

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

UNIVERSITÉ MOHAMED KHIDER, BISKRA

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

DÉPARTEMENT DE MATHÉMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

DJEDDA Imad Eddine

Titre :

L'effets de bord pour l'estimation à noyau de la densité

Membres du Comité d'Examen :

Dr. CHINE Amel	UMKB	Président
Dr. KHEIREDDINE Souraya	UMKB	Encadreur
Dr. BENBRAIKA Ghazlane	UMKB	Examineur

Juin 2021

DÉDICACE

Je dédie ce travail :

A deux personnes qui m'ont donné leur confiance et leur soutien tout au long de mes études, mon père et ma mère.

A mes très chers frères et mes très chères soeurs qui m'ont encouragé sur le long de mon parcours universitaire.

Je dédie ce travail à tout mes chères amis .

A tous les membres de ma famille, petits et grands, veuillez trouvez dans ce modeste travail l'expression de mon affection.

A tous mes collègues de Master 2 . Et enfin, à tous ceux qui me sont chers.

REMERCIEMENTS

Tout d'abord, nous tenons à remercier Dieu le tout puissant de nous avoir accordé patience, courage et volonté afin de réaliser mener à terme ce modeste travail. Je tiens à remercier tout particulièrement mon encadreur, Dr. Kheireddine Souraya pour ses conseils, sa grande disponibilité et sa générosité. La pertinence de ses questions et toujours me memotiver et me diriger. Je voudrais également remercier tous les membres de jury d'avoir accepté d'évaluer et d'examiner ce travail, merci pour toutes leurs remarques et critiques. Je remercie chaleureusement toute ma famille, qui m'a soutenu, encouragé et poussé durant toutes mes années d'étude. J'exprime mes remerciements à tous mes enseignants du département de Mathématiques qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour, ainsi que le personnel de l'administration. Ce travail de mémoire de master est la première expérience dans l'activité de recherche .Il n'aurait pas été aussi fructueux sans l'aide de plusieurs personnes . Je vais donc m'essayer à trouver les mots justes pour exprimer spécifiquement ma reconnaissance à tous ceux qui ont contribué de près en de loin à ce travail.

Résumé

Dans ce travail nous avons étudié une méthode non paramétrique d'estimation des densités. Ensuite nous avons vu la méthode d'estimation par l'histogramme et ses propriétés et on a étudié les propriétés de l'estimateur à noyau. Enfin, nous avons essayé de simuler un estimateur de densité simple par la méthode des histogrammes et la méthode d'estimation du noyau.

Mots-clés : Estimation à noyau , densité, noyaux, Paramètre de lissage, MISE.

الملخص

درسنا في هذا العمل طريقة غير بارامترية لتقدير الكثافات. ثم رأينا طريقة تقدير المدرج التكراري وخصائصها ودرسنا خصائص مقدر النواة. و أخيرًا ، حاولنا محاكاة مقدر كثافة بسيط بطريقة الرسم البياني وطريقة تقدير النواة .

Abstract

In this work we have studied a non-parametric method for estimating densities. Then we saw the histogram estimation method and its properties and we studied the properties of the kernel estimator. Finally, we tried to simulate a simple density estimator by the histogram method and the kernel estimation method.

Table des matières

Dédicace	i
Remerciements	ii
Résumé	iii
Table des matières	iv
Liste des figures	vii
Liste des tableaux	ix
Introduction	1
1 Estimation non paramétrique d'une densité.	4
1.1 Définitions et Critères :	5
1.2 Exemples de noyaux :	9
1.3 Estimation par la methode de l'histogramme	12
1.3.1 Propriétés de l'estimateur par histogramme :	13
1.4 Estimation par la méthode des noyaux :	16
1.4.1 Estimation par noyaux associés unidimensionnelles :	17
1.4.2 Estimation par noyaux associés multidimensionnels :	21
1.4.3 Estimation par noyaux associés mixtes :	22

1.4.4	Estimateur à noyau dans le cas multivarié :	23
1.5	Propriétés de l'estimateur à noyau :	24
1.5.1	Etude du biais :	24
1.5.2	Etude de la variance :	25
1.6	Le choix optimal du noyau :	26
1.7	Choix théorique optimal du paramètre de lissage :	28
1.8	Simulation des données :	30
2	Estimation à noyau de la densité et correction des effets de bord.	34
2.1	Introduction :	34
2.2	Cadres d'études :	35
2.3	Estimation à noyau classique de la densité et de ses dérivées :	36
2.4	Amélioration à l'aide de noyaux réducteurs de biais :	39
2.5	Application à une méthode de correction des effets de bord :	41
2.6	Preuves :	43
2.6.1	Preuve du Lemme 2.3.1 :	43
2.6.2	Preuve du Lemme 2.3.2 :	44
2.6.3	Preuve du Théorème 2.3.1 :	45
2.6.4	Preuve du Théorème 2.3.2 :	45
2.6.5	Preuve du Lemme 2.4.1 :	46
2.6.6	Preuve du Théorème 2.3.3 :	46
2.6.7	Preuve du Théorème 2.4.1 :	46
2.7	Conclusion :	48
3	Simulation	50
3.1	Un estimateur simple de la densité : l'histogramme :	50
3.2	Code R et illustration graphique du choix du nombre de classes :	51
3.3	Estimateurs à noyaux :	54

3.4 Illustration graphique et code R	55
3.5 Codes des simulations :	57
3.5.1 Estimateur à noyau décomposé :	57
3.5.2 Fonction densité :	58
3.5.3 Sortie de densité :	58
3.5.4 Application de la méthode du noyau :	59
3.5.5 Package sm :	59
3.5.6 Variation du paramètre de lissage :	60
3.5.7 Comparaison de 3 densités :	60
Conclusion	62
Bibliographie	64
Annexe A : Abréviations et Notations	71
Annexe B : Abréviations et Notations	73

Table des figures

1.1	Estimation de densité avec python, fonction "gaussian_kde" du package "scipy.stats.kde".	6
1.2	Les courbes des noyaux les plus communs.	10
1.3	Estimateur 'à noyau basé sur 7 observations (h=4)	11
1.4	Exemples de noyaux.	11
1.5	Effet de la variation de h sur l'estimateur à noyau.	11
1.6	Différents histogrammes associés à un même ensemble de données	
	avec différents paramètres de lissages.	15
1.7	Exemples noyaux associés discrets	19
1.8	Densité d'un noyau d'Epanechnikov dans le cas multivarié.	24
1.9	Le "trad-off" biais-variance en fonction de h.	26
1.10	Estimation de la densité avec différents noyaux.	27
1.11	Illustration d'un phénomène de sous-lissage lors de l'estimation	
	d'une densité .	29
1.12	Illustration d'un phénomène de sur-lissage lors de l'estimation d'une densité .	29
1.13	Illustration d'une estimation idéale	30
1.14	Lissages par des estimateurs à noyaux continus de la distribution d'un	
	échantillon	32
1.15	Lissages par des estimateurs à noyaux continus de la distribution d'un	
	échantillon	32

1.16 Comparaison des lissages par l'estimateur à noyau continu d'Epanchnikov	
en faisant	33
3.1 la densité par un histogramme (par bibliothèque ggplot2)	52
3.2 la densité par un histogramme (par bibliothèque ggplot2) tel' que nbr de	
classes=3/10/100	53
3.3 la densité par un histogramme (par bibliothèque ggplot2) tel' que nbr de	
classes=30/100	54
3.4 représentation de l'estimateur à noyau (par la la version de ggplot) pour	
h=0.1/0.5/0.8/1.2	56
3.5 L'estimation de densité par la méthode de Sheather et Jones (SJ) et la	
méthode de la validation croisée	57

Liste des tableaux

1.1 Quelques noyaux associés discrets	19
1.2 Quelques noyaux et leurs efficacités	27

Introduction

La théorie de l'estimation est une des préoccupations majeures des statisticiens. On trouve dans la littérature deux types d'approches d'estimations de la densité de probabilité : l'approche paramétrique et l'approche non-paramétrique. L'approche paramétrique suppose que les données sont issues d'une loi de probabilité de forme connue dont seuls les paramètres sont inconnus. Son objectif est de connaître la vraie valeur du paramètre ou plus généralement une fonction de cette valeur. Le principal inconvénient de cette approche est qu'elle nécessite une connaissance préalable du phénomène aléatoire considéré.

Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique propose de laisser parler les données, sans spécifier au préalable de forme sur f . L'outil d'estimation non paramétrique nous est fourni par l'histogramme : une fois les données regroupées en classes de valeurs, les fréquences empiriques sont représentées par des aires rectangulaires dont les bases correspondent aux classes elles mêmes. L'histogramme convient bien pour des analyses relativement grossières. Néanmoins, ses discontinuités n'apparaissent pas très naturelles et, ce qui est plus grave, les points tombant près des bords d'une lieu ne sont pas différenciés, ceci explique la variabilité des interprétations statistiques que l'on peut faire d'un histogramme suivant le choix de l'origine et des classes. Pour des densités raisonnablement lisses, l'histogramme apparaît donc comme un estimateur sévèrement limité.

Il existe d'autres méthodes non paramétriques plus robustes que la méthode par l'histogramme : la méthode d'estimation par les séries orthogonales et la méthode du noyau.

Nous regarderons brièvement en quoi consiste la méthode d'estimation par l'histogramme et la méthode par les séries orthogonales et en détail l'estimateur par la méthode du noyau vu sa souplesse d'utilisation et ses propriétés de convergence. Le succès rencontré par l'estimateur à noyau auprès de la communauté des utilisateurs peut essentiellement s'expliquer en trois points :

- D'abord l'expression théorique de l'estimateur est extrêmement simple puisque il s'écrit sous la somme de n variables aléatoires indépendantes et identiquement distribuées.
- Ensuite l'estimateur converge vers la densité f en de nombreux sens, en particulier au sens L_1 . d'autres part, il est convergent dans tous les modes : en probabilité, en moyenne, presque sûrement et presque complètement.
- Enfin, l'estimateur à noyau est flexible, dans la mesure où il laisse à l'utilisateur une grande latitude non seulement dans le choix du noyau K , mais encore dans le choix du paramètre de lissage h . C'est Rosenblatt [52] en 1956, suivi de Parzen [50] en 1962, qui ont proposé une classe d'estimateurs à noyau d'une densité univariée. Les estimateurs à noyau sont des fonctions de deux paramètres K , appelé noyau, et h dit paramètre de lissage (largeur de fenêtre). Rosenblatt reprenait l'idée de Fix et Hodges en 1951, qui consistait à estimer la densité en un point, en comptant le nombre d'observations situées dans l'intervalle de longueur $2h$ et centré en ce point. Avant de construire les estimateurs à noyaux de la densité, d'en mesurer les performances théoriques et, le cas échéant, d'identifier le meilleur, il est nécessaire de spécifier un critère d'erreur qui puisse être éventuellement optimisé. Citons par exemple l'erreur quadratique moyenne intégrée MISE et l'intégrale de l'erreur quadratique ISE. Les propriétés de convergence de l'estimateur à noyau ont été établies par Parzen [50], Silverman [53] et Nadaraya [47]. Les théorèmes relatifs à l'erreur quadratique moyenne et l'erreur quadratique intégrée moyenne ont été obtenus sous forme élémentaire par Parzen [50]. Enfin, c'est Epanechnikov en 1969 [27] qui s'est rendu compte de l'existence d'un noyau asymptotiquement optimal K_e . Mais l'erreur quadratique moyenne asymptotiquement intégrée varie peu en fonction du choix de K . Si le choix

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

Le mémoire est structuré comme suit :

Une introduction générale pour situer notre étude , Un premier chapitre où nous présentons les différentes méthodes non paramétriques d'estimation de la densité tel que la méthode des histogrammes et la méthode du noyau qui sera étudiée en détail. Le deuxième chapitre traitera les principales méthodes de sélection du paramètre de lissage par la méthode du noyau. Le troisième chapitre regroupera les résultats de simulation des différentes méthodes de sélection du paramètre de lissage. Nous terminerons par une conclusion générale.

Chapitre 1

Estimation non paramétrique d'une densité.

L'approche classique pour estimer une densité est de supposer un modèle paramétrique : par exemple, en dimension 1, on représente les données par un histogramme, et si la courbe est en cloche avec des queues légères, on conclut qu'il y a de fortes chances que le modèle suive une loi gaussienne. Il n'y a alors plus qu'à estimer la moyenne et la variance (μ, σ^2) , c'est-à-dire un paramètre de dimension 2. On peut aussi se trouver dans un cas où on a des connaissances a priori sur les données, nous amenant à poser encore une loi paramétrique (ex typique : nombre de voitures passant par un carrefour par jour, représenté en général par une loi de poisson). Il y a plusieurs problèmes possibles avec cette approche : en dimension supérieure à 2 il sera difficile de représenter les données et d'intuiter une loi connue, parfois on n'a pas de connaissances a priori sur le sujetetc.

De plus, si on se trompe de modèle, on arrivera à une interprétation erronée des données. Un modèle non paramétrique est moins rigide, et fait moins de suppositions a priori sur les données. Evidemment, comme pour le cas des tests, si on a des connaissances a priori fiables sur les données nous indiquant un modèle paramétrique, il faut utiliser le modèle paramétrique. Autrement dit, si le modèle paramétrique choisi est correct, ou plus pré-

cisément suffisamment proche de la réalité, alors le modèle paramétrique sera en général meilleur qu'un modèle non paramétrique.

1.1 Définitions et Critères :

Définition 1.1.1 (*Statistique non paramétrique*)

La statistique paramétrique est le cadre "classique" de la statistique. Le modèle statistique y est décrit par un nombre fini de paramètres. Typiquement $M = \{P_\theta, \theta \in R^p\}$ est le modèle statistique qui décrit la distribution des variables aléatoires observées.

* Observations réelles avec un seul mode : $\{N(\mu, \sigma^2), \mu \in R, \sigma^2 \in R^{+*}\}$, modèle Gaussien.

* Observations réelles avec plusieurs modes :

$M_K = \left\{ \sum_{i=1}^K p_i N(\mu_i, \sigma^2), (p_1, \dots, p_K) \in (0, 1)^K, \sum_i p_i = 1, (\mu_1, \dots, \mu_K) \in R^K, \sigma^2 \in R^{+*} \right\}$, modèle de mélange Gaussien.

* Observations de comptage : $M = \{P_\lambda, \lambda \in R^{+*}\}$ modèle loi Poisson.

* $M = \{P \text{ à support dans } S \text{ fini}\} \simeq [0, 1]^{|S|^{-1}}$.

Par opposition, en statistique *non paramétrique*, le modèle n'est pas décrit par un nombre fini de paramètres. Divers cas de figures peuvent se présenter, comme par exemple :

* On s'autorise toutes les distributions possibles, *i.e.* on ne fait aucune hypothèse sur la forme/nature/type de la distribution des variables aléatoires.

* On travaille sur des espaces fonctionnels, de dimension infinie.

* Le nombre de paramètres du modèle n'est pas fixe et varie (augmente) avec le nombre d'observations.

* Le support de la distribution est discret et varie (augmente) avec le nombre d'observations.

les densités continues sur $[0; 1]$, ou les densités monotones sur \mathbb{R} .

* Le nombre de paramètres du modèle n'est pas fixe et varie (augmente) avec le nombre d'observations.

* Le support de la distribution est discret et varie (augmente) avec le nombre d'observations.

Définition 1.1.2 (Estimation de densité)

On observe toujours X_1, \dots, X_n , n variables réelles de loi P . Mais on suppose en plus que P est absolument continue par rapport à la mesure de Lebesgue et on souhaite estimer sa densité f . En général, la dérivée de \widehat{F}_n n'est pas une bonne solution.

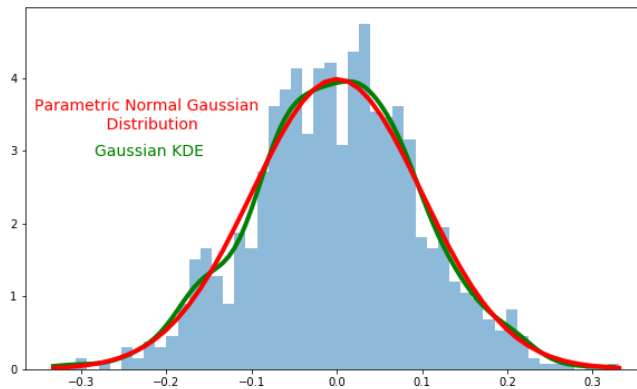


FIG. 1.1 – Estimation de densité avec python, fonction "gaussian_kde" du package "scipy.stats.kde".

Définition 1.1.3 (Régression non-paramétrique)

On observe une suite de couples $((X_i, Y_i))_{1 \leq i \leq n}$ obéissant au modèle $Y_i = f(X_i) + \epsilon_i$, $i = 1, \dots, n$

On cherche à estimer la fonction de régression f . On peut aussi considérer d'autres problèmes de statistique non-paramétrique qui ne sont pas directement de l'estimation.

Définition 1.1.4 (Estimateur à noyau)

*soit $K : R \rightarrow R$ intégrable telle que $\int K(u)du = 1$. Alors K est appelé *noyau*.

*Pour tout $h > 0$ petit (en fait $h = h_0 \xrightarrow{n \rightarrow +\infty} 0$) .on peut définir $\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K(\frac{X_i - x}{h})$.
 estimateur à noyau de f . On a $\int \widehat{f}_n(x)dx = 1$ et si $K > 0$ Alors \widehat{f}_n est une densité.

*Le paramètre $h > 0$ est appelé *fenêtre*. C'est un paramètre de lissage : plus h est grand, plus l'estimateur est régulier.

Remarque 1.1.1 *On considérera souvent des noyaux positifs et paires. mais ce n'est pas obligatoire.*

L'estimateur à noyau est probablement l'estimateur le plus utilisé et certainement le plus étudié mathématiquement, car il possède des propriétés qui le rendent fort intéressant. Un estimateur à noyau noté f_n de la fonction f est défini par :

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) \quad (1.13)$$

où $\{h_n \geq 1\}$ est une suite de réels positifs appelés **paramètres de lissage** ou **largeur** de la fenêtre, qui tend vers 0 quand n tend vers l'infini. Comme nous allons le voir par la suite, si le noyau K est une fonction de densité alors l'estimateur à noyau f_n est lui aussi une fonction de densité. De plus, ce dernier possède les propriétés de continuité et de différentiabilité. De sorte que si, par exemple, K est la densité normale alors f_n possède des dérivées de tout ordre.

Définissons maintenant plus généralement la notion d'estimateur à noyau :

Définition 1.1.5 • soit $K : \mathbb{R} \rightarrow \mathbb{R}$, on dit que K est un noyau si et si

$$\int K(u)du = 1$$

- K est dit *positif* si $K(u) \geq 0, \forall u$.
- K est dit *symétrique* si $K(u) = K(-u), \forall u$.

Propriété 1.1.1 *Un estimateur à noyau est une densité.*

Démonstration :
$$\int_{-\infty}^{+\infty} f_n(x)dx = \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K\left(\frac{x-X_i}{h_n}\right) dx = \frac{1}{nh_n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u)du$$
 (changement de variable $u = \frac{x-X_i}{h_n}$)

$$= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{+\infty} K(u)du = \frac{1}{n} n = 1.$$

Définition 1.1.6 (Noyau d'ordre r)

soit $r \geq 1$ un entier .On dit que qu'un noyaux K est d'ordre r si :

$$\forall j = 1, \dots, r. \int u^j K(u)du = 0 \text{ et } \int u^{r+1} K(u)du \neq 0.$$

Remarque 1.1.2 * si K est un noyau pair alors pair alors K est d'ordre au moins 1.

