



UNIVERSITE KASDI MERBAH
OUARGLA
Faculté des Mathématiques et des Sciences de la
Matière



DÉPARTEMENT DE MATHÉMATIQUES

Mémoire Master

Domaine : Mathématiques

Spécialité : Probabilité et Statistique

THÈME

*Estimation non paramétrique de la
fonction densité de probabilité avec un
noyau.*

Établi par :
Belahcene Intissar

Jury composé de :
Encadreur - Baheddi Aissa
Président - Meflah Mabrouk
Examineur - Agoune Rachid

Année universitaire : 2016/2017

DÉDICACE

Je dédie ce modeste travail :

*à l'esprit de ma cher mère **MABROUKA** Que Dieu la protège pour m' avoir encouragé et permis d' entreprendre mes études. Sans eux, je ne serais pas là .
, et mon chère père **AHMED** qui avait toujours souhaité ma réussite.*

à mes chers frères Ammer Louai. à mes chères sœurs Nesrine, Ferial, Aya, qui étaient toujours présents pour moi , pour ses disponibilités, ses conseils et pertinents, et son aides précieuse qui m 'a apporté tout au long de ce travail.

à ma fiancé M^{ed} Nadjibe ,pour sa dévouement exemplaire et les énormes sacrifices qu'il m'a consentis durant mes études

à mes oncles et mes tantes. à mes chers cousins et cousines maternel et paternel, ma belle mère, mon beau père, mes belles sœurs et beaux frères qui m ' ont toujours encouragés , ils ont toujours aimé me voir réussir. Je les remercie pour tout ce qu'ils mon fait

REMERCIEMENT

La louange est à Allah

Avant tout nous remercions ALLAH le tout puissant, de nous avoir donnés la force, le courage, la volonté et la patience pour terminer ce mémoire .

*Nous tenons à présenter nos sincères remerciements **Baheddi Aissa**, pour avoir accepté de nous encadrer et de nous avoir orientées, aidées et encouragées tout au long de la réalisation de ce travail. pour ses conseils permanents, et sa disponibilité . Nous avons beaucoup appris à ses côtés et nous sommes très honorés de l'avoir eu pour encadreur.*

*Nos remerciements vont également aux **membres du jury** pour l'intérêt qu'ils ont porté à notre travail. Nous remercions vivement Monsieur. **Meflah Mabrouk** qui nous a fait l'honneur de présider le jury de ce mémoire . Monsieur **Agoune Rachid** pour nous avoir honorés de sa présence en acceptant d'examiner ce travail.*

*Le professeur vertueux **Chacha Lamia** remercie aussi de me aider tout au long de la période de travail*

*Nous exprimons toute nos reconnaissance et gratitude à l'administration et à l'ensemble du corps enseignant de l'Université **KASDI MERBAH** pour leurs efforts à nous garantir la continuité et l'aboutissement de ce programme de Master*

Finally nous tenons à remercier nos familles respectives nos chères sœurs, et frères, pour leurs soutiens, conseils, et leur aide, ainsi que tous nos ami(e)s avec qui nous avons partagée les bons moments inoubliables de notre cursus universitaire.

TABLE DES MATIÈRES

Dédicace	i
Remerciement	ii
Notations	v
1	3
1.1 Introduction	3
1.2 Un peu d'histoire	3
1.3 La problématique :	4
1.3.1 Intuition du méthode du noyau	4
1.3.2 Estimation non-paramétrique par noyaux associés et données de panel en marketing.	5
1.3.3 Estimateurs à noyau et théorie des valeurs extrêmes.	5
1.3.4 Méthode non-paramétrique des noyaux associés mixtes et applications.	5
1.3.5 Estimation de la fonction de répartition.	6
1.4 Analyse des documents	6
1.5 Résumé	6
2	7
2.1 Introduction	7
2.1.1 La construction d'un estimateur à noyaux	8
2.2 Notion de noyaux	8
2.2.1 Exemples des noyaux continus symétriques	9
2.2.2 Propriétés d'un estimateur à noyau symétrique	10
2.2.3 Choix du paramètre de lissage h	14

2.2.4	Choix du Noyaux	16
2.3	Estimation de la densité avec un noyau continue asymétrique Gamma . . .	17
2.3.1	Introduction	17
2.3.2	Exemples de noyaux continues asymétriques	17
2.3.3	Propriétés d'un estimateur à noyau Gamma	21
2.3.4	Un autre estimateur à noyau asymétrique pour la densité	23
2.3.5	propriétés de l'estimateur $\hat{f}_2(x)$	24
2.3.6	Choix du paramétré de lissage h	24
3		27
3.1	Introduction	27
3.2	Estimation à noyau symétrique ,asymétrique avec des données i.i.d	28
3.2.1	L'estimation à noyau Gaussien :	28
3.2.2	L'estimation à noyau Gamma :	30
3.3	Influence de choix du paramètre de lissage h :	32
3.3.1	Cas symétrique (Gaussien) :	32
3.3.2	Cas asymétrique (Gamma) :	32
	Conclusion	34

NOTATIONS

- $\hat{F}_n(x)$:la fonction de répartition empirique
- $\hat{f}(x)$:l'estimateur à noyau de la densité de probabilité f .
- $\hat{f}_1(x)$:Un estimateur non paramétrique d'une densité de probabilité f à noyau Gamma .

INTRODUCTION GÉNÉRAL

Souvent en statistique à partir d'une suite des variables aléatoires 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 elle mêmes, donc au lieu de supposer une loi paramétrique et d'utiliser un échantillon pour estimer les différents paramètres, on utilise directement l'échantillon comme estimateur de la densité.

La méthode d'estimation non paramétrique la plus simple est celle de l'histogramme qui était introduit par John Graunt en 1662 [5], un problème vient du fait que l'histogramme donne une fonction qui n'est pas continue.

La méthode d'estimation non paramétrique du noyau fut introduit par Rosenblatt en 1956 [7], puis amélioré par Parzen en 1962 [06]. Mais ces estimateurs connaissent quelques inconvénients, parmi lesquels les problèmes de support.

Le premier cas celui des effets de bords dans l'estimateur de la densité à support bornée, comme les noyaux symétriques assignent des poids à l'extérieur du support de la densité.

Récemment, Certain auteurs ont proposé dans le cas d'une densité à support bornée, l'utilisation de noyaux dont le support coïncide avec celui de la densité estimée, ceci a efficacement résout le problème des effets de bords puisque les noyaux utilisées sont généralement asymétriques, et peuvent changer la forme selon la position du point d'estimation. c'est notamment le cas de Chen avec les noyau gamma pour estimer la densité respectivement dans $[0, \infty[$

CHAPITRE 1

1.1 INTRODUCTION

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

1.2 UN PEU D'HISTOIRE

la méthode d'estimation non paramétrique du noyau fut introduit par Rosenblatt en 1956, puis amélioré par Parzen en 1962. Mais ces estimateurs connaissent quelques inconvénients, parmi lesquels les problèmes de support, Lorsque le support de f n'est symétrique les estimateurs classiques provoquent des effets aux bornes. Certains auteurs ont proposé pour ce problème, l'utilisation du noyau dont le support coïncide avec celui de la densité estimée. c'est notamment le cas de Chen avec les noyaux Béta et Gamma pour estimer la densité respectivement dans $[0,1]$ et $[0, \infty[$

1.3 LA PROBLÉMATIQUE :

Comment on estime la densité de probabilité avec la méthode du noyaux en utilisant des données i. i. d ?

Ces problèmes locaux :

1. Quelle est l'influence de paramètre de lissage h sur l'estimateur \hat{f} ?
2. Comment en choisit la taille d'échantillon n ?

Les hypothèses

1. le paramètre de lissage h a une grande influence sur d'estimateur .plus que le h est petit l'estimateur est bon.
2. Plus que la taille d'échantillon n augmente on obtient un bon estimateur.

les motivations

Souvent en statistique à partir d'une suite de variables aléatoires on cherche à estimer la densité de probabilité avec des méthodes paramétrique ou non paramétrique. Mon mémoire est consacré à l'étude d'estimation non paramétrique avec une noyaux asymétrique Gamma

Les obstacles

Il est impossible dans certain cas d'observer une phénomène statistique pour trouver tout les valeurs qui présentent l'échantillon. Ce qui provoque parfois le manque des variables.

1.3.1 Intuition du méthode du noyau

La méthode de Parzen est une généralisation de la méthode d'estimation par histogramme. Dans un histogramme, la densité en un point x est estimée par la proportion d'observations x_1, x_2, \dots, x_n qui se trouvent à proximité de x . Pour cela, on trace une boîte en x et dont la largeur est gouvernée par un paramètre de lissage h ; on compte ensuite le nombre d'observations qui appartiennent à cette boîte. Cette estimation, qui dépend du

paramètre de lissage h , présente de bonnes propriétés statistiques mais est par construction non-continue. La méthode du noyau consiste à retrouver la continuité : pour cela, on remplace la boîte centrée en x et de largeur h par une gaussienne centrée en x . Plus une observation est proche du point de support x plus la courbe en cloche lui donnera une valeur numérique importante. À l'inverse, les observations trop éloignées de x se voient affecter une valeur numérique négligeable. L'estimateur est formé par la somme (ou plutôt la moyenne) des courbes en cloche.

après liée les documents

1.3.2 Estimation non-paramétrique par noyaux associés et données de panel en marketing.

le thème est Estimation Non-paramétrique par Noyaux Associés et Données de Panel en Marketing (Tunis)

l'auteur est Imen Ben Khalifa

le problème global est Quel est les méthodes statistiques associées pour l'estimation non-paramétrique

les résultats est :estimé la densité de probabilité par le méthode de noyau.

1.3.3 Estimateurs à noyau et théorie des valeurs extrêmes.

le thème est Estimateurs à Noyau et théorie des valeurs extrêmes.

l'auteur est ETIENNE DOUCET

le problème global est Comment estime la fonction de densité de probabilité ?

les résultats est :L'estimation par maximum de vraisemblance est une méthode paramétrique et L'estimation par noyau est une méthode non paramétrique.

1.3.4 Méthode non-paramétrique des noyaux associés mixtes et applications.

le thème est Méthode non-paramétrique des noyaux associés mixtes et applications

l'auteur est :Francial Giscard Baudin LIBENGUÉ DOBÉLÉ-KPOKA

le problème global est :Quel est Estimateur de densités mixtes ?

les résultats est : l'estimateur non-paramétrique par noyaux associés mixtes, pour les densités à supports partiellement continus et discrets.

