

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 Melle. SELLAMI Chahrazad

Titre :

Estimation de la densité par noyaux asymétriques et son
application

Devant le Jury :

Mr.	CHERFAOUI Mouloud	Prof.	U. Biskra	Président
Mr.	YAHIA Djabrane	Prof.	U. Biskra	Rapporteur
Mme.	KHEIREDDINE Souraya	MCB	U. Biskra	Examinatrice

Soutenu Publiquement le 27 Juin 2022

Dédicace

Je dédie ce modeste travail :

À mon cher père.

À ma cher mère.

À mes soeurs et frères.

À mes amis qui j'ai passé des moments mémorables et agréables.

À tous mes professeurs.

À tous ceux qui m'aiment.

À tous ceux que j'aime.

Remerciements

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

J'adresse mes sincères remerciements, mon appréciation et mon respect à mon encadreur Prof. YAHIA Djabrane.

Je tiens à remercier les membres du comité d'examen qui m'ont fait l'honneur de bien vouloir étudier avec attention mon travail : Prof. Charfaoui Mouloud en qualité de président et Mme Kheireddine Souraya en qualité d'examinatrice.

Je remercie chaleureusement toute ma famille et mes amis, qui m'a soutenu, encouragé et poussé durant toutes mes années d'étude.

A tous ceux qui ont contribué à la réalisation de ce mémoire :

Merci.

Notations et symbols

$E(\cdot)$:	Espérance mathématique.
$var(\cdot)$:	Variance mathématique.
$biais(\cdot)$:	Biais d'un estimateur.
$\hat{f}_n(\cdot)$:	Estimateur de la fonction de densité.
$F(\cdot)$:	Fonction de répartition.
$F_n(\cdot)$:	Fonction de répartition empirique.
$K(\cdot)$:	Fonction de noyau.
h	:	Paramètre de lissage.
h_{opt}	:	Fenêtre h optimale locale.
h_{opt}^*	:	Fenêtre h optimale globale.
MSE	:	Erreur moyenne quadratique.
$AMSE$:	Erreur moyenne quadratique asymptotique.
$MISE$:	Erreur quadratique moyen intégré.
$AMISE$:	Erreur quadratique moyen intégré asymptotique.
v.a	:	Variable aléatoire.
$B(a, b)$:	Loi Béta de paramètres a et b .
$G(a, b)$:	Loi gamma de paramètres a et b .
\aleph	:	Support de x
$\aleph_{x,h}$:	Support de noyau

Table des matières

Dédicace	i
Remerciements	i
Notations et symbols	iii
Table des matières	iv
Table des figures	vi
Liste des tableaux	vii
Introduction	1
1 Estimation non paramétrique de la densité	3
1.1 Estimation de la densité par histogramme	3
1.1.1 Propriétés asymptotiques de l’histogramme	4
1.1.2 Choix de la fenêtre optimale de l’histogramme	7
1.2 Estimation de la densité par noyau	7
1.2.1 Exemples de noyaux symétriques classiques	10

1.2.2	Propriétés asymptotiques	10
1.2.3	Sur le choix de la fenêtre et du noyau	12
2	Estimation de la densité par noyau asymétrique	15
2.1	Noyau asymétrique	16
2.1.1	Estimateur à noyau associé continu et asymétrique	16
2.1.2	Propriétés de l'estimateur à noyau asymétrique	17
2.1.3	Exemples de noyaux associés continus asymétriques	19
2.2	Noyau bêta	19
2.2.1	Propriétés d'un estimateur à noyau bêta	21
2.3	Noyau gamma	25
2.3.1	Propriétés d'un estimateur à noyau Gamma	27
	Conclusion	33
	Bibliographie	34

Table des figures

1.1	Estimation par histogramme basé sur un échantillon de taille $n =$	
	500, $X \rightarrow N(0, 1)$.	5
1.2	Représentation graphique de quelques noyaux symétriques	10
1.3	Estimation d'une densité normale par différents noyaux.	14
1.4	Choix de la fenêtre h , cas de la densité normale.	14
2.1	Estimateur à noyau symétrique K est $N(0, 1)$ d'une densité $Exp(1)$.	15
2.2	Estimateur \hat{f}_1 à noyau bêta pour $h = 0.05$, $n = 100, 500, 1000$.	21
2.3	Estimateur \hat{f}_3 à noyau gamma, $h = 0.025$, $n = 50, 100, 500$.	27

Liste des tableaux

1.1	Quelques noyaux symétriques classiques.	10
1.2	Efficacité relative de quelques noyaux classiques.	13
2.1	Quelques noyaux asymétriques.	19

Introduction

Souvent en statistique à partir d'une suite des variables aléatoires (v.a's) nous cherchons à estimer la densité de probabilité avec des méthodes paramétriques tel que la méthode de maximum de vraisemblance, ou des méthodes non paramétriques. Le principe d'estimation non paramétrique est de laisser les données telles qu'elles sont sans aucune hypothèse à priori. Alors, au lieu de supposer une loi paramétrique et d'utiliser un échantillon pour estimer les différents paramètres, on utilise directement seulement l'échantillon pour construire un estimateur pour la densité.

La méthode d'estimation non paramétrique la plus simple est celle de l'histogramme qui était introduit par Graunt en (1662), mais avec un problème (inconvenient) qui vient du fait que l'histogramme donne une fonction qui n'est pas continue (en escalier donc non lisse).

C'est ainsi que la méthode d'estimation non paramétrique du noyau introduite par Rosenblatt en (1956) pour estimer des densités de probabilité, puis amélioré par Parzen en (1962), constitue une généralisation de la méthode d'histogramme et qui possède d'excellentes propriétés statistiques.

Mais avec ça, ces estimateurs connaissent quelques inconvénients, parmi lesquels les problèmes du support. En effet, les fonctions noyaux utilisées sont à support

symétrique ($] -1, 1[$, R , ...) ce qui pose un problème quand la densité à estimer est à support asymétrique (comme le cas de loi Gamma, Exponentielle, log normale, Beta,...).

Dans ce mémoire, nous proposons donc d'étudier des estimateurs à noyau asymétrique de la densité. Ce travail est constitué de deux chapitres comme suit :

Dans le premier chapitre, nous regarderons en quoi consiste la méthode d'estimation non paramétrique de la densité par histogramme et celle du noyau. Nous nous intéressons à l'estimateur à noyau continu symétrique d'une densité de probabilité f inconnue sur R . Plus précisément, nous présentons cet estimateur et nous donnons ses propriétés fondamentales telles que l'erreur quadratique moyenne et intégrée, les méthodes de sélection du noyau et de la fenêtre.

Dans le deuxième chapitre, nous donnons la définition d'un noyau associé continu asymétrique. Nous présentons aussi quelques estimateurs développés dans ce cas et leurs propriétés asymptotiques. En s'appuyant sur des exemples de données simulées avec le logiciel R.

Chapitre 1

Estimation non paramétrique de la densité

Dans ce chapitre, nous allons présenter l'estimateur de la fonction densité par histogramme et par noyau symétrique. Nous présentons leurs propriétés statistiques et nous donnons quelque exemples explicatifs.