* pour $j = 0$, on a $\int u^j K(u)du = \int K(u)du = 1.$

*on sait contruire des noyaux d'ordre r pour tout entier $r \geq 0.$

Définition 1.1.7 (Biais des estimateurs à noyau)

Le risque quadratique ponctuel :

$$R_x(f_n; f) = E f(\hat{f}_n(x) - f(x))^2 = \text{Biais}^2 f(\hat{f}_n(x)) + \text{Var}_f(\hat{f}_n(x)).$$

Etude du biais :

$$E_f(\hat{f}_n(x)) = E_f \left(\frac{1}{h} K\left(\frac{X-x}{h}\right) \right) = \int \left(\frac{1}{h} K\left(\frac{u-x}{h}\right) \right) f(u)du = \int K(d) f(x+hd)du.$$

Si f est une fonction dérivable au voisinage de x , alors on peut écrire

$$f(x+hd) = f(x) + hdf'(x+\varepsilon hd), \varepsilon \in]0; 1[.$$

D'où

$$E_f(\hat{f}_n(x)) = \int K(d)[f(x) + hdf'(x+hd)]dv = f(x) + h \int vK(v) f'(x+hv)dv .$$

Si de plus $\|f\| < +\infty$ et $\int dK(d)j\partial d < \infty$, alors on obtient que $E_f(f_n(x)) = f(x)+O(h)$, lorsque $h \rightarrow 0$.

Controle du biais :

Dans ce cas, on a montré que le biais $\left| E_f(\widehat{f}_n(x)) - f(x) \right|$ converge vers 0 lorsque $h \rightarrow 0$. Plus généralement, si on suppose que f appartient à une classe de fonctions su samment régulières, on va pouvoir montrer une décroissance du terme de biais vers 0.

Variance des estimateurs a noyaux :

Proposition 1.1.1 *Si f est une densité bornée sur \mathbb{R} . (i.e. $\|f\| < \infty$) et si K est un noyau tel que $\int K^2(u)du < +\infty$,*

alors pour tout $x \in \mathbb{R}$, pour tout $h > 0$ et tout $n \geq 1$, on a $Var_f(\widehat{f}_n(x)) = \frac{\|f\|_\infty (\int K^2(u)du)}{nh}$. Si de plus,

$f(x) > 0$ et f continue au voisinage de x et $\int K(u)du < +\infty$, alors

$$Var_f(\widehat{f}_n(x)) = \frac{f(x)}{nh} (\int K^2(u)du)(1 + o(1)). \text{ lorsque } h \rightarrow 0.$$

1.2 Exemples de noyaux :

Voici quelques exemples de noyaux les plus communément utilisés :

- le noyau rectangulaire : $K(u) = \frac{1}{2}I_{[-1,1]}(u)$. C'est celui qui donne l'estimateur de type histogramme appelé noyau de Rosenblatt.
- le noyau triangulaire : $K(u) = (1 - |u|)I_{[-1,1]}(u)$
- le noyau d'Epanechnikov : $K(u) = \frac{3}{4}(1 - u^2)I_{[-1,1]}(u)$
- le noyau de Tukey ou biweight : $K(u) = \frac{15}{16}(1 - u^2)^2I_{[-1,1]}(u)$
- le noyau gaussien : $K(u) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{u^2}{2}}, u \in \mathbb{R}$

Les deux premiers ont l'avantage d'être simples, le noyau triangulaire étant continu partout et conduisant à une estimation \widehat{f}_n continue. Le troisième doit sa notoriété à une propriété d'optimalité théorique mais sans grand intérêt pratique. Le quatrième est, à notre sens, le plus intéressant car donnant une estimation dérivable partout, tout en étant simple à mettre en œuvre. En fait il s'agit du noyau le plus simple parmi les noyaux de forme polynomiale dérivables partout. Ainsi il assure le lissage local de la fonction \widehat{f}_n . Ce noyau est d'une forme très proche du noyau Gaussien et il est donc préférable. Notons que plus

la valeur de h est élevée plus on élargit la fenêtre, ce qui donne un effet de lissage globale de f_n plus important.

Voici quelques courbes de noyaux usuels présentées ci-dessous :

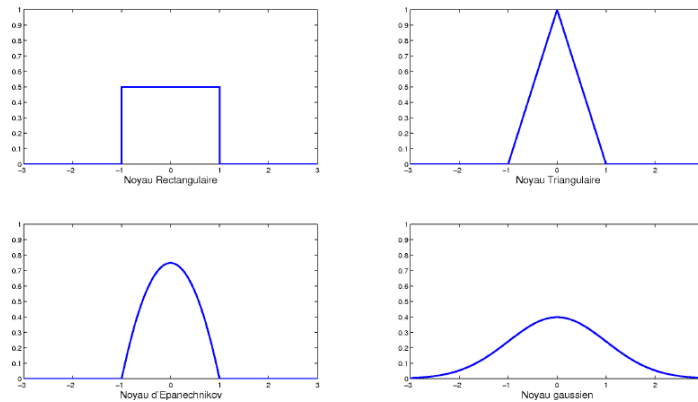


FIG. 1.2 – Les courbes des noyaux les plus communs.

Pour mieux saisir l'intuition derrière l'estimateur à noyau, nous avons construit cet estimateur en utilisant un ensemble de données constitué seulement de 7 observations. Le noyau K a été choisi comme étant la densité d'une loi normale de moyenne 0 et de variance 1 et le paramètre de lissage h égale à 4. On centre d'abord un noyau individuel sur chacune des 7 observations et la valeur de l'estimateur à noyau $f(z)$ au point z est simplement la somme des ordonnées de chacun des 7 noyaux individuels à ce point x comme représentée à la Fig [L.3](#). Dans une région où l'on a plusieurs observations, la vraie densité a une valeur relativement grande et l'estimateur de la densité, par la méthode du noyau, nous donne effectivement une valeur relativement grande ce qui est observé dans la Fig [L.3](#).

Par exemple si $x = 5$ on a $\hat{f}(x) = 0.03$ qui est égale a la somme des densités des 7 noyaux gaussiens au même point $x = 5$.autre exemple : (Rosenblatt ou noyau rectangulaire , Noyau triangle , Epanechnikov.Beiweight)

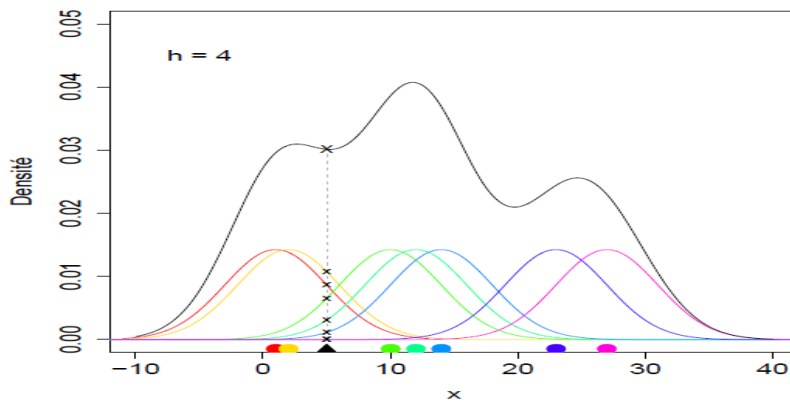


FIG. 1.3 – Estimateur ‘à noyau bas’ e sur 7 observations ($h=4$)

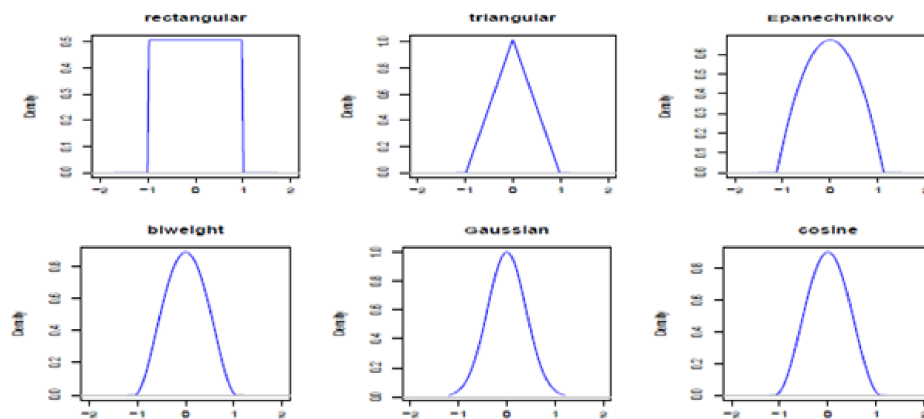


FIG. 1.4 – Exemples de noyaux.

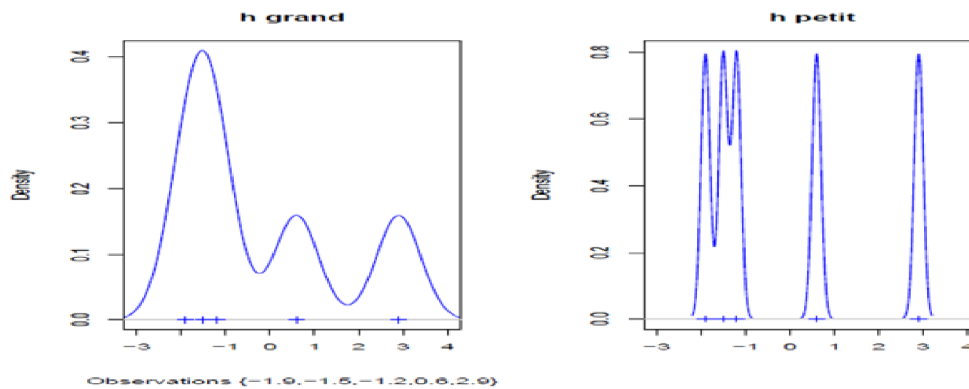


FIG. 1.5 – Effet de la variation de h sur l’estimateur à noyau.

Mise en perspective histogrammes/noyaux :

- * Dans l'estimateur par histogramme, on calcule la fréquence des observations dans des intervalles *fixés à l'avance*,
- * Dans l'estimateur à noyau *rectangulaire*, on calcule la fréquence des observations dans une fenêtre *glissante*.
- * Dans l'estimateur à noyau *gaussien*, toutes les observations sont prises en compte : celles qui sont proches du point x ou on estime la densité ont *un poids plus important* que les autres.

1.3 Estimation par la méthode de l'histogramme

La probabilité qu'une observation de X appartienne à l'intervalle $\{m_j - \frac{h}{2}, m_j + \frac{h}{2}\}$ et donnée par :

$$P(X \in \left\{m_j - \frac{h}{2}, m_j + \frac{h}{2}\right\}) = \int_{m_j - \frac{h}{2}}^{m_j + \frac{h}{2}} f(u) du$$

L'estimation de la densité par l'histogramme au point x est donnée par :

$$f_n(x) = \frac{1}{nh} I \left\{ \left[m_j - \frac{h}{2}, m_j + \frac{h}{2} \right] \right\}.$$

Avant de construire l'histogramme, on doit choisir le positionnement des intervalles ainsi que le paramètre de lissage h . Chacun de ces choix peut avoir un effet significatif sur l'histogramme résultant, ce qui constitue une difficulté. Une autre difficulté de la méthode de l'histogramme est que l'on estime toutes les densités par des fonctions étagées ce qui n'est pas toujours le cas. De plus, cet estimateur de la densité est une fonction discontinue, donc l'application de certaines opérations sur l'estime, comme par exemple une dérivée ou une intégration, devient impossible ou très difficile à effectuer.

1.3.1 Propriétés de l'estimateur par histogramme :

On étudiera quelques propriétés statistiques de l'estimation par la méthode de l'histogramme. On commence par la présentation du concept de l'erreur quadratique moyenne qui est définie comme suit :

$MSE(f_n(x)) = E[f_n - f(x)]^2$ qui s'écrit aussi sous la forme suivante :

$$MSE(f_n) = Var(f_n(x)) + Biases^2(f_n(x)) \quad \text{où}$$

$$Var(f_n) = E[(f_n(x))^2] - (E[f_n(x)])^2 \quad \text{et}$$

$$Biases(f_n(x)) = E[f_n(x) - f(x)].$$

Etude du biais :

Supposons que l'on dispose d'un échantillon d'observations X_1, \dots, X_n , provenant d'une distribution univariée continue et possédant pour fonction de densité la fonction f que l'on désire estimer. Pour $x \in B_j$ fixe, on a :

$$E(f_n(x)) = \frac{1}{nh} \sum E[X_i \in B_j] = \frac{1}{n} E[I_{\{X_i \in B_j\}}]$$

On note que $I_{\{X_i \in B_j\}} \sim B(p)$ tel que $p = \int_{(j-1)h}^{jh} f(u)du$

$$I(X_1 \in B_j) = \begin{cases} 1, \text{ avec la probabilité } p \\ 0, \text{ avec la probabilité } 1 - p \end{cases}$$

on a $E[I_{\{X_i \in B_j\}}] = \int_{(j-1)h}^{jh} f(u)du$

ce qui veut dire que f_n est biaisé et son biais est :

$$Biases = \frac{1}{h} \int_{(j-1)h}^{jh} f(u)du - f(x) = \frac{1}{h} \int_{(j-1)h}^{jh} f(u) - f(x)du$$

En appliquant le développement de Taylor pour $f(u) - f(x)$ au point x on a :

$$Biases = f'(x) \left(h(j - \frac{1}{2}) - x \right) + o(h). \quad (1.1)$$

Etude de la variance :

Calculons la variance de l'estimateur f_n

$$\begin{aligned} \text{var}(f_n(x)) &= \frac{1}{n^2 h^2} \sum_{i=1}^n \text{var} I_{\{X_i \in B_j\}} = \frac{1}{n h^2} \text{var} \{ I_{\{X_i \in B_j\}} \} = \frac{1}{n h^2} \int_{(j-1)h}^{jh} f(u) du \left(1 - \int_{(j-1)h}^{jh} f(u) du \right) \\ \text{var}(f_n(x)) &= \frac{1}{n h} f(x) + o\left(\frac{1}{n h}\right) \end{aligned} \quad (1.2)$$

On remarque que la variance diminue si h augmente et inversement pour le biais.

L'erreur quadratique moyenne (MSE) :

En utilisant (1.1) et (1.2) on trouve

$$MSE(f_n(x)) = f'(x)^2 \left(h(j - \frac{1}{2}) - x \right)^2 + \frac{1}{n h} f(x) + o(h) + o\left(\frac{1}{n h}\right) \quad (1.3)$$

On conclut que le MSE converge si $h \rightarrow 0$ et $n h \rightarrow \infty$.

L'erreur quadratique moyenne intégrée (MISE) :

$$MISE = \int_{-\infty}^{+\infty} MSE(f_n) dx$$

En utilisant (1.3), on peut écrire :

$$MISE = \int_{-\infty}^{+\infty} \frac{1}{n h} + \int_{-\infty}^{+\infty} \sum_j I(x \in B_j) f'(x) \left(h(j - \frac{1}{2}) - x \right)^2 dx \quad (1.4)$$

$$= \frac{1}{n h} \sum_j \int_{B_j} f'(x)^2 \left(h(j - \frac{1}{2}) - x \right)^2 dx \quad (1.5)$$

$$= \frac{1}{n h} + \sum_j \int_{B_j} f'(x)^2 dx \int_{B_j} \left(h(j - \frac{1}{2}) - x \right)^2 dx \quad (1.6)$$

$$= \frac{1}{n h} + \frac{h^2}{12} \int_{B_j} f'(x)^2 dx \quad (1.7)$$

En posant $C(f) = \int f'(x)^2 dx$. on a :

$$MISE(f_n) = \frac{1}{n h} + \frac{h^2 C(f)}{12}$$

Le paramètre de lissage optimal est celui qui minimise le MISE et qui est donné par :

$$\frac{dMISE(f_n)}{dh} = \frac{-1}{n h^2} + \frac{1}{6} C(f) = 0 \quad \text{Par conséquent : } h_{opt} = \left(\frac{6}{C(f)} \right)^{\frac{1}{3}} n^{-\frac{1}{2}} \quad (1.8)$$

Exemple 1.3.1 *On dispose des données de 709 modèles d'avion du 20^e siècle. On désire estimer la densité*

de l'envergure de ces avions par la méthode de l'histogramme en utilisant différents paramètres de lissage :

Remarque 1.3.1 *Dans ce travail on utilise le langage R pour la simulation des données. Accéder aux données*

data (aircarf) avec la commande `provide.data(aircarf)` nécessite l'installation du package (sm) :

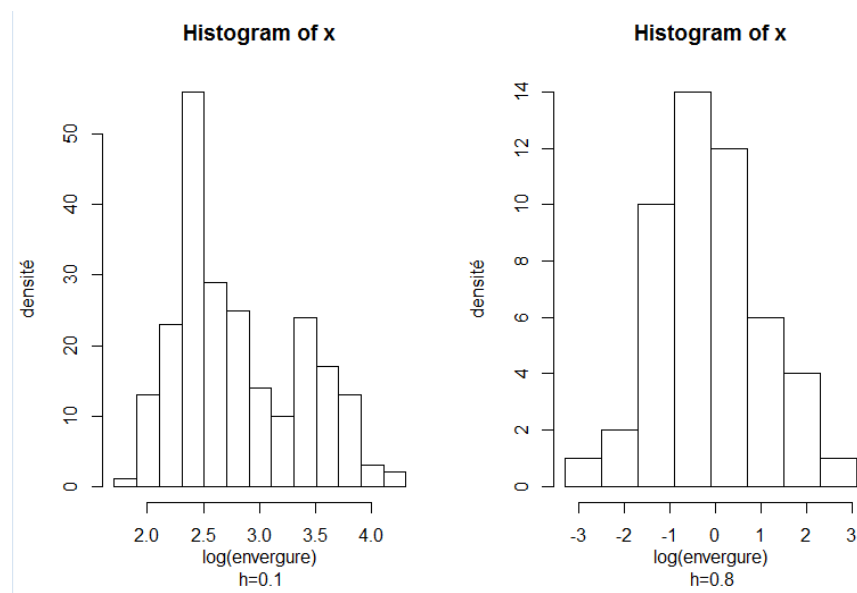


FIG. 1.6 – Différents histogrammes associés à un même ensemble de données avec différents paramètres de lissages.

La Fig [1.6](#) nous montre deux histogrammes basés sur le même ensemble de données et avec deux paramètres de lissage petit et grand. Un paramètre de lissage trop petit conduit à un histogramme plus découpé, tandis que l'autre donne un histogramme plus lissé comme le montre la Fig [1.6](#). A l'exception du choix du paramètre de lissage, ces difficultés peuvent

être évitées si l'on utilise la méthode d'estimation de la densité par le noyau. C'est la raison pour laquelle nous utiliserons cet estimateur plutôt que l'histogramme dans ce travail. Une étude approfondie de celui-ci est présentée dans la section suivante.