1.3.5 Estimation de la fonction de répartition.

le thème est Estimation de la fonction de répartition.

l'auteur est Remi Servien

le problème global est : Quel est la construction d'un estimateur à noyau ?

les résultats est : la densité de probabilité f est égale à la dérivée de la fonction de répartition F .

1.4 ANALYSE DES DOCUMENTS

Après avoir lu les documents on remarque que les documents on généralement concentré sur Estimation Non-paramétrique par Noyaux Associés et Données de Panel en Marketing et Estimation de la fonction de répartition et Méthode non-paramétrique des noyaux associés mixtes et applications et Estimateur à noyau et théorie des valeurs extrêmes la même mais on remarque la déferrent sur :

l'estimateur à noyau continu symétrique \hat{f} d'une densité de probabilité f inconnue sur \mathbb{R} et les noyaux associés continus et discrets pour définir les noyaux associés mixtes

L'estimation par maximum de vraisemblance est une méthode paramétrique. Le statisticien doit donc décider d'une loi ou d'une famille de lois paramétriques avant d'utiliser la méthode du maximum de vraisemblance pour estimer les paramètres de la loi choisie.

1.5 RÉSUMÉ

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

CHAPITRE 2

2.1 INTRODUCTION

Supposons que nous observons n variables aléatoires indépendantes identiquement distribuées (i.i.d.) X_1, X_2, \dots, X_n de densité de probabilité inconnue f de \mathbb{R} dans $[0, \infty[$. Soit $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X . la fonction de répartition empirique est estimée par :

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(X_i \leq x)} \quad (2.1)$$

La loi fort des grands nombres nous donne

$$\forall x \in \mathbb{R} : \hat{F}_n(x) \xrightarrow{P} F(x) \quad \text{si } n \longrightarrow +\infty$$

Rosenblatt (1956) est le premier qui a donné un exemple d'estimateur à partir de $\hat{F}_n(x)$ pour $h > 0$ est petit

$$\hat{f}(x) \simeq \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h}$$

2.1.1 La construction d'un estimateur à noyaux

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

$$\begin{aligned} f(x) &= \lim_{h \rightarrow 0} \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \\ &= \lim_{h \rightarrow 0} \frac{P(x-h < X \leq x+h)}{2h} \end{aligned}$$

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

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2h} \frac{P(x-h < X_i \leq x+h)}{n} \\ &= \frac{1}{2hn} \sum_{i=1}^n I\{x-h < X_i \leq x+h\} \\ &= \frac{1}{2h} \sum_{i=1}^n I\left\{-1 \leq \frac{x-X_i}{h} < +1\right\}. \end{aligned}$$

Notons que cet estimateur peut encore s'écrire comme

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

Où

$$K_0(y) = \begin{cases} 1/2 & \text{si } y \in [-1, 1[\\ 0 & \text{sinon} \end{cases}$$

Avec $K_0(\cdot)$ La densité de probabilité uniforme sur l'intervalle $[-1, 1[$ est appelée noyau de Rosenblatt[07] . Cet estimateur peut être généralisé en remplaçant la fonction de poids $K_0(\cdot)$ par une fonction de poids plus générale K par exemple une densité de probabilité quelconque (Normale, Gamma, Bêta... etc)

2.2 NOTION DE NOYAUX

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

Définition 2.2.1 Soit $K : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction intégrable tel que $\int_{\mathbb{R}} K(u) du = 1$ K est dit noyau.

pour $n \in \mathbb{N}^*$, on appelle $h_n > 0$ la fenêtre ou paramétré de lissage. et $\hat{f}(x)$ d'estimateur à

noyau de la densité de probabilité f (estimateur de Parzen -Rozenblatt)[07] définit pour tout $x \in \mathbb{R}$ par

$$\hat{f}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{X_i - x}{h_n}\right) \quad (2.2)$$

On note $h_n = h$

Dans cette partie, nous présentons d'estimateur d'une densité de probabilité 'a noyau symétrique , puis nous donnons les différentes propriétés de cet estimateur tel que le Biais , la Variance, le MSE ,et le MISE. Puis nous précisons le choix de paramètre de lissage h .

Définition 2.2.2 Un noyau est dit symétrique si, pour tout u dans son ensemble de définition $K(u) = K(-u)$, $\forall u$ dans sa domaine de définition.

ce qui implique l'égalité suivante :

$$\int_{\mathbb{R}} uK(u)du = 0 \quad (2.3)$$

De plus, elle est de carré intégrable

$$\int_{\mathbb{R}} K^2(u)du < +\infty \quad (2.4)$$

et nous avons aussi la variance de K finie

$$\int_{\mathbb{R}} u^2K(u)du < +\infty \quad (2.5)$$

2.2.1 Exemples des noyaux continus symétriques

Voici quelques exemples de noyaux symétriques les plus utilisés :

Noyaux	Supports	Densités
Biweight	$[-1, 1]$	$K(u) = \frac{15}{16}(1 - u^2)^2$
Epanechnikov	$[-1, 1]$	$K(u) = \frac{3}{4}(1 - u^2)$
Gaussien	\mathbb{R}	$K(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$
Rectangulaire	$[-1, 1]$	$K(u) = \frac{1}{2}$
Triangulaire	$[-1, 1]$	$K(u) = 1 - u 1_{\{ u \leq 1\}}$

TABLE 2.1 : Exemples des noyaux symétriques

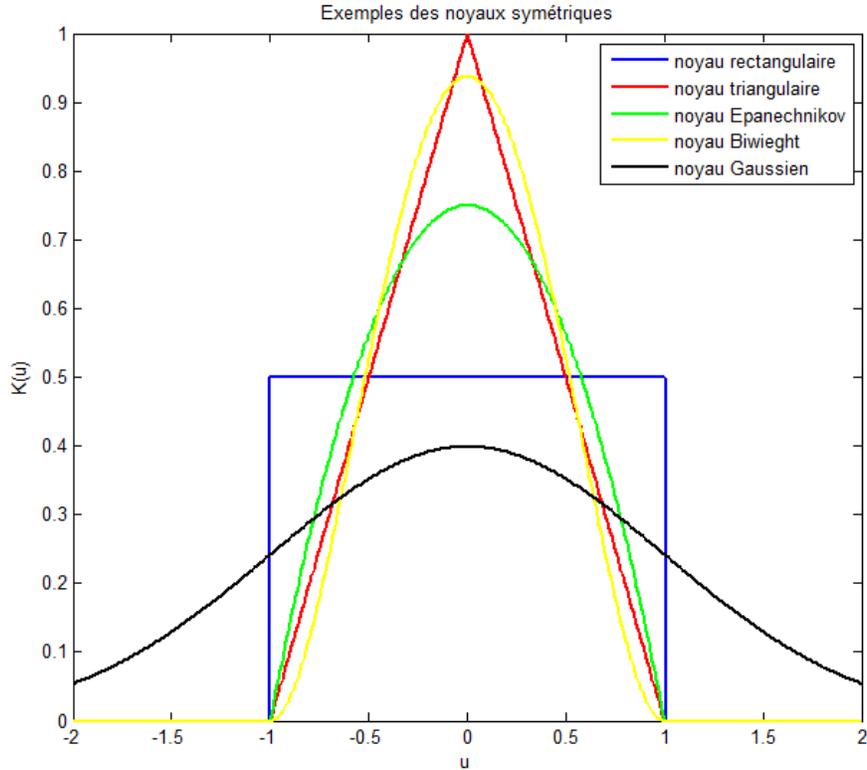


FIGURE 2.1 : Exemples de noyaux symétriques

2.2.2 Propriétés d'un estimateur à noyau symétrique

La fonction $x \rightarrow \hat{f}(x)$ est une densité de probabilité.

Démonstration Soit $\hat{f}(x)$ un estimateur non paramétrique à noyau symétrique défini par :

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

$\hat{f}(x)$ vérifie $\int_{\mathbb{R}} \hat{f}(x) dx = 1$. En effet

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

Car les X_i sont i.i.d

$$= \int_{-\infty}^{+\infty} K(u)du = 1 \quad \text{en posant} \quad u = \frac{X - x}{h}$$

Biais ponctuel

Le biais ponctuel mesure la différence entre la valeur moyenne de l'estimateur $\hat{f}(x)$ et la valeur de la fonction inconnue f en un point x .

Biais ($\hat{f}(x)$) = $\mathbb{E}(\hat{f}(x)) - f(x)$

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

$$\text{Biais}(\hat{f}(x)) = \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2)$$

Les variables aléatoires sont i.i.d nous avons donc :

$$\begin{aligned} 1. E(\hat{f}(x)) &= \frac{1}{nh} E\left(\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)\right) \\ &= \frac{1}{h} E\left(K\left(\frac{X - x}{h}\right)\right) \\ &= \frac{1}{h} \int_{-\infty}^{+\infty} K\left(\frac{y - x}{h}\right) f(y) dy \\ &= \int_{-\infty}^{+\infty} K(u) f(x + uh) du. \quad \text{en posant} \quad u = \frac{y - x}{h} \end{aligned}$$

En utilisant le développement de Taylor de f au voisinage de x on obtient :

$$\begin{aligned} f(x + uh) &= f(x) + huf'(x) + \frac{(uh)^2 f''(x)}{2} + o(h^2 u^2) \\ E(\hat{f}(x)) &= \int_{-\infty}^{+\infty} K(u) \left[f(x) + f'(x)uh + \frac{1}{2} f''(x)(uh)^2 + o(h^2) \right] du \\ &= f(x) \int_{-\infty}^{+\infty} K(u) du + hf'(x) \int_{-\infty}^{+\infty} uK(u) du + f''(x) \frac{h^2}{2} \int_{-\infty}^{+\infty} u^2 K(u) du \end{aligned}$$

Sous les conditions précédents on a :

$$E(\hat{f}(x)) = f(x) + \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2)$$

Ainsi

$$\begin{aligned} \text{Biais}(\hat{f}(x)) &= E(\hat{f}(x)) - f(x) \\ &= \frac{h^2}{2} f''(x) \int_{-\infty}^{+\infty} u^2 K(u) du + o(h^2) \end{aligned}$$

