé la méthode de transformation (transformation method en anglais). Elle se compose de trois étapes. Tout d'abord,

une transformation g est choisie dans une famille paramétrique de telle sorte que la densité des variables $Y_i = g(X_i)$ ait une dérivée première qui est approximativement égale à 0. Ensuite, un estimateur à noyau de la réflexion est appliqué à Y_i . Enfin, cet estimateur est converti par un autre changement de variables pour obtenir une estimation de f .

Müller(1991) [8] a proposé d'utiliser les noyaux optimaux aux bornes. La solution la plus récente est de remplacer le noyau symétrique par un noyau asymétrique, qui n'assigne pas un poids en dehors du support de la densité qu'on veut estimer. Cette idée due à *Chen* (1999) [1] pour des noyaux bêta et *Chen* (2000) [2] pour les noyaux gamma, dans le cas de densités à support $[0, 1]$ et $[0, +\infty[$ respectivement.

En pratique, le choix du noyau associé dépend du support de la distribution des données dont on se propose d'estimer la densité (on choisit le noyau K selon le support des données) c-à-d.

- Si le support des données est $]-\infty, +\infty[$, on peut prendre le noyau Gaussien (2.11), ou le noyau d'Epanechnikov(2.13)
- Si le support des données est $[0, +\infty[$, on peut prendre le noyau Gamma.
- Si le support est $[0, 1]$, on peut prendre le noyau bêta.

Conclusion

Dans ce mémoire nous avons essayé d'étudier le problème du choix du noyau dans l'estimation de la densité. Cette méthode non paramétrique est une généralisation de celle d'estimation par histogramme et elle basée sur deux quantités, le noyau K et le paramètre de lissage h .

La fenêtre h à un rôle important, car elle détermine la qualité d'estimation et contrôle le lissage (sur lissage, sous lissage) d'une façon très sensible. En effet, des faibles valeurs de h implique un sous-lissage et quand h grandit, la variance devient petite et c'est le biais qui augmente, cela implique un sur-lissage de l'estimateur de f .

Mais malgré qu'en estimation à noyau le choix de la fonction K est de faible effet. Nous étudions dans ce mémoire, cette fonction en mettons l'accent sur les types du noyau existent, leurs construction et propriétés, le problème du choix optimal d'une telle fonction K , et l'influence de ce choix sur la qualité d'estimation, ainsi que le problème des effets de bord rencontré dans les cas d'estimation des densités à supports compacts.

Notons finalement, que la fonction noyau est-elle même une densité. Une question importante (perspective de recherche) est donc d'essayer de l'estimer d'une façon non paramétrique, afin d'obtenir une fonction noyau idéale.

Bibliographie

- [1] Chen, S.X. (1999). Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* 31, 131-145.
- [2] Chen, S.X.(2000). Gamma kernel estimators for density functions. *Annals of Institute for Statistics and Mathematics* 52, 471-480.
- [3] Comte, F. (2017). Estimation non-paramétrique. Cours de master. Deuxième édition.
- [4] Deheuvels, p.(1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de statistique appliquée*, 25(3), 5-42.
- [5] Epanechnikov, V. (1969). Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications* 14, 153–158.
- [6] Hodges, J. et Lehmann. E. (1956)., The efficiency of some nonparametric competitors of the t-test., *Annals Mathematiques statistiques* 27(1), 324-335.
- [7] Marron, J.S. & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society* 4 , 653-671.
- [8] Müller, H.G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* (78), 521–530.

- [9] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* 33, 1065–1076.
- [10] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* 27, 832–837.
- [11] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communication in Statistics Theory and Methods* 5, 1123-1136.
- [12] Silverman, B. W. (1986) *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability, London. Chapman and Hall.
- [13] Tsybakov, A. B. (2003). *Introduction à l'estimation non paramétrique* (Vol. 41). Springer Science & Business Media.
- [14] Wand, M.P., Jones, M .C. (1995). *Kernel Smoothing*. Monographs on Statistics and Applied Probability, London. Chapman and Hall.