1.4 Estimation par la méthode des noyaux :

Le concept de noyau d'abord a été introduit par ROSENBLATT(1956), mais c'est CACOULOS (1966) qui a été le premier à utiliser le terme "noyau" pour désigner la fonction que l'on utilise dans les méthodes non paramétriques. En hydrologie, c'est YAKOTZ(1983) et FELUCH(1983) qui ont introduit indépendamment la méthode des noyaux lors d'une conférence de l'AGU à l'automne 1983.

Dans la méthode des noyaux, une fonction $K(x)$ est associée à chaque observation de l'échantillon. La seule véritable restriction concernant le noyau K est que son intégration sur tout le domaine de définition de x doit être égale à un. On rencontre parfois d'autres restrictions théoriques qui sont appliquées à K , comme la symétrie ou la positivité sur tout le domaine de définition du noyau (ADAMOWSKI,1989). Toutefois, ces restrictions sont surtout introduites afin de simplifier les développements théoriques. L'estimation non paramétrique de la fonction de densité peut se voir comme le cumul des fonctions K de chaque observation sur tout le domaine : Supposons que nous observons n variables aléatoires i.i.d X_1, \dots, X_n de densité f . L'objectif de notre étude est la construction d'un estimateur de f en un point fixe x . Notons $F(x) = P(X_1 \leq x)$ la fonction de répartition de la loi de X_1 . La densité est la dérivée de la fonction de répartition, ce qui permet d'écrire pour tout x :

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

Considérons la fonction de répartition empirique :

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

La loi des grands nombres permet d'affirmer que F_n est un estimateur de F , c'est-à-

dire

$$F_n(x) \xrightarrow{P} F(x). \quad (1.9)$$

De plus, le théorème de Glivenko-Cantelli nous donne :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s} 0 \quad (1.10)$$

Il est même possible d'obtenir des intervalles de confiance et de tester l'adéquation des données à différentes lois. Néanmoins, il n'est pas évident d'utiliser F_n pour estimer f .

Une des premières idées intuitives est de considérer pour fixé "petit" :

$$f_n(x) = \frac{F_n(x+h) - F_n(x-h)}{2h} = \frac{1}{2nh} \sum_{i=1}^n I_{\{-h \leq X_i - x \leq h\}}. \text{ On a alors :}$$

$$\begin{aligned} E[f_n(x)] &= \frac{1}{2n} (E[F_n(x+h)] - E[F_n(x-h)]) \\ &= \frac{1}{2} (F_n(x+h) - F_n(x-h)). \end{aligned}$$

$E[f_n(x)]$ tend vers $f(x)$ quand $h \rightarrow 0$. Il faut donc faire dépendre h de la taille de l'échantillon, et le faire tendre vers 0 quant $n \rightarrow \infty$, de sorte que Cet estimateur appelé estimateur de Rosenblatt, est le premier exemple d'estimateur à noyau construit à l'aide du noyau $f_n(x)$ soit un estimateur asymptotiquement sans biais de $f(x)$. L'estimateur f_n reste une fonction en escalier. Pour obtenir quelque chose de plus lisse, on peut remarquer que :

$$f_n(x) = \frac{1}{2nh} \sum_{i=1}^n I_{|x-h_n, x+h_n|}(X_i) = \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} I_{\{x-h_n < X_i \leq x+h_n\}} \quad (1.11)$$

$$= \frac{1}{nh_n} \sum_{i=1}^n \frac{1}{2} I_{[-1,1]}\left(\frac{x-X_i}{h_n}\right) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \quad (1.12)$$

Cet estimateur appelé estimateur de Rosenblatt, est le premier exemple d'estimateur à noyau construit à l'aide du noyau $k(u) = \frac{1}{2} I_{\{-1 < u \leq 1\}}$.

1.4.1 Estimation par noyaux associés unidimensionnelles :

Estimation par noyaux classiques univariés :

Définition 1.4.1 Une fonction K de support S est dite **noyau** si elle est une densité de probabilité symétrique (i.e $K(-u) = K(u)$), de moyenne K nulle ($\mu = \int_S uK(u)du = 0$), de variance σ_K^2 finie ($\sigma_K^2 = \int_S u^2K(u)du < +\infty$) et de carré intégrable

($\int_S K^2(u)du < +\infty$). Précisons ici qu'en tant que densité de probabilité, le noyau K est positif et de masse totale égale à 1 (i.e pour tout élément u de S , $K(u) > 0$ et $\int_S K(u)du = 1$).

Définition 1.4.2 Soit $h_n > 0$ la fenêtre de lissage et K la fonction noyau vérifiant la Fig 1.4.1 l'estimateur à noyau continu (classique) de f est défini en un point $x \in T$ par :

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

Alors la moyenne des valeurs des fonctions noyaux centrés sur chaque observation X_i de l'échantillon est l'estimation par noyau en un point donné $x \in T$.

Estimation par noyaux associés discrets univariés :

Définition 1.4.3 On appelle type de noyau discret toute fonction de masse de probabilité (f.m.p) K_θ , paramétrée par

$\theta \in \phi \subseteq \mathbb{R}^2$, de support $S_\theta \subseteq \mathbb{Z}$ et de carré sommable. Donnons maintenant la définition améliorée du noyau associé discret.

Définition 1.4.4 Soit $x \in T \subseteq \mathbb{Z}$ et $h > 0$ avec T le support de la fonction de masse de probabilité f à estimer. Une densité paramétrée $K_{x,h}$ de support $S_{x,h} \subseteq \mathbb{Z}$ est appelée noyau associé discret lorsque les conditions suivantes sont vérifiées :

$$x \in S_{x,h} \tag{1.13}$$

$$E(Zx; h) = x + A(x; h) \tag{1.14}$$

$$Var(Zx; h) = B(x; h) \tag{1.15}$$

avec $A(x; h)$ et $B(x; h)$ tendent vers zéro quand h tend vers 0, et $Z_{x,h}$ une variable aléatoire discrète de loi $K_{x,h}$.

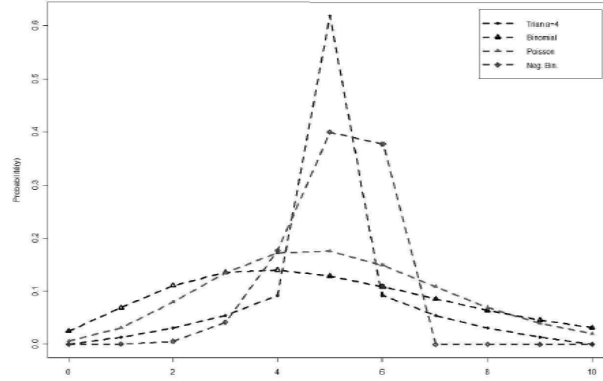


FIG. 1.7 – Exemples noyaux associés discrets

Noyau associé	$K_{x;h}(u)$	T
Aitchison et Aitken	$(1-h)\prod_{\{u \neq x\}} + \frac{h}{c-1}\prod_{\{u=x\}}$	$\{0; 1; \dots; c-1\}$
Binomial	$\frac{(x+1)!}{u!(x+1-u)!} \left(\frac{x+h}{x+1}\right)^u \left(\frac{1-h}{x+1}\right)^{x+1-u}$	$\{0; 1; \dots; x+1\}$
Binomial Négatif	$\frac{(x+u)!}{u!x!} \left(\frac{x+h}{2x+1+h}\right)^u \left(\frac{x+1}{2x+1+h}\right)^{x+1}$	\mathbb{N}
Poisson	$\frac{(x+h)^u e^{-(x+h)}}{u!}$	\mathbb{N}
Wang et Van Ryzin	$\frac{(m+1)^h - u-x ^h}{P(m,h)}$	$\{0; \pm 1; \dots; \pm k\}$
Triangulaire	$(1-h)\prod_{\{u \neq x\}} + \frac{1}{2}(1-h)h^{ u-x } \prod_{\{u-x \geq 1\}}$	\mathbb{Z}

TAB. 1.1 – Quelques noyaux associés discrets.

avec $P(m; h) = (2m+1)(m+1)^h - 2 \sum_{k=0}^m k^m$

La définition suivante présente l'estimateur à noyau associé discret.

Estimation à noyaux associés continus univariés :

Définition 1.4.5 (i) Le mode $M(a; b)$ d'un type de noyau $K(a; b)$ appartient toujours à son support $S(a; b)$.

(ii) Le mode a la meilleure probabilité que le point moyen.

(iii) Lorsque la dispersion autour du mode tend vers zéro alors celle autour de la moyenne tend également vers zéro.

Définition 1.4.6 On considère $x \in T \subseteq \mathbb{R}$ et $h > 0$ avec T le support de la densité f , à estimer. Une fdp paramétrée $K_{x;h}$ de support $S_{x;h} \subseteq \mathbb{R}$ est appelée noyau associé continu lorsque les conditions suivantes sont satisfaites :

$$x \in S_{x;h} \tag{1.16}$$

$$E(Z_{x;h}) = x + A(x; h) \tag{1.17}$$

$$Var(Z_{x;h}) = B(x; h) \tag{1.18}$$

où $Z_{x;h}$ est une variable aléatoire de densité $K_{x;h}$ sur $S_{x;h}$ et $A(x; h); B(x; h)$ tendent vers 0 quand h tend aussi vers 0.

Dans la proposition suivante, nous montrons que tous les noyaux classiques sont des noyaux associés. Nous y donnons aussi la forme de leurs supports en tant que noyaux associés ainsi que leurs caractéristiques.

Proposition 1.4.1 Soit K un noyau classique. Pour un $x \in T = \mathbb{R}$ donné et $h > 0$, alors le noyau associé classique de support S_x ($S_{x;h} = x - h$ les bornes du S) et de plus :

$$E(Z_{x;h}) = x \text{ et } Var(Z_{x;h}) = h^2 \sigma_K^2 \tag{1.19}$$

En d'autres termes, (1.19) montre que les caractéristiques A et B du noyau associé classique $K_{x;h}$ sont :

$$A(x; h) = 0 \text{ et } B(x; h) = o(h^2) \tag{1.20}$$

Nous désignons tout au long de cette section par $X_1; X_2; \dots; X_n$ une suite de variables aléatoires (i.i.d) de densité f inconnue sur $T \subseteq \mathbb{R}$, par $Z_{\theta(x;h)}$ une variable aléatoire de densité $K_{\theta(x;h)}$ (où $K_{\theta(x;h)}$ est un noyau associé construit par la méthode mode-dispersion).

Définition 1.4.7 Soit f la densité à estimer sur $T \subseteq \mathbb{R}$; $h > 0$ et $K_{\theta(x;h)}$ un noyau

associé continu. L'estimateur f_n de f issu de $K_{\theta(x;h)}$ est défini par :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{\theta(x;h)}(X_i); x \in T \subseteq \mathbb{R}.$$

1.4.2 Estimation par noyaux associés multidimensionnels :

Définition 1.4.8 (*Noyau associé discret multivarié*)

Soient $x \in T_d \subseteq Z^d$ et H une matrice des fenêtres, avec T_d est le support de la fmp f à estimer. Une fmp $K_{x;H}(\cdot)$ de support $S_{x;H} \subseteq Z^d$ est appelée noyau associé discret multivarié si

$$x \in S_{x;H} \tag{1.21}$$

$$E(Z_{x;H}) = x + A(x; H) \tag{1.22}$$

$$Cov(Z_{x;H}) = B(x; H) \tag{1.23}$$

où $A(x; H)$ et $B(x; H)$ tendent vers le vecteur nul et la matrice nulle respectivement quand $H \rightarrow 0^d$ (0^d est la matrice carrée nulle d'ordre d) et $Z_{x;H}$ est un vecteur de variables aléatoires discrètes de loi $K_{x;H}$.

Définition 1.4.9 (*Noyau associé discret produit*)

Soient $T_1^{[j]}$ le support des marges univariés de f ; $x = (x_1; x_2; \dots; x_d)^t$ est le vecteur cible h_j avec $j = 1; \dots; d$ sont les fenêtres de lissage univariés et $K_{x_j;h_j}^{[j]}$ est le j -ème noyau associé discret univarié de support $S_{x_j;h_j}$. "Le noyau associé discret multivarié produit" est défini comme suit :

$$K_{x;H}(\cdot) = \prod_{j=1}^d K_{x_j;h_j}^{[j]}(\cdot); \forall x_j \in T_1^{[j]} \subseteq Z \tag{1.24}$$

Estimateur à noyau associé discret multivarié :

Définition 1.4.10 Soient $x_1; \dots; x_n$ des vecteurs aléatoires (*i.i.d*) de fmp multivariée commune inconnue f à estimer sur T_d . L'estimateur à noyau associé discret multivarié \hat{f}_n de f est de la forme :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x;H}(X_i); x \in T_d \tag{1.25}$$

où $K_{x;H}(\cdot)$ est le noyau associé discret multivarié dépendant du vecteur cible x et de la matrice des fenêtres H symétrique définie positive, tend vers la matrice nulle (0_d) quand n tend vers ∞ .

Définition 1.4.11 (*L'estimateur à noyau associé produit*)

Selon [1.4.10](#) le produit des noyaux associés discrets univariés (1.28) conduit à une version utile de (1.25) que nous appelons "estimateur à noyau associé produit" :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{x_j; h_j}^{[j]}(X_{ij}); x_j \in T_1^{[j]} \subseteq Z \quad (1.26)$$

où $T_1^{[j]}$ est le support de la marge univariée de f pour $j = 1; \dots; d; x = (x_1; \dots; x_d)^t \in X_{j=1}^d \cdot T_1^{[j]}$, $X_i = (X_{i1}; \dots; X_{id})^t$

pour $i = 1; \dots; n$ et $h_1; \dots; h_d$ sont les paramètres de lissage unidimensionnels. La fonction $K_{x_j; h_j}^{[j]}$ est le j^{eme} noyau associé discret univarié de support $S_{x_j; h_j} \subseteq Z$.

1.4.3 Estimation par noyaux associés mixtes :

Définition 1.4.12 *Un type de noyau mixte K de support S est une combinaison convexe des types de noyaux continus et discrets $K^{[j]}$ de support $S^{[j]}$. On écrit alors $K_\theta(u) = \sum_j \beta_j K_\theta^{[j]}(u) 1_{S^{[j]}}(u)$; avec $\beta_j > 0$, tels que $\sum_j \beta_j = 1$.*

Nous précisons ici que le support S du type de noyau mixte est la réunion des supports $S^{[j]}$ des types de noyaux continus et discrets dont il est issu. On le note généralement par :

$$S_\theta = \cup_j S^{[j]} \quad (1.27)$$

Sans perte de généralité, nous désignerons par S_{θ_1} le support d'un type de noyau continu et par S_{θ_N} celui d'un type de noyau discret. Par conséquent, le support S de K sera la réunion de S_{θ_1} et S_{θ_N} . Nous allons maintenant définir le noyau associé mixte.

Définition 1.4.13 *Soit $T = \cup_j T_j$. le support de la densité mixte f à estimer. Un noyau associé mixte $K_{\theta(x;h)}$ est une densité de probabilité paramétrée par le point d'estimation x et le paramètre de lissage h , et constitué à la fois de noyaux associés continus et discrets. Il s'écrit :*

$$K_{\theta(x;h)}(\cdot) = \sum_{j=1}^p \beta_j K_{x;h}^{[j]}(\cdot) 1_{S_{x;h}^{[j]}}(x) 1_{T_j}(\cdot) \quad (1.28)$$

où les β_j sont des réels positifs tels que $\sum_j \beta_j = 1$. Il a pour support $S_{\theta(x;h)}$ défini par :

$$S_{\theta(x;h)} = \bigcup_{j=1}^p S_{x;h}^{[j]} \quad (1.29)$$

avec $S_{x;h}^{[j]}$ supports de $K_{x;h}^{[j]}(\cdot)1_{S_{x;h}^{[j]}}(x)1_{T_j}(\cdot)$.

On vérifie facilement que $K_{\theta(x;h)}(\cdot)$ récrire de la forme :

$$x \in S_{\theta(x;h)} \quad (1.30)$$

$$E(Z_{x;h\beta}) = x + A_{\theta}(x; h) \quad (1.31)$$

$$Var(Z_{x;h\beta}) = B_{\theta}(x; h) \quad (1.32)$$

où $A_{\theta}(x; h) = \sum_{j=1}^p A_j(x; h)1_{S_{x;h}^{[j]}}(x)$ et $B_{\theta}(x; h) = \sum_{j=1}^p B_j(x; h)1_{S_{x;h}^{[j]}}(x)$ tendent vers 0 lorsque h tend vers 0 et $Z_{\theta}(x; h)$ est une variable aléatoire de loi $K_{\theta}(x; h)$. Proposons maintenant les

estimateurs à noyaux associés mixtes et quelques unes de leurs propriétés fondamentales.

$$\hat{f}_n(x) = \sum_{j=1}^p \frac{\beta_j}{n_j} \sum_{i=1}^{n_j} K_{x;h}^{[j]} 1_{S_{x;h}^{[j]}}(x) 1_{T_j}(X_i) \quad (1.33)$$

où n_j le nombre d'observations tombant dans T_j et les $K_{x;h}^{[j]}(\cdot)1_{S_{x;h}^{[j]}}(x)1_{T_j}(\cdot)$ sont des noyaux associés de support $S_{x;h}^{[j]}$. Cet estimateur peut être écrit encore sous la forme :

$$\hat{f}(x) = \sum_{j=1}^p \beta_j \hat{f}_j(x) 1_{T_j}(x) \quad (1.34)$$

où les \hat{f}_j sont les estimateurs à noyaux associés $K_{x;h}^{[j]}(\cdot)1_{S_{x;h}^{[j]}}(x)1_{T_j}(\cdot)$ de support $S_{x;h}^{[j]}$, pondérés par les poids j (connus) sur les composantes T_j de T .

Nous signalons ici que l'estimateur à noyaux associés mixtes hérite simultanément quelques une des propriétés élémentaires des noyaux associés discrets et continus le constituant.

1.4.4 Estimateur à noyau dans le cas multivarié :

On suppose que X_i est un q - vecteur aléatoire et on veut estimer la densité

$f(x) = f(x_1, \dots, x_q)$. L'estimateur à noyau multivarié est donné par :

$$\hat{f}_n(x) = \frac{1}{n|H|} \sum_{i=1}^n K(H^{-1}(X_i - 1)) \quad (1.35)$$

Ou $K(u)$ est la fonction du noyau multivarié qui dépend du vecteur

$$H = (h_1, \dots, h_q) \text{ et } |H| = h_1, h_2, \dots, h_q. \text{ avec } \int k(u)(du) = \int k(u)du_1 du_2 \dots du_q = 1$$

Typiquement, $k(u)$ est : $k(u) = K(u_1)K(u_2)\dots K(u_q)$

Quand $K(u)$ est le produit des noyaux alors la densité marginale de $f_n(x)$ est égale a la

densité de l'estimateur à noyau dans le cas univarié avec k la fonction du noyau et h_j le paramètre de lissage.