1.1 Estimation de la densité par histogramme

En statistique, l'histogramme est une présentation graphique de la répartition d'une variable aléatoire (v.a) X (Pearson, 1895). Supposons que f est à support compact inclus dans $[0, 1]$, soit C_1, \dots, C_m une partition uniforme de $[0, 1[$:

$$C_k = \left[\frac{k-1}{m}, \frac{k}{m} \right[\quad k = 1, \dots, m,$$

Estimer cette densité f par la méthode de l'histogramme revient à approcher f par une fonction en escalier, constantes par morceaux sur les intervalles C_j . Posons

$h = \frac{1}{m}$. On approche f par la fonction

$$f_h(x) = \sum_{k=1}^m \frac{p_k}{h} \mathbf{1}_{C_k(x)},$$

tel que

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

avec

$$p_k = \int_{C_k} f(x) dx = E(\mathbf{1}_{C_k(x)}) = P(C_k(x)),$$

d'autre part, Pearson a estimé p_k par

$$\hat{p}_k = \hat{E}(\mathbf{1}_{C_k(x)}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{C_k(X_i)}$$

On observe que chaque \hat{p}_k représente la proportion des observations X_i se trouvant dans l'intervalle C_k . Nous définissons donc l'estimateur de f par histogramme à m classes comme suit

$$\hat{f}_h(x) = \sum_{k=1}^m \frac{\hat{p}_k}{h} \mathbf{1}_{C_k(x)}. \quad (1.1)$$

1.1.1 Propriétés asymptotiques de l'histogramme

– L'espérance : pour tout $x \in C_k$:

$$\hat{f}_h(x) = \frac{\hat{p}_k}{h} = \frac{1}{nh} \sum_{i=1}^n \mathbf{1}_{C_k(X_i)} = \frac{Z_k}{nh} \quad (1.2)$$

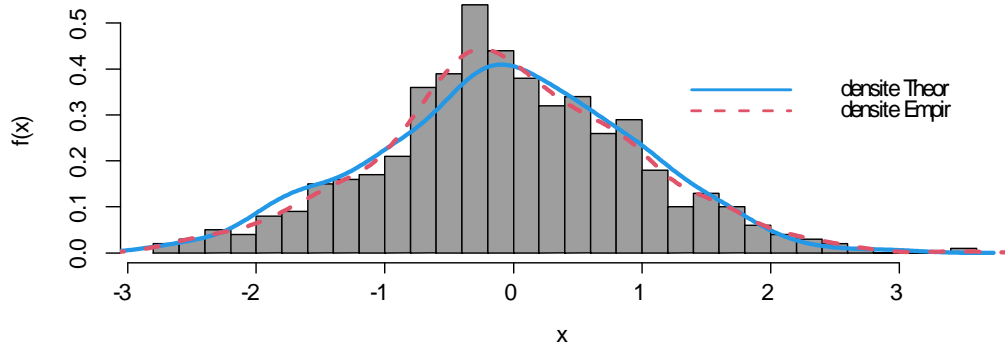


FIG. 1.1 – Estimation par histogramme basé sur un échantillon de taille $n = 500$, $X \rightarrow N(0, 1)$.

avec $Z_k = \sum_{i=1}^n \mathbf{1}_{C_k}(X_i)$ suit une loi Binomiale $B(p, n)$, de paramètre n et $p = P(\mathbf{1}_{C_k} = 1) = p_k$. Alors,

$$E(\hat{f}_h(x)) = \frac{1}{nh} E(Z_k) = \frac{p_k}{h}.$$

– La variance : de même, on a

$$\text{var}(\hat{f}_h(x)) = \frac{p_k(1-p_k)}{nh^2}.$$

– Le biais :

$$\text{biais}(\hat{f}_h(x)) = E(\hat{f}_h(x)) - f(x) = \frac{p_k}{h} - f(x) \neq 0.$$

– Erreur quadratique moyenne (MSE) :

$$\begin{aligned} MSE\left(\hat{f}_h(x)\right) &= \text{biais}^2\left(\hat{f}_h(x)\right) + \text{var}\left(\hat{f}_h(x)\right) \\ &= \left(\frac{p_k}{h} - f(x)\right)^2 + \frac{p_k(1-p_k)}{nh^2}. \end{aligned}$$

– Erreur quadratique moyenne intégrée ($MISE$) :

$$\begin{aligned} MISE\left(\hat{f}_h(x)\right) &= \int MSE\left(\hat{f}_h(x)\right) dx \\ &= \int \text{biais}^2\left(\hat{f}_h(x)\right) + \int \text{var}\left(\hat{f}_h(x)\right) \\ &= \int f^2(x) dx + \frac{1}{nh} - \frac{1}{h} \left(1 + \frac{1}{n}\right) \sum_{k=1}^m p_k^2. \end{aligned}$$

Théorème 1.1.1 *Supposons que la densité f de X est deux fois continûment différentiable, sous la condition $h \rightarrow 0$ quand $n \rightarrow \infty$, alors*

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

Remarque 1.1.1 *Cette erreur quadratique moyenne intégrée tend vers zéro quand h tend vers zéro et $nh \rightarrow \infty$ quand $n \rightarrow \infty$. Ce critère d'erreur quadratique moyenne est un critère global. L'approximation asymptotique de la $MISE$ est donnée par :*

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

1.1.2 Choix de la fenêtre optimale de l'histogramme

On a d'après le théorème précédent :

$$\begin{aligned} h_{opt}^* &= \arg \min_h AMISE \left(\hat{f}_h(x) \right) \\ &= \arg \min_h \left(\frac{h^2}{12} \int (f'(x))^2 dx + \frac{1}{nh} \right) \\ &= cn^{-\frac{1}{3}}. \end{aligned}$$

En remplaçant, la valeur de h_{opt}^* dans l'expression $AMISE$:

$$AMISE \left(\hat{f}_{h_{opt}^*} \right) = \frac{(h_{opt}^*)^2}{12} \int (f'(x))^2 dx + \frac{1}{n(h_{opt}^*)} = c'n^{-\frac{2}{3}}.$$

c et c' sont deux constantes (indépendantes de n).

Remarque 1.1.2 *L'un des inconvénients majeurs de l'estimation par histogramme est que cette méthode ne donne que des estimateurs non lisse de f (en escalier par construction). Ce qui nous conduit à d'autre méthodes utilisant des fonctions plus lisses. C'est le sujet de la section suivante.*

1.2 Estimation de la densité par noyau

Soit X_1, \dots, X_n un n échantillon d'une v.a X (il s'agit d'une suite de v.a's indépendantes identiquement distribuées (i.i.d.) issus de X), de densité de probabilité inconnue f par rapport à la mesure de Lebesgue sur R . La fonction de distribution correspondante est

$$F(x) = \int_{-\infty}^x f(t) dt,$$

dont la fonction de distribution empirique est

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

où $\mathbf{1}(A)$ désigne la fonction indicatrice d'indice A . Par la loi forte des grands nombres, on a

$$F_n(x) \rightarrow F(x), \quad \forall x \in R,$$

presque sûrement quand $n \rightarrow \infty$. Donc $F_n(x)$ est un estimateur consistant de $F(x)$ pour tout $x \in R$. Pour h suffisamment petit, nous pouvons écrire l'approximation

$$f(x) = \frac{F(x+h) - F(x-h)}{2h}$$

En remplaçant F par l'estimation F_n nous définissons

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

La fonction \hat{f}_n est un estimateur de f appelé estimateur de Rosenblatt. On peut l'écrire aussi sous la formule :

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

où $K_0(t) = \frac{1}{2} \mathbf{1}_{(|t| \leq 1)}$.

Une simple généralisation de l'estimateur de Rosenblatt est donnée par

$$\hat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right). \tag{1.3}$$

avec $h = h_n$ est le paramètre de lissage (fenêtre) choisi en fonction de n telle que

$$\lim_{n \rightarrow \infty} h = 0$$

et K est une fonction définie comme suit :

Définition 1.2.1 Soit $K : R \rightarrow R$ une fonction intégrable et telle que

$$\int_R K(t) dt = 1$$

alors K est appelé noyau (kernel en anglais) :

1) si $K(x) \geq 0 \forall x \in R$, et $\int_R K(x) dx = 1$, alors \hat{f}_n est une densité

$$\int_R \hat{f}_n(x) dx = 1.$$

2) Un noyau est dit symétrique, si pour tout t dans son ensemble de définition

$$K(t) = K(-t).$$

Nous supposons que le noyau symétrique K vérifie les conditions suivantes :

A1) $K(t) \geq 0, \forall t \in R$.

A2) $\int_R tK(t)dt = 0$.

A3) $\int_R t^2K(t)dt < \infty$

On dit dans ce cas que K est un noyau d'ordre 2.

Noyaux	Densité	Support
Epanechnikov	$\frac{3}{4}(1-x^2)$	$[-1, 1]$
Gaussien	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$	R
Uniforme (rectangulaire)	$\frac{1}{2}$	$[-1, 1]$
Quartique (Biweight)	$\frac{15}{16}(1-x^2)^2$	$[-1, 1]$

TAB. 1.1 – Quelques noyaux symétriques classiques.

