

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

UNIVERSITE MOHAMED KHIDER, BISKRA

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

DEPARTEMENT DE MATHEMATIQUES



Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : **Statistique**

Par

Rim Ghelani

Titre :

**Estimation à noyau d'une densité de
probabilité conditionnelle**

Membres du Comité d'Examen :

Dr. BENAMEUR Sana	UMKB	Présidente.
Dr. CHERFAOUI Mouloud	UMKB	Encadreur.
Dr. DHIABI Samra	UMKB	Examinatrice.

Juin 2021

DÉDICACE

Je dédie ce mémoire

*À.....la plus belle créature que Dieu a créée sur terre ,, à cet source de tendresse, de
patience et de générosité,,*

à.....ma mère Djamila !

À..... la joie de ma vie, celui qui m'a implanté le courage et l'amour du travail.

à.....mon père LAich

À.....mes chers frères " Oussama, Younes, Saif" et ma belle soeur Wafa.

À.....toute ma famille.

À..... tous mes amis.

À.....tous ceux qui m'aiment.

REMECIEMENTS

*Je veux tout d'abord à remercier **ALLAH** le tout puissant et miséricordieux, qui m'a donné la force et la patience d'accomplir ce Modeste travail*

*Je tiens à remercier vivement mon encadreur, **Dr. CHERFAOUI Mouloud**, pour sa gentillesse, sa disponibilité et sa contribution générale à l'élaboration de ce travail.*

Mes plus profonds remerciements vont à mes parents. Tout au long de mon cursus, ils m'ont toujours soutenu, encouragé et aidé. Ils m'ont donné toutes les chances pour réussir. Qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de mon plus affectueuse gratitude pour eux.

*Mes remerciements s'adressent également à tous les membres du jury : **Dr. BENA-MEUR SANA** et **Dr. DHIABI SAMRA** pour l'honneur qu'elles me font en acceptant de juger mon travail.*

*Je tiens aussi à remercier mon professeur vertueux **Pr. FATAH BENATIA** pour ses aides sans failles et ses efforts sans limites depuis de nombreuses années.*

Enfin, je tiens à remercier ma chère sœur, mes frères pour leurs sacrifices, mes très chères amies "K, D, A. Talbi, A. Naili, R. Firas, S, R, W, N" pour leurs soutien moral tout au long de la préparation de ce mémoire et tous mes camarades de promotion 2021.

Sans oublier tous mes professeurs de l'université de Biskra.

Table des matières

Dédicace	i
Remerciements	ii
Table des matières	iii
Liste des figures	vi
Liste des tableaux	vii
Introduction générale	1
1 Estimation à noyau de la densité de probabilité univariée	3
Introduction	3
1.1 Estimateur à noyau d'une densité univariée	3
1.2 Propriétés de l'estimateur	4
1.2.1 Espérance, Biais et Variance de l'estimateur	5
1.2.2 Comportement asymptotique de l'estimateur à noyau	5
1.2.3 Vitesse de convergence	6
1.3 Choix du noyau	7
1.4 Choix du paramètre de lissage	8
1.4.1 Choix optimal	9
1.4.2 Estimateur Rule Of Thumb (règle de référence)	11

1.4.3	Estimateur de Sheather et Jones	12
1.4.4	Validation croisée non biaisée (UCV)	14
1.4.5	Validation croisée biaisée (BCV)	16
	Conclusion	17
2	Estimation à noyau d'une fonction de densité conditionnelle	18
	Introduction	18
2.1	Notion de la norme (distance)	18
2.1.1	Définition de la norme	19
2.1.2	Exemples de distances classiques	19
2.2	Estimateur à noyau d'une densité conditionnelle	21
2.2.1	Cas de variables explicative réelles	21
2.2.2	Cas de variables explicatives multivariées (vectorielles)	22
2.2.3	Cas de variables explicatives fonctionnelles	24
2.3	Propriétés asymptotiques de l'estimateur	24
2.3.1	Biais et variance	24
2.3.2	Convergence en norme L^1 , pour x_0 fixé	25
2.3.3	Convergence en norme L^2	26
2.4	Choix du noyau et du paramètre de lissage	27
2.4.1	Choix optimal	28
2.4.2	La règle de référence	28
2.4.3	Validation croisée	30
	Conclusion	31
3	Exemple d'illustration numérique	32
	Introduction	32
3.1	Présentation de l'application et ses paramètres	32
3.2	Résultats numérique et graphique	34

3.3 Discussion des résultats	37
Conclusion	37
Conclusion générale	39
Bibliographie	41
Annexe A : Abréviations et Notations	44

Table des figures

3.1	Exemples de la variation de la densité $f(y/x)$ en fonction de $\ x\ _2$	34
3.2	Exemples d'un échantillon $x_i(t)$ de taille $n = 10$ issu de la variable X	34
3.3	Variation du h^* et du \overline{ISE}_{h^*} en fonction de la taille de l'échantillon.	36
3.4	Variation du (a^*, b^*) et du $\overline{ISE}_{(a^*, b^*)}$ en fonction de la taille de l'échantillon.	36

Liste des tableaux

1.1	Noyaux usuels et leurs supports.	7
2.1	Distances utilisées dans le cas $\mathbb{E} = \mathbb{R}^q$ avec $q \in \mathbb{N}/\{0\}$ et $p \in \mathbb{R}_*^+$	20
2.2	Normes pour le calcul de distances entre deux fonctions.	20
3.1	Résultats de simulation en fonction de la norme utilisée dans la définition de l'estimateur $\hat{f}(Y/x)$	35

Introduction générale

L'une des fonctions de densité la plus usuel dans différents champs des statistiques est bien que la fonction de densité conditionnelle.

Soient $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$, n observations équipondérées issues d'une variable aléatoire réelle (Y, X) de densité de probabilité réelle $f(y/x)$ **entièrement inconnue**.

Comment obtenir une estimation de f au point x à partir de la seule information contenue dans l'échantillon ?

La question que nous avons posée signifie que nous sommes intéressé plus à l'estimation non paramétrique, précisément l'*estimation non paramétrique de la densité de probabilité conditionnelle*.

Dans la littérature, plusieurs méthodes sont été dédiées à l'estimation de la densité de probabilité en générale et d'une densité conditionnelle en particulier. Mais ça reste que la méthode du noyau, qui peut être vue comme une extension de la méthode d'estimation par histogramme, est la plus populaire. Cette popularité de l'estimateur à noyau peut s'expliquer par, au moins trois raisons : la simplicité de sa forme, ses modes de convergence multiples et sa flexibilité qui s'interprété par la liberté de l'utilisateur dans le choix du noyau K et du paramètre de lissage h .

Historiquement, l'estimation à noyau d'une densité conditionnelle a été considéré pour la première fois par Rosenblatt (1969) [13]. Il a étudié le problème d'estimation de la densité $f(y/x)$ où X et Y sont des variables aléatoires univariées. Ensuite par plusieurs auteurs dont Hyndman et al. (1996) [8], Youndjé et al. (1993) [?], Fan et Yim (2004) [?]....