On peut donc montrer que quand $K(u)$ prend la forme d'un produit, le biais de l'estimateur est :

$$\text{Biais}(\hat{f}(x)) = \frac{K_d(k)}{d!} \sum_{j=1}^q \frac{\delta^d}{\delta x_j^d} f(x) h_j^d + o(h_1^d + \dots + h_1^d)$$

et la variance est donnée par : $\text{Var}(\hat{f}(x)) = \frac{f(x)R(k)^q}{n|H|} + o(\frac{1}{n})$ où $R(k) = \int k^q(u)du$

Par conséquent le AMISE est :

$$\text{AMISE}(\hat{f}(x)) = \frac{K_d^2(k)}{(d!)^2} \int \left(\sum_{j=1}^q \frac{\delta^d}{\delta x_j^d} f(x) h_j^d \right)^2 (dx) + \frac{R(k)^q}{n|H|}.$$

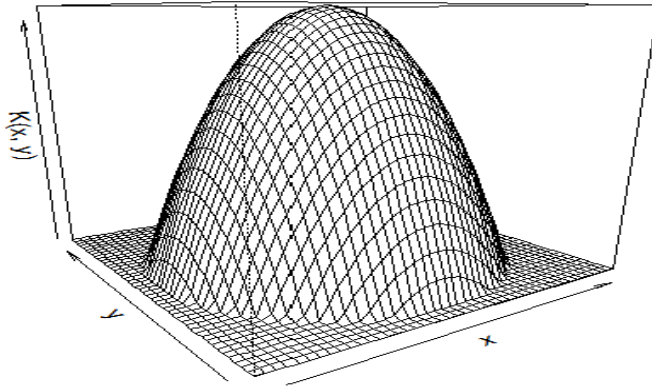


FIG. 1.8 – Densité d'un noyau d'Epanechnikov dans le cas multivarié.

1.5 Propriétés de l'estimateur à noyau :

Nous allons maintenant donner quelques propriétés statistiques élémentaires de l'estimateur de la densité

à noyau ainsi que différentes méthodes pour choisir le paramètre de lissage.

1.5.1 Etude du biais :

Supposons que l'on dispose d'un échantillon X_1, \dots, X_n issu d'une *v.a* X d'observations pour fonction de densité la fonction f que l'on désire estimer. On suppose que f_n est

l'estimateur à noyau obtenu en utilisant le noyau K et le paramètre de lissage h et f_n .

Supposons que :

$$K(u) \geq 0, \int K(u)du = 1, \int K(u)udu = 0, \int K(u)du < \infty$$

et en supposant que la densité de probabilité f admet les deux premières dérivées (continues) nécessaire.(1.36)

$$E[f_n(x)] = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} E \left[K \left(\frac{x-X_i}{h} \right) \right] = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{x-t}{h} \right) f(t) dt = \frac{1}{h} \int_{-\infty}^{+\infty} K \left(\frac{x-t}{h} \right) f(t) dt$$

La transformation $z = \frac{x-t}{h}$, i.e $t = -hz + x$, $\left| \frac{dz}{dt} \right| = \frac{1}{h}$

$$E[f_n(x)] = \int_{-\infty}^{+\infty} K(z) f(x-hz) dz$$

Un développement de Taylor de $f(x-hz)$ nous donne :

$$f(x-hz) = f(x) - hzf'(x) + \frac{1}{2}(hz)^2 f''(x) + o(h^2)$$

$$\begin{aligned} E[f_n(x)] &= \int_{-\infty}^{+\infty} K(z) f(x) dz - \int_{-\infty}^{+\infty} K(z) hzf'(x) dz + \int_{-\infty}^{+\infty} K(z) \frac{(hz)^2}{2} f''(x) dz + o(h^2) \\ &= f(x) \int_{-\infty}^{+\infty} K(z) dz - hf'(x) \int_{-\infty}^{+\infty} zK(z) dz + \frac{(hz)^2}{2} f''(x) \int_{-\infty}^{+\infty} z^2 K(z) dz + o(h^2) \\ &= f(x) + \frac{h^2}{2} K_2 f''(x) + o(h^2) \end{aligned}$$

$$Biais(f_n(x)) \approx \frac{h^2}{2} K_2 f''(x) + o(h^2) \tag{1.36}$$

Le biais dépend de :

- h : paramètre de lissage.
- K_2 : la variance du noyau.
- $f''(x)$ la seconde dérivée de la fonction de densité au point x .

1.5.2 Etude de la variance :

La variance de $f_n(x)$ est donnée par :

$$Var(f_n(x)) = var \left(\frac{1}{nh} \sum_{i=1}^n K \left(\frac{x-X_i}{h} \right) \right) = \frac{1}{n^2 h^2} \sum_{i=1}^n var \left(K \left(\frac{x-X_i}{h} \right) \right)$$

car les X_i , $i = 1, \dots, n$, sont *i.i.d*

$$\begin{aligned} Var \left(K \left(\frac{x-X_i}{h} \right) \right) &= E \left[K \left(\frac{x-X_i}{h} \right)^2 \right] - E \left[K \left(\frac{x-X_i}{h} \right) \right]^2 \\ &= \int K \left(\frac{x-t}{h} \right)^2 f(t) dt - \left(\int K \left(\frac{x-t}{h} \right) f(t) dt \right)^2 \end{aligned}$$

$$\text{Var}(f_n(x)) = \frac{1}{n} \int \frac{1}{h^2} K\left(\frac{x-t}{h}\right)^2 f(t) dt - \frac{1}{n} f(x) + \text{Biais}(f_n(x))^2$$

En effectuant le changement de variable suivant $z = \frac{x-t}{h}$, obtient :

$$\text{Var}(f_n(x)) = \frac{1}{nh} \int K(x)^2 f(x-hz) dz - \frac{1}{n} (f(x) + o(h^2))^2$$

et en effectuant un développement limité à l'ordre 2, il vient :

$$\text{Var}(f_n(x)) = \frac{1}{nh} \int K(x)^2 (f(x) - hzf'(x) + o(h)) dz - \frac{1}{n} (f(x) + o(h^2))^2$$

• Discussion du comportement du biais et de la variance :

- _ Le biais décroît si h diminue mais la variance augmente.
- _ La variance diminue si h augmente mais le biais augmente.
- _ Pour que la variance tende vers zéro, il faut que $nh \rightarrow \infty$.
- _ Plus la courbure de la densité est haute en x , plus le biais est grand.
- _ La variance est plus grande pour des valeurs plus grandes de la densité.

La figure suivante nous permet de mieux voir le comportement du biais et de la variance.

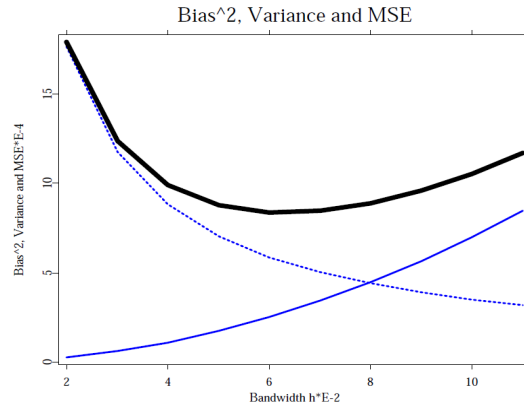


FIG. 1.9 – Le "trad-off" biais-variance en fonction de h .

La variance est représentée par la courbe en pointillé et le biais par la courbe fine, la courbe en gras représente le MSE.

1.6 Le choix optimal du noyau :

Le tableau suivant donne quelques noyaux et leurs efficacités respectives :

Kernel K	Eff (K)
Epanchinkov $\frac{3}{4}(1 - u^2)I(u \leq 1)$	1.000
Quartie (biweight) $\frac{15}{16}(1 - u^2)^2I(u \leq 1)$	0.994
Triweight $\frac{35}{32}(1 - u^2)^3I(u \leq 1)$	0.987
Gaussien $\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$	0.951
Uniforme $\frac{1}{2}I(u \leq 1)$	0.930
Triangle $(1 - u)I(u \leq 1)$	0.986

TAB. 1.2 – Quelques noyaux et leurs efficacités

Exemple 1.6.1 Nous considérons les données *data* (*faithful*) qui se trouvent dans le package *datasses* et on représente dans la figure suivante les différentes densités des quatre noyaux différents en utilisant le paramètre de lissage optimale.

Remarque 1.6.1 D'après la Fig [1.10](#), on voit que les densités estimées par les quatre noyaux sont pratiquement les mêmes sauf le noyau rectangulaire qui donne un estimateur moins lisse que les autres, ce qui veut dire que le choix du noyau n'a pas impact sur l'estimateur à noyau.

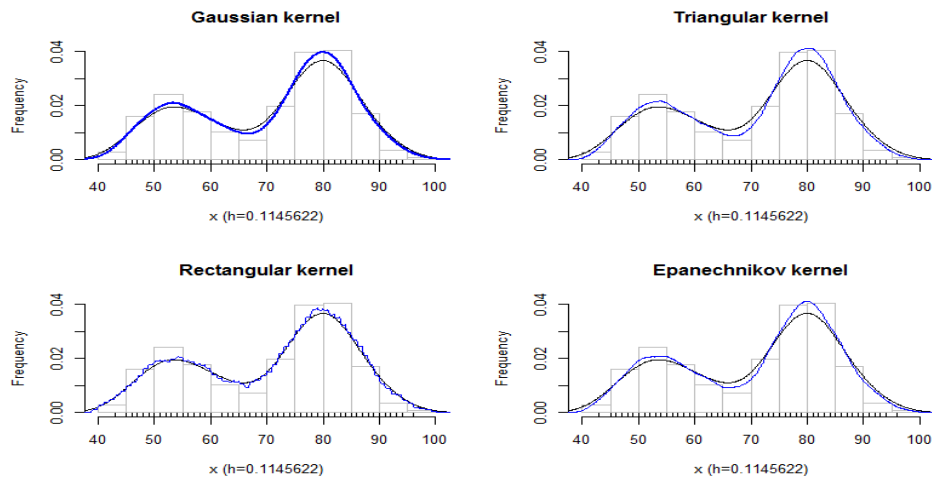


FIG. 1.10 – Estimation de la densité avec différents noyaux.

1.7 Choix théorique optimal du paramètre de lissage :

Pour le paramètre de lissage on fait la distinction entre

- h paramètre de lissage constant (ou global).
- $h(x)$ paramètre de lissage variable (local).

Ces choix différents du paramètre de lissage résultent des estimateurs à noyau suivants :

$$\begin{aligned}
 f_n(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \\
 f_{n,L}(x) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h(x)} K\left(\frac{x-X_i}{h(x)}\right) \\
 \frac{dMISE(f_n)}{dh} &= h^3 K_2^2 \beta(f) - \frac{1}{nh^2} j_2. \\
 h_{opt} &= \left(\frac{1}{n} \frac{\gamma(K)}{\beta(f)}\right)^{\frac{1}{5}} \tag{1.37}
 \end{aligned}$$

Un critère approprié pour sélectionner un paramètre de lissage variable (local) $h(x)$ est la mesure de performance locale MSE.

$$\frac{dMISE}{dh} = h^3 K_2^2 f''(x)^2 - \frac{1}{nh^2} f(x) j_2$$

d'où

$$h_{opt}(x) = \left(\frac{1}{n} \frac{f(x)\gamma(K)}{\beta(f)}\right)^{\frac{1}{5}} \tag{1.38}$$

Remarque 1.7.1 Les choix h_{opt} et $h_{opt}(x)$ sont des choix théoriques, qui ne sont pas utilisables en pratique car

ils d'épandent des quantités inconnues f et f'' .

Exemple 1.7.1 on fait varier le paramètre de lissage pour estimer la densité par la méthode du noyau, et on

obtient les résultats représentés dans les trois figures suivantes :

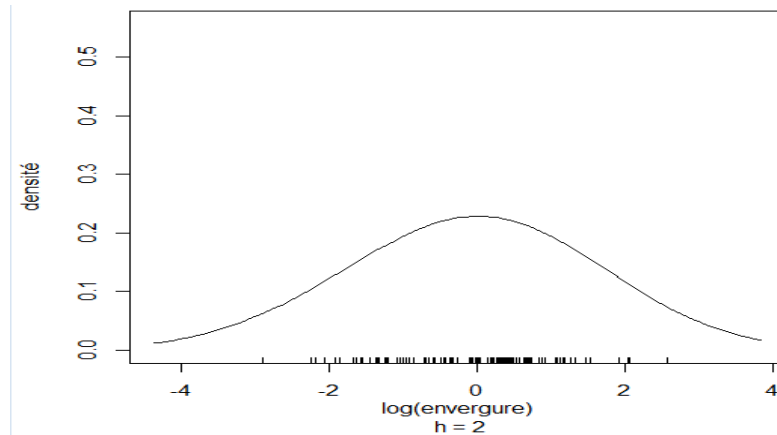


FIG. 1.11 – Illustration d'un phénomène de sous-lissage lors de l'estimation d'une densité .

Le choix d'une valeur de h trop grande conduit à une courbe trop lisse. La courbe estimée ne traduit pas suffisamment les variations de la vraie distribution. Fig [1.11](#)

Par contre, en choisissant un paramètre de lissage très petit, l'allure de la distribution change. Il s'agit d'une distribution estimée Fig [1.12](#)

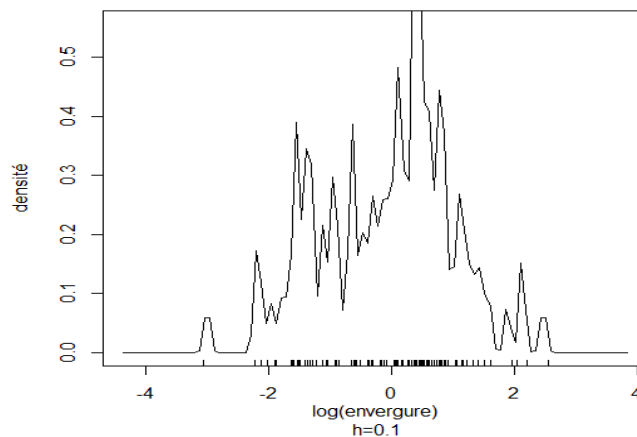


FIG. 1.12 – Illustration d'un phénomène de sur-lissage lors de l'estimation d'une densité .

L'estimation de la densité nécessite également le choix adéquat de la fenêtre h , et pour cette valeur idéale du paramètre de lissage, nous obtenons une allure qui suit parfaitement la vraie distribution. FIG [1.13](#)

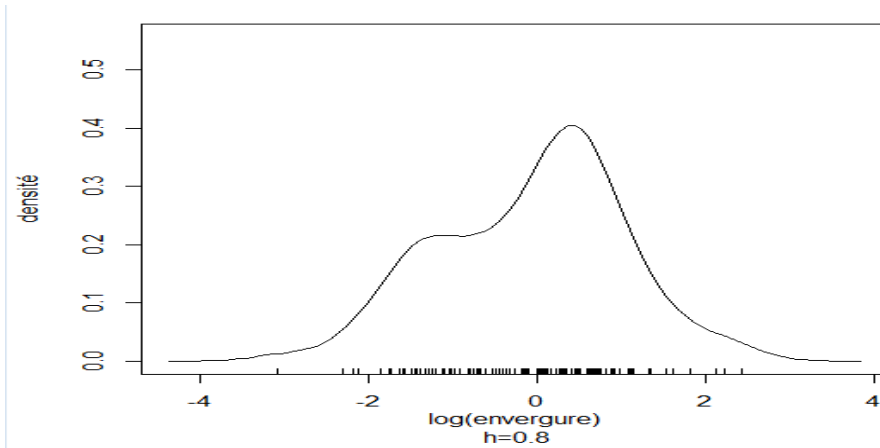


FIG. 1.13 – Illustration d'une estimation idéale

Les courbes obtenues dans cette exemple illustrent à quel point les formes estimées sont différentes en fonction de l'ordre de grandeur du paramètre de lissage. La principale difficulté repose sur le choix optimal de la fenêtre

1.8 Simulation des données :

Dans cette partie, nous illustrons certains estimateurs à noyaux continus symétriques à savoir le noyau d'Epanechnikov, le noyau Gaussien, le noyau Biweight et le noyau triangulaire. Nous simulons un échantillon de taille $n = 100$ de la loi normale centrée et réduite. Pour chaque noyau fixé, la fenêtre optimale est choisie par les méthodes de validation croisée par moindres carrés et de Plug-in. Nous obtenons les résultats suivants :

Nous obtenons dans la Fig [1.14](#) pratiquement des estimations similaires pour chaque noyau continu utilisé ; ceci s'explique par le fait que les noyaux continus symétriques possèdent tous des efficacités proches l'une de l'autre. Pour chaque noyau continu symétrique fixé, la Fig [1.15](#) présente la fenêtre optimale $h_{cv} = 0.1636$.

Pour cette valeur de h , les estimations des différentes densités sont pratiquement similaires. Nous comparons différentes estimations en faisant varier la valeur de la fenêtre pour le même noyau continu. Nous choisissons le noyau optimal d'Epanechnikov. Les simulations

effectuées dans la Fig [1.16](#) mettent en lumière le fait que les performances pratiques des estimateurs à noyaux continus symétriques considérés dépendent fortement du choix de la fenêtre h . Par conséquent, ce choix est plus crucial que le choix du noyau. Les valeurs de h sont celles choisies par Plug-In ($hPi = 0.338$), validation croisée par vraisemblance ($h_{cv} = 0.429$) et deux autres valeurs arbitraires tel que $h = 0.5$ et $h = 1$.

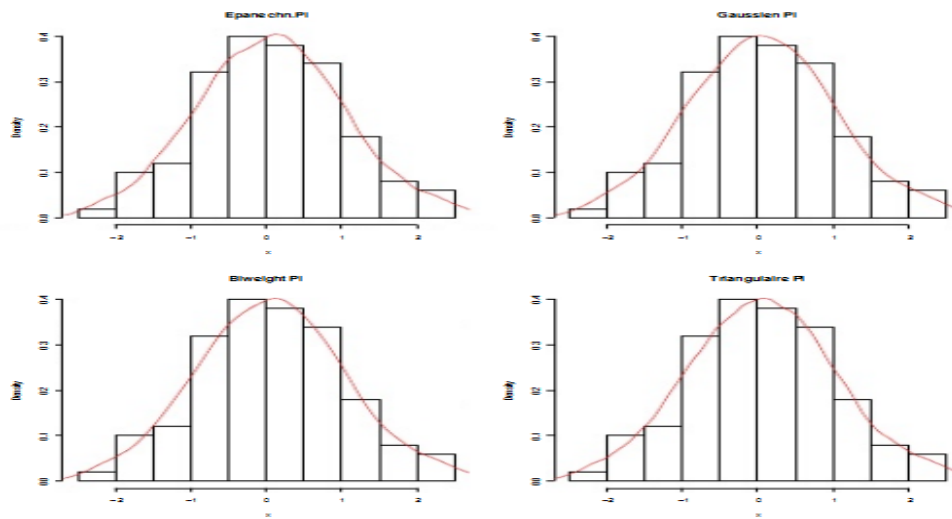


FIG. 1.14 – Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon

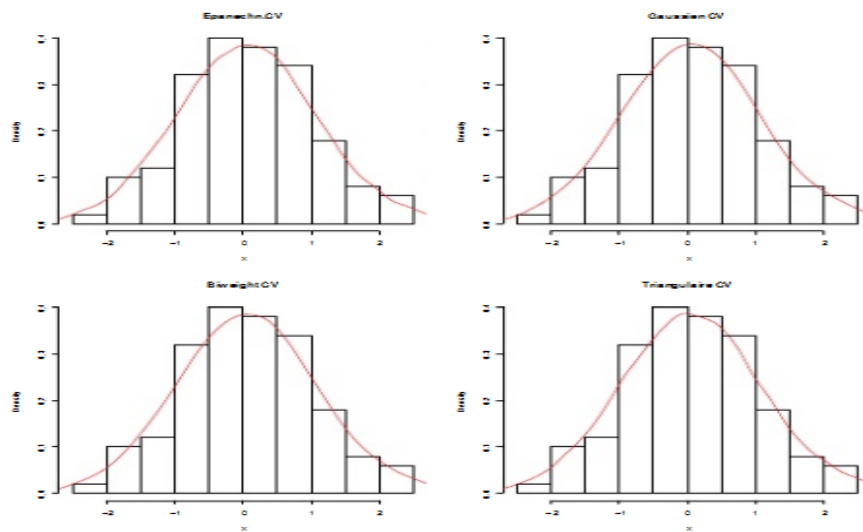


FIG. 1.15 – Lissages par des estimateurs à noyaux continus de la distribution d'un échantillon

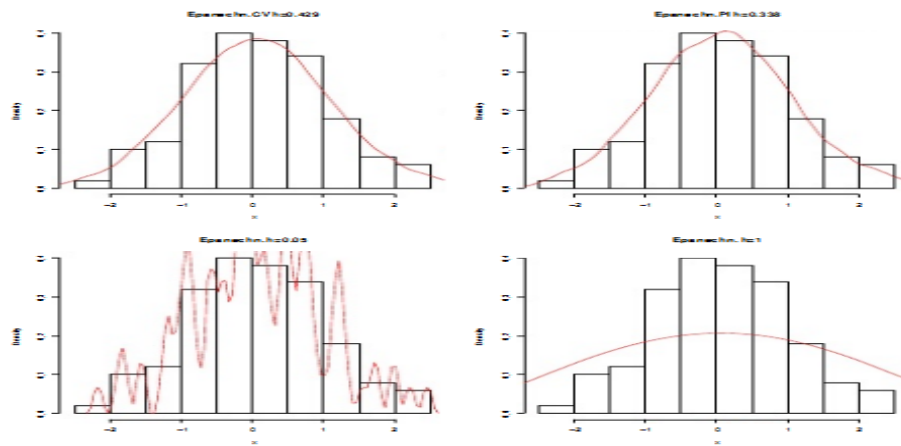


FIG. 1.16 – Comparaison des lissages par l'estimateur à noyau continu d'Epanchnikov en faisant

Chapitre 2

Estimation à noyau de la densité et correction des effets de bord.