1.2.1 Exemples de noyaux symétriques classiques

La représentation graphique de ces noyaux symétriques donnés dans le tableau précédent est donnée par la figure [1.2](#) suivante :

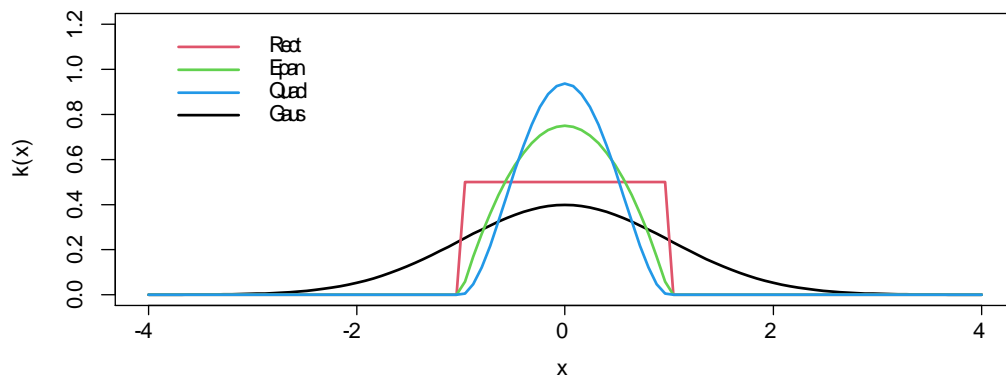


FIG. 1.2 – Représentation graphique de quelques noyaux symétriques

1.2.2 Propriétés asymptotiques

Nous donnons dans la suite quelques propriétés asymptotiques de l'estimateur à noyau de la densité :

– Biais :

$$\begin{aligned} \text{biais} \left(\hat{f}_n(x) \right) &= E \left(\hat{f}_n(x) \right) - f(x) \\ &= E \left(\frac{1}{nh} \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) \right) - f(x), \end{aligned}$$

On utilise le développement de Taylor de f au voisinage de x à l'ordre 2 :

$$f(x + th) = f(x) + \frac{th}{1!} f'(x) + \frac{(th)^2}{2!} f''(x) + O(h^2).$$

Alors,

$$\text{biais} \left(\hat{f}_n(x) \right) = \frac{1}{2} h^2 f''(x) u_2(K) + O(h^2), \quad (1.4)$$

où $u_2(K) = \int_{\mathbb{R}} u^2 K(u) du$.

– Variance :

$$\text{var} \left(\hat{f}_n(x) \right) = \text{var} \left(\frac{1}{nh} \sum_{i=1}^n K \left(\frac{X_i - x}{h} \right) \right) = \frac{1}{nh} f(x) R(K) + O\left(\frac{1}{nh}\right) \quad (1.5)$$

où $R(K) = \int K^2(u) du$.

Remarque 1.2.1 *Pour f une densité bornée dont la dérivée seconde est bornée, alors on a :*

$$\text{biais} \left(\hat{f}_n(x) \right) \leq c_1 h^2 \quad \text{et} \quad \text{var} \left(\hat{f}_n(x) \right) \leq \frac{c_2}{nh}$$

où c_1 et c_2 sont des constantes données.

– Erreur quadratique moyenne (MSE) : D'après les résultats (1.4) et (1.5), l'expression de la MSE en un point x fixé est

$$MSE \left(\hat{f}_n(x) \right) = \frac{h^4}{4} u_2^2(K) (f''(x))^2 + \frac{f(x)}{nh} R(K) + O(nh^{-1}) + O(h^4) \quad (1.6)$$

La $AMSE$ est donnée par :

$$AMSE\left(\hat{f}_h(x)\right) = \frac{h^4}{4}u_2^2(K)(f''(x))^2 + \frac{f(x)}{nh}R(K). \quad (1.7)$$

– Erreur quadratique moyenne intégrée ($MISE$) :

$$\begin{aligned} MISE\left(\hat{f}_n(x)\right) &= \int MSE\left(\hat{f}_n(x)\right) dx \\ &= \frac{1}{nh}R(K) + \frac{h^4}{4}u_2^2(K)R(f''(x)) + O(nh^{-1}) + O(h^4). \end{aligned} \quad (1.8)$$

où $R(g) = \int_{\mathbb{R}} g^2(y)dy$.

– La $MISE$ Asymptotique notée $AMISE$ est donnée par :

$$AMISE\left(\hat{f}_h(x)\right) = \int MSE\left(\hat{f}_h(x)\right) dx = \frac{1}{nh}R(K) + \frac{h^4}{4}u_2^2(K)R(f''(x)). \quad (1.9)$$

1.2.3 Sur le choix de la fenêtre et du noyau

Par minimisation en h de la $AMSE$, on trouve la fenêtre optimale locale, notée

h_{opt} :

$$h_{opt} = \arg \min_h AMSE\left(\hat{f}_n(x)\right)$$

en dérivant l'expression (1.7) par rapport h :

$$\frac{\partial AMSE}{\partial h}(h) = 0 \Rightarrow h_{opt} = \left(\frac{R(K)f(x)}{u_2^2(K)f''(x)^2} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

De plus, par minimisation en h de la $AMISE$, on trouve la fenêtre optimale globale, notée h_{opt}^* :

$$h_{opt}^* = \arg \min_h AMISE\left(\hat{f}_n(x)\right)$$

En dérivant l'équation (1.9) par rapport h :

$$\frac{\partial AMISE}{\partial h}(h) = 0 \Rightarrow h_{opt}^* = \left(\frac{R(K)}{u_2^2(K)R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}.$$

Concernant le choix du noyau, on pose $h = h_{opt}^*$. La minimisation de la $AMISE$:

$$\begin{aligned} AMISE(\hat{f}_n(x)) &= \frac{1}{nh_{opt}^*}R(K) + \frac{(h)^4}{4}u_2^2(K)R(f''(x)) \\ &= \frac{5}{4}(u_2^2(K)R(f'')R^4(K))^{\frac{1}{5}}n^{-\frac{4}{5}} \end{aligned}$$

par rapport à K , donne comme solution le noyau dite d'Epanechnikov (Epanechnikov, 1969) :

$$K_{opt}(t) = \frac{3}{4}(1 - t^2)\mathbf{1}_{(|t| \leq 1)}$$

Définition 1.2.2 L'efficacité relative d'un noyau K par rapport à K_{opt} est donnée par :

$$eff(K) = \frac{AMISE(K_{opt})}{AMISE(K)} = \left(\frac{u_2^2(K_{opt})R^4(K_{opt})}{u_2^2(K)R^4(K)} \right)^{\frac{1}{5}} \leq 1.$$

Dans le tableau suivant nous donnons l'efficacité relative de quelques noyaux usuels :

Noyaux	Efficacité relative
Epanechnikov	1
Gaussien	0.946
Uniforme (rectangulaire)	0.930
Quartique (Biweight)	0.994

TAB. 1.2 – Efficacité relative de quelques noyaux classiques.

Il est clair que le choix du noyau n'influe pas beaucoup sur la qualité d'estimation (la figure 1.3, montre que le choix du noyau est de faible effet).