Variance ponctuelle

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

$$Var(\hat{f}(x)) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + o\left(\frac{1}{nh}\right)$$

$$\begin{aligned} 2. Var(\hat{f}(x)) &= V \left\{ \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \right\} \\ &= \frac{1}{n} \left\{ \frac{1}{h^2} E \left(\left[K\left(\frac{X - x}{h}\right) \right]^2 \right) - \frac{1}{h} E^2 \left(K\left(\frac{X - x}{h}\right) \right) \right\} \\ &= \frac{1}{nh} \int_{-\infty}^{+\infty} (K(u))^2 f(x + uh) du - \frac{1}{n} \left[\int_{-\infty}^{+\infty} K(u) f(x + uh) du \right]^2 \end{aligned}$$

Le terme

$$\frac{1}{n} \left[\int_{-\infty}^{+\infty} K(u) f(x + uh) du \right]^2 \rightarrow 0 \quad \text{lorsque } n \rightarrow \infty$$

Donc

$$Var(\hat{f}(x)) = \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + o\left(\frac{1}{nh}\right)$$

Ainsi

$$\text{Biais} \quad \left(\hat{f}(x) \right) = \frac{1}{2} f''(x) \mu_2 h^2 + o(h^2) \quad \text{où} \quad \mu_2 = \int K(u) u^2 du$$

$$\text{Var} \quad \left(\hat{f}(x) \right) = \frac{1}{nh} f(x) R(K) + o\left(\frac{1}{nh}\right) \quad \text{Où} \quad R(K) = \int K^2(u) du$$

- Pour un h petit le *Biais* $\left(\hat{f}(x) \right)$ dépend de $f''(x)$ et du moment d'ordre 2 du noyau, le Biais est de signe de $f''(x)$.

- Si $h = h_n \rightarrow 0$; quand $n \rightarrow \infty$, alors *Biais* $\left(\hat{f}(x) \right) \rightarrow 0$

- Si $h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$, Alors *Var* $\left(\hat{f}(x) \right) \rightarrow 0$

•

$$\begin{cases} \text{Si } h \text{ diminue} & \longrightarrow (\text{Biais})^2 \text{ diminue et la Variance augmente} \\ \text{Si } h \text{ augmente} & \longrightarrow (\text{Biais})^2 \text{ augmente et la Variance diminue} \end{cases}$$

Erreur quadratique moyenne (MSE)

L'erreur quadratique moyenne (en anglais "Mean squared Error") est donnée par :

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

Propriétés 6

$$\begin{aligned} MSE(x) &= \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + o\left(\frac{1}{nh}\right) \\ &+ \frac{h^4}{4} \{f''(x)\}^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + o(h^4) \end{aligned}$$

démonstration

$$\begin{aligned} 1. MSE(x) &= E \left(\hat{f}(x) - f(x) \right)^2 \\ &= E \left(\hat{f}(x) - E \left(\hat{f}(x) \right) + E \left(\hat{f}(x) \right) - f(x) \right)^2 \\ &= \text{Var} \left(\hat{f}(x) \right) + \left(\text{Biais}(\hat{f}(x)) \right)^2 \\ &= \frac{1}{nh} f(x) \int_{-\infty}^{+\infty} K^2(u) du + o\left(\frac{1}{nh}\right) \\ &+ \frac{h^4}{4} (f''(x))^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + o(h^4) \end{aligned}$$

Erreur quadratique moyenne intégrée (MISE) :

L'erreur quadratique moyenne intégrée (en anglais "Mean Integrated Squared Error") est donnée par :

$$MISE(n, h) = \int_{-\infty}^{+\infty} MSE(x) dx$$

Propriété 7

$$\begin{aligned} MISE(n, h) &= \frac{1}{nh} \int_{-\infty}^{+\infty} \left(f(x) \left[\int_{-\infty}^{+\infty} K^2(u) du \right] + o\left(\frac{1}{nh}\right) \right) dx \\ &+ \frac{h^4}{4} \int_{-\infty}^{+\infty} \left((f''(x))^2 \left[\int_{-\infty}^{+\infty} u^2 K(u) du \right]^2 + o(h^4) \right) dx \end{aligned}$$

• $h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ quand $n \rightarrow \infty$ Alors $MSE(x) \rightarrow 0$

On dit que $\hat{f}(x)$ converge en probabilité vers $f(x)$.

2.2.3 Choix du paramètre de lissage h

Le paramètre de lissage h a une grande influence sur la performance de l'estimateur $\hat{f}(x)$. Il y a essentiellement deux approches pour trouver une largeur de bande optimale. La première consiste à trouver le paramètre qui minimise l'erreur quadratique moyenne de $\hat{f}(x)$, c'est-à-dire

$$\text{ArgMin} \left[E \left(\hat{f}(x) - f(x) \right)^2 \right]$$

On obtient donc un paramètre de lissage optimal, qui varie en fonction du x où l'on veut estimer la fonction de densité f .

La seconde approche nous donne un paramètre de lissage optimal globale, qui ne dépend pas de x . Pour ce faire, on cherche le h qui minimise l'erreur quadratique moyenne intégrée (MISE), c'est-à-dire

$$h_{\text{optimal}} = \text{ArgMin}[MISE(n, h)] = \text{ArgMin} \left[\int_{\mathbb{R}} E \left\{ \hat{f}(x) - f(x) \right\}^2 dx \right]$$

On suppose que la densité à estimer f et le noyau K sont des fonctions de carré intégrable, de sorte que le MISE est finie.

$$MISE(n, h) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh} R(K) + o \left(\frac{1}{nh} + h^4 \right)$$

l'approximation asymptotique de la MISE est donné par

$$AMISE(n, h) = \frac{h^4}{4} R(f'') (\mu_2(K))^2 + \frac{1}{nh} R(K)$$

On dérive ce AMISE par rapport à h et on égale à 0, on obtient

$$\frac{\partial}{\partial h} AMISE(n, h) = h^3 R(f'') (\mu_2(K))^2 - \frac{R(K)}{nh^2}$$

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

$$\implies h_{\text{optimal}} = \left[\frac{R(K)}{R(f'') (\mu_2(K))^2} \right]^{\frac{1}{5}} n^{-1/5} \quad (2.6)$$

Cas particuliers

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires de densité de probabilité f , supposons que f appartient à une famille de distributions normales, $\mathcal{N}(\mu; \sigma^2)$

$$\text{Si } \begin{cases} X_i \sim \mathcal{N}(\mu; \sigma^2) \\ K \sim \mathcal{N}(0, 1) \end{cases} \quad \text{Alors } h_{\text{optimal}} = 1.06 \hat{\sigma} n^{-1/5}$$

Si $f \sim \mathcal{N}(\mu; \sigma^2)$ alors $f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$, avec $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$,
 et

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

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

$$\begin{aligned} R(f'') &= \int_{-\infty}^{+\infty} [f''(x)]^2 dx \\ &= \frac{1}{\sigma^6} \int_{-\infty}^{+\infty} \left\{ \varphi''\left(\frac{x-\mu}{\sigma}\right) \right\}^2 dx \\ &= \frac{1}{\sigma^5} \int_{-\infty}^{+\infty} \{\varphi''(v)\}^2 dv \end{aligned}$$

Nous avons :

$$\begin{aligned} \varphi(v) &= \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \\ \Rightarrow \varphi'(v) &= -\frac{v}{\sqrt{2\pi}} e^{-v^2/2} \\ \Rightarrow \varphi''(v) &= \frac{1}{\sqrt{2\pi}} (v^2 - 1) e^{-v^2/2}. \end{aligned}$$

$$\begin{aligned} R(f'') &= \frac{1}{\sigma^5} \int_{-\infty}^{+\infty} \left\{ \frac{1}{\sqrt{2\pi}} (v^2 - 1) e^{-v^2/2} \right\}^2 dv \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ \int_{-\infty}^{+\infty} v^4 e^{-v^2} dv - 2 \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} v^2 e^{-v^2} dv + \int_{-\infty}^{+\infty} e^{-v^2} dv \right\} \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{2} \int_{-\infty}^{+\infty} \frac{u^2}{2} e^{-u^2/2} \frac{1}{\sqrt{2}} du + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2}} e^{-u^2/2} du \right\} \quad \text{avec } u = \sqrt{2}v \\ &= \frac{1}{\sigma^5} \frac{1}{2\pi} \left\{ -\frac{1}{4} \sqrt{\pi} + \sqrt{\pi} \right\} \\ &= \frac{1}{\sigma^5} \frac{3}{8\sqrt{\pi}}. \end{aligned}$$

Donc, l'expression du paramètre de lissage optimal devient

$$h_{\text{optimal}} = \left[\frac{8\sqrt{\pi} R(K)}{3 [\mu_2(K)]^2} \right]^{\frac{1}{5}} \hat{\sigma} n^{-1/5}$$

où $\hat{\sigma}$ est un estimateur de σ , tel que

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

On a $K \rightarrow N(0, 1)$ alors

$$\begin{aligned}
 R(K) &= \int_{-\infty}^{+\infty} [K(u)]^2 du \\
 &= \int_{-\infty}^{+\infty} \left[\frac{1}{\sqrt{2\pi}} e^{-u^2/2} \right]^2 du \\
 &= \int_{-\infty}^{+\infty} \frac{1}{2\pi} e^{-u^2} du \\
 &= \frac{1}{2\pi} \sqrt{\pi} = \frac{1}{2\sqrt{\pi}}
 \end{aligned}$$

Nous avons $\mu_2(K) = \int_{-\infty}^{+\infty} u^2 K(u) du = 1$.

Nous remplaçons dans l'équation (1.3) nous obtenons :

$$h_{optimal} = \left(\frac{4}{3}\right)^{\frac{1}{5}} \hat{\sigma} n^{-1/5} = 1.06 \hat{\sigma} n^{-1/5}$$

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires de densité de probabilité f , supposons que f appartient à une famille de distributions normales $\mathcal{N}(\mu; \sigma^2)$, soit K est un noyau d'Epanechnikov.

$$\text{Si } \begin{cases} X_i & \sim \mathcal{N}(\mu; \sigma^2) \\ K(u) & = \frac{3}{4}(1-u^2), 1_{|u| \leq 1} \end{cases} \quad \text{Alors} \quad h_{optimal} = 2.34 \hat{\sigma} n^{-1/5}$$

2.2.4 Choix du Noyaux

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

$$eff(K_1, K_2) = \frac{AMISE(K_1, n, h)}{AMISE(K_2, n, h)} < 1$$

Le choix du noyaux n'influe pas trop dans le cas du noyaux symétriques.