Les travaux cités ci-haut ont été limités au cas de variables explicatives univariées. Cependant, dans la pratique d'autres situations peuvent être envisagées et se présentent. En effet, la variable explicative peut être multivariée (un vecteur aléatoire) qui est le cas de la régression multiple où la variable réponse est définie par $E(y/x)$ dont la base est une densité conditionnelle. La variable explicative peut être une fonction aléatoire de telle situation est souvent rencontrée dans l'analyse d'une réponse dépendant d'un processus aléatoire.

L'estimation de la densité conditionnelle dans cette dernière situation, c'est-à-dire l'estimation à noyau de la fonction de densité conditionnelle lorsque la variable explicative est fonctionnelle, ont été introduites par Ferraty et al.(2005). Ces auteurs ont obtenu la convergence presque complète dans le cas i.i.d. Depuis cet article, une littérature abondante s'est développée sur l'estimation de la densité conditionnelle, ses dérivées et ses applications à d'autres champs (régression, taux de hasard,...), entre autres on peut citer : Ferraty et al.(2006), Ezzahrioui et Ould-Saïd (2005,2006), Laksaci (2007), Laksaci et al.(2010)...

Le présent travail porte sur l'estimation à noyau d'une densité conditionnelle lorsque la variable explicative est : univariée (réelle), multivariée (un vecteur aléatoire) ou une fonction aléatoire. Pour atteindre notre objectif nous avons organisé le présent comme suit :

- Le premier chapitre sera consacré à un rappel sur l'estimation à noyau d'une densité en sa version basique c'est-à-dire l'estimation d'une densité univariée.
- Dans le deuxième chapitre, nous allons aborder dans un premier temps la notion de l'estimation à noyau d'une densité conditionnelle lorsque la variable expliquée et la variable explicatives sont univariées. Dans un second lieu, les formes de l'estimateur à noyau de $f(y/x)$ associées aux variables explicatives multivariées et fonctionnelles seront exposées. Enfin, le problème du choix du paramètre de lissage et du noyau sera exposé également.
- Avant de conclure, dans le chapitre 3, une étude numérique illustrative sera présentée pour le cas d'estimation de $f(y/x)$ lorsque x est une fonction aléatoire.

Chapitre 1

Estimation à noyau de la densité de probabilité univariée

Introduction

Dans le présent chapitre, nous intéressons à l'estimation à noyau d'une densité de probabilité univariée. Nous allons d'abord exposer la définition de l'estimateur à noyau d'une densité de probabilité univariée. Par la suite, nous allons nous intéresser aux propriétés (biais, variance, convergence,...) de l'estimateur en question. En fin, avant de conclure nous allons aborder le problème du choix des paramètres de cet estimateur, à savoir : le choix du noyau et du paramètre de lissage.

1.1 Estimateur à noyau d'une densité univariée

Parzen (1962) [11] a étudié les propriétés fondamentales de l'estimateur à noyau de la densité, juste après son introduction par Rosenblatt (1956) [12]. Depuis, cet estimateur est devenu un objet classique étudié par les statisticiens. L'estimateur de la densité de probabilité par la méthode du noyau est le plus répandu aujourd'hui, car il répond au problème du choix des différents paramètres dans l'estimation à histogramme et possède

de bonnes propriétés. L'idée de construction de cet estimateur consiste à évaluer f au point x , en comptant le nombre d'observations tombées dans un certain voisinage de $x \in \mathbb{R}$.

Définition 1.1.1 Soit x_1, \dots, x_n un n -échantillon de densité inconnue f sur \mathbb{R} . On appelle *estimateur à noyau* ou *estimateur de Parzen–Rosenblatt de la densité f* associé à x_1, \dots, x_n , la fonction aléatoire $\hat{f} : \mathbb{R} \rightarrow \mathbb{R}_+$ définie, pour tout $x \in \mathbb{R}$, par :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad (1.1)$$

où $h = h(n)$ est appelé *paramètre de lissage* et la fonction K est appelée *noyau satisfaisant les conditions suivantes* :

$$\int_{\mathbb{R}} K(u) du = 1, \quad \int_{\mathbb{R}} uK(u) du = 0 \quad \text{et} \quad \sigma_K^2 = \int_{\mathbb{R}} u^2 K(u) < \infty. \quad (1.2)$$

1.2 Propriétés de l'estimateur

Cette section est consacrée aux principaux résultats théoriques obtenus dans la littérature sur les propriétés de l'estimateur à noyau définie dans (1.1), à savoir :

- Le biais et la variance, de l'estimateur à noyau, et leurs comportements asymptotiques.
- La convergence en moyenne quadratique et en moyenne quadratique intégrée.
- La convergence uniforme (en probabilité, presque complète).
- La convergence en norme L_1 .

1.2.1 Espérance, Biais et Variance de l'estimateur

Les expressions de l'espérance, biais et de la variance de l'estimateur à noyau sont données respectivement par (pour plus de détails voir Silverman [18]) :

$$\begin{aligned}
 \mathbb{E} \left(\hat{f}(x) \right) &= f(x) + \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2), \\
 \text{Biais} \left(\hat{f}(x) \right) &= \mathbb{E} \left(\hat{f}(x) \right) - f(x) = \frac{h^2}{2} f''(x) \mu_2(K) + o(h^2), \\
 \text{Var} \left(\hat{f}(x) \right) &= \frac{f(x)}{nh} \int_{-\infty}^{\infty} K^2(y) dy - \frac{f'(x)}{n} \int_{-\infty}^{\infty} y K^2(y) dy - \frac{1}{n} \left(f(x) + \text{biais} \hat{f}(x) \right)^2,
 \end{aligned} \tag{1.3}$$

où $\mu_2(K) = \int_{-\infty}^{\infty} y^2 K(y) dy$.

1.2.2 Comportement asymptotique de l'estimateur à noyau

Parzen [11] a élaboré les conditions de plusieurs types de convergence de l'estimateur à noyau ainsi que la convergence de ses propriétés. Les principaux résultats obtenus, par l'auteur, sont résumés dans le Théorème suivant :

Théorème 1.2.1 *Sous les conditions suivantes :*

1. $\lim_{n \rightarrow +\infty} h(n) = 0$ et $\lim_{y \rightarrow +\infty} |yK(y)| = 0$,
2. $\sup_y |K(y)| < \infty$ et $\int_{-\infty}^{\infty} |K(y)| dy < \infty$,
3. $\int_{-\infty}^{\infty} K(y) dy = 1$.

on a :

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\hat{f}(x) \right) = f(x) \text{ et } \lim_{n \rightarrow \infty} nh \text{Var} \left(\hat{f}(x) \right) = f(x) \int_{-\infty}^{\infty} K^2(y) dy.$$

Si de plus, $\lim_{n \rightarrow \infty} nh(n) = \infty$, alors

- $\lim_{n \rightarrow \infty} \text{MSE}(\hat{f}(x), f(x)) = 0$, pour tout point x pour lequel la densité f est continue,
- $\lim_{n \rightarrow \infty} \text{MISE}(\hat{f}, f) = 0$, $\forall f \in \mathbb{L}^p$.
- $\hat{f}(x) \xrightarrow{\text{loi}} \mathcal{N} \left(\mathbb{E} \left(\hat{f}(x) \right), \text{Var} \left(\hat{f}(x) \right) \right)$.