Contrairement à la fenêtre h , dont le choix est très sensible (les petites valeurs de h donne une sur-estimation ou grande variance et les grandes valeurs de h donne

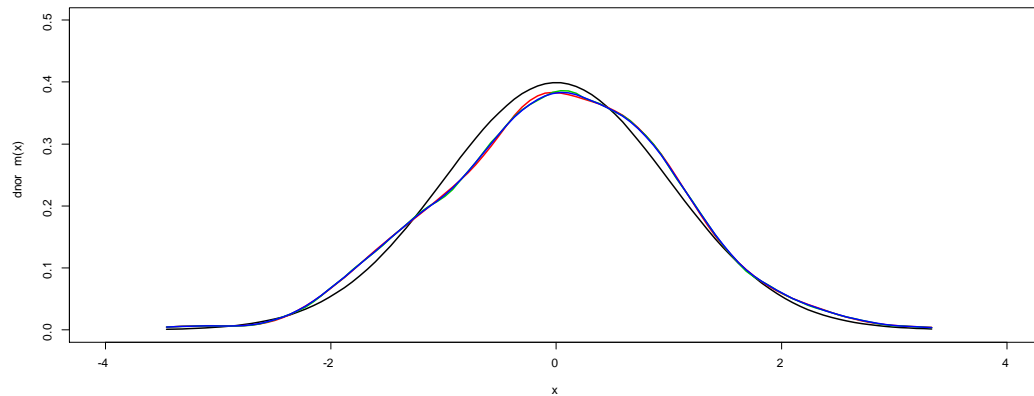


FIG. 1.3 – Estimation d'une densité normale par différents noyaux.

une sous-estimation ou un grand biais), comme cela est constaté dans la figure

1.4 suivante :

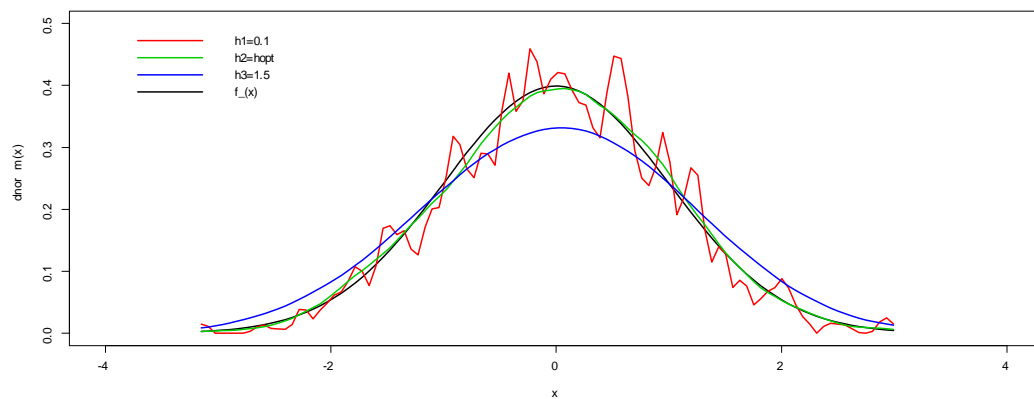


FIG. 1.4 – Choix de la fenêtre h , cas de la densité normale.

Chapitre 2

Estimation de la densité par noyau asymétrique

Si le support de f est $[0, +\infty[$, les noyaux symétriques donnent une estimation de plus en plus biaisée lorsque x est proche de 0. D'où le problème du biais aux bords (voir la figure [2.1](#), dont la variable X suit une loi $Exp(1)$ et le noyau K est $N(0, 1)$), ce qui nous conduit à l'utilisation de noyaux asymétriques et c'est le sujet de ce deuxième chapitre.

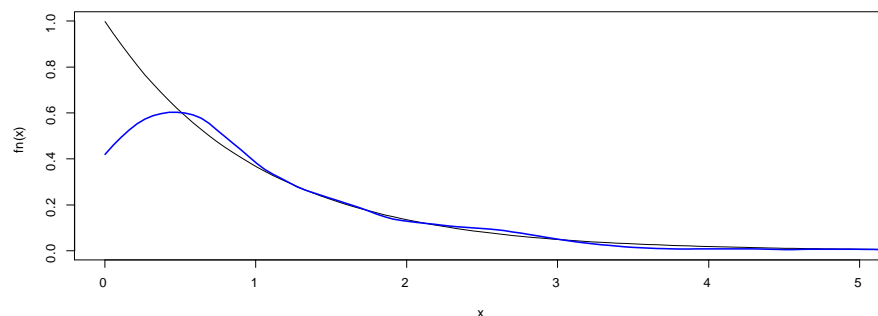


FIG. 2.1 – Estimateur a noyau symétrique K est $N(0, 1)$ d'une densité $Exp(1)$.

2.1 Noyau asymétrique

Les définitions suivantes donnent les notions du noyau associé, et de l'estimateur à noyau associé pour la fonction de densité f inconnue sur le support noté \aleph .

Définition 2.1.1 *Soit $x \in \aleph \subseteq R$ et $h > 0$. On appelle noyau associé toute densité de probabilité d'une variable aléatoire continue ou discrète $k_{x,h}$ de support $\aleph_{x,h}$ vérifiant les conditions suivantes :*

$$\begin{aligned} \aleph_{x,h} \cap \aleph &\neq \phi \quad \text{et} \quad \aleph_{x,h} \supseteq \aleph. \\ \lim_{h \rightarrow 0} E(K_{x,h}) &= x \quad \text{et} \quad \lim_{h \rightarrow 0} \text{var}(K_{x,h}) = 0. \end{aligned}$$

2.1.1 Estimateur à noyau associé continu et asymétrique

Soit X_1, X_2, \dots, X_n un n -échantillon aléatoire i.i.d. de densité de probabilité inconnue f sur l'ensemble $\aleph \subseteq R$. Chen (1999, 2000) et plus tard Scaillet (2004) introduisent des noyaux continus asymétriques. Ces derniers sont appropriés pour l'estimation d'une densité continue à support compact ou borné d'un seul côté tel que $\aleph = [a, b]$ avec $a \in R$ et $b \in \bar{R}$, par exemple $[0, 1]$ ou $[0, +\infty[$. Pour $x \in \aleph$, les estimateurs à noyau associé continu asymétrique sont définis aussi de la même forme que l'estimateur à noyau classique :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \tag{2.1}$$

où $h > 0$ est le paramètre de lissage et $K_{x,h}(t) = \frac{1}{h} K\left(\frac{t-x}{h}\right)$ est le noyau associé continu.

2.1.2 Propriétés de l'estimateur à noyau asymétrique

■ L'espérance :

$$\begin{aligned} E[\hat{f}_n(x)] &= E\left[\frac{1}{n}\sum_{i=1}^n K_{x,h}(X_i)\right] = E[K_{x,h}(X_1)] = \int_{\mathbb{R} \cap \mathbb{R}_{x,h}} K_{x,h}(t) f(t) dt \\ &= E[f(K_{x,h})]. \end{aligned}$$

On utilise le développement de Taylor de $f(K_{x,h})$ au point moyen $E(K_{x,h}) = u_{x,h}$ à l'ordre 2, nous obtenons

$$f(K_{x,h}) = f(u_{x,h}) + \frac{(K_{x,h} - u_{x,h})}{1!} f'(x) + \frac{(K_{x,h} - u_{x,h})^2}{2!} f''(x) + O(h^2).$$

Alors l'espérance de l'estimateur présenté dans (2.1) est

$$\begin{aligned} E[f(K_{x,h})] &= f(u_{x,h}) + \frac{1}{2} E\{(K_{x,h} - u_{x,h})^2\} f''(x) \\ &= f\{E(K_{x,h})\} + \frac{1}{2} var(K_{x,h}) f''(x). \end{aligned}$$

■ Le biais : pour un x fixé, on a

$$\begin{aligned} \text{biais} [\hat{f}_n(x)] &= E[\hat{f}_n(x)] - f(x) \\ &= [f\{E(K_{x,h})\} - f(x)] + \frac{1}{2} var(K_{x,h}) f''(x). \end{aligned} \tag{2.2}$$

■ La variance : pour un x fixé,

$$\begin{aligned}
 \text{var} \left[\hat{f}_n(x) \right] &= \text{var} \left[\frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \right] \\
 &= \frac{1}{n} [\text{var} (K_{x,h}(X_1))] \\
 &= \frac{1}{n} [E(K_{x,h}(X_1))^2] - \frac{1}{n} [E(K_{x,h}(X_1))]^2 \\
 &= \frac{1}{n} \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} K_{x,h}^2(t) f(t) dt - \frac{1}{n} \left\{ \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} K_{x,h}(t) f(t) dt \right\}^2 \\
 &= \frac{1}{n} \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} K_{x,h}^2(t) f(t) dt - \frac{1}{n} \left\{ \text{biais}(\hat{f}_n(x)) + f(x) \right\}^2 \quad (\text{b3})
 \end{aligned}$$

■ Erreur quadratique moyenne intégrée ($MISE$) :

$$MISE \left(\hat{f}_n(x) \right) = \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} E \left[\hat{f}_n(x) - f(x) \right] dx,$$