2.3 ESTIMATION DE LA DENSITÉ AVEC UN NOYAU CONTINUE ASY- MÉTRIQUE GAMMA

2.3.1 Introduction

Si le support de f est $[0, +\infty[$, les noyaux symétriques donnent une estimation de plus en plus biaisée lorsque x approche de 0. D'où le problème de biais aux bords . plusieurs auteurs ont proposé d'utiliser des noyaux asymétriques.

Dans cette section, nous présentons l'estimateur de la densité de probabilité à noyau non symétrique . Avant cela nous donnons quelques exemples de noyaux asymétriques à support dans \mathbf{R}^+

2.3.2 Exemples de noyaux continus asymétriques

Noyaux	Supports	Densités
Gamma	\mathbb{R}^+	$K_{\frac{1}{xh}-1, \frac{1}{h}}(u) = \frac{h^{1-1/xh}}{\Gamma(\frac{1}{xh}-1)} u^{-1/xh} \exp\left(-\frac{1}{hu}\right)$
Inverse Gamma	\mathbb{R}^+	$K_{\frac{1}{xh}-1, \frac{1}{h}}(u) = \frac{h^{1-1/xh}}{\Gamma(\frac{1}{xh}-1)} u^{-1/xh} \exp\left(-\frac{1}{hu}\right)$
Béta	$[0, 1]$	$K_{\frac{x}{h}+1, \frac{(1-x)}{h}+1}(u) = \frac{u^{x/h}(1-u)^{(1-x)/h}}{\beta\left(\frac{x}{h}+1, \frac{(1-x)}{h}+1\right)}$
Log normal	\mathbb{R}^+	$K_{\log(x)+h^2, h}(u) = \frac{\exp\left\{\frac{-1}{2h^2}[\log(u)-\log(x \exp(h^2))]\right\}}{uh\sqrt{2\pi}}$

TABLE 2.2 : Exemples de noyaux non symétriques

REMARQUE 4.

Nous simulons sous Matlab chacun des noyaux précédents

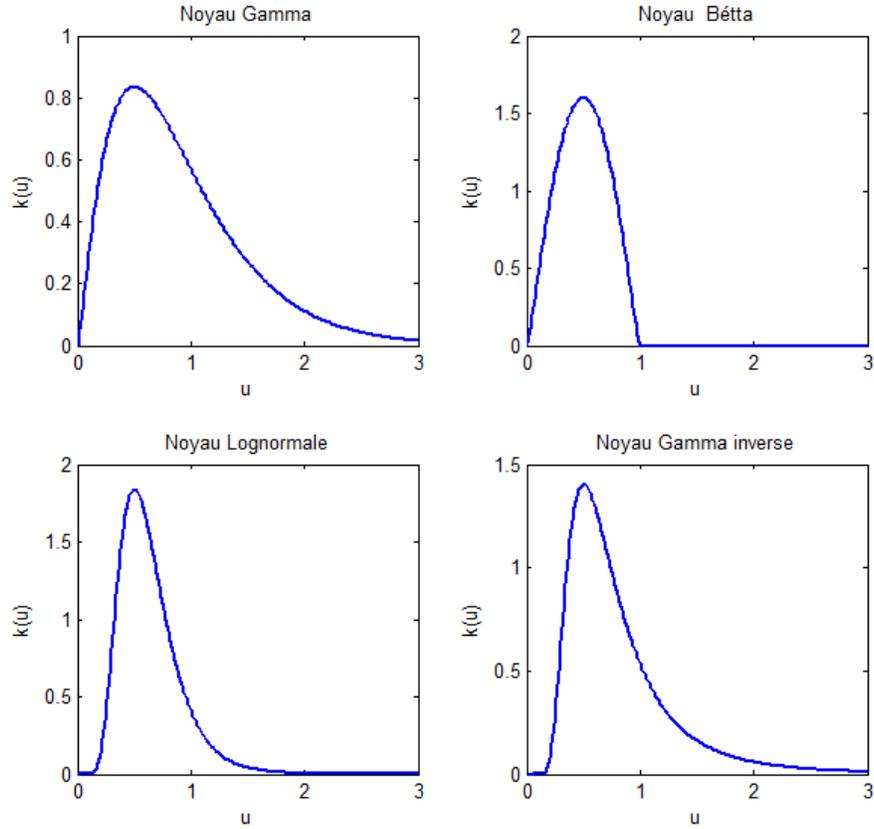


FIGURE 2.2 : Exemples de noyaux asymétriques

Dans la suite, nous présentons d'estimateur à noyau asymétrique de la densité, en particulier d'estimateur à Gamma proposé par Chen en 2000 [08]

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires i.i.d de densité de probabilité f inconnue à support dans R_+ Un estimateur non paramétrique d'une densité de probabilité f à noyau Gamma est défini par

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

avec

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

Où $K_{\frac{x}{h}+1,h}(t)$ c'est la fonction de densité d'une loi gamma de paramètre de forme $a = \frac{x}{h} + 1$ et paramètre d'échelle $b=h$

Nous rappelons les différentes propriétés d'une variable aléatoire X qui suit la loi Gamma de paramètre a et b .

Si $X \sim \Gamma(a, b)$ avec $a, b > 0$ on a

$$-f_X(x) = \frac{x^{a-1} e^{-\frac{x}{b}}}{b^a \Gamma(a)} \quad \text{pour } x > 0,$$

$$-\Gamma(\frac{1}{2}) = \sqrt{\pi}, \Gamma(1), \Gamma(n) = (n - 1)! \text{ pour tout entier } n > 0$$

$$-\Gamma(x) = (x - 1)\Gamma(x - 1) \text{ pour } x > 1.$$

$$- \text{Si } X \sim \Gamma(a, b), \text{ alors } E[X] = a/b, \text{ Var}[X] = a/b^2$$

$$- \text{Si } a = 1, \text{ alors } X \sim \text{Exp}(b).$$

La figure suivante représente les graphes de quelques densités de loi Gamma pour différents paramètres de forme a et paramètre d'échelle b : On remarque que l'allure de la

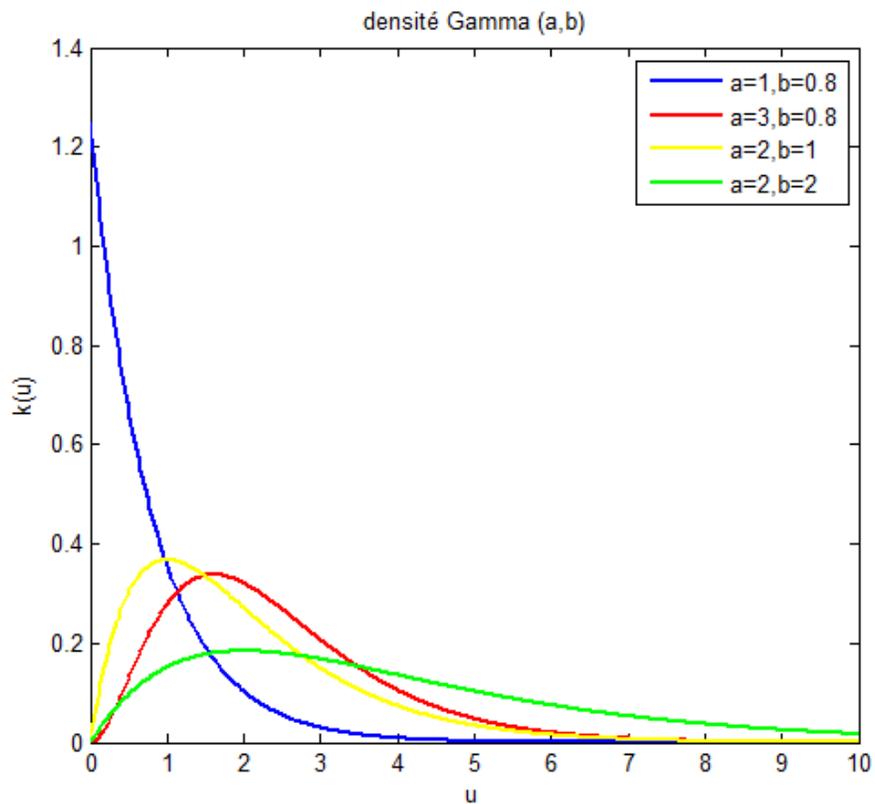


FIGURE 2.3 : Quelques densités de loi Gamma

densité Gamma change en fonction de a , et b .

Nous donnons maintenant des différents graphes pour des densités de loi $\Gamma(\frac{x}{h} + 1, h)$ pour différentes valeurs de x .

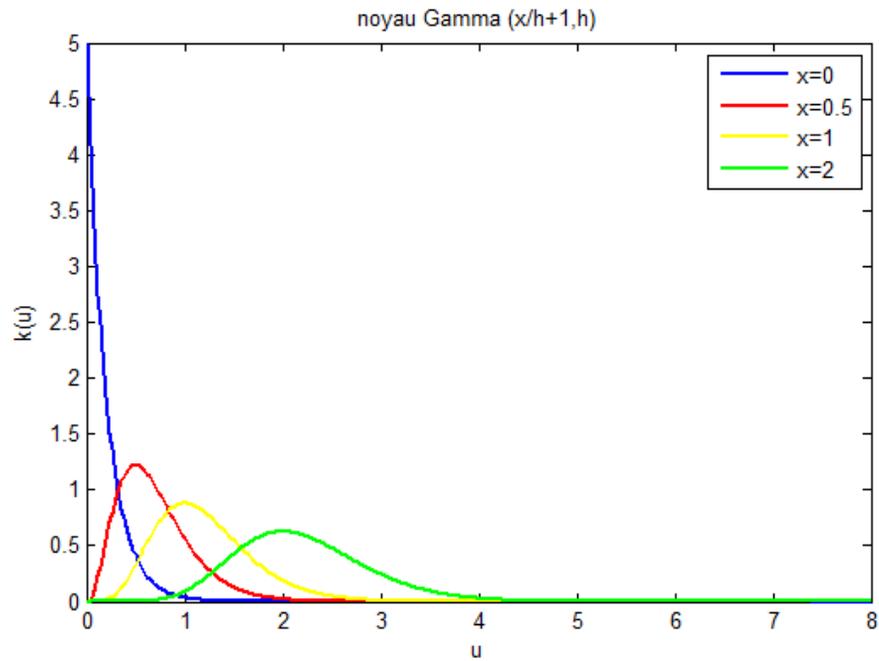


FIGURE 2.4 : Noyaux gamma, pour différentes valeurs de x , $x=0,0.5,1$ et 2 avec $h=0.2$

On remarquons que l'allure de la densité Gamma change en fonction de x , et h .