- $\forall \epsilon > 0, P \left(\sup_{x \in \mathbb{R}} |\hat{f}(x) - f(x)| < \epsilon \right) = 1$, si la transformée de Fourier $\tilde{K}(z) = \int_{-\infty}^{\infty} \exp(-izy)K(y)dy$ est absolument intégrable.

On note par \mathbb{L}^p : l'ensemble des fonctions f définies sur \mathbb{R} , telle que $\int |f(x)|^p dx < \infty$.

Le Théorème suivant donne le résultat élaboré par Nadaraya [9] concernant la convergence uniforme presque complète de l'estimateur à noyau classique.

Théorème 1.2.2 *Si h et K satisfaisant les conditions suivantes :*

1. K est un noyau positif à variation bornée,
2. f est uniformément continue,
3. $\lim_{n \rightarrow \infty} h(n) = 0$,
4. $\sum_{n=1}^{\infty} \exp(-\gamma nh(n)^2) < \infty, \forall \gamma > 0$,

alors :

$$\sup_x |\hat{f}(x) - f(x)| \longrightarrow 0, \text{ avec une probabilité 1.}$$

Pour la convergence en L_1 presque complète, Devroye [5] a dégagé des conditions de convergence, qui sont résumées dans le Théorème suivant :

Théorème 1.2.3 *Si,*

$$\lim_{n \rightarrow \infty} h(n) = 0, \quad \lim_{n \rightarrow \infty} nh(n) = \infty,$$

alors,

$$\forall f \in \mathcal{F}, \quad \lim_{n \rightarrow \infty} \int |\hat{f}(x) - f(x)| dx = 0, \text{ Presque Complètement,}$$

où \mathcal{F} est l'ensemble des densités de probabilité.

1.2.3 Vitesse de convergence

Wahba [20] a montré qu'on ne peut pas améliorer indéfiniment la convergence d'un estimateur \hat{f} vers f , même pour la fonction la plus régulière possible (indéfiniment dérivable,

bornée). C'est-à-dire, $MSE(\hat{f}(x), f(x))$ ne peut tendre vers 0 que d'un ordre $\frac{c}{n}$, où c est une constante.

1.3 Choix du noyau

Un noyau approprié aide à surmonter les problèmes des bosses (multimodes) et de la discontinuité de la densité estimée. Par exemple, si K est une distribution gaussienne, alors la fonction de densité estimée \hat{f} sera lisse et admet des dérivées de toutes ordres.

Dans la littérature, plusieurs fonctions ont été proposées pour jouer le rôle d'un noyau, la Table 1.1 résume les noyaux les plus usuels dans la pratique.

TABLE 1.1: Noyaux usuels et leurs supports.

Nom	Expression	Domaine
Noyau Uniforme (Rosenblatt)	$K(u) = \frac{1}{2}$	$ u \leq 1$
Noyau Box (boite)	$K(u) = \frac{1}{2\sqrt{3}}$	$ u \leq \sqrt{3}$
Noyau Triangulaire	$K(u) = (1 - u)$	$ u \leq 1$
Noyau Cosine	$K(u) = \frac{\pi}{4} \cos\left(\frac{\pi u}{2}\right)$	$ u \leq 1$
Noyau Gaussien	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$	$u \in \mathbb{R}$
Noyau Biweight (Tukey)	$K(u) = \frac{15}{16} (1 - u^2)^2$	$ u \leq 1$
Noyau Triweight	$K(u) = \frac{35}{32} (1 - u^2)^3$	$ u \leq 1$
Noyau Epanechnikov	$K_E(u) = \frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5}\right)$	$ u \leq \sqrt{5}$

Des résultats théorique montre que l'impact du changement du noyau sur les caractéristique de l'estimateur est faible de plus l'efficacité des différents noyaux au sens de l'AMISE, est pratiquement la même. Par conséquent, le choix du noyau dans l'estimation d'une densité n'est pas important.

1.4 Choix du paramètre de lissage

D'après la formule (1.1), on constate que l'estimateur $\hat{f}(x)$ de $f(x)$ ne dépend pas seulement du noyau K , mais aussi du paramètre de lissage h . Dans la pratique, l'étape critique dans l'estimation d'une densité par la méthode du noyau est le choix du paramètre h , qui contrôle le lissage de l'estimateur \hat{f} [11, 18]. En effet, le choix de h est une étape importante lors de l'estimation par la méthode du noyau, dans le sens où h optimal nécessite de connaître la densité que l'on cherche à estimer, ce qui n'est pas le cas en pratique (pour plus de détails voir la suite de la présente section). De plus, une petite perturbation de ce dernier est suffisante pour que les caractéristiques de \hat{f} changent complètement (performances numériques et/ou graphiques). Par ailleurs, si h est trop petit, le biais de l'estimateur devient petit devant sa variance et l'estimateur sera trop fluctuant. Pour cela, on obtient un phénomène de *sous-lissage*. Dans le cas contraire, lorsque h est trop grand, le biais prend l'ascendant sur la variance et l'estimateur varie peu, alors on obtient un phénomène de *sur-lissage*.

D'ailleurs, c'est la raison pour laquelle on trouve, dans la littérature, plusieurs travaux consacrés au choix de ce paramètre plutôt que sur le choix du noyau. Il existe différentes méthodes de sélection de ce paramètre que l'on peut regrouper en deux familles :

- méthodes de plug-in (réinjection)
- méthodes de Cross-Validation (Validation-croisée).

La multitude des méthodes de sélection de paramètre de lissage et leurs diversités du point de vue de leurs principes, est due au fait que ces méthodes restent incomplètes. Autrement dit, ces méthodes ont toujours des inconvénients [22], soit au sens de la qualité numériques de l'estimateur \hat{f} , par rapport à une norme d'erreur bien déterminée, ou au sens de la qualité graphique de l'estimateur (l'allure graphique de la courbe de \hat{f} est sur-lissée ou sous-lissée).

La section suivante présente quelques méthodes de sélection les plus usuelles.

1.4.1 Choix optimal

La décision du choix du paramètre de lissage h suppose la spécification d'un critère d'erreur qui puisse être optimisé. L'optimalité est intimement liée aux choix du critère, qui peut faire intervenir à la fois la densité inconnue f et l'estimateur \hat{f} . La décision du choix du paramètre de lissage (h) repose sur la spécification d'un critère d'erreur qui puisse être optimisé. Bien que l'optimalité n'est pas un concept absolu, de fait que cette optimalité est étroitement liée au choix d'un critère, qui fait intervenir à la fois la densité inconnue f et l'estimateur \hat{f} (c'est-à-dire le paramètre h et le noyau K).