2.1 Introduction

L'estimateur à noyau de la densité a été introduit et étudié par Rosenblatt [52], Parzen [50] et Watson et Leadbetter [32]. Plus tard, Schuster [57] a prouvé la convergence uniforme des dérivées de la densité.

Lorsque le support d'estimation est compact, les estimateurs à noyau sont sujets à des effets de bord. De nombreuses méthodes ont vu le jour afin d'y remédier. Gasser et Müller [63] ont modifié le noyau aux bords du support, tout comme Gasser, Müller et Mammitzsch puis Müller [64]. Schuster [57] et Silverman [53] ont adopté une méthode de réflexion (et même négative pour le second) autour des bords du support alors que Hall et Wehrly [51] ont employé une méthode géométrique alliant la réflexion et la modification du noyau aux bords du support. Cowling et Hall ont eu l'idée de générer des pseudodata et de les intégrer à l'estimateur à noyau.

Dans ce chapitre, nous nous concentrons sur l'estimation à noyau classique de la densité et de ses dérivées successives. Nous améliorons tout d'abord le biais des estimateurs à l'aide

de noyaux d'ordre supérieur. Nous développons ensuite une méthode de correction des effets de bord à partir des dérivées successives de la fonction estimée, ce qui justifie une fois de plus l'utilité d'estimer les dérivées successives d'une fonction.

2.2 Cadres d'études :

Soit x un réel fixé.

(C.1)Hypothèse (H1)

F_n est un estimateur consistant de F .

(C.2)Hypothèse (H2)

$\Psi(x, \cdot)$ est continue par rapport à la norme sup.

Nous donnons plus de détails sur l'hypothèse (H2). Elle signifie que pour toute suite de fonctions de répartition F_n vérifiant $\sup x \in \mathbb{R}$

$|F_n(x) - F(x)| \rightarrow 0$, nous obtenons $\sup x \in \mathbb{R} |\Psi(x, F_n) - \Psi(x, F)| \rightarrow 0$. Désormais, F_n désignera la fonction de répartition empirique classique basée sur les observations X_1, \dots, X_n :

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

Alors l'hypothèse (H1) est satisfaite et puisque nous supposons (H2),

(C.3)Hypothèse (H3)

$$(i) \int_{-1}^1 u^i K_r^{(m)}(u) du = \begin{cases} m! & , i = 0 & , \\ 0 & , i \neq m \text{ et } i \in \{1, \dots, r\}, \\ C_{r+1}, & i > r & , \end{cases}$$

avec C_{r+1} une constante finie ,

$$(ii) C = \sup_{u \in [-1, 1]} K_r^{(m)}(u) < +\infty,$$

(iii) Le support de $K_r^{(m)}(u)$ est $[-1, 1]$.

(C.4)Hypothèse (H4)

(h_n) est une suite décroissante de nombres réels strictement positifs telle que $h_n \rightarrow 0$

lorsque $n \rightarrow +\infty$.

Ainsi, l'expression $\widehat{\Psi}_n^{(m)}(x) = \frac{1}{h_n^m} \Psi(z, F_n) \frac{1}{h} K_r^{(m)}\left(\frac{z-x}{h}\right) dz$

de l'estimateur de $\Phi^{(m)}(x, F)$ est légèrement modifiée et devient :

$$\widehat{\Psi}_n^{(m)}(x) = \int \Psi(z, F_n) \frac{1}{h_n^{m+1}} K_r^{(m)}(u)(z-x) dz.$$

2.3 Estimation à noyau classique de la densité et de ses dérivées :

Soient $r \in \mathbb{N}^*$ et m un entier tels que $1 \leq m \leq r$. Notons X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées de même fonction de répartition F

Soit $t \in \mathbb{R}$. Nous faisons suite au (C.2) et nous nous tournons maintenant vers le but principal de l'estimation à noyau, qui est d'estimer les dérivées de la fonction de répartition, $F^{(m)}(t)$ pour

$m = 1, \dots, r$. Pour cela, nous avons besoin de poser l'hypothèse suivante :

Hypothèse (H1)

F est $(m+2)$ fois continûment différentiable en tout point t . Introduisons un noyau k tel que :

Hypothèse (H2)

(i) k est m fois continûment différentiable.

(ii) Le support de $k(m)$ est $[-1, 1]$

(iii) $\sup_{u \in [-1, 1]} k^{(m)}(u) < +\infty$,

puis nommons (h_n) une suite de fenêtres réelles strictement positives vérifiant

$h_n \rightarrow 0$ lorsque $n \rightarrow +\infty$.

Par commodité de notation, nous travaillons avec la densité $F = f$ comme fonction initiale, considérons l'estimateur de $f^{(m)}(t), m = 0, \dots, r-1$:

$$\widehat{f}_n^{(m)}(t) = \frac{1}{n} \sum_{i=1}^n g_n^{(m)}(t - X_i). \quad (2.1)$$

Les propriétés statistiques de l'estimateur $\widehat{f}_n^{(m)}(t)$ sont simplement dérivées car il correspond à une moyenne de copies *i.i.d.* d'une variable aléatoire.

$$g_n^{(m)}(t-X) = \frac{1}{h_n^{m+1}} K^{(m)}\left(\frac{t-X}{h_n}\right). \quad (2.2)$$

Nous énonçons deux lemmes concernant respectivement le biais et la variance de l'estimateur $\widehat{f}_n^{(m)}(t)$.

Lemme 2.3.1 *Si t est un réel et m est un entier compris entre 0 et $r-1$, alors sous les hypothèses (H1)–(H2), nous avons : $E\widehat{f}_n^{(m)}(t) = f^{(m)}(t) + O(h_n)$*

Lemme 2.3.2 *Soient $t \in R$ et m un entier compris entre 0 et $r-1$. Si (H1) et (H2) sont satisfaites, alors*

$$\text{var } \widehat{f}_n^{(m)}(t) = \frac{1}{nh_n^{2m+1}}(b_m(t) + O(h_n)) - \frac{(f^{(m)}(t) + O(h_n))^2}{n}. \quad (2.4)$$

$$\text{avec } b_m(t) = f(t) \int_{-1}^1 K^{(m)}(u)^2 du. \quad (2.5)$$

Si $m < r-1$, nous distinguons à quel point l'avantage de l'estimation polynomiale locale est bénéfique. En effet, l'ordre du biais dans (2.3) est supérieur à celui dans (2.6). Les expressions (2.5) et (2.18) intervenant dans la variance des estimateurs ne sont pas tout-à-fait similaires. La première dépend d'un noyau $k^{(m)}(\cdot)$ et dans la seconde, le terme $\int_{-1}^1 K^{(m+1)}(u) du$ (qui représente un lissage du noyau $K_r^{(m+1)}$) joue son rôle. Rencontrer $K_r^{(m+1)}$ pour estimer la dérivée m^e de la densité ne doit pas paraître déroutant puisque le noyau $K_r^{(m)}$ est utilisé pour l'estimation de la dérivée m^e de $\Phi(\cdot, F) = F(\cdot)$. Nous pouvons maintenant prouver deux résultats sur les propriétés asymptotiques de l'estimateur lorsque $n \rightarrow +\infty$. Le premier concerne la consistance.

Théorème 2.3.1 *Soient un réel t et un entier m compris entre 0 et $r-1$ et supposons les hypothèses (H1) – (H2) satisfaites. Si lorsque $n \rightarrow +\infty$,*

$$h_n \rightarrow 0 \quad (2.6)$$

$$\text{suffisamment lentement tel que } nh_n^{2m+1} \rightarrow +\infty. \quad (2.7)$$

alors l'estimateur $\widehat{f}_n^{(m)}(t)$ est convergent en moyenne quadratique

$$E \left[\widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right]^2 \rightarrow 0. \quad (2.8)$$

donc également convergent en probabilité,
$$\widehat{f}_n^{(m)}(t) \xrightarrow{P} f^{(m)}(t). \quad (2.9)$$

Nous devrions chercher des conditions plus faibles que (2.7) pour la convergence en probabilité, et pour cela il faudrait consulter . En quête d'un résultat sur la normalité asymptotique, le point de départ naturel est la différence normalisée entre l'estimateur et la valeur à estimer .

$$\frac{\widehat{f}_n^{(m)}(t) - f^{(m)}(t)}{\sqrt{\text{Var} \widehat{f}_n^{(m)}(t)}} = Zn + \delta n. \quad (2.10)$$

$$Zn = \frac{\widehat{f}_n^{(m)}(t) - E\widehat{f}_n^{(m)}(t)}{\sqrt{\text{Var} \widehat{f}_n^{(m)}(t)}}. \quad (2.11)$$

est une variable aléatoire normalisée se prêtant aux théorèmes de la limite centrale, et

$$\delta n = \frac{E\widehat{f}_n^{(m)}(t) - f^{(m)}(t)}{\sqrt{\text{Var} \widehat{f}_n^{(m)}(t)}}. \quad (2.12)$$

est un terme de biais qui peut tendre vers 0 avec un choix de fenêtre approprié.

Théorème 2.3.2 *Posons $t \in R$ et m un entier tel que $0 \leq m \leq r - 1$ et admettons les hypothèses (H1) – (H2).*

Si lorsque $n \rightarrow +\infty$, suffisamment lentement tel que $nh_n \rightarrow 0$ (2.13)

et suffisamment vite tel que

$$nh_n^{2m+3} \rightarrow 0 \quad (2.14)$$

avec

$$\sqrt{\frac{nh_n^{2m+1}}{b_m(t)}} \left(\widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right) \xrightarrow{d} N(0, 1). \quad (2.15)$$

avec $b_m(t)$ donné par (2.5).

Corollaire 2.3.1 *(Au théorème 2.3.2)*

Un intervalle de confiance asymptotique pour $f^{(m)}(t)$ au niveau de confiance $1 - e$ est donné par les bornes

$$\widehat{f}_n^{(m)}(t) \pm z_{\frac{1-e}{2}} \sqrt{\frac{b_m(t)}{nh_n^{2m+1}}}. \quad (2.16)$$

où z_{1-e} est le quantile d'ordre $1 - e$ de la loi normale centrée réduite. Sa longueur tend vers 0 si h_n tend vers 0 suffisamment lentement tel que la condition de consistance (2.7) est satisfaite :

$$nh_n^{2m+1} \rightarrow +\infty.$$

Les conditions de consistance (2.6) et (2.7) sont intuitivement acceptables : la première est la même pour tout m , alors que la seconde est plus restrictive quand m augmente.

Théorème 2.3.3 *Soient x un réel fixé et m un entier compris entre 0 et $r - 1$. Sous (H3l) – (H3) et si lorsque $n \rightarrow +\infty$, $h_n \rightarrow +\infty$ et suffisamment lentement tel que $h_n^{2m+1} \rightarrow +\infty$*

Les conditions (2.13) et (2.14) pour la normalité asymptotique sont moins intuitives : la première estn identique pour tout m , alors que la seconde devient moins stricte quand m augmente. Ces observations servent à illustrer que les deux modes de convergence en moyenne quadratique et en loi sont très déférents. Le Corollaire résout ce puzzle en indiquant que la normalité asymptotique est utile seulement si la condition de consistance est satisfaite.

Notons que les conditions de consistance et de normalité asymptotique dépendent de m donc si l'on s'intéresse à l'estimation conjointe des dérivées, on ne peut appliquer la même fenêtre pour tous les ordres de dérivation. Si $m < r - 1$, alors la condition $nh_n^{2r+1} \rightarrow 0$. est plus souple que la condition (2.14). Cela est dû à la réduction de biais, qui est l'avantage de la méthode d'approximation locale polynomiale par rapport à la méthode de convolution. Mais nous expliquons dans la section suivante comment modifier le cadre afin d'améliorer la condition (2.14).

2.4 Amélioration à l'aide de noyaux réducteurs de biais :

Soit d un entier non nul et posons a présent deux nouvelles hypothèses :

Hypothèse (H1')

F est $(m + d + 1)$ fois continument différentiable en tout point t .

Hypothèse (H3)

k est un noyau d'ordre d , c'est-à-dire tel que

$$(i) \int u^i K(u) du = 0 \text{ pour tout } i = 1, \dots, d-1$$

$$\text{et } (ii) \int \mathbf{u}^d K(u) du = 1.$$

L'ordre d'un noyau est toujours pair, donc d'ordre 2 ou plus, c'est pourquoi d désigne par la suite un nombre *pair*. Notons que les noyaux positifs symétriques sont d'ordre 2. Quant aux noyaux d'ordre supérieur à 2, ils possèdent des parties négatives et ne sont pas des densités de probabilité.

Le lemme suivant montre que l'ordre d'un noyau détermine sa capacité à réduire le biais de l'estimateur.

Lemme 2.4.1 *Si t est un réel et m est un entier compris entre 0 et $r-1$, alors sous les hypothèses (H1')–(H2)–(H3), nous avons*
$$E \left[\widehat{f}_n^{(m)}(t) \right] = f^{(m)}(t) + O(h_n^d), \quad (2.17)$$

L'estimateur est asymptotiquement sans biais et le biais est d'ordre h_n^d , ce qui améliore le résultat du lemme

2.3.2 puisque $d \geq 2$. Le **théorème 2.3.2** peut être ainsi renforcé comme suit.

Théorème 2.4.1 *Soient t un réel et m un entier puis supposons que les hypothèses (H1'); (H2); (H3) sont satisfaites.*

$$\text{Si } h_n \rightarrow 0 \text{ suffisamment lentement tel que } nh_n \rightarrow +\infty \quad (2.18)$$

et suffisamment vite tel que :

$$nh_n^{2(m+d)} + 1 \rightarrow 0 \quad (2.19)$$

avec

$$\sqrt{\frac{nh_n^{2(m+d)}}{b_m(t)}} \left(\widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right) \xrightarrow{e} N(0, 1). \quad (2.20)$$

avec $b_m(t)$ donné par (2.5). Si $d \geq 2$, la condition (2.19) est moins stricte que (2.14) tandis que les conditions (2.13) et (2.18) sont similaires.

2.5 Application à une méthode de correction des effets de bord :

Soient m, p et r des entiers, a, b, c et h des réels tels que $a \leq b$ et $h > 0$. Dans le cadre de l'estimation de la densité, les effets de bord surviennent lorsque f a un support compact $[a, b]$. Les fonctions $f^{(m)}(b)$ représentent des dérivées à gauche et on suppose qu'elles existent. Avec une fenêtre donnée $h < \frac{b-a}{2}$, les estimateurs $f_n^{(m)}(t)$ sont uniquement définis sur l'intervalle $[a+h, b-h]$. Pour $t \in [a, a+h] \cup [b-h, b]$, ils doivent donc être modifiés. La méthode standard de correction des effets de bord pour l'estimation de la densité est l'application d'un noyau ayant un support tronqué, par exemple $[-1, \frac{b-t}{h}]$ pour h fixé et $t \in [b-h, b]$. Cependant, une méthode alternative basée sur un développement de Taylor peut être proposée, à condition que des estimateurs soient connus pour les dérivées successives de la densité f . Considérons l'estimation de $f^{(m)}(b)$, pour $m < r-1$. Grâce à la formule de Taylor-Lagrange de $f^{(m)}$ à l'ordre $p \leq r-m$ en un point $b-c$ tel que $0 < c < b-a$, nous écrivons :

$$f^{(m)}(b) = \sum_{j=0}^r \frac{1}{j!} f^{(m+j)}(b-c)c^j + O(c^{p+1}) \quad (2.21)$$

où

$$O(c^{p+1}) = \frac{1}{(p+1)!} f^{(m+p+1)}(b-c')c^{p+1}$$

avec c' compris entre 0 et c . Un estimateur naturel de $f^{(m)}(b)$ est obtenu en injectant les estimateurs à noyau standards (2.1) des dérivées $f^{(m+j)}(b-c)$ apparaissant dans la somme à droite de (2.21) :

$$\widehat{f}_n^{(m)}(b) = \sum_{j=0}^p \frac{1}{j!} \widehat{f}_n^{(m+j)}(b-c)c^j, \quad (2.22)$$

Pourvu que les deux hypothèses suivantes soient satisfaites :

Hypothèse (H1'')

F est $(m+p+2)$ fois continûment différentiable en tout point t .

Hypothèse (H2')

(i) K est $(m+p)$ fois continûment différentiable.