Donc,

$$MISE \left(\hat{f}_n(x) \right) = \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} \text{biais}^2 \left(\hat{f}_n(x) \right) dx + \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} \text{var} \left(\hat{f}_n(x) \right) dx. \quad (2.3)$$

Remarque 2.1.1 *Le biais tend vers 0 quand $h = h_n$ tend vers 0 et n tend vers $+\infty$. La variance tend vers 0 quand*

$$\frac{1}{n} \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} K_{x,h}^2(t) f(t) dt \rightarrow 0 \quad \text{et} \quad h \rightarrow 0.$$

De plus, l'erreur quadratique moyenne intégrée converge vers 0 quand

$$\frac{1}{n} \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} \int_{\mathbb{R} \cap \mathbb{N}_{x,h}} K_{x,h}^2(t) f(t) dt dx \rightarrow 0 \quad \text{et } h \rightarrow 0.$$

2.1.3 Exemples de noyaux associés continus asymétriques

Dans le tableau suivant, nous donnons quelques noyaux asymétriques, avec leurs supports et formes de densités :

Noyaux	Supports	Densités
Gamma(a,b)	$[0, +\infty[$	$\frac{1}{\Gamma(a)} b^{-a} t^{a-1} \exp(-\frac{t}{b})$
Gamma inverse	$]0, +\infty[$	$\frac{1}{\Gamma(a)} b^a t^{-(a+1)} \exp(-\frac{b}{t})$
Gaussienne-inverse(a,b)	$]0, +\infty[$	$\frac{\sqrt{b}}{\sqrt{2\pi} t^3} \exp\left\{-\frac{b}{2a} \left(\frac{t}{a} - 2 + \frac{a}{t}\right)\right\}$
Gaussien-inverse-réciproque	$[0, +\infty[$	$\frac{\sqrt{b}}{\sqrt{2\pi} t} \exp\left\{-\frac{b}{2a} \left(at - 2 + \frac{1}{at}\right)\right\}$
Bêta(a,b)	$[0, 1]$	$\frac{1}{\beta(a,b)} t^{a-1} (1-t)^{b-1}$
Log-normal	$[0, +\infty[$	$\frac{1}{tb\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{\log(t)-a}{b}\right)^2\right\}$

TAB. 2.1 – Quelques noyaux asymétriques.

Remarque 2.1.2 Scaillet (2004) a introduit les noyaux associés gaussien-inverse et gaussien-inverse réciproque définis sur $[0, +\infty[$ et Jin and Kawczak (2003) a étudié la densité à noyau log-normal.

2.2 Noyau bêta

La densité de probabilité d'une loi bêta est définie et continue sur $[0, 1]$ telle que :

$$f_{B(a,b)}(t) = \frac{1}{\beta(a,b)} t^{a-1} (1-t)^{b-1} I_{[0,1]}(t)$$

où $a > 0, b > 0$, et $\beta(a, b)$ est la fonction bêta standard définie par

$$\beta(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt.$$

Si X est une variable aléatoire qui suit la loi bêta, alors

$$E(X) = \frac{a}{a+b} \text{ et } var(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Le noyau associé bêta a été introduit par Chen (1999) pour l'estimation des densités à support $[0, 1]$. IL a utilisé la loi bêta pour construire des noyaux associés continus asymétriques. Soit X_1, \dots, X_n un échantillon de v.a i.i.d sur $\mathfrak{N} = [0, 1]$, de densité de probabilité continue asymétrique inconnue f . Nous considérons l'estimateur \hat{f}_1 de f à noyau bêta tel que

$$\hat{f}_1(x) = \frac{1}{n} \sum_{i=1}^n K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(X_i)$$

avec

$$K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(t) = \frac{1}{\beta(\frac{x}{h}+1, \frac{1-x}{h}+1)} t^{\frac{x}{h}} (1-t)^{\frac{1-x}{h}}, \quad t \in [0, 1]$$

où $K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(t)$ est la fonction de densité d'une loi bêta de paramètre de forme $a = \frac{x}{h} + 1$ et paramètre d'échelle $b = \frac{1-x}{h} + 1$ avec $x \in [0, 1]$ et h le paramètre de de lissage.

Remarque 2.2.1 *Un autre estimateur à noyau bêta donné par Chen(1999), est défini comme suit :*

$$\hat{f}_2(x) = \frac{1}{n} \sum_{i=1}^n K_{B(x,h)}(X_i)$$

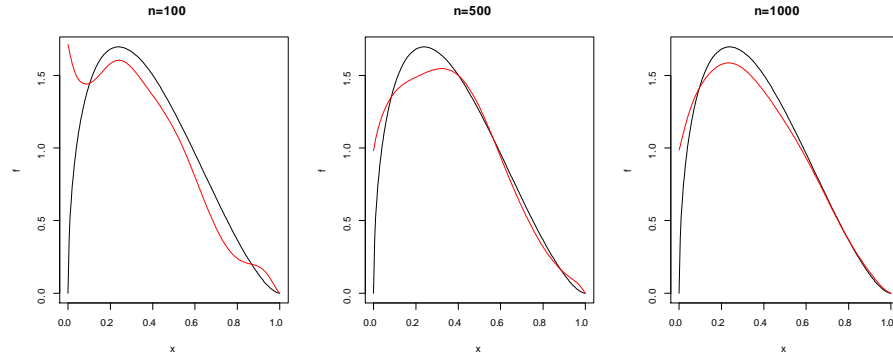


FIG. 2.2 – Estimateur \hat{f}_1 à noyau bêta pour $h = 0.05$, $n = 100, 500, 1000$.

avec

$$K_{B(x,h)}(t) = \begin{cases} K_{B(\frac{x}{h}, \frac{1-x}{h})}(t) & \text{si } x \in [2h, 1 - 2h] \\ K_{B(\rho(x), \frac{1-x}{h})}(t) & \text{si } x \in [0, 2h[\\ K_{B(\frac{x}{h}, \rho(1-x))}(t) & \text{si } x \in]1 - 2h, 1] \end{cases}$$

où

$$\rho(x, h) = 2h^2 + 2.5 - \sqrt{4h^4 + 6h^2 + 2.25 - x^2 - \frac{x}{h}}.$$

2.2.1 Propriétés d'un estimateur à noyau bêta

Nous donnons les différentes propriétés d'un estimateur d'une densité de probabilité à noyau bêta en se basant sur le travail de Chen(1999).

Propriétés de l'estimateur \hat{f}_1

1. Le biais : On a

$$\text{biais}[\hat{f}_1(x)] = E[\hat{f}_1(x)] - f(x),$$

avec

$$\begin{aligned} E[\hat{f}_1(x)] &= E\left(\frac{1}{n} \sum_{i=1}^n K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(X_i)\right) \\ &= \int_0^1 f(y) K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(y) dy \end{aligned}$$

Si on considère que $K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}$ est la densité d'une v.a $\xi_x \sim \text{beta}\left(\frac{x}{h} + 1, \frac{1-x}{h} + 1\right)$,
donc

$$E[\hat{f}_1(x)] = E[f(\xi_x)] = f(x) + f'(x)E((\xi_x - x)) + \frac{1}{2}f''(x)E((\xi_x - x)^2)$$

où $E((\xi_x - x)) = h(1 - 2x) + O(h^2)$ et

$$E((\xi_x - x)^2) = hx(1 - x) + h^2(1 - 2x)^2 + O(h^2).$$

Alors

$$E[f(\xi_x)] = f(x) + f'(x)h(1 - 2x) + \frac{1}{2}f''(x)hx(1 - x) + O(h),$$

donc le biais est donné par

$$\text{biais}[\hat{f}_1(x)] = f'(x)h(1 - 2x) + \frac{1}{2}f''(x)hx(1 - x) + O(h).$$