Supposons que nous nous intéressons au choix du paramètre de lissage h qui minimise l'Erreur Quadratique Intégrée Moyenne ($MISE$), c'est-à-dire, la quantité suivante :

$$\arg \min_h MISE(f, \hat{f}) = \arg \min_h \int \mathbb{E} \left(\hat{f}(x) - f(x) \right)^2 dx.$$

Afin de déterminer le h optimal, au sens du $MISE$, nous avons besoin du résultat suivant :

Théorème 1.4.1 (*Scott [15]*)

Si f a une dérivée seconde absolument continue, $f^{(3)} \in \mathbb{L}^2$, le noyau $K \in \mathbb{L}^2$ et une densité de probabilité continue, symétrique de variance $\sigma_K^2 > 0$, alors, sous les conditions $h(n) \rightarrow 0$ et $nh(n) \rightarrow \infty$, on a le développement asymptotique :

$$MISE = \frac{h^4}{4} \sigma_K^4 \int (f''(x) dx)^2 + \frac{\int K^2(x) dx}{nh} + o\left(h^5 + \frac{1}{n}\right), \quad (1.4)$$

où, \mathbb{L}^2 est l'ensemble des fonctions f définies sur \mathbb{R} , telles que $\int |f(x)|^2 dx < \infty$.

En exploitant le résultat de Théorème (1.4), on peut définir la quantité suivante :

$$AMISE = \frac{h^4}{4} \sigma_K^4 \int (f''(x))^2 dx + \frac{\int K^2(x) dx}{nh}, \quad (1.5)$$

appelée l'Erreur Quadratique Intégrée Moyenne Asymptotique.

On remarque que le premier terme du membre à droite du développement (1.5) est un terme de biais, alors que le second est un terme de variance. De plus, on constate que, le terme du biais est une fonction croissante en h , alors que le terme de la variance est une fonction décroissante en h . C'est-à-dire, les deux termes varient dans le sens inverse par rapport à h . Une largeur de fenêtre h trop importante entraînera une augmentation du biais et une diminution de la variance, alors qu'une largeur de fenêtre trop petite provoquera une augmentation de la variance et une diminution du biais. De ce fait, le paramètre de lissage h^* optimal au sens du critère de l'*AMISE*, devra réaliser un compromis entre les valeurs de la variance et celle du biais.

Par ailleurs, pour obtenir le paramètre de lissage h^* qui minimise l'Erreur Quadratique Intégrée Moyenne Asymptotique, il suffit de résoudre le système suivant :

$$\begin{cases} \frac{dAMISE}{dh} = 0, \\ \frac{d^2AMISE}{dh^2} > 0. \end{cases} \quad (1.6)$$

à partir de l'expression (1.5), on aura :

$$h^* = \left[\frac{R(K)}{\sigma_K^4 R(f'')} \right]^{1/5} n^{-1/5}, \quad (1.7)$$

avec

$$R(g) = \int (g(x))^2 dx.$$

Notons que h^* est une quantité déterministe qui dépend du nombre d'observations n .

La valeur du *AMISE* optimale ($AMISE^* = AMISE(h^*)$) est donnée par :

$$AMISE^* = \frac{5}{4} [\sigma_K R^4(K) R(f'')]^{1/5} n^{-4/5}. \quad (1.8)$$

En plus de sa nature asymptotique, la largeur de fenêtre optimale h^* dépend de la densité inconnue f au travers du paramètre $R(f'')$. Cette largeur de fenêtre "idéale" (relativement

au critère d'erreur retenu) n'est donc pas directement calculable dans la pratique. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R(f'')$ par un estimateur approprié ou une quantité bien définie, d'où le principe des méthodes plug-in.

1.4.2 Estimateur Rule Of Thumb (règle de référence)

L'estimateur Rule Of Thumb du paramètre de lissage, noté h_{rot} , suppose que nous utilisons le noyau gaussien pour estimer une densité f d'une distribution normale centrée (la moyenne égale à 0) et de variance σ^2 . De ce fait, la quantité $R(f'')$ est définie par :

$$R(f'') = \int (f''(x))^2 dx = \frac{3}{8} \sqrt{\pi} \sigma^{-5}. \quad (1.9)$$

En substituant $R(f'')$ et K par leurs formules dans (1.7), on obtient :

$$\begin{aligned} h_{rot} &= (4\pi)^{-1/10} \left[\frac{3}{8} \pi^{-1/2} \sigma \right] n^{-1/5} = \left(\frac{4}{3} \right)^{1/5} \sigma n^{-1/5} \\ &= 1.06 \sigma n^{-1/5}. \end{aligned} \quad (1.10)$$

Il suffit donc d'estimer σ à partir des données et de substituer cet estimateur dans la formule ci-dessus. D'après Silverman (1986) [18], la formule (1.10) donnera de bons résultats si la population est réellement normalement distribuée. Par contre, elle peut donner une distribution trop lissée si la population est multimodale. Dans ce cas, de meilleurs résultats peuvent être obtenus, si on utilise une mesure plus robuste de l'étendue. Au lieu d'utiliser σ , on utilise plutôt l'interquartile IQ définie par :

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

où, $X_{(n/4)}$ et $X_{(3n/4)}$ sont respectivement le premier et le troisième quartile de l'échantillon observé. Or, dans le cas où X suit une loi normale, l'écart interquartile est $IQ = 1.394$

alors h_{rot} de l'équation (1.10) devient

$$h_{rot} = 1.06IQn^{-1/5}.$$

Cette dernière formule peut aussi donner une distribution trop lissée si la vraie densité est multimodale et parfois elle donne aussi des résultats moins bons que si l'on avait utilisé l'écart type, d'où le meilleur des deux méthodes peut être obtenu en utilisant un estimateur adaptatif de l'étendue. C'est à dire, en utilisant A au lieu de σ dans la formule (1.10) où A est défini par $A = \min(\sigma, IQ)$, donc la formule pour h_{rot} devient alors :

$$h_{rot} = 1.06An^{-1/5}. \quad (1.11)$$

Cette correction est insuffisante dans de nombreux cas. Par exemple, si la vraie densité est multimodale.

1.4.3 Estimateur de Sheather et Jones

Sheather et Jones (1991) [17], recommandent l'utilisation de l'estimateur naturel $\hat{R}_a(f'')$, en faisant observer que le terme de biais $\frac{R(K'')}{na^5}$ est positif et peut donc servir à annuler le terme de biais (négatif) de l'erreur quadratique moyenne entre $\hat{R}_a(f'')$ et $R(f'')$. Afin de faire disparaître quelques effets indésirables du terme du biais. Les deux auteurs sont contraints de mettre en place une procédure de type plug-in en trois étapes.