2.3.3 Propriétés d'un estimateur à noyau Gamma

D'après les propriétés d'une densité de loi Gamma, nous donnons les différentes propriétés d'un estimateur d'une densité de probabilité à noyau Gamma en se basant sur les études de Chen en 2000 [08].

Propriété 8.

$$1. \text{Biais} \left(\hat{f}_1(x) \right) = h \left(f'(x) + \frac{1}{2} x f''(x) \right) + o(h)$$

$$2. \text{Var} \left(\hat{f}_1(x) \right) \sim \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} x^{-1/2} f(x) + o(n^{-1} h^{-1/2}) & \text{si } \frac{x}{h} \rightarrow \infty \\ \frac{\Gamma(2k+1)}{2^{2k+1} \Gamma^2(k+1)} n^{-1} h^{-1} f(x) + o(n^{-1} h^{-1}) & \text{si } \frac{x}{h} \rightarrow K \end{cases}$$

Démonstration

$$1. \text{Biais} \left(\hat{f}_1(x) \right) = E \left(\hat{f}_1(x) \right) - f(x)$$

$$E \left(\hat{f}_1(x) \right) = E \left(K_{\frac{x}{h}+1, h}(X) \right) = \int_0^{\infty} K_{\frac{x}{h}+1, h}(y) f(y) dy = E \left(f(\xi_x) \right)$$

avec $\xi_x \sim \Gamma\left(\frac{x}{h} + 1, h\right)$. Alors $\mu_x = E(\xi_x) = x + h, \text{Var}(\xi_x) = xh + h^2$

Supposons que f soit de classe C^∞ sur $I \in \mathbb{R}$. par la formule de Taylor de $f(\xi_x)$ au voisinage de point μ (Voir Annexe pour la formule de Taylor) on obtient :

$$\begin{aligned} E \left(\hat{f}_1(x) \right) &= E \left(f(\mu_x) \right) + \frac{1}{2} f''(\mu_x) E(\xi_x - \mu_x)^2 + o(h) \\ &= f(\mu_x) + \frac{1}{2} f''(\mu_x) \text{Var}(\xi_x) + o(h) \\ &= f(x + h) + \frac{1}{2} f''(x + h) (xh + h^2) + o(h) \\ &= f(x) + hf'(x) + \frac{1}{2} f''(x) + o(h) \end{aligned}$$

Alors

$$\begin{aligned}
\text{Biais} \left(\hat{f}_1(x) \right) &= E \left(\hat{f}_1(x) \right) - f(x) = h \left((f'(x) + \frac{1}{2}x f''(x)) \right) + o(h) \\
2 \cdot \text{Var} \left(\hat{f}_1(x) \right) &= \frac{1}{n} \text{Var} \left(K_{\frac{x}{h}+1, h}(X_i) \right) \\
&= \frac{1}{n} \left(E(K_{\frac{x}{h}+1, h}^2(X)) \right) - \frac{1}{n} \left(E(K_{\frac{x}{h}+1, h}(X)) \right)^2 \\
&\leq \frac{1}{n} E \left(K_{\frac{x}{h}+1, h}^2(X) \right) \\
&\leq \int_0^\infty K_{\frac{x}{h}+1, h}^2(t) f(t) dt \\
&\leq \int_0^\infty \left(\frac{t^{x/h} e^{-\frac{t}{h}}}{h^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)} \right)^2 f(t) dt \\
&\leq \frac{h^{(\frac{2x}{h})+1} \Gamma((\frac{2x}{h}) + 1)}{2^{(\frac{2x}{h})+1} h^{2(\frac{x}{h}+1)} \Gamma^2(\frac{x}{h} + 1)} \int_0^\infty \frac{2^{(\frac{2x}{h})+1} t^{(\frac{2x}{h})} e^{-\frac{2t}{h}}}{h^{(\frac{2x}{h})+1} \Gamma(\frac{2x}{h} + 1)} f(t) dt \\
&\leq B_h(x) E \{ f(\eta_x) \}
\end{aligned}$$

avec η_x suit la loi $\Gamma \left((\frac{2x}{h}) + 1, h \right)$ et $B_h(x) = \frac{h^{-1} \Gamma(\frac{2x}{h} + 1)}{2^{(\frac{2x}{h})+1} \Gamma^2(\frac{x}{h} + 1)}$ on pose

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

alors

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

$$\text{Var} \left\{ \hat{f}_1(x) \right\} = n^{-1} B_h(x) E \{ f(\eta_x) \} = n^{-1} B_h(x) f(x)$$

Propriété 9.

$$1.MSE(x) = \left[h \left(f'(x) + \frac{1}{2} x f''(x) \right) \right]^2 + \frac{1}{2\sqrt{\pi}} h^{-1/2} x^{-1/2} + o(h^2).$$

$$2.MISE(n, h) = h^2 \int_0^{\infty} \left\{ x f'(x) + \frac{1}{2} x^2 f''(x) \right\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{\infty} x^{-1/2} f(x) dx + o\left(\frac{1}{n\sqrt{h}} + h^2\right)$$

Démonstration

A partir des expressions de la variance et de Biais de $\hat{f}(x)$ nous obtenons :

$$\begin{aligned} 1.MSE\left(\hat{f}_1(x)\right) &= \left[\text{Biais}\left(\hat{f}_1(x)\right) \right]^2 + \text{Var}\left(\hat{f}_1(x)\right) \\ &= \left[h \left(f'(x) + \frac{1}{2} x f''(x) \right) \right]^2 + \frac{1}{2\sqrt{\pi}} h^{-1/2} x^{-1/2} + o(h^2) \end{aligned}$$

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

2.3.4 Un autre estimateur à noyau asymétrique pour la densité

L'estimateur \hat{f}_1 est asymptotiquement sans biais, Ce biais vient de $f(x+h)$ et $\frac{1}{2}f''(x+h)(xh+h^2)$. Comme solution, Chen en 2000 a défini l'estiméteur suivante :

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

Avec

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

2.3.5 propriétés de l'estimateur $\hat{f}_2(x)$

Propriété 10.

$$1. E(\hat{f}_2(x)) = \begin{cases} f(x) + \frac{1}{2}hx f''(x) + o(h) & \text{si } x \geq 2h \\ f(x) + \xi_h h f'(x) + o(h) & \text{si } x \in [0, 2h[\end{cases}$$

$$2. \text{Biais}(\hat{f}_2(x)) = \begin{cases} \frac{1}{2}hx f''(x) + o(h) & \text{si } x \geq 2h \\ \xi_h(x) h f'(x) + o(h) & \text{si } x \in [0, 2h[\end{cases}$$

où $\xi_h(x) = (1-x)(\rho_h(x) - \frac{x}{h}) / (1 + h\rho_h(x) - x)$

$$3. \text{MSE}(x) = \frac{1}{4} \{x f''(x)\}^2 h^2 + \frac{1}{2\sqrt{\pi}} n^{-1} (hx)^{-1/2} f(x) + o(n^{-1}h^{-1/2})$$

$$4. \text{MISE}(n, h) = \frac{h^2}{4} \int_0^{+\infty} \{x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{+\infty} x^{-1/2} f(x) dx + o(n^{-1}h^{-1/2})$$

2.3.6 Choix du paramétré de lissage h

Soit $h_i^* = \text{ArgMin}(MISE(n, h_i))$ avec $i = 1, 2$.

on a :

$$\begin{cases} MISE(n, h_1) = h^2 \int_0^{\infty} \{x f'(x) + \frac{1}{2}x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{\infty} x^{-1/2} f(x) dx \\ \quad + o(n^{-1}h^{-1/2}) \\ MISE(n, h_2) = \frac{h^2}{4} \int_0^{+\infty} \{x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{+\infty} x^{-1/2} f(x) dx \\ \quad + o(n^{-1}h^{-1/2}) \end{cases}$$

les approximations asymptotiques pour les MISE sont donné par :

$$\begin{cases} AMISE(n, h_1) = h^2 \int_0^{\infty} \{x f'(x) + \frac{1}{2}x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{\infty} x^{-1/2} f(x) dx \\ AMISE(n, h_2) = \frac{h^2}{4} \int_0^{+\infty} \{x f''(x)\}^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-1/2} \int_0^{+\infty} x^{-1/2} f(x) dx \end{cases}$$

On dérive par rapport à h_i et on égale à 0 ,on obtient :

$$\left\{ \begin{array}{l} \frac{\partial}{\partial h_1} AMISE(n, h_1) = 2h \int_0^\infty \left\{ x f'(x) dx + \frac{1}{2} x f''(x) \right\}^2 - \frac{n^{-1}}{4\sqrt{\pi}} h^{-3/2} \int_0^\infty x^{-1/2} f(x) dx \\ \frac{\partial}{\partial h_2} AMISE(n, h_2) = \frac{h}{2} \int_0^\infty \{x f''(x)\}^2 - \frac{n^{-1}}{4\sqrt{\pi}} h^{-3/2} \int_0^\infty x^{-1/2} f(x) dx \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{\partial}{\partial h_1} AMISE(n, h_1) = 0 \\ \frac{\partial}{\partial h_2} AMISE(n, h_2) = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} h_1^* = \left(\frac{\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx}{4 \int_0^\infty \left\{ x f'(x) + \frac{1}{2} x f''(x) \right\}^2 dx} \right)^{2/5} n^{-2/5} \\ h_2^* = \left(\frac{\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx}{\int_0^\infty (x f''(x))^2 dx} \right)^{2/5} n^{-2/5} \end{array} \right.$$

Si les 2 intégrales $\int_0^\infty \{f'(x)\}^2 dx$ et $\int_0^\infty \{f''(x)\}^2 dx$ existent.