2. La variance :

$$\begin{aligned} \text{var}[\hat{f}_1(x)] &= \text{var}\left[\frac{1}{n} \sum_{i=1}^n K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(X_i)\right] \\ &= \frac{1}{n} \left[E\left[\left(K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(X)\right)^2\right] - \left[E\left(\hat{f}_1(x)\right)\right]^2 \right] \end{aligned}$$

avec

$$E \left[\left(K_{B(\frac{x}{h}+1, \frac{1-x}{h}+1)}(X) \right)^2 \right] = A_h(x) E[f(\gamma_x)]$$

où $\gamma_x \sim \beta(\frac{2x}{h} + 1, \frac{2(1-x)}{h} + 1)$ et pour h assez petit

$$A_h(x) = \begin{cases} \frac{1}{2\sqrt{\pi}} \frac{h^{-\frac{1}{2}}}{[x(1-x)]^{\frac{1}{2}}} & \text{si } \frac{x}{h} \rightarrow \infty \text{ et } \frac{1-x}{h} \rightarrow \infty \\ \frac{\Gamma(2k+1)}{2^{1+2k}\Gamma^2(k+1)} h^{-1} & \text{si } \frac{x}{h} \rightarrow k \text{ et } \frac{1-x}{h} \rightarrow k \end{cases}$$

donc la variance est donnée par

$$\text{var}[\hat{f}_1(x)] = \begin{cases} \frac{1}{2\sqrt{\pi}} \frac{n^{-1}h^{-\frac{1}{2}}}{[x(1-x)]^{\frac{1}{2}}} [f(x) + O(n^{-1})] & \text{si } \frac{x}{h} \rightarrow \infty \text{ et } \frac{1-x}{h} \rightarrow \infty \\ \frac{\Gamma(2k+1)}{2^{1+2k}\Gamma^2(k+1)} (nh)^{-1} [f(x) + O(n^{-1})] & \text{si } \frac{x}{h} \rightarrow k \text{ ou } \frac{1-x}{h} \rightarrow k \end{cases}$$

3. Erreur quadratique moyenne intégrée (*MISE*) :

$$\begin{aligned} \text{MISE}[\hat{f}_1(x)] &= h^2 \int_0^1 \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 dx \\ &\quad + \frac{1}{2n\sqrt{h\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx + O\left(\frac{1}{n\sqrt{h}} + h^2\right). \end{aligned}$$

Pour minimiser la *MISE* par rapport à la fenêtre h , nous avons

$$\begin{aligned} \text{AMISE}[\hat{f}_1(x)] &= h^2 \int_0^1 \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 dx \\ &\quad + \frac{1}{2n\sqrt{h\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx. \end{aligned}$$

On dérive donc *AMISE* par rapport à h

$$\frac{\partial \text{AMISE}}{\partial h}(h) = 0$$

nous obtenons la fenêtre optimale h_{opt1}^* :

$$h_{opt1}^* = \frac{1}{4^{\frac{2}{5}}} \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{\left[\int_0^1 (1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) dx \right]^{\frac{2}{5}}} n^{-\frac{2}{5}} = O\left(n^{-\frac{2}{5}}\right).$$

$$\begin{aligned} MISE_{h_{opt1}^*}(\hat{f}_1(x)) &= \frac{5}{4^{\frac{4}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^1 \frac{f(x)}{\{x(1-x)\}^{\frac{1}{2}}} dx \right]^{\frac{4}{5}} \\ &\quad \times \left[\int_0^1 \left[(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right]^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \end{aligned}$$

Propriétés de l'estimateur \hat{f}_2

1. Le biais :

$$\text{biais}[\hat{f}_2(x)] = \begin{cases} \frac{1}{2}x(1-x)f''(x)h + O(h^2) & \text{si } x \in [2h, 1-2h] \\ \xi(x)hf'(x) + O(h) & \text{si } x \in [0, 2h] \\ -\xi(1-x)hf'(x) + O(h) & \text{si } x \in [1-2h, 1] \end{cases}$$

2. La variance :

$$\text{var}[\hat{f}_2(x)] = \frac{1}{2n\sqrt{h\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx$$

3. Erreur quadratique moyenne intégrée ($MISE$) :

$$\begin{aligned} MISE[\hat{f}_2(x)] &= \frac{1}{4}h^2 \int_0^1 [x(1-x)f''(x)]^2 dx + \frac{1}{2n\sqrt{h\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx \\ &\quad + O\left(\frac{1}{n\sqrt{h}} + h^2\right). \end{aligned}$$

4. La fenêtre optimale h_{opt2}^* :

$$h_{opt2}^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{\left[\int_0^1 [x(1-x)f''(x)]^2 dx \right]^{\frac{2}{5}}} n^{-\frac{2}{5}} = O\left(n^{-\frac{2}{5}}\right).$$

Donc la *MISE* associée est

$$\begin{aligned} MISE_{h_{opt2}^*}[\hat{f}_1(x)] &= \frac{5}{4^{\frac{3}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^1 [x(1-x)]^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \\ &\quad \times \left[\int_0^1 [x(1-x)f''(x)]^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \end{aligned}$$

Remarque 2.2.2 *On remarque que*

$$\int_0^1 \left[(1-2x) f'(x) + \frac{1}{2} x(1-x) f''(x) \right]^2 dx \geq \int_0^1 [x(1-x)f''(x)]^2 dx.$$

On déduit que

$$MISE_{h_{opt1}^*}[\hat{f}_1(x)] \geq MISE_{h_{opt2}^*}[\hat{f}_2(x)] \text{ et } h_{opt1}^* \geq h_{opt2}^*$$

Alors \hat{f}_2 est le meilleur estimateur pour f .

2.3 Noyau gamma

La densité de probabilité d'une loi gamma est définie et continue sur $[0, +\infty[$ telle que :

$$f_{G(a,b)}(t) = \frac{1}{\Gamma(a)} b^{-a} t^{a-1} \exp\left(-\frac{t}{b}\right)$$

avec

$$\Gamma(a) = \int_{R_+} t^{a-1} e^{-t} dt$$

Si X est une variable aléatoire qui suit la loi gamma, alors

$$E(X) = ab \text{ et } var(X) = ab^2.$$

Chen (2000) est le premier a utilisé la loi gamma pour construire des noyaux associés continus asymétriques.

Pour $a = \frac{x}{h} + 1$ et $b = h$, on définit un noyau associé gamma par :

$$K_{G(\frac{x}{h}+1,h)}(t) = \frac{1}{h^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)} t^{\frac{x}{h}} e^{-\frac{t}{h}}, \quad t \in [0, +\infty[$$

de loi gamma à support $\aleph_{x,h} = [0, +\infty[$, d'espérance $E\left(K_{G(\frac{x}{h}+1,h)}\right) = x + h$ et de variance $var\left(K_{G(\frac{x}{h}+1,h)}\right) = xh + h^2$.

Soit X_1, \dots, X_n un échantillon de v.a i.i.d sur $\aleph = [0, +\infty[$, de densité de probabilité continue asymétrique inconnue f deux fois dérivables. Nous considérons l'estimateur \hat{f}_3 de f à noyau gamma tel que

$$\begin{aligned} \hat{f}_3(x) &= \frac{1}{n} \sum_{i=1}^n K_{G(\frac{x}{h}+1,h)}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)} X_i^{\frac{x}{h}} e^{-\frac{X_i}{h}} \end{aligned}$$

La figure [2.3](#) donne l'estimateur \hat{f}_3 à noyau gamma pour $h = 0.025$ et n varié ($n = 50, 100$ et 500) :

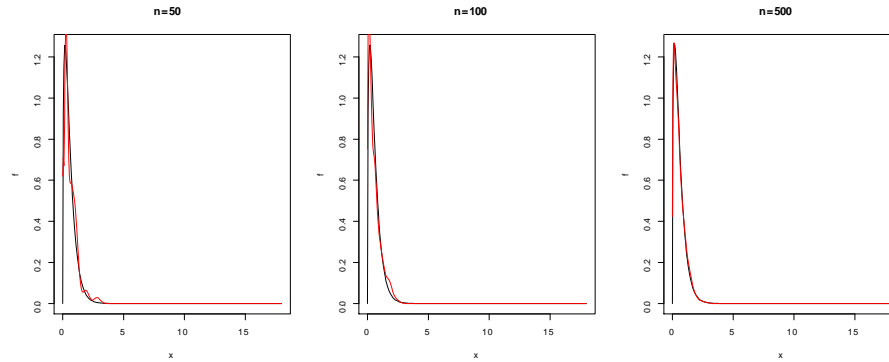


FIG. 2.3 – Estimateur \hat{f}_3 à noyau gamma, $h = 0.025$, $n = 50, 100, 500$.