Sheather et Jones [17], choisissent d'estimer $R(f'') = \int_{-\infty}^{+\infty} (f'')^2 dx$ par :

$$S(a) = \frac{1}{n(n-1)a^5} \sum_{i=1}^n \sum_{j=1}^n L^{(4)}\left(\frac{x_i - x_j}{a}\right), \quad (1.12)$$

où $L^{(4)}$ désigne la dérivée quatrième du noyau suffisamment lisse L et a est un nouveau paramètre de lissage appelé paramètre pilote. Les deux auteurs choisissent le noyau gaussien

suivant :

$$K(u) = L(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

L'estimateur $S(a)$ est obtenu en écrivant $\hat{R}_a(f'') = R(f''_a)$. Sous les conditions de régularité suffisantes on a :

$$\int_{\mathbb{R}} f''^2(x) dx = \int_{\mathbb{R}} f^{(4)}(x) f(x) dx = \mathbb{E}(f^{(4)}(x)),$$

alors, $R(f'')$ peut être estimé par :

$$S(a) = R(f''_a) = \frac{1}{n} \sum_{i=1}^n f_a^{(4)}(x_i).$$

En utilisant la dérivée quatrième du noyau L , $S(a)$ peut s'écrire comme suit :

$$S(a) = \frac{1}{n(n-1)a^5} \sum_{i=1}^n \sum_{j=1}^n L^{(4)}\left(\frac{x_i - x_j}{a}\right).$$

Ainsi, le paramètre de lissage optimal est estimé par :

$$\hat{h} = \left(\frac{R(K)}{\sigma_K^4 S(a)}\right)^{1/5} n^{-1/5}, \quad (1.13)$$

avec

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

Sheather et Jones [17], affirment que le paramètre de lissage \hat{a} qui minimise l'Erreur Quadratique Moyenne $\mathbb{E}(S(a) - R(f''))^2$ admet la représentation asymptotique suivante :

$$\hat{a} = \left[\frac{-2K^{(4)}(0)}{\sigma_K^2 n \int_{\mathbb{R}} f^{(6)}(x) f(x) dx} \right]^{1/7}. \quad (1.14)$$

En combinant les équations (1.14) et (1.13), on peut avoir une expression reliant les deux

largeurs de fenêtre \hat{a} et \hat{h} .

$$\hat{\alpha}(h) = 1.357 \left(\frac{S(a)}{T(b)} \right)^{1/7} h^{5/7}, \quad (1.15)$$

avec

$$T(b) = \frac{1}{n(n-1)b^7} \sum_{i=1}^n \sum_{j=1}^n L^{(6)} \left(\frac{x_i - x_j}{b} \right),$$

est l'estimateur de $\int_{\mathbb{R}} f^{(6)}(x) f(x) dx$.

$$a = 0.920 \hat{\lambda} n^{-1/7} \quad \text{et} \quad b = 0.912 \hat{\lambda} n^{-1/9},$$

où, $\hat{\lambda}$ (estimateur de λ) représente une mesure d'échelle de f (par exemple son écart interquartile). Le paramètre de lissage final est la solution de l'équation suivante :

$$\hat{h} = \left(\frac{R(K)}{\sigma_K^4 S(\hat{\alpha}(\hat{h}))} \right)^{1/5} n^{-1/5}, \quad (1.16)$$

qui peut être réécrite sous la forme :

$$\left(\frac{1}{2\sqrt{\pi}} \right)^{1/5} S(\hat{\alpha}(\hat{h})) (f'')^{-1/5} n^{-1/5} - \hat{h} = 0. \quad (1.17)$$

1.4.4 Validation croisée non biaisée (UCV)

Cette méthode a été proposée par Rudemo (1982) [14] et Bowman (1984) [3]. Elle consiste à choisir le paramètre de lissage qui minimise un estimateur convenable de :

$$\begin{aligned} UCV(h) &= \int_{\mathbb{R}} [\hat{f}(x) - f(x)]^2 dx - \int_{\mathbb{R}} f^2(x) dx \\ &= \int_{\mathbb{R}} \hat{f}^2(x) dx - 2 \int_{\mathbb{R}} \hat{f}(x) f(x) dx. \end{aligned}$$

Puisque $\int_{\mathbb{R}} f^2(x)dx$ ne dépend pas du paramètre de lissage h . On peut choisir alors le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

$$\int_{\mathbb{R}} \hat{f}^2(x)dx - 2 \int_{\mathbb{R}} \hat{f}(x)f(x)dx.$$

Maintenant, on veut trouver un estimateur de $\int_{\mathbb{R}} \hat{f}(x)f(x)dx$. Pour cela, remarquons que

$$\int_{\mathbb{R}} \hat{f}(x)f(x)dx = \mathbb{E} \left(\hat{f}(x) \right).$$

L'estimateur empirique de $\int_{\mathbb{R}} \hat{f}(x)f(x)dx$, est alors

$$\frac{1}{n} \sum_{i=1}^n f_{h,i}(x_i),$$

et le critère à optimiser est :

$$UCV(h) = \int_{\mathbb{R}} \hat{f}^2(x)dx - \frac{2}{n} \sum_{i=1}^n f_{h,i}(x_i), \quad (1.18)$$

où,

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

est l'estimateur de la densité construit à partir de l'ensemble de points sauf au point x_i .

Nous notons par h_{ucv} l'estimateur de h qui minimise $UCV(h)$.

La popularité de cette méthode est due à la motivation intuitive et au fait que cet estimateur est asymptotiquement optimal sous de faibles conditions. L'optimalité asymptotique de la validation croisée non biaisée a été obtenue par Stone [19]. Cependant, cette méthode présente deux problèmes majeurs. D'une part, son manque de robustesse par rapport aux changements de la taille de l'échantillon c'est-à-dire, le résultat de simulation peut se révéler extrêmement variable d'un échantillon à un autre. D'autre part, la fonctionnelle à minimiser a souvent tendance à présenter plusieurs minimums locaux [7]. Pour d'autres études voir Hall [6], Burman [4], Scott et Terrell [16].

1.4.5 Validation croisée biaisée (BCV)

Le critère de validation croisée biaisée, a été introduit par Scott et Terrell (1987) [16] pour remédier aux problèmes de la méthode "validation croisée non biaisée". Il s'agit d'introduire un biais dans le UCV afin de réduire sa variance. Rappelons que l'Erreur Quadratique Intégrée Moyenne Asymptotique s'écrit sous la forme :

$$AMISE = \frac{h^4}{4} \sigma_K^4 R(f'') + \frac{R(K)}{nh}. \quad (1.19)$$

Le paramètre de lissage basé sur la méthode de validation croisée biaisée est la valeur de h qui minimise un estimateur du $AMISE$. à partir de la formule (1.19), il est clair que afin d'estimer l' $AMISE$, il suffit d'estimer $R(f'')$. Un estimateur naturel de ce dernier terme est donné par $R(\hat{f}'')$ (\hat{f} est l'estimateur de la densité f obtenu par la méthode du noyau). Finalement, Scott et Terrell [16] ont proposé la forme de l'estimateur de $AMISE$ à minimiser (critère de $BCV(h)$), qui se résume comme suit :

Proposition 1.4.1 (*Scott et Trell [16]*)

Soit X_1, X_2, \dots, X_n un n -échantillon *i.i.d* issu d'une variable aléatoire X de fonction de densité f . Pour un noyau K , on obtient :

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{\mu_2^2(K)}{4n^2} \sum_i \sum_{j, j \neq i} K_h^{(2)} K_h^{(2)}(X_i - X_j). \quad (1.20)$$

Des résultats de simulations ont été obtenus pour la méthode de validation croisée biaisée dans le travail de Park et Marron [10]. Les auteurs ont constaté que la méthode validation croisée biaisée présente le même point faible que celui de la méthode validation croisée non biaisée. Cette méthode nous fournit plusieurs minimums locaux pour la fonctionnelle cible à minimiser.