On remarque que

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

$$\Rightarrow MISE_{h_1^*}(\hat{f}_1(x)) \leq MISE_{h_2^*}(\hat{f}_2(x))$$

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

Cas particulier

Supposons que $f \sim \exp(\frac{1}{\theta})$. c'est-à-dire $f(x) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}$ alors :

$$1. \int_0^\infty x^{-1/2} f(x) dx = \theta^{-1/2} \sqrt{\pi}$$

$$2. f''(x) = \frac{1}{\theta^3} e^{-x/\theta}$$

$$3. \int_0^\infty (x f''(x))^2 dx = \frac{1}{4\theta^3} x^{-1/2}$$

Nous remplaçons dans la formule du h_2^* on trouve

$$h_{exp}^* = \theta 2^{2/5} n^{-2/5}$$

puis en remplace θ par son estimateur de maximum de vraisemblance, à la fin on obtient :

$$h_{exp}^* = \bar{x} (2/n)^{2/5}$$

La Figure 3.5 dans le chapitre 3 représente le graphe d'une densité théorique de loi exp avec son estimateur en utilisant cette formule de paramètre de lissage h_{exp}^*

On a vérifié dans le cas d'estimation d'une densité à noyau symétrique que $\int_{-\infty}^{+\infty} \hat{f}(x)dx = 1$, Ce qui n'est pas le cas pour un estimateur à noyau asymétrique Gamma .Pour monter ça nous calculons 4 estimateurs pour une densité théorique $N(0, 1)$ à noyau Gaussien et 4 estimateurs à noyau Gamma pour une densité théorique $\exp(2)$. puis nous calculons l'intégrale de chaque estimateur (respectivement $\int_{-10}^{10} \hat{f}(x)dx$ et $\int_0^{10} \hat{f}_1(x)dx$), en utilisant la formule de trapaze généralisé avec un programme sous langage Matlab. En utilisant 4 échantillons de variables aléatoires de taille $n = 50$, $n = 100$, $n = 500$, $n = 5000$ et $h = 0.2$, nous obtenons les résultats dans le tableau suivante :

		n=50	n=100	n=500	5000
Noyau Gaussien	$\int_{-10}^{10} \hat{f}(x)dx$	1.0000	1.0000	1.0000	1.0000
Noyau Gamma	$\int_0^{10} \hat{f}(x)dx$	0.9101	0.9122	0.9543	0.9762

TABLE 2.3 : Quelques valeurs d'intégral des estimateurs à noyau Gaussien et Gamma pour une densité théorique $N(0,1)$, et $\exp(2)$

On remarquons que tous les intégrales des estimateurs à noyau Gaussien sont égaux à 1, mais les intégrales des estimateurs à noyau Gamma sont proches de 1. dans ce dernier cas l'intégrale de l'estimateur s'approche de 1 lorsque la taille d'échantillon n augmente

CHAPITRE 3

3.1 INTRODUCTION

Dans ce chapitre nous effectuons plusieurs simulations sous Matlab pour illustrer les résultats théoriques établis dans le deuxième chapitre précédents concernant l'estimation de la densité à noyau avec des données indépendantes.

- Dans la 1^{ère} étape nous calculons plusieurs estimateurs à noyau symétrique Gaussien pour une densité théorique de loi $N(0, 1)$ et différentes valeurs de h indépendantes de la taille d'échantillon n , puis nous calculons l'estimateur avec une valeur de h qui dépend de n ($h = h(n) = 1.06 * (n)^{-1/5}$), après nous comparons visuellement les différents graphes .
par la suite nous calculons l'estimateur à noyau Gaussien pour différentes densités théoriques à support dans \mathbb{R}^+
- Dans la 2^{ème} étape nous calculons des estimateurs à noyau asymétrique Gamma pour une densité théorique Gamma pour différentes valeurs de h , puis nous calculons cet estimateur avec la formule de h optimal, après nous comparons entre les graphes.
- En dernier le but de savoir l'influence du nombre de données générées n ainsi que le paramètre de lissage h , nous varions chaque fois la taille n et on a calculé le MSE, Biais, MISE, de l'estimateur à noyau Gaussien puis à noyau Gamma. Après en comparant les résultats obtenus.

3.2 ESTIMATION À NOYAU SYMÉTRIQUE ,ASYMÉTRIQUE AVEC DES DONNÉES I.I.D

3.2.1 L'estimation à noyau Gaussien :

- nous générons 100 observations de loi normale centrée réduite $X_i \sim N(0, 1)$, $i=1\dots n$

En se basant sur les données X_i , nous calculons quatre estimateurs à noyau Gaussien, avec quatre valeur différentes de paramètre de lissage h . pour $x \in [-3, 3]$. après nous prenons pour un cinquième estimateur $h = h_{N(0,1)}^* = 1.06n^{-1/5}$ fonction de la taille n .

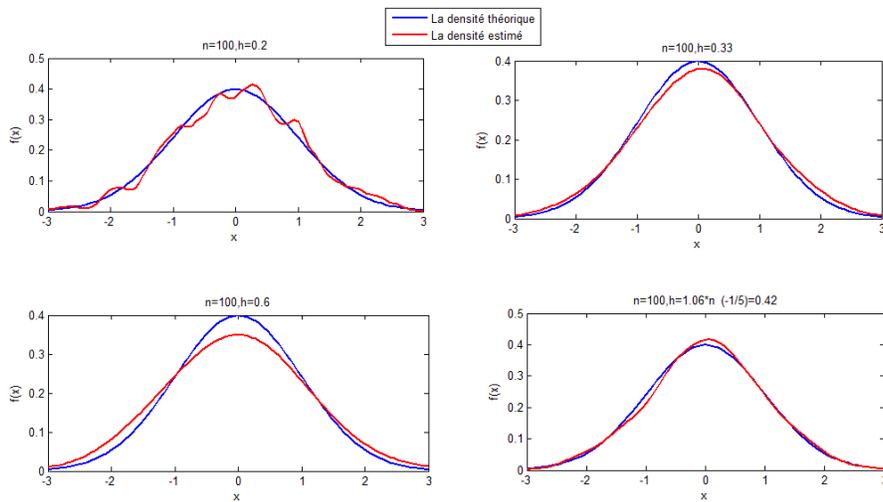


FIGURE 3.1 : La densité théorique $N(0,1)$ et la densité estimé à noyau Gaussien et $n = 100$, $h = 2$, $h = 0.33$, $h = 0.6$, $h = 0.8$

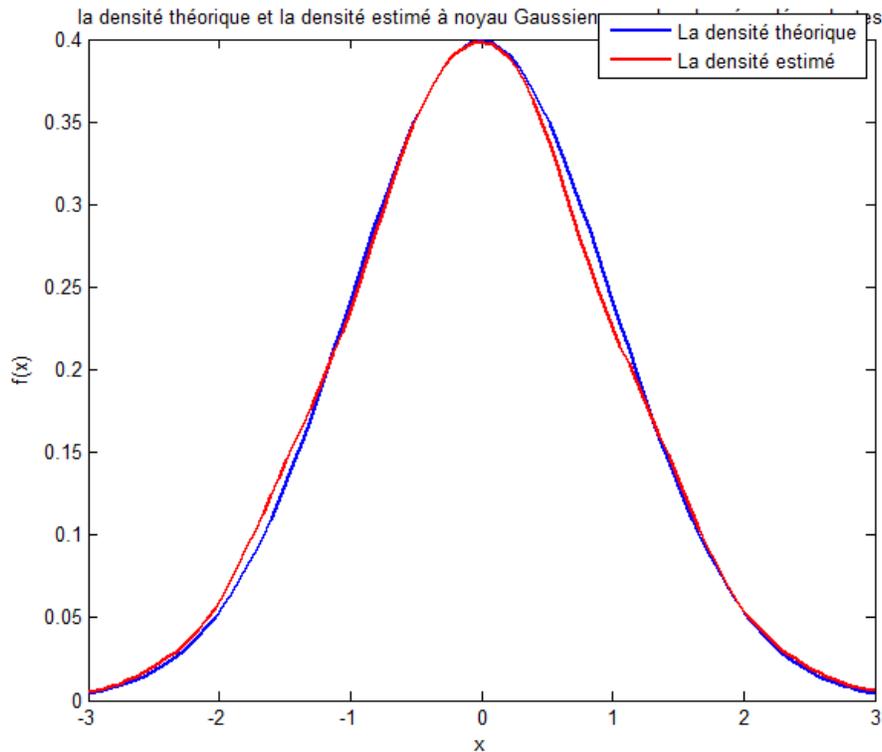


FIGURE 3.2 : suit la loi normale densité théorique $N(0,1)$ et la densité estimé à noyau Gaussien et $n = 100, h = h_{optimal}$

Commentaire

On peut voir l'influence de la largeur de la fenêtre sur l'estimateur de la densité , dans le cas où h est indépendante de n , nous avons des courbes qui sont sous-lissage , par contre dans le cas où h dépend de n la courbe est sur -lissage.

- Nous générons 100 observations $X_i, i=1...n$ de loi de $\exp(2)$, Weibul $W(2, 2)$, $\Gamma(2, 1)$ et Log Normale $LN(0, 1)$. puis nous calculons l'estimateur à noyau Gaussien $\hat{f}_n(x)$ sur l'intervalle $[0, 10]$. nous traçons les graphes des densités théoriques avec les densité estimées.

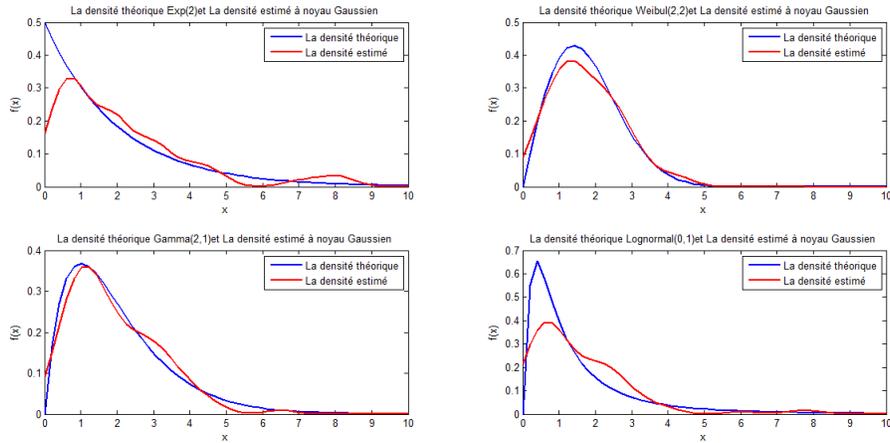


FIGURE 3.3 : Les densités théoriques $Exp(1/2), W(2, 2), \Gamma(2, 1), LN(0, 1)$, et la densité estimé à noyau Gaussien avec $n = 100, h = 0.42$

Commentaire

Nous remarquons dans les quarts graphes que l'écart entre les densités estimés et théoriques ($Biais(\hat{f}_n(x))$) est très grand lorsque x est au voisinage de zéro d'ou l'effet de bord.