(ii) Le support de $K^{(m+j)}$ est $[-1, 1]$ pour $j = 0, \dots, p$,

(iii) $\sup_{u \in [-1, 1]} |K^{(m+j)}(u)| < +\infty$ pour $j = 0, \dots, p$. Alors l'estimateur $f_n^{(m)}(b)$ est proche de avec un terme d'erreur d'ordre $O(C_p + 1)$. pour n assez grand, d'après la consistance des estimateurs $\widehat{f}_n^{(m+j)}(b - c)$ En quête de résultats asymptotiques plus raffinés, nous remplaçons la constante c par une suite (C_n) convergeant vers 0. Il est désormais nécessaire d'expliquer comment les fenêtres a l'origine des estimateurs $\widehat{f}_n^{(m+j)}$ sont déterminées. Rappelant le **théorème 2.3.1**, le **théorème 2.3.2** ainsi que son Corollaire, un choix naturel de fenêtre intervenant dans l'expression des estimateurs $\widehat{f}_n^{(m+j)}$ est

$$h_{m+j,n} = n^{-\frac{1}{2(m+j+1)}} h, \quad (2.23)$$

où $h > 0$ est une constante choisie convenablement. En effet, nous avons $h_{m+j,n} \rightarrow 0$ et $nh_{m+j,n}^{2(m+j)+1} \rightarrow +\infty$, assurant la consistance, et $nh_{m+j,n}^{2(m+j)+3} \rightarrow 0$, garantissant la normalité asymptotique et le rétrécissement des intervalles de confiance lorsque la taille de l'échantillon n croît. Insérons $c = c_n$ dans (2.21) et (2.22), alors :

$$\begin{aligned} \widehat{f}_n^{(m)}(b) - f^{(m)}(b) &= \sum_{j=0}^p \frac{1}{j!} \left(\widehat{f}_n^{(m+j)}(b - c_n) - f^{(m+j)}(b - c_n) \right) c_n^j + O(c_n^{p+1}) \\ &= \sum_{j=0}^p b_{m+j} \frac{\sqrt{(b-c_n)}}{j!} a_{j,n} Z_{m+j,n} + O(c_n^{p+1}), \end{aligned} \quad (2.24)$$

où

$$a_{j,n} = (nh_{m+j,n}^{2(m+j)+1})^{-\frac{1}{2}} c_n^j. \quad (2.25)$$

et

$$z_{m+j,n} = \sqrt{\frac{nh_{m+j,n}^{2(m+j)+1}}{b_{m+j}(b-c_n)}} (\widehat{f}_n^{(m+j)}(b - c_n) - f^{(m+j)}(b - c_n)). \quad (2.26)$$

La décomposition (2.24) est primordiale pour démontrer la consistance de $\widehat{f}_n^{(m)}(b)$ l'estimateur corrige des effets de bord des dérivées d'ordre supérieur de la densité..

Théorème 2.5.1 *Si les hypothèses (H1'') et (H2I) sont vérifiées alors l'estimateur $\widehat{f}_n^{(m)}(b)$*

est convergent en probabilité $\widehat{f}_n^{(m)}(b) \xrightarrow{P} f^{(m)}(b), \quad (2.27)$

avec le choix de fenêtre (2.23). Or la suite (C_n) doit être telle que C_n soit le plus petit possible, et puisque

$h_{m+j,n}$, n augmente avec j , nous devrions prendre

$$c_n = h_{m,n} = n^{-\frac{1}{2(m+1)}} h. \quad (2.28)$$

lorsque nous effectuons le développement de Taylor jusqu'à l'ordre p . Avec un tel choix pour C_n , nous établissons la normalité asymptotique de l'estimateur $\widehat{f}_n^{(m)}(b)$.

Théorème 2.5.2 *Sous les hypothèses (H1'') et (H2'), si la suite (C_n) vérifie (2.28), le choix de fenêtre (2.23)*

conduit à :

$$\frac{p!}{\sqrt{b_{m+p}(b)}} n^{\frac{1}{4}(\frac{1}{m+p+1} + \frac{2p}{m+1})} h^{m+\frac{1}{2}} (\widehat{f}_n^{(m)}(b) - f^{(m)}(b)) \xrightarrow{L} N(0, 1). \quad (2.29)$$

Avec

$$b_{m+p}(b) = f(b) \int_{-1}^1 K^{(m+p)}(u)^2 du. \quad (2.30)$$

Pour $p = 0$, le résultat est conforme à celui obtenu dans le Théorème [2.3.2](#).

$$h_n = n^{-\frac{1}{2(m+1)}} h.$$

Corollaire 2.5.1 *(Au théorème [2.4.1](#))*

Un intervalle de confiance pour au niveau de confiance $1 - e$ est donné par les bornes

$$\widehat{f}_n^{(m)}(b) \pm z_{1-e/2} \frac{\sqrt{b_{m+p}(b)}}{p! h^{m+\frac{1}{2}} n^{\frac{1}{4}(\frac{1}{m+p+1} + \frac{2p}{m+1})}}. \quad (1.31)$$

Le résultat est valable seulement si $b_{m+p}(b) > 0$. La longueur de l'intervalle de confiance est égale à $2 \sqrt{b_{m+p}(b)}$ fois un facteur qui décroît avec p : l'exposant $\frac{1}{4} \frac{1}{m+p+1} + \frac{2p}{m+1}$. augmente avec p , il est supérieur à $\frac{1}{4(m+1)}$ (pour $p = 0$) et converge vers $+\infty$ (lorsque $p \rightarrow +\infty$).

2.6 Preuves :

Remarque 2.6.1 *Dans toutes les preuves, t représente un réel et m correspond à un entier compris entre 0 et $r - 1$. v est un nombre pair supérieur ou égal à 2.*

2.6.1 Preuve du Lemme [2.3.1](#) :

Nous avons mis en lumière précédemment le fait que l'estimateur $\widehat{f}_n^{(m)}(t)$ est une moyenne de copies *i.i.d.* d'une variable aléatoire

$$g_n^{(m)}(t - X) = \frac{1}{h_n^{m+1}} k^{(m)}\left(\frac{t-X}{h_n}\right)$$

Ainsi, l'espérance de l'estimateur $f_n^{(m)}(t)$ est égale au premier moment de $g_n^{(m)}(t - X)$, que nous calculons plus bas.

$$E \left[g_n^{(m)}(t - X) \right] = \int g_n^{(m)}(t - s) f(s) ds = \int f^{(m)}(t - s) g_n(s) ds,$$

d'après l'identité entre (C.3) et (C.4). En utilisant un développement de Taylor-Lagrange avec s' compris entre 0 et s , nous obtenons

$$\begin{aligned} E \left[g_n^{(m)}(t - X) \right] &= \int_{-h_n}^{h_n} (f^{(m)}(t) - f^{(m+1)}(t - s')s) g_n(s) ds \\ &= f^{(m)}(t) + O(h_n) \end{aligned} \quad (2.32)$$

$$|O(h_n)| \leq 2h_n \sup\{|f^{(m+1)}(t-s)|; s \in [-h_n, h_n]\}. \quad (2.33)$$

D'où

$$E f_n^{(m)}(t) = f^{(m)}(t) + O(h_n). \quad (2.34)$$

2.6.2 Preuve du Lemme 2.3.2 :

La formule (2.1) utilisée conjointement avec l'indépendance et l'équidistribution des Variables $g_n^{(m)}(t - X)$ entraîne que

$$\text{Var} \widehat{f}_n^{(m)}(t) = \text{Var} \frac{g_n^{(m)}(t-X)}{n}. \quad (2.35)$$

Déterminer le second moment de $g_n^{(m)}(t - X)$ s'avère donc crucial pour disposer d'une expression de la variance exacte asymptotique de l'estimateur $\widehat{f}_n^{(m)}(t)$. En appliquant successivement la définition de l'espérance, le changement de variable $u = \frac{s-t}{h_n}$ puis la formule de Taylor-Lagrange pour u' compris entre 0 et $h_n u$, avec $u \in [-1, 1]$, on a :

$$\begin{aligned} E \left[g_n^{(m)}(t - X) \right]^2 &= E \left[\frac{1}{h_n^{m+1}} k^{(m)}\left(\frac{t-X}{h_n}\right) \right]^2 = \frac{1}{h_n^{2m+1}} \int \frac{1}{h_n} k^{(m)}\left(\frac{t-s}{h_n}\right)^2 f(s) ds \\ &= \frac{1}{h_n^{2m+1}} \int_{-1}^1 k^{(m)}(-u)^2 f(t + h_n u) du \\ &= \frac{1}{h_n^{2m+1}} \left(f(t) \int_{-1}^1 k^{(m)}(u)^2 du + O(h_n) \right). \end{aligned}$$

Grâce à (2.32), il vient que :

$$\text{Var} \left[g_n^{(m)}(t - X) \right] = \frac{1}{h_n^{2m+1}} (b_m(t) + O(h_n)) - (f^{(m)}(t) + O(h_n))^2, \quad (2.36)$$

avec

$$b_m(t) = f(t) \int_{-1}^1 k^{(m)}(u)^2 du . \quad (2.37)$$

En injectant (2.36) dans (2.35), nous obtenons

$$Var \widehat{f}_n^{(m)}(t) = \frac{1}{nh_n^{2m+1}} (b_m(t) + O(h_n)) - \frac{f^{(m)}(t) + O(h_n)^2}{n} . \quad (2.38)$$

2.6.3 Preuve du Théorème 2.3.1 :

L'erreur quadratique moyenne de $\widehat{f}_n^{(m)}(t)$ est

$$\begin{aligned} E \left[\widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right]^2 &= Var \widehat{f}_n^{(m)}(t) + [E \widehat{f}_n^{(m)}(t) - f^{(m)}(t)]^2 \\ &= \frac{1}{nh_n^{2m+1}} (b_m(t) + O(h_n)) - \frac{(f^{(m)}(t) + O(h_n))^2}{n} + O(h_n^2) \end{aligned}$$

Par (2.34), le second terme est $O(h_n^2)$, qui tend vers 0 sous la condition (2.6). Le premier terme est donné par (2.38), qui tend vers 0 si la condition (2.7) est satisfaite.

Cela prouve (2.8), et ensuite (2.9) découle de l'inégalité de Markov 3.5.7 pour $p = 2$:

$$P \left[\left| \widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right| > \varepsilon \right] \leq \frac{IE \left[\widehat{f}_n^{(m)}(t) - f^{(m)}(t) \right]^2}{\varepsilon^2} .$$

2.6.4 Preuve du Théorème 2.3.2 :

On remarque la décomposition suivante :

$$\sqrt{\frac{nh_n^{2m+1}}{b_m(t)}} (\widehat{f}_n^{(m)}(t) - f^{(m)}(t)) = \sqrt{\frac{Var \widehat{f}_n^{(m)}(t)}{\frac{b_m(t)}{nh_n^{2m+1}}}} (Z_n + \delta_n), \quad (2.39)$$

avec Z_n et δ_n donnés respectivement par (2.11) et (2.12). Analysons d'abord le terme Z_n , candidat pour être asymptotiquement normal. En insérant les expressions (2.1), (2.2), et (2.4), on a

$$Z_n = \frac{\frac{1}{nh_n^{m+1}} \sum_{i=1}^n (k^{(m)}(\frac{t-X_i}{h_n}) - Ek^{(m)}(\frac{t-X_i}{h_n}))}{\sqrt{\frac{1}{nh_n^{2m+1}} (b_m(t) + O(h_n)) - \frac{(f^{(m)}(t) + O(h_n))^2}{n}}} .$$

Nous pensons immédiatement à utiliser le théorème de la limite centrale de Lindesberg-Feller 3.5.7 avec $k_n = n$ et $Y_{n,i}$, des copies i.i.d. de

$$Y_n = \frac{k^{(m)}(\frac{t-X}{h_n})}{\sqrt{nh_n(b_m(t) + O(h_n))}} .$$

Comme $Var Z_n = 1$ pour tout n , on en déduit que $\lim_{n \rightarrow +\infty} Var \widetilde{Z}_n = 1$ est satisfaite. Le

noyau $k^{(m)}$ est borne selon (H2(iii)) donc, si on suppose (2.13), alors la condition de Lindesberg-Feler (iii) $\forall \varepsilon > 0, nE[Y_n^2 1[|Y_n| > \varepsilon]] \rightarrow 0$

est vérifiée. Ceci entraîne que \tilde{Z}_n suit une loi normale centrée réduite, donc y compris Z_n Concentrons-nous ensuite sur le terme de biais (2.12). En injectant (2.3) et (2.4), nous obtenons :

$$\delta_n = \frac{O(h_n)}{\sqrt{1nh_n^{2m+1}(b_m(t)+O(h_n)) - \frac{(f^{(m)}(t)+O(h_n))^2}{n}}} = O(\sqrt{nh_n^{2m+3}}),$$

qui converge vers 0 sous la condition (2.14). Finalement, observons que

$$\frac{Var \hat{f}_n^{(m)}(t)}{b_m(t)nh_n^{2m+1}} P \rightarrow 1 .$$

2.6.5 Preuve du Lemme 2.4.1 :

L'hypothèse (H1') nous permet d'effectuer un développement de Taylor de $f^{(m)}$ à l'ordre $d - 1$ au point t et par suite, nous pouvons remplacer (2.32) par

$$\int_{-h_n}^{h_n} \left(\sum_{j=0}^{d-1} f^{(m+j)}(t) \frac{(-s)^j}{j!} + f^{(m+d)}(t - st) \frac{s^d}{d!} \right) g_n(s) ds$$

pour s' compris entre 0 et s et obtenir

$$E \hat{f}_n^{(m)}(t) = f^{(m)}(t) + O(h_n^d). \tag{2.40}$$

2.6.6 Preuve du Théorème 2.3.3 :

Pour démontrer le **théorème 2.3.3**, on s'inspire de la preuve du **théorème 2.3.2**. Toutes les étapes doivent être reproduites à l'identique, exceptée celle concernant le calcul du terme de biais dans (2.12). Grâce à (2.40), celui-ci a pour expression :

$$\delta_n = \frac{O(h_n^v)}{\frac{1}{nh_n^{2m+1}}(b_m(t)+O(h_n))} = O(\sqrt{nh_n^{2(m+v)+1}}),$$

et converge vers 0 lorsque $n \rightarrow +\infty$ sous la condition (2.19).

2.6.7 Preuve du Théorème 2.4.1 :

Reprenons l'expression (2.24)

$$\tilde{f}_n^{(m)}(b) - f^{(m)}(b) = \sum_{j=0}^p \frac{\sqrt{b_{m+j}(b-c_n)}}{j!} a_{j,n} Z_{m+j,n} + O(c_n^{p+1}), \quad (2.41)$$

avec $a_{j,n}$ et $Z_{m+j,n}$ données respectivement par (2.25) et (2.26). Par inspection de la preuve du Théorème 2.3.2, avec $b - c_n$ dans le rôle de t , il vient que :

$$Z_{m+j,n} \xrightarrow{L} N(0, 1) \quad (2.42)$$

pour tout $j = 0, \dots, p$.

Puisque $nh_{m+j,n}^{2(m+j)+1} \rightarrow +\infty$, $cn \rightarrow 0$, et $b_{m+j}(b - c_n) \rightarrow b_{m+j}(b)$, les règles de Cramér 3.5.7 entraînent que

$\tilde{f}_n^{(m)}(b)$ est un estimateur consistant de $f^{(m)}(b)$.

Preuve du Théorème.5 :

Dans le but d'obtenir un résultat sur la distribution asymptotique de $\tilde{f}_n^{(m)}(b)$ nous avons besoin d'identifier le terme $a_{j,n}$ dominant dans l'expression à droite de (2.24). Pour cela, développons le coefficient $a_{j,n}$ donné par (2.25) à l'aide de (2.23) et (2.28),

$$\begin{aligned} a_{j,n} &= \left(n \left(n^{-\frac{1}{2(m+j+1)}} h \right)^{2(m+j)+1} \right)^{-\frac{1}{2}} \left(n^{-\frac{1}{2(m+1)}} h \right)^j \\ &= n^{-\frac{1}{4} \left(\frac{1}{m+j+1} + \frac{2j}{m+1} \right)} h^{-(m+\frac{1}{2})}. \end{aligned} \quad (2.43)$$

Le terme $a_{j,n}$ dominant étant celui de plus petit ordre, nous sommes amenés à déterminer

$$\operatorname{argmax}_{j=0, \dots, p} g(j),$$

où

$$g : \mathbb{R}^+ \rightarrow \mathbb{R} \text{ est la fonction définie par } g(x) = \frac{1}{m+x+1} + \frac{2x}{m+1}.$$

Puisque

$$g''(x) = \frac{2}{(m+x+1)^3} > 0,$$

la fonction g est *convexe*, donc elle atteint son maximum en $x = 0$ ou en $x = p$.

Le dernier cas se produit si $g(0) < g(p)$, c'est-à-dire si $0 < \frac{1}{m+p+1} + \frac{2p-1}{m+1}$,

ce qui équivaut à $0 < (2p + 2m + 1)p$. Cette inégalité est toujours vérifiée donc le terme $a_{j,n}$ dominant dans la somme à droite de (2.24) est le p ème terme $a_{p,n}$. Une conséquence directe est :

$$\frac{a_{j,n}}{a_{p,n}} \rightarrow 0 \quad \text{pour } j = 1, \dots, p-1. \quad (2.44)$$

De plus, avec (2.28) et (2.43), on a :

$$\begin{aligned} \frac{c_n^{p+1}}{a_{p,n}} &= \left(n^{-\frac{1}{2(m+1)}} h \right)^{p+1} n^{\frac{1}{4}} \left(\frac{1}{m+p+1} + \frac{2p}{m+1} \right) h^{m+12} \\ &= n - \frac{1}{4} \left(\frac{2}{m+1} - \frac{1}{m+p+1} \right) h^{m+p+\frac{3}{2}}, \end{aligned}$$

d'où

$$\frac{c_n^{p+1}}{a_{p,n}} \rightarrow 0. \quad (2.45)$$

En effet,

$$\frac{2}{m+1} - \frac{1}{m+p+1} = \frac{2p+m+1}{(m+1)(m+p+1)} > 0.$$

En multipliant par $\frac{p!}{\sqrt{b_{m+p}(b)a_{p,n}}}$

dans (2.24), il vient que :

$$\frac{p!}{\sqrt{b_{m+p}(b)a_{p,n}}} \left(\widehat{f}_n^{(m)}(b) - f^{(m)}(b) \right) = \sum_{j=0}^p \frac{p!}{\sqrt{b_{m+p}(b)}} \frac{b_{m+j}(b-c_n)}{j!} \frac{a_{j,n}}{a_{p,n}} Z_{m+j,n} + O\left(\frac{c_n^{p+1}}{a_{p,n}}\right)$$

donc d'après (2.44) et (2.45) et la règle de Cramér 3.5.7 (2) $\frac{p!}{\sqrt{b_{m+p}(b)a_{p,n}}} \left(\widehat{f}_n^{(m)}(b) - f^{(m)}(b) \right)$ a la même loi asymptotique que $\frac{\sqrt{b_{m+p}(b-c_n)}}{\sqrt{b_{m+p}(b)}} Z_{m+p,n}$.

Faisons tendre n vers $+\infty$ dans cette dernière expression puis utilisons le fait que $c_n \rightarrow 0$,

$b_{m+p}(b-c_n) \rightarrow b_{m+p}(b)$ et (2.42). La règle de Cramér 3.5.7 (ii) puis la forme de $a_{p,n}$

obtenue à l'aide de (4.43) pour $j = p$ nous permet d'arriver à

$$\frac{p!}{\sqrt{b_{m+p}(b)}} n^{\frac{1}{4}} \left(\frac{1}{m+p+1} + \frac{2p}{m+1} \right) h^{m+\frac{1}{2}} \left(\widehat{f}_n^{(m)}(b) - f^{(m)}(b) \right) \xrightarrow{L} N(0, 1), \quad (2.46)$$

avec

$$b_{m+p}(b) = f(b) \int_{-1}^1 k^{(m+p)}(u)^2 du. \quad (2.47)$$

2.7 Conclusion :

Ce chapitre s'est concentré sur l'estimation à noyau classique de la densité et de ses dérivées successives. Afin de réduire le biais de ces estimateurs, nous avons montré l'avantage d'employer des noyaux d'ordre supérieur. La convergence en probabilité et la normalité asymptotique de ces nouveaux estimateurs ont été prouvées. Ce chapitre a ensuite motivé

l'estimation à noyau des dérivées successives de la densité (ou d'une autre fonction) en montrant qu'elle est utile pour remédier aux effets de bord de l'estimateur de toute dérivée de la densité (ou d'une autre fonction). Nous avons développé une méthode de correction des effets de bord apparaissant notamment lorsque le support d'estimation est compact et qui pourrait être appliquée en analyse de survie.