Conclusion

Dans ce chapitre, nous avons présenté les principales notions d'estimation d'une densité de probabilité univariée par la méthode du noyau. Une autre densité très importantes et très usitée dans la statistique est bien que la fonction de densité conditionnelle. L'estimation de cette dernière densité par la méthode du noyau fera l'objet du chapitre suivant et ceci toute en considérons différentes types de variables explicative (réelles, vectorielles, fonctionnelle).

Chapitre 2

Estimation à noyau d'une fonction de densité conditionnelle

Introduction

L'une des fonctions de densité la plus usuel dans différents champs des statistiques est bien que la fonction de densité conditionnelle. Dans ce chapitre nous sommes intéressés au problème d'estimation non-paramétrique d'une densité conditionnelle par la méthode du noyau. A cet effet, après la présentation de l'estimateur à noyau de cette densité, nous citons quelques de ses propriétés tel : le biais, la variance, la convergence.... dans le cas où la variable explicative est réelle. Ensuite, l'expression de l'estimateur en question correspondante aux variables explicatives vectorielles et fonctionnelles sera exposée. Enfin, le problème du choix du noyau et de paramètre de lissage sera abordé.

2.1 Notion de la norme (distance)

Avant de présenter l'estimateur à noyau d'une densité conditionnelle, il est intéressant de rappeler quelques notions de bases liées aux normes qui sont utiles pour la compréhension de certaines notions que nous allons exposer par la suite du présent mémoire.

2.1.1 Définition de la norme

En mathématiques, une distance est une application qui formalise l'idée intuitive de distance, c'est-à-dire la longueur qui sépare deux points. C'est par l'analyse des principales propriétés de la distance usuelle que Fréchet introduit la notion d'espace métrique, développée ensuite par Hausdorff. Elle introduit un langage géométrique dans de nombreuses questions d'analyse et théorie des nombres.

Définition 2.1.1 *On appelle distance sur un ensemble \mathbb{E} toute application définie sur le produit $\mathbb{E}^2 = \mathbb{E} \times \mathbb{E}$ et à valeurs dans l'ensemble \mathbb{R}^+ des réels positifs ou nuls,*

$$d : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{R}^+$$

vérifiant les propriétés suivantes :

1. **Symétrie** : $\forall (\mathbf{a}, \mathbf{b}) \in \mathbb{E}^2, \mathbf{d}(\mathbf{a}, \mathbf{b}) = \mathbf{d}(\mathbf{b}, \mathbf{a})$
2. **Séparation** : $\forall (\mathbf{a}, \mathbf{b}) \in \mathbb{E}^2, \mathbf{d}(\mathbf{a}, \mathbf{b}) = \mathbf{0} \iff \mathbf{a} = \mathbf{b}$
3. **Inégalité triangulaire** : $\forall (\mathbf{a}, \mathbf{b}, \mathbf{c}) \in \mathbb{E}^3, \mathbf{d}(\mathbf{a}, \mathbf{c}) \leq \mathbf{d}(\mathbf{a}, \mathbf{b}) + \mathbf{d}(\mathbf{a}, \mathbf{c})$

Remarque 2.1.1

- Un ensemble muni d'une distance s'appelle un espace métrique.
- Généralement une autre notation pour la distance est utilisée :

$$\forall (x, y) \in \mathbb{E} \times \mathbb{E} \quad d(x, y) = \|y - x\| .$$

2.1.2 Exemples de distances classiques

Dans cette section nous présentons les distances les plus usitées dans la pratique pour mesurer la distance entre deux points vectoriels (deux vecteurs) et deux points fonctionnels (deux fonctions).

Cas de deux vecteurs

Une expression particulière de la distance est lorsque $\mathbb{E} = \mathbb{R}^q$ (avec $q \in \mathbb{N}/\{0\}$), on peut définir de plusieurs manières la distance entre deux points, bien qu'elle soit généralement donnée par la distance euclidienne (ou 2-distance). Soit deux points de \mathbb{R}^q , (x_1, x_2, \dots, x_q) et (y_1, y_2, \dots, y_q) , on exprime les différentes distances ainsi :

Nom	Paramètre	Expression
Distance Manhattan	1-distance	$\sum_{i=1}^q x_i - y_i $
Distance Euclidienne	2-distance	$\left(\sum_{i=1}^q (x_i - y_i)^2 \right)^{\frac{1}{2}}$
Distance de Minkowski	p -distance	$\left(\sum_{i=1}^q x_i - y_i ^p \right)^{\frac{1}{p}}$
Distance de Tchebychev	∞ -distance	$\lim_{p \rightarrow \infty} \left(\sum_{i=1}^q x_i - y_i ^p \right)^{\frac{1}{p}} = \sup_{1 \leq i \leq n} x_i - y_i $

TABLE 2.1: Distances utilisées dans le cas $\mathbb{E} = \mathbb{R}^q$ avec $q \in \mathbb{N}/\{0\}$ et $p \in \mathbb{R}_*^+$.

Cas de deux fonctions

Un autre cas particulier qui peut se présenter dans la pratique est bien que l'ensemble \mathbb{E} est un ensemble de fonctions. Dans ce cas les normes présentées dans la table 2.1, ne sont plus valables. Soit deux fonctions $x(t)$ et $y(t)$ définies de $\mathbb{E} \rightarrow \mathbb{F}$. Ainsi on exprime les différentes distances entre ces deux fonctions par l'une des distances suivantes :

Nom	Paramètre	Expression
Distance Manhattan	1-distance	$\int_E x(t) - y(t) dt$
Distance Euclidienne	2-distance	$\left(\int_E (x(t) - y(t))^2 dt \right)^{\frac{1}{2}}$
Distance de Minkowski	p -distance	$\left(\int_E x(t) - y(t) ^p dt \right)^{\frac{1}{p}}$
Distance de Tchebychev	∞ -distance	$\lim_{p \rightarrow \infty} \left(\int_E x(t) - y(t) ^p dt \right)^{\frac{1}{p}} = \sup_t x(t) - y(t) $

TABLE 2.2: Normes pour le calcul de distances entre deux fonctions.

Remarque 2.1.2

- La 2-distance permet de généraliser l'application du théorème de Pythagore à un espace de dimension q . C'est la distance la plus "intuitive".
- La ∞ -distance présente la particularité amusante de permettre la définition en toute rigueur de sphères cubiques (voir oxymore).
- La 1-distance permet de définir des sphères octaédriques.
- La p -distance est rarement utilisée en dehors des cas $p = 1, 2$ ou ∞ .