3.2.2 L'estimation à noyau Gamma :

- Nous générons 100 observations de loi exponentielle $X_i \sim Exp(1/2)$ puis nous calculons l'estimateur à noyau Gamma $\hat{f}_1(x)$ pour $x \in [0, 10]$. avec $n = 50, n = 100, n = 500, n = 5000$, et un paramètre de lissage $h_{exp}^* = \bar{x} (2/n)^{2/5}$

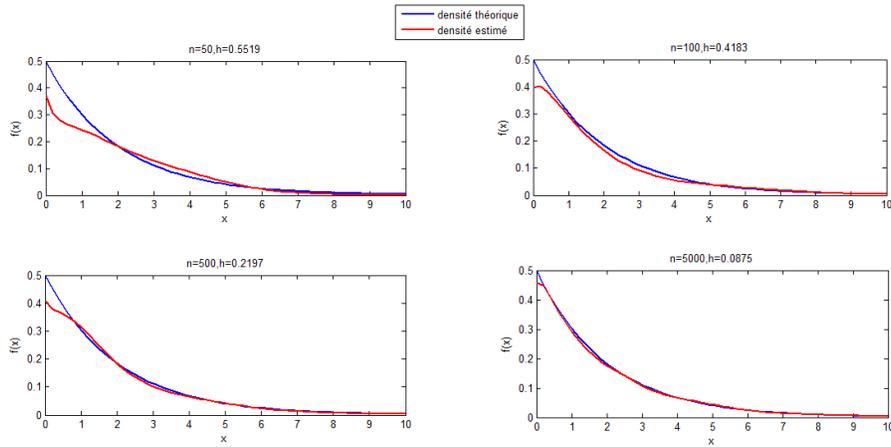


FIGURE 3.4 : la densité théorique avec les densités estimés à noyau Gamma pour différents valeurs de n

Nous traçons maintenant les estimateurs à noyau Gamma notés \hat{f}_1 et \hat{f}_2

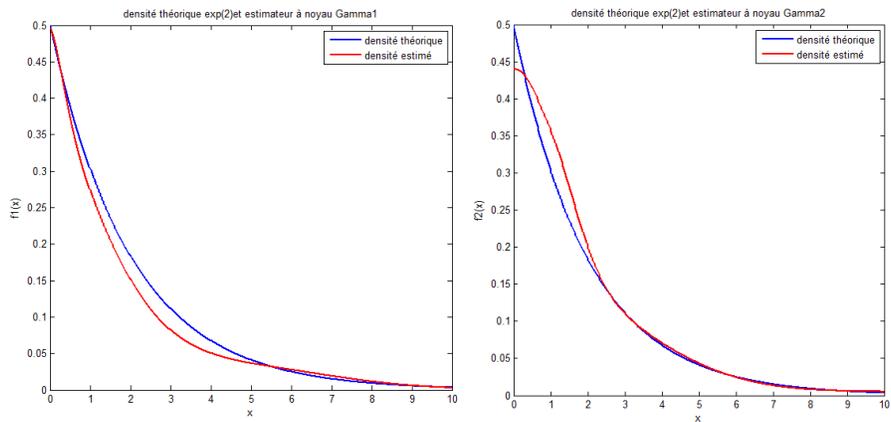


FIGURE 3.5 : La densité théorique $\exp(1/2)$ et les deux estimateurs à noyau Gamma 1 et Gamma 2

Commentaire

- L'estimateur de $f(x)$ à noyau Gamma 1 est très aplatie que l'estimateur à noyau Gamma 2.

3.3 INFLUENCE DE CHOIX DU PARAMÈTRE DE LISSAGE H :

3.3.1 Cas symétrique (Gaussien) :

- Nous calculons le *MISE*, *MSE*, *Biais* d'estimateur à noyau Gaussien pour différentes valeurs de n. quelque soit la valeur de $n \in \{ 50, 100, 500, 5000 \}$ on aura n valeurs de MSE et n valeurs de Biais, mais une seul valeur de MISE, on prend donc les moyennes des MSE , Biais les résultats sont dans le tableau suivante :

	n	h_{opt}	n	h_{opt}	n	h_{opt}	n	h_{opt}
	50	0.4847	100	0.4220	500	0.3059	5000	0.1930
$MSE(\hat{f}(x))$	0.0011		0.0008		0.0003		0.0001	
$MISE(\hat{f}(x))$	0.6885		0.4672		0.2103		0.0313	
$Biais(\hat{f}(x))$	0.2044		0.0156		-0.0141		-0.0240	

TABLE 3.1 : Les performances des estimateurs à noyau Gaussien pour des densités théoriques $N(0, 1)$ et $h = 1.06 * n^{-1/5}$

Commentaire

Les résultats obtenus dans tableau montrent que :

- Si la taille d'échantillon n augmente ,le paramètre de lissage optimal diminue. ($n \nearrow \implies h \searrow$).
- L'erreur quadratique moyenne et l'erreur quadratique moyenne intégré de $\hat{f}(x)$ (MSE , MISE) tendent vers zéro si h tends vers zéro.

\implies Alors pour réduire le MSE MISE ,il faut choisir un petit paramètre de lissage

3.3.2 Cas asymétrique (Gamma) :

- avec le même principe nous calculons le *MISE*, *MSE*, *Biais* de chaque estimateur(pour les deux estimateurs $\hat{f}_1(x)$ et $\hat{f}_2(x)$). les résultats sont dans les tableaux suivantes :

	n	h_{opt}	n	h_{opt}	n	h_{opt}	n	h_{opt}
	50	0.8278	100	0.6274	500	0.3296	5000	0.1312
$MSE(\hat{f}_1(x))$	0.0027		0.0019		0.0005		0.0001	
$MISE(\hat{f}_1(x))$	1.5855		1.1090		0.2877		0.0774	
$Biais(\hat{f}_1(x))$	-1.6589		-2.0178		-2.9697		-1.9697	

TABLE 3.2 : Les performances des estimateurs à noyau Gamma $\hat{f}_1(x)$ pour des densités théoriques $exp(2)$ et $h = \bar{x} * (2/n)^{2/5}$

	n	h_{opt}	n	h_{opt}	n	h_{opt}	n	h_{opt}
	50	0.8278	100	0.6274	500	0.3296	5000	0.1312
$MSE(\hat{f}_2(x))$	0.0022		0.0015		0.0003		0.0001	
$MISE(\hat{f}_2(x))$	1.3242		0.8717		0.2022		0.0334	
$Biais(\hat{f}_2(x))$	0.7144		-2.0766		-1.9081		-1.8800	

TABLE 3.3 : Les performances des estimateurs à noyau Gamma $\hat{f}_2(x)$ pour des densités théoriques $exp(2)$ et $h = \bar{x} * (2/n)^{2/5}$

Commentaire

- Pour les deux estimateurs $\hat{f}_1(x)$ et $\hat{f}_2(x)$ Si $n \nearrow \implies h \searrow$.

- $\forall n \in \{ 50, 100, 500, 5000 \} \implies \begin{cases} MSE(\hat{f}_2(x)) \leq MSE(\hat{f}_1(x)) \\ MISE(\hat{f}_2(x)) \leq MISE(\hat{f}_1(x)) \end{cases}$

CONCLUSION

Dans ce mémoire nous avons essayé d'étudier les propriétés d'un estimateur non paramétrique pour un fonction densité de probabilité avec la méthode de noyaux en utilisant des données indépendants . en premier nous avons présenté la méthode d'estimation à noyau symétrique proposé par Rosenblatt en 1956 puis amélioré par Parzen en 1962. ensuite nous avons passé aux estimation non paramétrique avec un noyau asymétrique Gamma avec des données indépendant, proposé par Chen en 1999-2000.

Les résultats obtenu sur l'estimation d'une densité de probabilité à noyau continu asymétrique avec des données complètes indépendantes peuvent être généraliser au cas d'une suite de variables aléatoires indépendantes Dans le cas de données censurées, Blum et Susarla (1980) ont introduit un estimateur à noyau d'une densité de probabilité f .

L'estimateur à noyau d'une densité de probabilité peut être utilisé pour des densités multidimensionnelles.

BIBLIOGRAPHIE

- [1] Bouezmarni,T.and Scaillet,O.2004."Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data".Econometric Theory, 21, 390 – 412.
- [2] Etienne Doucet , "Estimateur à noyau théorie des valeurs extrêmes".
- [3] Francial Giscard Baudin Libengué Dobélé Kopoka , " Méthode non paramétrique des noyaux associés mixtes et application ".
- [4] Imen Ben Khalifa, " Estimation Non-paramétrique par Noyaux Associés et Données de Panel en Marketing"
- [5] Remi Servien," Estimation de la fonction de répartition"
- [6] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics 27, 832–837.
- [7] SONG XI CHEN (2000). Probability density function estimation using gamma kernels. Annals of the Institute of Statistical Mathematics 54, 471–480.
- [8] Parzen, E. 1962. "On estimation of a probability density function and mode", The Annals of Mathematical Statistics, vol.33, no.3, p. pp. 1065-1076

ANNEXE

- **A1 . La convergence en moyenne quadratique**

On dit que X_n converge en moyenne quadratique vers la (v.a) X si les (v.a) X_1, X_2, \dots ont un moment d'ordre 2 et si

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

et on note

$$X_n \xrightarrow[n \rightarrow \infty]{m.q} X$$

- **A2 . Rappel de la méthode vraisemblance**

la méthode du maximum de vraisemblance est une méthode d'estimation paramétrique permettant de donner un estimateur d'une loi de probabilité inconnue dont on observe des réalisations i.i.d.