Chapitre 3

Simulation

3.1 Un estimateur simple de la densité : l'histogramme

Supposons pour simplifier qu'on soit en dimension 1 et que les variables de l'échantillon soient à valeurs dans $[0, 1]$ donc $f : [0, 1] \rightarrow \mathbb{R}^+$.

On se donne un découpage de $[0, 1]$ en un certain π nombre de classes $]a_1, a_2], \dots,]a_p, a_{p+1}]$. Pour simplifier encore, on suppose que les classes sont de même longueur $a_{i+1} - a_i = a_i - a_{i-1}$. Cette longueur est notée h . Estimer f par la méthode de l'histogramme consiste simplement à estimer f par une fonction constante sur chaque classe, cette constante étant liée à la proportion de X_i tombant dans cette classe. Plus exactement on pose, pour

$$t \in]a_j, a_{j+1}], \quad \hat{f}_n(t) = \frac{1}{nh} \text{Card}\{i : X_i \in]a_j, a_{j+1}]\}$$

Pour voir très exactement d'où vient cette formule : on a, si f est égale à une constante c_j constante sur $]a_j, a_{j+1}]$,

$$F(a_{j+1}) - F(a_j) = \int_{a_j}^{a_{j+1}} f(t) dt = c_j h$$

La performance de cet estimateur dépend fortement du nombre de classes.

3.2 Code R et illustration graphique du choix du nombre de classes.

On va illustrer l'importance de bien choisir le nombre de classes par un exemple faisant intervenir une densité bimodale. On va pour cela simuler un mélange de deux lois gaussiennes : la densité simulée est

$$f(x) = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \left(\exp\left(-\frac{(x-2)^2}{2}\right) + \exp\left(-\frac{(x-6)^2}{2}\right) \right)$$

On devrait donc, si l'approximation par l'histogramme est bien faite, se retrouver avec deux "cloches" qui se chevauchent un petit peu (écart-type=1) et qui sont centrées en 2 et 6 respectivement.

Simulation d'un échantillon de taille n=500 de loi de densité f :

```
f=function(x){0.5*dnorm(x,mean=2)+0.5*dnorm(x,mean=6)}
```

```
sim=function(n){
```

```
  X=rnorm(n,2,1)
```

```
  Y=rnorm(n,6,1)
```

```
  ber=rbinom(n=n,size=1,prob=0.5)
```

```
  return(ber*X+(1-ber)*Y)}
```

```
Z=sim(500)
```

On estime la densité par un histogramme (on utilise ici la bibliothèque ggplot2) et on rajoute la vraie densité f en rouge :

```
library(ggplot2)
```

```
p<
```

```
p1<-p+ geom_histogram(aes(y=..density..),color="black",fill="white")+
```

```
stat_function(fun=f,col='red')+
```

```
labs(title="nb de classes= 30")
```

```
p1
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

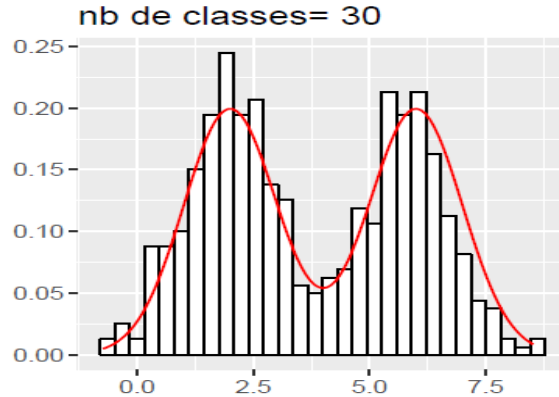


FIG. 3.1 – la densité par un histogramme (par bibliothèque ggplot2)

La fonction `histogram` dans `ggplot` calcule un histogramme avec 30 classes par défaut (ce qu'il signale d'ailleurs). Ce n'est donc pas la valeur optimale en général. Essayons avec d'autres valeurs du nombre de classes (=bins).

```
p1<-p+geom_histogram(aes(y=..density..),bins=3,color="black",fill="white")+stat_function(fun=f,
4,12))+labs(title="nb de classes = 3")
```

```
p2<-p+geom_histogram(aes(y=..density..),bins=10,color="black",fill="white")+stat_function(fun=f,
4,12))+labs(title="nb de classes = 10")
```

```
p3<-ggplot(data.frame(x=Z),aes(x))+geom_histogram(aes(y=..density..),bins=100,color="black",fill=
4,12))+labs(title="nb de classes = 100",x="",y="")
```

```
library(gridExtra)#pour faire apparaitre les trois figures en même temps
```

```
grid.arrange(p1,p2,p3,nrow=1)
```

On peut aussi indiquer le pas h (binwidth) plutôt que le nombre de classes (bins).

On constate donc que, avec une fenêtre h trop petite, c'est-à-dire avec un trop grand nombre de classes, on fait apparaître trop de variations souvent insignifiantes (variance trop grande). Au contraire avec une fenêtre h trop grande, on a une approche trop grossière (biais trop grand) et une distribution peu discriminante : en particulier ici on ne voit même plus qu'il s'agit d'une distribution bimodale. On voit qu'il faut trouver un compromis entre le biais (au carré) et la variance, compromis qu'on va illustrer plus en détail plus loin, par le calcul.

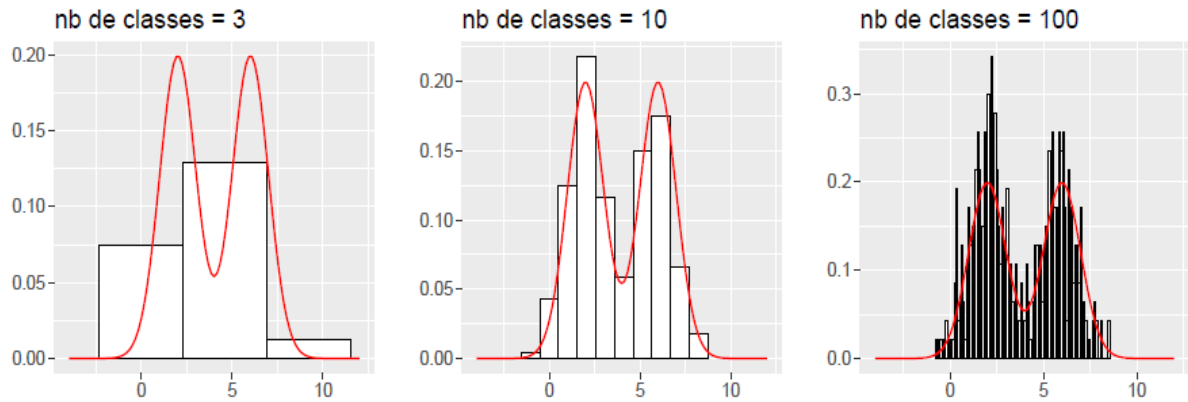


FIG. 3.2 – la densité par un histogramme (par bibliothèque ggplot2) tel' que nbr de classes=3/10/100

Il existe d'ailleurs dans R des estimations de la taille optimale du pas h , cf l'aide en ligne ou la page wikipedia sur l'histogramme. L'estimateur par histogramme étant présenté ici essentiellement à titre illustratif, nous ne donnons pas plus de détails sur le sujet. Des détails plus précis seront donnés pour l'estimateur qui nous intéresse vraiment : l'estimateur à noyau.

Evidemment le nombre optimal de classes dépend de n . Illustrons ceci en changeant la taille de l'échantillon : on passe de 500 à 50000.

```
Z=sim(50000)
```

```
p<-ggplot(data.frame(x=Z),aes(x))+labs(x="",y="")
```

```
p1<-p+geom_histogram(aes(y=..density..),color="black",fill="white")+stat_function(fun=f,col='red',
de classes=30")
```

```
p2<-ggplot(data.frame(x=Z),aes(x))+geom_histogram(aes(y=..density..),bins=100,color="black",fill=
labs(title="nb de classes=100")
```

```
grid.arrange(p1,p2,nrow=1)
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

On voit donc qu'avec un nombre de classes égal à 100, on a, contrairement à précédemment, un très bon choix.

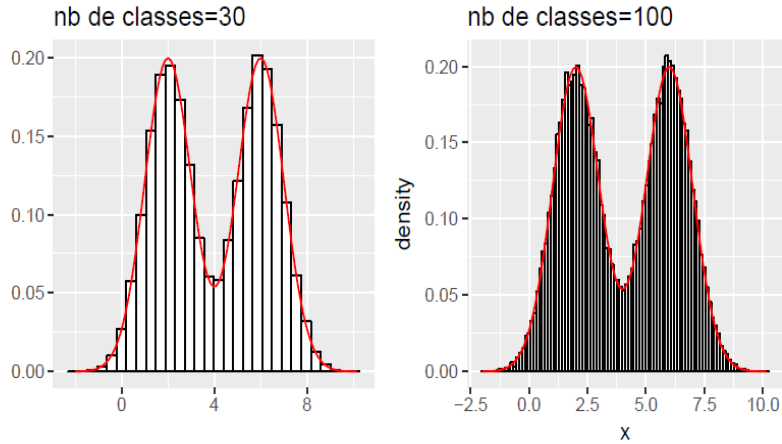


FIG. 3.3 – la densité par un histogramme (par bibliothèque ggplot2) tel' que nbr de classes=30/100

La taille optimale du nombre de classes est croissante avec n , autrement dit, le pas h optimal décroît avec n , ce que l'on va illustrer plus tard avec l'estimateur à noyau de fenêtre h .

Remarquez que l'on fait deux approximations successives : une première approximation quand on approche la densité par une fonction constante par morceaux, et ensuite une deuxième approximation quand on approche chaque constante à l'aide des données.

3.3 Estimateurs à noyaux

Un inconvénient de l'estimateur par histogramme précédent est que la fonction de densité résultante \hat{f}_n n'est pas régulière : il s'agit d'une fonction constante par morceau, qui a donc des sauts aux extrémités de chaque classe. En général, la densité à estimer est plus lisse, au moins continue. L'estimation par noyau a pour but de répondre à cet écueil.

Principe : Si f est continue en x (ce qui va être le cas pour les classes de fonctions qu'on va considérer) alors

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

L'idée est donc d'utiliser l'approximation suivante, pour h petit,

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

Pour estimer la densité f on peut donc passer par un estimateur \widehat{F}_n de la cdf F . Voyons ce qui se passe si on choisit comme estimateur la fonction de répartition empirique F_n .

(On rappelle que $F_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{X_i \leq x}$) On choisit un $h > 0$ petit pour que l'approximation ci-dessus soit valable, et on pose

$$\widehat{f}_n(x) = \frac{F(x+h) - F(x-h)}{2h} = \frac{1}{n} \sum_{i=1}^n \frac{1}{2h} 1_{X_i \in]x-h, x+h]}$$

K_0 est appelé le noyau de Rosenblatt. Cet estimateur a le même inconvénient d'irrégularité que l'estimateur par histogramme. On a donc l'idée d'utiliser des noyaux plus réguliers.

Définition 3.3.1 *Etant donné K un noyau et $h > 0$, on pose $\forall x \in \mathbb{R}, \widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{X_i - x}{h}\right)$*

Remarque 3.3.1 — *La plupart des noyaux sont symétriques, positifs et sont décroissants sur \mathbb{R}^+ comme le noyau Gaussien : plus y est proche de 0, plus $K(y)$ est grand. Donc, pour un $x \in \mathbb{R}$ donné, plus une observation X_i est proche de x , plus $K\left(\frac{X_i - x}{h}\right)$ est grand. Donc $\widehat{f}_n(x)$ est d'autant plus grand que x est proche de beaucoup d'observations X_i (somme de beaucoup de grandes valeurs $K\left(\frac{X_i - x}{h}\right)$).*

— L'estimateur est somme de fonctions $K(X_i - x/h)$ qui sont continues si K est continu. Donc \widehat{f}_n est continu si K est continu.

— $\int \widehat{f}_n(x) dx = 1$, donc, si $K(x) \geq 0, \forall x \in \mathbb{R}$, alors \widehat{f}_n est une densité.

— Le paramètre $h > 0$ est appelé fenêtre (bandwidth). C'est un paramètre de lissage : plus h est grand, plus l'estimateur est régulier. Comme dans le cas de l'estimateur à histogramme, le choix de h est délicat, la fenêtre h optimale devant réaliser un équilibre biais/variance (cf section suivante).

— Dans la pratique, le choix du noyau est peu influent, contrairement au choix de la fenêtre !

3.4 Illustration graphique et code R

On va utiliser le même exemple de distribution bimodale que précédemment. L'estimation par noyaux peut se faire avec différentes méthodes. On peut utiliser la fonction `density`

du package `stat`. Cette procédure n'estime que des densités à une seule variable. Pour des fonctions multivariées, on peut utiliser par exemple la fonction `kde` du package `ks` (de 1 à 6 variables).

Par défaut le noyau utilisé est le noyau gaussien, il est possible de changer de noyau avec l'option `kernel`. On va en fait utiliser la version de `ggplot` pour représenter l'estimateur à noyau. La fonction qui permet de dessiner l'estimateur à noyau est `geom_density`

Le paramètre représentant le fenêtre `h` s'appelle `bw` (comme *bandwidth*). On illustre l'influence du choix de la fenêtre. On tire les mêmes conclusions que pour l'histogramme.

```
p<-ggplot(data.frame(x=Z),aes(x))+labs(x="",y="")
p1<-p+geom_density(bw=0.1)+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("h=0.1")
p2<-p+geom_density(bw=0.5)+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("h=0.5")
p3<-p+geom_density(bw=0.8)+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("h=0.8")
p4<-p+geom_density(bw=1.2)+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("h=1.2")
grid.arrange(p1,p2,p3,p4,nrow=2,ncol=2)
```

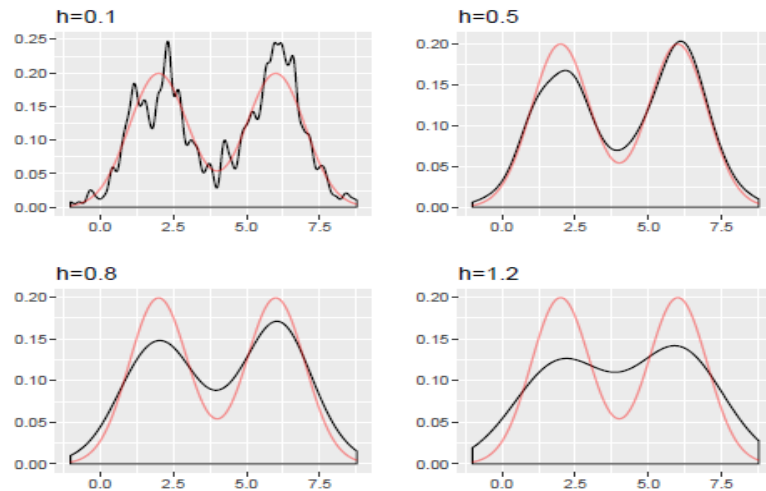


FIG. 3.4 – représentation de l'estimateur à noyau (par la la version de `ggplot`) pour $h=0.1/0.5/0.8/1.2$

Pour finir, on illustre le choix de deux fenêtres calculées à partir des données. L'une est la méthode de Sheather et Jones (SJ) et l'autre est basée sur la validation croisée, qui sera

vue en fin de chapitre (ucv=unbiased cross-validation). Pour d'autres méthodes, consultez la documentation liée à bw.

```
p<-ggplot(data.frame(x=Z),aes(x))+labs(x="",y="")
```

```
p5<-p+geom_density(bw="ucv")+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("ucv")
```

```
p6<-p+geom_density(bw="SJ")+stat_function(fun=f,col='red',alpha=0.4)+ggtitle("SJ")
```

Il existe une version en dimension 2 de cette fonction dans ggplot2 qui s'appelle `geom_density_2d`.

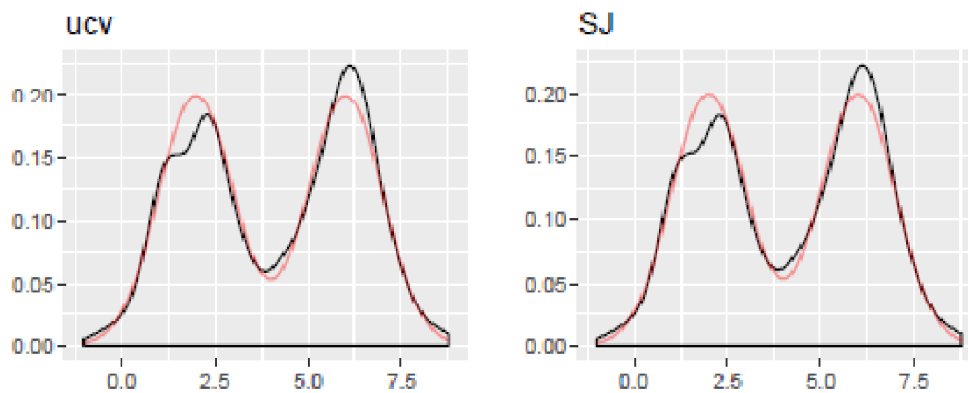


FIG. 3.5 – L'estimation de densité par la méthode de Sheather et Jones (SJ) et la méthode de la validation croisée

Il existe une version en dimension 2 de cette fonction dans ggplot2 qui s'appelle `geom_density_2d`

3.5 Codes des simulations :

3.5.1 Estimateur à noyau décomposé :

Tout estimateur à noyau est composé d'une somme de n fonctions, soit une pour chaque observation.

```
> x = c(-1.111, -0.257, 1.797, 2.163, -2.2264, -0.949)
```

```
> K = fonction(x,a) dnorm(x,a,1.09)/6
```

```
# noyau gaussien, n = 6, h = 1.09 (règle de Silverman)
```

```
> b = seq(-6,6,.01) # valeurs où l'on calcule K
> for (i in 1 :6) {plot(b,K(b,x[i]),type="l",ylab=" ",
xlim=c(-7,7),ylim=c(0,.06),col="blue");
par(new=T)}
#chaque observation détermine un graphique
#xlim, ylim permettent une super position parfaite des graphiques#
> title("Estimateur du noyau décomposé")
```

3.5.2 Fonction densité :

Forme générale de la fonction densité

```
density(x, bw = "nrd0", adjust = 1, kernel =c("gaussian", "epanechnikov", "rectangu-
lar","triangular", "biweight", "cosine", "optcosine"),
```

```
weights =NULL, window = kernel, width, give.Rkern = FALSE,
```

```
n =512, from, to, cut = 3, na.rm = FALSE, ...)
```

```
x #vecteur de données univariées
```

```
bw #paramètre de lissage
```

```
bw="nrd0" #règle de Silverman
```

```
bw="nrd" #règle de Scott
```

```
bw="ucv" #règle de la validation croisée,
```

```
bw="SJ-dpi" #règle de Sheather-Jones
```

```
kernel #type de noyau, par défaut "gaussian"
```

```
n=512 #n. de points équidistants où f est estimée
```

3.5.3 Sortie de densité :

Liste de longueur 7 dont les principales composantes sont :

```
[[1]] x #vecteur des points où f est estimée
```

[[2]] y #vecteur des valeurs de f estimées Par défaut n = 512 points ou valeurs. Si n > 512, le n. de points ou de valeurs est la puissance de 2 supérieure à 512.