2.2 Estimateur à noyau d'une densité conditionnelle

Les méthodes du noyau sont utilisées de manière intensive par la communauté des non-paramétrés car elles constituent un moyen utile de faire une pondération locale. Dans cette section nous allons rappeler quelle est la pondération locale du noyau dans les cas de variables explicatives réels et multivariés avant de l'étendre au contexte fonctionnel.

2.2.1 Cas de variables explicative réelles

On cherche l'estimateur de la densité de probabilité conditionnelle de Y sachant que $X = x$ (on note $Y/(X = x)$) où X et Y sont des variables aléatoires univariées. On définit la densité conditionnelle de $Y/(X = x)$ comme suit :

$$f(y/x) = \frac{g(x, y)}{m(x)}, \tag{2.1}$$

où $g(x, y)$ est la densité joint de (X, Y) et $m(x)$ est la densité marginal de X .

Hyndman et al. (1996) [8] ont considéré la forme modifiée de l'estimateur de Rosenblatt pour définir l'estimateur de la densité conditionnelle, donnée par :

$$\hat{f}(y/x) = \frac{\hat{g}(x, y)}{\hat{m}(x)}, \tag{2.2}$$

où \hat{g} et \hat{m} sont les estimateurs respectifs de Parzen-Rosenblatt de g et m , définis respectivement par :

$$\hat{g}(x, y) = \frac{1}{nab} \sum_{j=1}^n K\left(\frac{x - X_j}{a}\right) K\left(\frac{y - Y_j}{b}\right), \quad (2.3)$$

$$\hat{m}(x) = \frac{1}{na} \sum_{j=1}^n K\left(\frac{x - X_j}{a}\right), \quad (2.4)$$

avec K étant un noyau définie sur \mathbb{R} , a est un paramètre de lissage dans la direction de x , b est un paramètre de lissage dans la direction de y et $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ est un n -échantillon issue de la variable aléatoire (X, Y) .

En substituant $\hat{g}(x, y)$ et $\hat{m}(x)$ données respectivement dans (2.3) et (2.4), l'estimateur de la densité conditionnelle s'écrit comme suit :

$$\hat{f}(y/x) = \frac{\frac{1}{nab} \sum_{j=1}^n K\left(\frac{x - X_j}{a}\right) K\left(\frac{y - Y_j}{b}\right)}{\frac{1}{na} \sum_{j=1}^n K\left(\frac{x - X_j}{a}\right)} = \frac{1}{b} \sum_{j=1}^n \omega_j(x) K\left(\frac{y - Y_j}{b}\right), \quad (2.5)$$

avec

$$\omega_j(x) = \frac{K\left(\frac{x - X_j}{a}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{a}\right)},$$

Supposons que le noyau $K(\cdot)$ est une fonction réelle, non négative, symétrique et deux fois intégrable, c'est-à-dire :

$$\int_{\mathbb{R}} K(u) du = 1, \quad \int_{\mathbb{R}} u K(u) du = 0 \quad \text{et} \quad \sigma_K^2 = \int_{\mathbb{R}} u^2 K(u) < \infty.$$

2.2.2 Cas de variables explicatives multivariées (vectorielles)

Dans des situations multivariées, on observe n vecteurs aléatoires x_1, x_2, \dots, x_n issue d'une variable x évalué en \mathbb{R}^q . La pondération locale précédente du noyau peut être facilement étendue à cette situation. A cette fin, il suffit de considérer un noyau multivarié K^* qui

sera une fonction de \mathbb{R}^q dans \mathbb{R} . Le premier moyen (naturel) de le faire est de définir K^* comme un produit de q fonctions de noyau réel K_1, K_2, \dots, K_q :

$$\forall u = t(u_1, u_2, \dots, u_q) \in \mathbb{R}^q, K^*(u) = K_1(u_1) \times K_2(u_2) \times \dots \times K_q(u_q).$$

Ainsi, dans cette situation, l'estimateur (2.5) doit être réécrit sous la forme suivante :

$$\hat{f}(y/x) = \frac{\sum_{j=1}^n K^* \left(\frac{x-X_j}{A} \right) K \left(\frac{y-Y_j}{b} \right)}{b \sum_{j=1}^n K^* \left(\frac{x-X_j}{A} \right)} \quad (2.6)$$

$$= \frac{\sum_{j=1}^n \prod_{i=1}^q K_i \left(\frac{x^{(i)}-X_j^{(i)}}{a_i} \right) K \left(\frac{y-Y_j}{b} \right)}{b \sum_{j=1}^n \prod_{i=1}^q K_i \left(\frac{x^{(i)}-X_j^{(i)}}{a_i} \right)} \quad (2.7)$$

avec $A = (a_1, a_2, \dots, a_q)$ dont la composante a_i est le paramètre de lissage dans la direction de la $i^{\text{ème}}$ composante du vecteur x .

Une deuxième façon consiste à combiner une fonction de noyau réelle K^* avec une norme (désignée par $\|\cdot\|_p$) dans \mathbb{R}^q comme suit :

$$\forall u \in \mathbb{R}^q, K^*(u) = K \left(\|u\|_p \right),$$

avec $\|u\|_p$ est l'une des distance présentée dans la table 2.1 et dans ce cas, l'estimateur (2.5) doit être réécrit sous la forme suivante :

$$\hat{f}(y/x) = \frac{\sum_{j=1}^n K \left(\frac{\|x-X_j\|_p}{a} \right) K \left(\frac{y-Y_j}{b} \right)}{b \sum_{j=1}^n K \left(\frac{\|x-X_j\|_p}{a} \right)}, \quad (2.8)$$

avec $\|\cdot\|_p$ est p -distance donnée dans la table 2.1.

Notez que si $K_1, K_2, \dots, K_q = 1_{[-1;1]}$ et si $\|\cdot\|$ est la norme suprême, les deux approches

(l'estimateur (2.6) et (2.8)) coïncident en prenant $K = 1_{[-1;1]}$.

En outre, parce que $\|u\|_p$ est toujours une quantité positive, le noyau réel K devrait avoir un support positif cela conduit à utiliser des fonctions asymétriques pour le noyau K .

2.2.3 Cas de variables explicatives fonctionnelles

Soit x_1, x_2, \dots, x_n n v.a.f (variables aléatoire fonctionnelles) évalué dans E et soit x un élément fixe de E . Dans de telles situation l'estimateur de l'estimateur est similaire a celle décrite dans (2.8) excepter la norme qui sera définie cette fois-ci dans un cadre continue. C'est-à-dire, l'estimateur de $f(y/x)$ dans ce cas sera défini par :

$$\hat{f}(y/x) = \frac{\sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{a}\right) K\left(\frac{y-Y_j}{b}\right)}{b \sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{a}\right)}, \quad (2.9)$$

où $\|\cdot\|_p$ est une la p -distance définie sur E (voir la table 2.2), K est un noyau réel (asymétrique).