Un autre estimateur à noyau gamma donné par Chen(2000) est défini comme suit :

$$\hat{f}_4(x) = \frac{1}{n} \sum_{i=1}^n K_{G(\rho_h(x), h)}(X_i),$$

tel que

$$\rho_h(x) = \begin{cases} \frac{x}{h} & \text{si } x \geq 2h \\ \frac{1}{4} \left(\frac{x}{h}\right)^2 + 1 & \text{si } x \in [0, 2h[. \end{cases}$$

2.3.1 Propriétés d'un estimateur à noyau Gamma

Nous donnons les différentes propriétés d'un estimateur d'une densité de probabilité à noyau Gamma en se basant sur l'étude de Chen(2000).

Propriétés de l'estimateur \hat{f}_3

1. Le biais :

$$\text{biais}[\hat{f}_3(x)] = E[\hat{f}_3(x)] - f(x),$$

Ainsi,

$$E[\hat{f}_3(x)] = \int_0^{+\infty} K_{G(\frac{x}{h}+1, h)}(t) f(t) dt = E[f(\xi_x)]$$

avec $\xi_x \sim \Gamma(\frac{x}{h} + 1, h)$, $\mu_x = E(\xi_x) = x + h$ et $var(\xi_x) = xh + h^2$. On utilise le développement de Taylor de $f(\xi_x)$ au voisinage de point moyen μ_x , nous obtenons

$$\begin{aligned} E[f(\xi_x)] &= E[f(\mu_x)] + \frac{1}{2}f''(\mu_x)E(\xi_x - \mu_x)^2 + O(h) \\ &= f(\mu_x) + \frac{1}{2}f''(\mu_x)var(\xi_x) + O(h) \\ &= f(x) + hf'(x) + \frac{h}{2}xf''(x) + O(h) \end{aligned}$$

alors

$$biais[\hat{f}_3(x)] = hf'(x) + \frac{h}{2}xf''(x) + O(h).$$

2. La variance :

$$\begin{aligned} var[\hat{f}_3(x)] &= \frac{1}{n} var \left[K_{G(\frac{x}{h}+1, h)}(X_i) \right] \\ &= \frac{1}{n} \left[E \left(K_{G(\frac{x}{h}+1, h)}(X) \right)^2 \right] - \frac{1}{n} \left[E \left(K_{G(\frac{x}{h}+1, h)}(X) \right) \right]^2 \\ &\leq \frac{1}{n} \left[E \left(K_{G(\frac{x}{h}+1, h)}(X) \right)^2 \right] \\ &\leq \int_0^\infty K_{G(\frac{x}{h}+1, h)}^2(t) f(t) dt = \int_0^\infty \left(\frac{t^{\frac{x}{h}} e^{-\frac{t}{h}}}{h^{\frac{x}{h}+1} \Gamma\left(\frac{x}{h} + 1\right)} \right)^2 f(t) dt \\ &\leq \int_0^\infty \left(\frac{t^{\frac{2x}{h}} e^{-\frac{2t}{h}}}{h^{2(\frac{x}{h}+1)} \Gamma^2\left(\frac{x}{h} + 1\right)} \right) f(t) dt \\ &\leq \frac{2^{\frac{2x}{h}+1} h^{\left(\frac{2x}{h}+1\right)} \Gamma\left(\frac{2x}{h} + 1\right)}{2^{\frac{2x}{h}+1} h^{\left(\frac{2x}{h}+1\right)} \Gamma\left(\frac{2x}{h} + 1\right)} \int_0^\infty \frac{t^{\frac{2x}{h}} e^{-\frac{2t}{h}}}{h^{2(\frac{x}{h}+1)} \Gamma^2\left(\frac{x}{h} + 1\right)} f(t) dt \\ &\leq \frac{h^{\left(\frac{2x}{h}+1\right)} \Gamma\left(\frac{2x}{h} + 1\right)}{2^{\frac{2x}{h}+1} h^{2\left(\frac{x}{h}+1\right)} \Gamma^2\left(\frac{x}{h} + 1\right)} \int_0^\infty \frac{2^{\frac{2x}{h}+1} t^{\frac{2x}{h}} e^{-\frac{2t}{h}}}{h^{\left(\frac{2x}{h}+1\right)} \Gamma\left(\frac{2x}{h} + 1\right)} f(t) dt \\ &\leq B_h(x) E[f(\eta_x)], \end{aligned}$$

tel que

$$B_h(x) = \frac{h^{-1}\Gamma\left(\frac{2x}{h} + 1\right)}{2^{\frac{2x}{h}+1}\Gamma^2\left(\frac{x}{h} + 1\right)}$$

et η_x suit la loi $\Gamma\left(\frac{2x}{h} + 1, h\right)$. Posons

$$R(z) = \frac{\sqrt{2\pi}e^{-z}z^{z+\frac{1}{2}}}{\Gamma(z+1)}, \quad z \geq 0$$

donc

$$\begin{aligned} B_h(x) &= \frac{h^{\frac{1}{2}}x^{-\frac{1}{2}}R^2\left(\frac{x}{h}\right)}{2\sqrt{\pi}R\left(\frac{2x}{h}\right)} \\ &= \begin{cases} \frac{1}{2\sqrt{\pi}}h^{-\frac{1}{2}}x^{-\frac{1}{2}} & \text{si } \frac{x}{h} \rightarrow \infty \\ \frac{\Gamma(2k+1)}{2^{2k+1}\Gamma^2(k+1)}h^{-1} & \text{si } \frac{x}{h} \rightarrow k \end{cases} \end{aligned}$$

Alors,

$$\text{var}[\hat{f}_3(x)] = \begin{cases} \frac{1}{2\sqrt{\pi}}n^{-1}h^{-\frac{1}{2}}x^{-\frac{1}{2}}f(x) & \text{si } \frac{x}{h} \rightarrow \infty \\ \frac{\Gamma(2k+1)}{2^{2k+1}\Gamma^2(k+1)}n^{-1}h^{-1}f(x) & \text{si } \frac{x}{h} \rightarrow k \end{cases}$$

3. Erreur quadratique moyenne intégrée (*MISE*) :

$$\begin{aligned} \text{MISE}[\hat{f}_3(x)] &= \int_0^{+\infty} \text{MSE}[\hat{f}_3(x)]dx = \int_0^{+\infty} [\text{biais}[\hat{f}_3(x)]]^2 + \int_0^{+\infty} \text{var}[\hat{f}_3(x)] \\ &= \int_0^{+\infty} \left[hf'(x) + \frac{h}{2}xf''(x)\right]^2 + \int_0^{+\infty} \frac{1}{2\sqrt{\pi}}n^{-1}h^{-\frac{1}{2}}x^{-\frac{1}{2}}f(x) + O(h^2). \end{aligned}$$

Alors

$$\begin{aligned} \text{MISE}[\hat{f}_3(x)] &= h^2 \int_0^{+\infty} \left[f'(x) + \frac{1}{2}xf''(x)\right]^2 dx + \frac{1}{2\sqrt{\pi}}n^{-1}h^{-\frac{1}{2}} \int_0^{+\infty} x^{-\frac{1}{2}}f(x)dx \\ &\quad + O\left(\frac{1}{n\sqrt{h}} + h^2\right). \end{aligned}$$

Pour minimiser la $MISE$ par rapport à la fenêtre h , nous avons

$$AMISE[\hat{f}_3(x)] = h^2 \int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^{+\infty} x^{-\frac{1}{2}} f(x) dx$$

On dérive par rapport à h

$$\frac{\partial AMISE}{\partial h}(h) = 0$$

on obtient :

$$2h \int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx - \frac{1}{2n} \frac{1}{2h^2 \sqrt{\pi}} \int_0^{+\infty} x^{-\frac{1}{2}} f(x) dx = 0.$$