[[3]] bw #paramètre de lissage h utilisé

[[4]] n #n. de points où b f est calculé (puissance de 2)

L'application de density ne produit aucun graphique. On obtient plutôt de l'information sur les valeurs estimées et sur h.

La fonction plot appliquée à l'objet créé par density produit le graphique.

3.5.4 Application de la méthode du noyau :

```
> par(mfrow=c(1,2))
```

```
> taux = scan("cdrate.dat")
```

Jeu des taux d'épargne. Bimodal : deux types d'institutions financières.

Paramètre de lissage par défaut

```
> bw.nrd0(taux) h = 0.1152792 (Silverman)
```

```
> bw.nrd(taux) h = 0.1357733 (Scott)
```

```
> bw.SJ(taux) h = 0.0687455 (Sheather-Jones)
```

```
> plot(density(taux, kernel="rectangular"), xlab="taux",  
ylab="Densité", main="noyau rectangulaire")
```

```
> plot(density(taux, kernel="gaussian"), xlab="taux",  
ylab="Densité", main="noyau normal")
```

#on utilise ici le paramètre de lissage par défaut (règle de Silverman)

3.5.5 Package sm :

*Fonction sm.density : > library(sm)

```
sm.density(x, h, model = "none", weights = NA, group=NA, ...) #estimation avec le noyau  
gaussien
```

*Produit le graphique de l'estimateur à *noyau gaussien*

*x #données uni- (vecteur), bi- ou tridimensionnelles (matrice, tableau); h #paramètre de lissage (*par défaut, règle de Scott*)

*Autres arguments intéressants (voir sm.options)

add=TRUE #ajoute à un graphique la densité estimée

display="se" #produit une bande de variabilité method #choix de h, ="normal" (Scott), "cv" (validation croisée),

"sj" (Sheather-Jones)

hmult #multiple du h, = 1 par défaut.

3.5.6 Variation du paramètre de lissage :

> library(sm)Jeu aircraft : 709 modèles d'avion du 20^e siècle,tableau à 8 variables, incluant Span et Period

> provide.data(aircraft)# commande propre à sm

> y = log(Span[Period==3])#var. Span 1956–1984

> par(mfrow=c(2,2))

sm.density(y,hmult=1/36,xlab="log(envergure)\n hm =1/36", ylab="densité")# *hmult*
facteur multipliant h optimal

sm.density(y,hmult=1/3,xlab="log(envergure)\n hm =1/3", ylab="densité")

sm.density(y,hmult=2,xlab="log(envergure)\n hm = 2",ylab="densité")

sm.density(y,hmult=6,xlab="log(envergure)\n hm =6",ylab="densité").

3.5.7 Comparaison de 3 densités :

> y1 = log(Span)[Period==1]#début du 20e siècle

> y2 = log(Span)[Period==2]#milieu du 20e siècle

> y3 = log(Span)[Period==3]#fin du 20e siècle

Superposition des graphiques

> sm.density(y3,xlab="log(envergure)",ylab="densité",col="red")

```
> sm.density(y2,add=T,lty=2,col="blue")
```

```
> sm.density(y1,add=T,lty=3,col="black")
```

Légende

```
> legend(3.15,1,c("1914-1935", "1936-1955", "1956-1984"),lty=3 :1)
```

#3 périodes du 20^e siècle couvertes par le jeu aircraft.

Conclusion

Dans ce travail, nous avons étudié une méthode non-paramétrique d'estimation de densité. Dans le premier chapitre, nous avons trouvé différentes définitions et critères d'estimation non paramétrique d'une densité, l'estimation à noyau et biais des estimateurs à noyau (Le risque quadratique ponctuel, contrôle du biais, variance des estimateurs à noyaux). Nous avons donné plusieurs exemples de noyaux. Ensuite, on a vu l'estimation par la méthode de l'histogramme et leurs propriétés. Nous avons étudié les propriétés de l'estimateur à noyau, le choix optimal du noyau et choix théorique optimal du paramètre de lissage.

Nous avons présenté l'état de l'art sur les estimateurs à noyaux continus (classiques) de Rosenblatt (1956) et Parzen (1962) de la densité univariée continue et ses propriétés statistiques, et des estimateurs à noyaux associés discrets de Kokonendji & Senga Kiessé (2011) et continus. Nous avons commencé par une définition pour les noyaux continus (classiques) qui jusqu'alors se déduit de celle de l'estimateur puis nous avons aussi amélioré la définition des noyaux associés discrets et continus. Nous avons introduit la forme de l'estimateur à noyau associé discret/continu multivarié. Enfin, nous illustrons certains estimateurs à noyaux continus symétriques à savoir le noyau d'Epanechnikov, le noyau Gaussien, le noyau Biweight et le noyau triangulaire. Nous simulons un échantillon de taille n de la loi normale centrée et réduite.

Le deuxième chapitre s'est concentré sur l'estimation à noyau classique de la densité et de ses dérivées successives. Afin de réduire le biais de ces estimateurs, nous avons montré l'avantage d'employer des noyaux d'ordre supérieur. La convergence en probabilité et

la normalité asymptotique de ces nouveaux estimateurs ont été prouvées. Ce chapitre a ensuite motivé l'estimation à noyau des dérivées successives de la densité en montrant qu'elle est utile pour remédier aux effets de bord de l'estimateur de toute dérivée de la densité . Nous avons développé une méthode de correction des effets de bord apparaissant notamment lorsque le support d'estimation est compact et qui pourrait être appliquée en analyse de survie.

Dans La troisième chapitre, nous avons essayé de simuler un estimateur simple de la densité par la méthode d'histogramme et par méthode d'estimateur à noyau.et nous avons donné l'illustration (graphique / code R) de chaque méthode.Depui,nous avons donné quelques Codes des simulations(Estimateur à noyau décomposé,Variation du paramètre de lissage,Comparaison de 3 densités,...etc).

Bibliographie

- [1] Adamowski K. Nonparametric kernel estimation of flood frequencies. 1985
- [2] Adamowski K. Plotting formula for flood frequency. 1981.
- [3] A. A. Borovkov., Mathematical statistics. Gordon and Breach Science Publishers, Amsterdam. 1998.
- [4] A. Sen, On Economic Inequality, Norton, New York, 1973.
- [5] A. B. Tsybakov. Introduction à l'estimation non-paramétrique, volume 41 of Mathématiques & Applications (Berlin) [Mathematics & Applications]. Springer-Verlag, Berlin, 2004.
- [6] Alexandre B. Tsybakov. Introduction to Nonparametric Estimation. Springer, New York ; London, 1st edition. 2nd printing. 2008 edition edition, November 2008.
- [7] Ann. Math. Statist., 27, 832-837, 1956.
- [8] Bahadur, R. R, A note on quantiles in large samples, Ann. Math. Stat., 37, 577-580. (1966).
- [9] Berkoun Youcef. statistiques inferentielle. Cours de 1ère année Master, U.M.M.T.O Tizi-ouzou, 2010.
- [10] [Bowman. A. W, An alternative method of cross-validation for the smoothing of density estimates. Biometrika, 71. 353-360. 1984.
- [11] Branko Miladinovic. Kernel density estimation of reliability with applications to extreme value distribution. 2008

- [12] Bowman, A. W, An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353-360.1984.
- [13] Bareche, A. & A ssani, D. (2008). Kernel density in the study of the strong stability of the $M=M=1$ queueing system. *Operations Research Letters*, 36, 535{538}.
- [14] Bowman, A. W. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* 71, 2 (1984), 353–360.
- [15] Burman, P. A date dependent approach to density estimation. *Zeitschrift Für Wahrscheinlichkeitstheorie and Verwandte Gebiete*, 69 (1985), 609–628.
- [16] B. W. Silverman, *Density estimation for statistics and data analysis*, London,Chapman and Hall, 1986.
- [17] Cao, R. Cuevas, A., and Gonzdez-Manteiga, W. A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis* 17(1) (1994), 153–176.
- [18] Coudret, R., Durrieu, G., and Saracco, J. Estimateurs a noyau bimodaux d'une densité bimodale et comparaison avec d'autres estimateurs non paramétriques. *Proc de la société Française de Statistique* (2012).
- [19] Cacoullous.T, Estimation of multivariate density. *Analns of the institue of statistical Mthématics* 18, 178.189. 1966.
- [20] Cybakov, A. B. *Introduction à l'estimation non paramétrique*, vol. 41. Springer Science & Business Media, 2003.
- [21] Chen, S.X. (1999) Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* 31, 131-145.
- [22] Deheuvels, P., and Hominal, P. Estimation non paramétrique de la densité comptetenu d'informations sur le support. *Revue de Statistique Appliquée*, 27, 47,68 (1979).
- [23] David, H.A, *Order Statistics*, 2nd Edition, New York : John Wiley. 1981.
- [24] Davidson.J, *Stochastic Limit Theory*. Oxford University Press, New York. 1994.

- [25] Duin R.P.W. On the choice of smoothing parameters for Parzen estimators of probability density function. 1976.
- [26] E. Parzen Nonparametric statistical data modeling, American Statistical Association, vol.74, pp.105-121. 1979.
- [27] Epanechnikov, V. Nonparametric estimation of a multidimensional probability density. *Teoriya Veroyatnostei i ee Primeneniya* 14, 1 (1969), 156–161.
- [28] Epanechnikov, V.A. (1969). Nonparametric estimation of a multivariate probability density, *Theory of Probability and Its Applications* 14, 153-158.
- [29] E. F. Schuster, Estimation of a probability density function and its derivatives, *Ann. Math. Statist.*, 40, 1187-1195, 1969.
- [30] E. F. Schuster, Incorporating support constraints into nonparametric estimators of densities, *Commun. Statist.*, 14, 1123-1126, 1985.
- [31] G. R. Shorack, J. A. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.
- [32] G. S. Watson, M. R. Leadbetter, On the estimation of the probability density II, *Ann. Math. Statist.*, 34, 480-491, 1963.
- [33] Hall, P. Cross - validation in density estimation. *Biometrika* ,69 (1982), 383–390.
- [34] Hebbema, J.D.F. Hermans, J., and Vandenbroek, K. A stepwise discriminant analysis program using density estimation. in *compstat*, ed. G.Bruckmann, Wien Physica-Verlag (1974), 101–110.
- [35] H. G. Müller, Boundary effects in nonparametric curve estimation models, *Compstat*, 84-89, 1984.
- [36] H. G. Müller, Nonparametric regression analysis of longitudinal data, *Lecture Notes in Statistics*, 46, Springer-Verlag, New York, 1988.
- [37] H. G. Müller, Smooth optimum kernel estimators near endpoints, *Biometrika*, 78,521-530, 1991.

- [38] J. Rice, M. Rosenblatt, Estimation of the log survivor function and hazard function, *Sankhya Ser. A*, 38, 60-78, 1976.
- [39] J. M. Sarabia, A general definition of the Leimkuhler curve, *J. Informetrics*, 2,156-163, 2008.
- [40] Jin, X. and Kawczak, J. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, 4 : 103-124, (2003).
- [41] Kerkyacharian, G. and Picard, D. Density estimation by kernel and wavelets methods : optimality of besov spaces. *Statistics and Probability Letters*, 18(4) :327-336, (1993).
- [42] Kokonendji, C.C. and Libengué, F. G. Méthode des noyaux associés continus et estimation de densité. *Journées de Statistique de la SFdS*, 6 pages, Tunis, (2011).
- [43] Kokonendji, C.C. & Senga Kiessé, T. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, 8, 497-516.
- [44] L. Wasserman. All of nonparametric statistics. Springer Texts in Statistics. Springer-Verlag, 2006.
- [45] M. Rosenblatt, Remarks on some nonparametric estimates of a density function,
- [46] Nadaraya, E. On non parametric estimation density function and regression. *Theory Probab P.P.L* (1965).
- [47] P. K. Sen, *Sequential nonparametrics*, Wiley, New York, 1981.
- [48] Park, B. U., and Marron, S. J. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85 (1990), 66–72.
- [49] Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 3 (09 1962), 1065–1076.
- [50] P. Hall, T. E. Wehrly, A geometrical method for removing edge effects from kernel type nonparametric regression estimators, *J. Amer. Stat. Assoc.*, 86, 665-672, 1991.

- [51] Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* 27, 3 (Sept. 1956), 832–837.
- [52] Silverman, B. W. *Density estimation for statistics and data analysis*, vol. 26. CRC press, 1986.
- [53] S. Zhang, R. J. Karunamuni, M. C. Jones, An improved estimator of the density function at the boundary, *J. Amer. Stat. Assoc.*, 448, 1231-1241, 1999.
- [54] Saadi, N. and Adjabi, S. On the estimation of the probability density by trigonometric series. *Communications in Statistics Theory and Methods*, 38(19) :3583-3595, (2009).
- [55] Strauss, O. Estimation imprécise de densité de probabilité par transfert imprécis de comptage. *GRETSI - CNRS, Saint Martin d’Hères (France)*(2005), pp. 165–168.
- [56] Schuster, E.F. Gregory, G. On the nonconsistency of maximum likelihood nonparametric density estimators. *Computer Science and Statistics : Poceeding of the 13th Symposium on the Interface*, ed. W.F.Eddy, New York Spring Verlag (1981), 295–298.
- [57] Scott, D. W. *Multivariate Density Estimation : Theory, Practice, and Visualization*. Wiley Interscience, New York, 1992.
- [58] Scott, D. W., and Factor, L. E. Monte carlo study of three data-based nonparametric probability density estimators. *Journal of the American Statistical Association* 76,373 (1981), 9–15.
- [59] Scott, D. W., Tapia, R. A., and Thompson, J. R. Nonparametric probability density estimation by discrete maximum penalized-likelihood criteria. *The annals of statistics* (1980), 820–832.
- [60] Scott, D. W., and Terrell, G. R. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association* 82, 400 (1987), 1131–1146.
- [61] Sheather, S.J. Jones, M. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, B53 (1991), 683–690.

- [62] T. Gasser, H. G. Müller, Kernel estimation of regression functions, In : Smoothing techniques for curve estimation, Lecture Notes in Mathematics, 757, 23-68, Springer-Verlag, Berlin, 1979.
- [63] T. Gasser, H. G. Müller, V. Mammitzsch, Kernels for nonparametric curve estimation, J. Roy. Statist. Soc. Ser. B, 47, 238-252, 1985.

Annexe A : Abréviations et Notations

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

$\mathbb{E}[X]$	espérance mathématique ou moyenne du v.a. X .
exp	exponentiel
F	fonction de répartition
F_n	fonction de répartition empirique
<i>iid</i>	indépendant et identiquement distribuée.
<i>MISE</i>	L'erreur quadratique moyenne intégrée.
<i>MSE</i>	L'erreur quadratique moyenne.
h_{opt}	La fenêtre optimale.
<i>CV</i>	La validation croisée.
<i>BCV</i>	La validation croisée biaisée.
<i>UCV</i>	La validation croisée non biaisée.
<i>LCV</i>	La validation croisée par le maximum de vraisemblance.

X	predicteur variable
Y	variable d'intérêt
$E(X)$	espérance ou moyenne de X
F	fonction de distribution fonction de distribution
f	densité marginale de X
\hat{f}	estimateur de f
h	largeur de bande
iid	indépendant et identiquement distribué
K	kernel fonction
$f^{(j)}$	la j^{th} dérivée
f' , f'' , f'''	la première, la deuxième et la troisième dérivées de f .

Annexe B : Théorèmes et Critères

[1] Règles de Cramer :

1. Soient X une variable aléatoire et c une constante. (X_n) et (Y_n) sont deux suites de variables aléatoires telles que $X_n \xrightarrow{L} X$ et $Y_n \xrightarrow{L} c$. Alors :

(i) $X_n + Y_n \xrightarrow{L} X + c$,

(ii) $X_n \cdot Y_n \xrightarrow{L} X \cdot c$,

(iii) $\frac{X_n}{Y_n} \xrightarrow{L} \frac{X}{c}$ si $c \neq 0$.

2. Soit X une variable aléatoire. (X_n) et (Y_n) sont deux suites de variables aléatoires telles que

$$X_n \xrightarrow{L} X \text{ et } |X_n - Y_n| \xrightarrow{P} 0. \text{ Alors } Y_n \xrightarrow{L} X.$$

[2] Théorème des fonctions continues :

Soit $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ une fonction continue en tout point d'un ensemble C tel que

$$P[X \in C] = 1. \text{ Alors :}$$

(i) Si $X_n \xrightarrow{L} X$, alors $g(X_n) \xrightarrow{L} g(X)$.

(ii) Si $X_n \xrightarrow{P} X$, alors $g(X_n) \xrightarrow{P} g(X)$.

(iii) Si $X_n \xrightarrow{p.s} X$, alors $g(X_n) \xrightarrow{p.s} g(X)$.

Grâce au théorème des fonctions continues, ces règles sont vraies en remplaçant la convergence en loi par la convergence en probabilité ou la convergence presque sûre.

2. Soit X une variable aléatoire. (X_n) et (Y_n) sont deux suites de variables aléatoires telles que $X_n \xrightarrow{L} X$ et $|X_n - Y_n| \xrightarrow{P} 0$. Alors $Y_n \xrightarrow{L} X$.

[3] Inégalité de Markov :

Soient X une variable aléatoire et $p \in \mathbb{N}^*$. Alors pour tout $\varepsilon > 0$, on a

$$P(|X| \geq \varepsilon) \leq \frac{E[|X|^p]}{\varepsilon^p}.$$

[4] Théorème de Lindeberg-Feller :

Pour tout $n = 1, 2, \dots$ soit $Y_{n_1}, \dots, Y_{n, k_n}$ des d -vecteurs aléatoires indépendants de carré intégrable satisfaisant les conditions :

$$(i) \forall \varepsilon > 0, \sum_{i=1}^{k_n} E [\|Y_{ni}\|^2 1_{[\|Y_{ni}\| > \varepsilon]}] \rightarrow 0 \quad (\text{Lindeberg-Feller})$$

$$(ii) \sum_{i=1}^{k_n} Cov Y_{ni} \rightarrow 1 \neq 0.$$

lorsque $n \rightarrow +\infty$. Alors le théorème de la limite centrale de Lindeberg-Feller pour tableaux triangulaires affirme que $\sum_{i=1}^{k_n} (Y_{ni} - EY_{ni}) \xrightarrow{L} N(0, 1)$.

Si $k_n = n$ et Y_{n_1}, \dots, Y_{n_n} sont des copies i.i.d. d'une variable aléatoire Y_n , alors (i) et (ii) se réduisent à :

$$(iii) \forall \varepsilon > 0, nE [\|Y_{ni}\|^2 1_{[Y_n > \varepsilon]}] \rightarrow 0.$$

et

$$(iv) nCov Y_n \rightarrow 1 \neq 0.$$

Nous donnons deux conditions suffisantes pour (iii) faciles à vérifier. La première s'obtient à l'aide du théorème de la convergence dominée.