Soient X_1, \dots, X_n n réalisations d'une v . a X de densité $f(x, \theta)$ on cherche à estimer θ en fonction des observations .on définit la fonction de log-vraisemblance

$$LogL(\theta) = \sum_{i=1}^n Log (f(x_i, \theta))$$

Un estimateur de θ est une fonction $\hat{\theta}$ dépendant des observations X_i qui est un maximum de $LogL(\theta)$ il s'agit de trouver $\hat{\theta}$ tel que :

$$\frac{\partial LogL}{\partial \theta} (\hat{\theta}) = 0$$

ANNEXE

- **A3. Rappel de la méthode d'histogramme**

On choisit un point d'origine t_0 et une longueur de classe h avec $h > 0$. Les classes sont définies par :

$$B_l = [t_l, t_{l+1}[\quad l \in Z$$

Avec

$$t_{l+1} = t_l + h$$

Un estimateur de f est donnée par

$$\hat{f}(x) = \frac{\nu_l}{nh}$$

Avec ν_l le nombre d'observations dans la classe B_l

ANNEXE

les programmes de sous Matlab

Figure3.1

```
clear all
n1=50;
h1= 0.2
z=randn(1,n1)
x = linspace(-3,3,n1)
s=zeros(size(x));
for i=1 :n1
k=exp(-(1/(2*h12)) * (x - z(i)).2)/sqrt(2 * pi)/h1
s = s + (k/n1)
y = normpdf(x, 0, 1)
end
plot(x, y, x, s, 'r')
xlabel('x')
ylabel('f(x)')
title('n = 50eth = 0.4847')
legend('La densit thorique',' La densit estim')
subplot(2, 2, 2)
clearall
n2 = 100;
h2 = 0.33
z = randn(1, n2)
x = linspace(-3, 3, 200)
```

ANNEXE

```
s=zeros(size(x));
for i=1 :n2 k=exp(-(1/(2*h2^2)) * (x - z(i)).^2)/sqrt(2 * pi)/h2
s = s + (k/n2)
y = normpdf(x,0,1)
end
plot(x,y,x,s,'r')
xlabel('x')
ylabel('f(x)')
title('n = 100eth = 0.4220')
legend('La densit thorique','La densit estim')
subplot(2,2,3)
clearall
n3 = 500;
h3 = 0.6;
z = randn(1,n3)
x = linspace(-3,3,200)
s = zeros(size(x));
fori = 1 : n3k = exp(-(1/(2 * h3^2)) * (x - z(i)).^2)/sqrt(2 * pi)/h3
s = s + (k/n3)
y = normpdf(x,0,1)
end
```

ANNEXE

```
plot(x,y,x,s,'r')
xlabel('x')
ylabel('f(x)')
title('n=500,et h=0.3019')
legend('La densité théorique','La densité estimé')
subplot(2,2,4)
clear all
n4=5000;
h4=0.42
z=randn(1,n4)
x = linspace(-3,3,200)
s=zeros(size(x));
for i=1 :n4
k=exp(-(1/(2*h42)) * (x - z(i)).2)/sqrt(2 * pi)/h4
s = s + (k/n4)
y = normpdf(x, 0, 1)
end
plot(x, y, x, s, 'r')
xlabel('x')
ylabel('f(x)')
title('n = 5000, eth = 0.1930')
legend('La densit thorique','La densit estim')
```

ANNEXE

Figure 3.2

```
clear all
n1=100;
h1= 1.06*(n1)(- 1/5)
z = randn(1, n1)
x = linspace(-3, 3, n1)
s = zeros(size(x));
for i = 1 : n1
k = exp(-(1/(2 * h12)) * (x - z(i)).2)/sqrt(2 * pi)/h1
s = s + (k/n1)
y = normpdf(x, 0, 1)
end
plot(x, y, x, s, 'r')
xlabel('x')
ylabel('f(x)')
title('n = 100eth = hopt')
legend('La densit thorique', 'La densit estim')
```

ANNEXE

Figure 3.3

```
clear all
n=100;
h=1.06*n^( - 1/5);
z = exprnd(2, n);
z1 = wblrnd(2, 2, n);
z2 = gamrnd(2, 1, n);
z3 = lognrnd(0, 1, n);
x = linspace(0, 10, 50);
s = zeros(size(x));
s1 = zeros(size(x));
s2 = zeros(size(x));
s3 = zeros(size(x));
```

ANNEXE

```
for i=1 :n
k=exp(-(1/(2*h^2)) * (x - z(i)).^2)/sqrt(2 * pi);
k1 = exp(-(1/(2 * h^2)) * (x - z1(i)).^2)/sqrt(2 * pi);
k2 = exp(-(1/(2 * h^2)) * (x - z2(i)).^2)/sqrt(2 * pi);
k3 = exp(-(1/(2 * h^2)) * (x - z3(i)).^2)/sqrt(2 * pi);
s = s + (k/(n * h));
s1 = s1 + (k1/(n * h));
s2 = s2 + (k2/(n * h));
s3 = s3 + (k3/(n * h));
y = exppdf(x, 2);
y1 = wblpdf(x, 2, 2);
y2 = gampdf(x, 2, 1);
y3 = lognpdf(x, 0, 1);
```

ANNEXE

```
end ;
figure
subplot(2,2,1)
plot(x,y,x,s,'r','LineWidth',2)
xlabel('x')
ylabel('f(x)')
legend('La densité théorique ','La densité estimé')
title('La densité théorique Exp(2)et La densité estimé à noyau Gaussien ')
subplot(2,2,2)
plot(x,y1,x,s1,'r','LineWidth',2)
xlabel('x')
ylabel('f(x)')
legend('La densité théorique ','La densité estimé')
title('La densité théorique Weibul(2,2)et La densité estimé à noyau Gaussien')
subplot(2,2,3)
plot(x,y2,x,s2,'r','LineWidth',2)
xlabel('x')
ylabel('f(x)')
legend('La densité théorique ','La densité estimé')
```

ANNEXE

```
title('La densité théorique Gamma(2,1)et La densité estimé à noyau Gaussien ')
subplot(2,2,4)
plot(x,y3,x,s3,'r','LineWidth',2)
xlabel('x')
ylabel('f(x)')
legend('La densité théorique ','La densité estimé')
title('La densité théorique Lognormal(0,1)et La densité estimé à noyau Gaussien ')

legend('La densité théorique','La densité estimé')
```

ANNEXE

Figure 3.4

```
subplot(2,2,1)
clear all
n1=50
x = linspace(0,10,100)
b1=(mean(x))*(2/n1)^(2/5)
a = (x/b1) + 1
z = gamrnd(1,2,1,n1)
s = zeros(size(x))
fori = 1 : n1
k = (exp((-z(i))/b1) * (z(i)).^(x./b1))./(gamma((x./b1) + 1) * (b1.^(x./b1) + 1))/b1
s = s + (k/n1 * b1)
end;
y = gampdf(x,1,2)
plot(x,y,x,s,'r')
xlabel('x')
ylabel('f(x)')
title('n = 50, h = 0.8278')
legend('La densit thorique','La densit estim')
subplot(2,2,2)
```

ANNEXE

```
clear all
n2=100
x = linspace(0,10,100)
b2=(mean(x))*(2/n2)^(2/5)
a = (x/b2) + 1
z = gamrnd(1, 2, 1, n2)
s = zeros(size(x))
for i = 1 : n2
k = (exp((-z(i))/b2) * (z(i)).^(x./b2))./(gamma((x./b2) + 1) * (b2.^(x./b2) + 1))/b2
s = s + (k/n2 * b2)
end;
y = gampdf(x, 1, 2)
plot(x, y, x, s, 'r')
xlabel('x')
ylabel('f(x)')
title('n = 100, h = 0.6274')
legend('La densit thorique', 'La densit estim')
```

ANNEXE

```
clear all
n3=500
x = linspace(0,10,100)
b3=(mean(x))*(2/n3)^(2/5)
a = (x/b3) + 1
z = gamrnd(1, 2, 1, n3)
s = zeros(size(x))
for i = 1 : n3
k = (exp((-z(i))/b3) * (z(i)).^(x./b3))./(gamma((x./b3) + 1) * (b3.^(x./b3) + 1))/b3
s = s + (k/n3 * b3)
end;
y = gampdf(x, 1, 2)
plot(x, y, x, s, 'r')
xlabel('x')
ylabel('f(x)')
title('n = 500, h = 0.3296')
legend('La densit thorique', 'La densit estim')
subplot(2, 2, 4)
```

ANNEXE

```
clear all
n4=5000
x = linspace(0,10,100)
b4=(mean(x))*(2/n4)^(2/5)
a = (x/b4) + 1
z = gamrnd(1,2,1,n4)
s = zeros(size(x))
for i = 1 : n4
k = (exp((-z(i))/b4) * (z(i)).^(x./b4))./(gamma((x./b4) + 1) * (b4.^(x./b4) + 1))/b4
s = s + (k/n4 * b4)
end;
y = gampdf(x,1,2)
plot(x,y,x,s,'r')
xlabel('x')
ylabel('f(x)')
title('n = 5000, h = 0.0.1312')
legend('La densit thorique','La densit estim')
```

الملخص:

في هذه المذكرة قمنا بتقديم تقدير اللامعلمية لكثافة الاحتمال مع اساليب النوى و اعطائنا بعض الخصائص لمقدر النواة المتناظر و الغير متناظر و ذلك باستخدام متغيرات مستقلة ذات توزيعات متماثلة.

الكلمات المفتاحية: مربع الخطأ المتكامل ، معلمة تمهيد ، الفرق ، تقدير اللامعلمية للطرق النوى.

Résumé :

Dans ce mémoire nous avons proposer l'estimation non paramétrique pour une densité de probabilité avec le méthode du noyau et nous donnons quelques propriétés asymptotiques pour un estimateur à noyau symétrique et asymétrique en utilisant des données complètes indépendantes identiquement distribuées .

Mots-clés : Erreur quadratique moyenne intégrée, Paramètre de lissage ,La Variance, Estimation non-paramétrique par noyau.

Abstract :

In this work ,we purpose the nonparametric estimation of an unknown density Of probability with the kernel method We give some asymptotic properties for a symmetric and asymmetric kernel estimator , by using identically distributed independent complete data

Key words: Mean integrated squared error, Parameter Smoothing, Variance , Nonparametric kernel estimation.