Alors, la la fenêtre optimale h_{opt3}^*

$$h_{opt3}^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^{+\infty} x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{4^{\frac{2}{5}} \left[\int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx \right]^{\frac{2}{5}}} n^{-\frac{2}{5}} = O\left(n^{-\frac{2}{5}}\right).$$

$$\begin{aligned} MISE_{h_{opt3}^*}(\hat{f}_3(x)) &= (h_{opt3}^*)^2 \int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx \\ &+ \frac{1}{2\sqrt{\pi}} n^{-1} (h_{opt3}^*)^{-\frac{1}{2}} \int_0^{+\infty} x^{-\frac{1}{2}} f(x) dx + o\left(\frac{1}{n\sqrt{h}} + h^2\right) \\ &= \frac{5}{4^{\frac{4}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^{+\infty} x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \\ &\times \left[\int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \end{aligned}$$

Propriétés de l'estimateur \hat{f}_4

1. Le biais :

$$E[\hat{f}_4(x)] = \begin{cases} f(x) + \frac{1}{2}hx f''(x) + O(h) & \text{si } x \geq 2h \\ f(x) + \xi_h(x)hf'(x) + O(h) & \text{si } x \in [0, 2h[\end{cases}$$

Alors,

$$\text{biais}[\hat{f}_4(x)] = \begin{cases} \frac{1}{2}hx f''(x) + O(h) & \text{si } x \geq 2h \\ \xi_h(x)hf'(x) + O(h) & \text{si } x \in [0, 2h[\end{cases}$$

$$\text{avec } \xi_h(x) = \frac{(1-x)(\rho_h(x) - \frac{x}{h})}{(1+h\rho_h(x)-x)}.$$

2. Erreur quadratique moyenne :

$$MSE[\hat{f}_4(x)] = \frac{h^2}{4} [x f''(x)]^2 + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} x^{-\frac{1}{2}} f(x) + O\left(h^2 + (nh)^{-\frac{1}{2}}\right).$$

3. Erreur quadratique moyenne intégrée :

$$MISE[\hat{f}_3(x)] = \frac{h^2}{4} \int_0^\infty [x f''(x)]^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx + O\left(h^2 + (nh)^{-\frac{1}{2}}\right)$$

Par minimisation de la *AMISE* par rapport à la fenêtre h :

$$AMISE[\hat{f}_4(x)] = \frac{h^2}{4} \int_0^\infty [x f''(x)]^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx,$$

la fenêtre optimale est h_{opt4}^* est telle que :

$$h_{opt4}^* = \frac{[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx]^{\frac{2}{5}}}{[\int_0^\infty [x f''(x)]^2 dx]^{\frac{2}{5}}} n^{-\frac{2}{5}} = O\left(n^{-\frac{2}{5}}\right).$$

La *MISE* associée est

$$\begin{aligned} MISE_{h_{opt4}^*}[\hat{f}_2(x)] &= \frac{5}{4^{\frac{4}{5}}} \left[\frac{1}{2\sqrt{\pi}} \int_0^{\infty} x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \\ &\times \left[\int_0^{\infty} [x f''(x)]^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}} \end{aligned}$$

Remarque 2.3.1 *On remarque que*

$$\int_0^{+\infty} \left[f'(x) + \frac{1}{2} x f''(x) \right]^2 dx \geq \int_0^{\infty} [x f''(x)]^2 dx$$

on déduit que

$$MISE_{h_{opt3}^*}[\hat{f}_3(x)] \geq MISE_{h_{opt4}^*}[\hat{f}_4(x)] \text{ et } h_{opt3}^* \geq h_{opt4}^*$$

Alors \hat{f}_4 est le meilleur estimateur pour f .

Conclusion

Dans ce mémoire nous avons essayé d'étudier les propriétés asymptotique d'un estimateur non paramétrique pour une fonction de densité de probabilité avec la méthode du noyau (symétrique et asymétrique).

Nous avons présenté dans premier chapitre la méthode d'estimation à noyau symétrique proposé par Rosenblatt en 1956 puis amélioré par Parzen en 1962. Ensuite, dans le deuxième chapitre, nous avons passé aux estimateurs à noyau asymétrique, où nous avons pris comme exemples les noyaux gamma et bêta proposés par Chen en (1999 – 2000).

En conclusion, pour l'estimation des densités à support asymétrique R^+ ou $[0, 1]$ par exemple, l'utilisation de noyau symétrique donne une estimation de plus en plus biaisée lorsque x est proche de 0. Ce qui montre l'importance de l'utilisation de noyau asymétrique, qui ont d'excellentes propriétés asymptotiques.

Bibliographie

- [1] Ben Khalifa, I. (2008). Estimation non-paramétrique par noyaux associés et données de panel en marketing. Projet de Fin d'Etude. Université du 7 Novembre à Carthage. Tunisie.
- [2] Belahcene, I. (2017) .Estimation non paramétrique de la fonction densité de probabilité avec un noyau. Mémoire master 2, Université Kasdi Merbah–Ouargla.
- [3] Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis*, 31(2), 131-145.
- [4] Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3), 471-480.
- [5] Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de statistique appliquée*, 25(3), 5-42.
- [6] Dobele-Kpoka, F.G.B.L. (2013). Méthode non-paramétrique des noyaux associés mixtes et applications. Thèse de Doctorat, Université Ouaga 1 (Ouagadougou, Burkina Faso).
- [7] Epanechnikov, V. A. (1969). Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1), 153-158.

- [8] Jeon, Y., Kim, J.H. (2013). A gamma kernel density estimation for insurance loss data. *Insurance : Mathematics and Economics*, 53(3), 569-579.
- [9] Graunt, J. (1662). *Natural and Political Observations made upon the Bills of Mortality*. T. Roycroft, London.
- [10] Kiessé, T.S. (2008). *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat, Université de Pau et des Pays de l'Adour. France.
- [11] Kokonendji, M.C.C. (2009). *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat, Université Toulouse. France.
- [12] Menacer, C.E. (2021). *Estimation non Paramétrique de la Densité de Probabilité*. Mémoire Master 2. Université Mohamed Khider Biskra. Algérie.
- [13] Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3), 1065-1076.
- [14] Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, 58, 347-352.
- [15] Rosenblatt, F. (1958). The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- [16] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. *Nonparametric statistics*, 16(1-2), 217-226.
- [17] Simonoff, J. S. (2012). *Smoothing methods in statistics*. Springer Science & Business Media.
- [18] Silverman, B.W. (2018). *Density estimation for statistics and data analysis*. Routledge.

- [19] Tsybakov, A.B. (2003). Introduction à l'estimation non paramétrique (Vol.41). Springer Science & Business Media.
- [20] TOUR, T. (2020). Sur l'estimation de la fonction de la densité. Mémoire Master 2. Université Mohamed Khider Biskra. Algérie.
- [21] Węglarczyk, S. (2018). Kernel density estimation and its application. In ITM Web of Conferences (Vol. 23, p. 00037). EDP Sciences.

ملخص

في هذه المذكرة، حاولنا دراسة خصائص التقارب لمقدرة لا معلمية لدالة كثافة الاحتمال باستخدام طريقة النواة (المتماثل وغير المتماثل). استنادًا إلى أمثلة بيانات تمت محاكاتها باستخدام البرنامج R، أظهرنا أيضًا أهمية استخدام النواة غير المتماثلة، والتي لها خصائص تقارب ممتازة مقارنة بالمتماثلة التي تعطي المزيد من المقدرات المتحيزة عند الحواف في حالة دالة الكثافة غير المتماثلة.

Résumé

Dans ce mémoire, nous avons essayé d'étudier les propriétés asymptotiques d'un estimateur non paramétrique pour une fonction de densité de probabilité avec la méthode du noyau (symétrique et asymétrique). En s'appuyant sur des exemples de données simulées avec le logiciel R, nous montrons aussi l'importance de l'utilisation des noyaux asymétriques, qui ont d'excellentes propriétés asymptotiques par rapport aux noyaux symétriques qui donnent d'estimateur de plus en plus biaisée au bords dans le cas de densité à support asymétrique.

Abstract

In this memory, we tried to study the asymptotic properties of a nonparametric estimator for a probability density function with (symmetric and asymmetric) kernel method. Based on examples of data simulated with the R software, we also show the importance of using asymmetric kernels, which have excellent asymptotic properties compared to symmetric kernels which give more and more biased estimators at the boundary in the case of density with asymmetric support.