2.3 Propriétés asymptotiques de l'estimateur

Dans cette section nous allons présenter quelques propriétés de l'estimateur à noyau d'une densité conditionnelle. Nous allons se limiter à la forme du biais et de la variance de l'estimateur, sa convergence ponctuelle en norme L^1 et sa convergence en norme L^2 (ponctuelle et globale).

2.3.1 Biais et variance

En se basant sur la dérivation et le développement de Taylor à l'ordre 2, Hyndman et al. [8] ont obtenu la forme asymptotique du biais et de la variance de l'estimateur et qui sont

données respectivement comme suit :

$$E \left[\hat{f}(y/x) \right] - f(y/x) = \frac{a^2 \sigma_K^2}{2} \left\{ 2 \frac{h'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} + \frac{b^2}{a^2} \frac{\partial^2 f(y/x)}{\partial y^2} \right\} - o(a^4) + o(b^4) + o(a^2 b^2) + o\left(\frac{1}{na}\right), \quad (2.10)$$

et

$$Var[\hat{f}(y/x)] = \frac{R(K)f(y/x)}{nabm(x)} [R(K) - bf(y/x)] + o\left(\frac{1}{n}\right) + o\left(\frac{b}{an}\right) + o\left(\frac{a}{bn}\right), \quad (2.11)$$

où $R(K) = \int K^2(u)du$, a condition que $a \rightarrow 0$ et $b \rightarrow 0$ lorsque $n \rightarrow \infty$.

Youndjé [21] a trouvé les conditions nécessaires sur les paramètres de lissage pour que l'estimateur \hat{f} soit consistant ponctuellement, et converge uniformément et en probabilité.

En effet, en fixant $(x, y) \in \mathbb{R}^2$ et $[\alpha, \beta]$ ($\alpha < \beta$) un intervalle de \mathbb{R} , il a trouvé des familles de lois de probabilité \mathcal{P}_{xy} et \mathcal{P}_α^β possédant les propriétés suivantes :

- $\hat{f}(y/x) \xrightarrow{p} f(y/x) \quad \forall \mu \in \mathcal{P}_{xy} \quad \text{Si} \quad a \rightarrow 0 \text{ et } nab \rightarrow +\infty \text{ quand } n \rightarrow +\infty.$
- $\sup_{(x,y) \in [\alpha,\beta] \times \mathbb{R}} \left| \hat{f}(y/x) - f(y/x) \right| \xrightarrow{p} 0 \quad \forall \mu \in \mathcal{P}_\alpha^\beta \quad \text{Si} \quad a \rightarrow 0 \text{ et } \frac{nab}{\ln(n)} \rightarrow +\infty.$

2.3.2 Convergence en norme L^1 , pour x_0 fixé

Soit $x_0 \in \mathbb{R}$ tel que $f(x_0) \neq 0$, nous savons d'une part que la fonction $y \mapsto f(y/x_0)$ est une densité. D'autre part ; l'espace naturel pour l'étude des densités est l'espace L^1 (voir Devroye 1987), de plus nous savons que la convergence ponctuelle presque partout en probabilité (resp. p.s.) implique la convergence L^1 en probabilité (resp. p.s.). Compte tenu de ces arguments et en s'inspirant des travaux de Devroye (1987), Youndjé [21] a obtenu des conditions suffisantes de convergence en L^1 (pour x_0 fixé) du type

$$\int \left| \hat{f}(y/x_0) - f(y/x_0) \right| dy \longrightarrow 0,$$

en probabilité, presque sûrement et presque complètement.

2.3.3 Convergence en norme L^2

Sachant que l'erreur quadratique moyenne (MSE) est la somme du carré du biais (2.10) avec la variance (2.11), alors l'erreur quadratique moyenne asymptotique ($AMSE$) est de la forme :

$$\begin{aligned}
 AMSE \left(\hat{f}(x), f(x) \right) &= \frac{a^4 \sigma_K^4}{4} \left[2 \frac{h'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} + \frac{b^2}{a^2} \frac{\partial^2 f(y/x)}{\partial y^2} \right]^2 \\
 &+ \frac{R(K)f(y/x)}{nabm(x)} [R(K) - bf(y/x)] + o\left(\frac{1}{n}\right) + o\left(\frac{b}{an}\right) \\
 &+ o\left(\frac{a}{bn}\right) + o(a^6) + o(b^6) + o(a^2b^4) + o(a^4b^2).
 \end{aligned} \tag{2.12}$$

On constate que cet estimateur est constitué à condition que $a \rightarrow 0$, $b \rightarrow 0$ et $nab \rightarrow \infty$ lorsque $n \rightarrow \infty$. Comme d'autres problèmes de lissage, les petites valeurs du paramètre de lissage donnent des petits biais et grandes variances, alors que les grands valeurs donnent des grands biais et petites variances. Les paramètres de lissage qui sont choisis pour minimiser (2.12) donnent un équilibre entre le biais et la variance.

L'erreur quadratique moyenne intégrée asymptotique ($AMISE$) est obtenue en faisant l'intégration par rapport à x et y de l' $AMSE$ pondéré, formé par le produit de (2.12) avec $m(x)$, sa forme est donnée comme suit :

$$MISE \approx \frac{c_1}{nab} - \frac{c_2}{na} + c_3a^4 + c_4b^4 + c_5a^2b^2, \tag{2.13}$$

où les constants c_1, c_2, c_3, c_4 et c_5 , qui dépendent du noyau K , la densité conditionnelle $f(y/x)$ et de la densité marginal $m(x)$, sont donnés par :

$$\left\{ \begin{array}{l}
 c_1 = \int R^2(K)dx, \\
 c_2 = \int \int R(K)f^2(y/x)dydx, \\
 c_3 = \int \int \frac{\sigma_K^4 m(x)}{4} \left\{ 2 \frac{h'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} \right\}^2 dydx, \\
 c_4 = \int \int \frac{\sigma_K^4 m(x)}{4} \left\{ \frac{\partial^2 f(y/x)}{\partial y^2} \right\}^2 dydx, \\
 c_5 = \int \int \frac{\sigma_K^4 m(x)}{2} \left\{ 2 \frac{h'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} \right\} \left\{ \frac{\partial^2 f(y/x)}{\partial y^2} \right\} dydx,
 \end{array} \right. \tag{2.14}$$

avec $R(g) = \int g^2(x)dx$.

2.4 Choix du noyau et du paramètre de lissage

Les choix plus communs du noyau sont définis en termes de fonction de densité de probabilité univariée et uni-modale, ce qui coïncide avec les choix qu'on a réalisés dans le premier chapitre. C'est-à-dire, le noyau K peut être sélectionné parmi les fonctions rangées dans la table 1.1. Alors il reste que le choix du paramètre de lissage. On s'intéresse à étudier les différentes méthodes du choix du paramètre de lissage dans le cas où l'échantillon est *i.i.d.* Pour évaluer et comparer leurs méthodes, Bashtannyk et Hyndman (2001) [2] ont utilisé les critères d'erreurs suivantes :

L'erreur moyenne quadratique intégrée :

$$MISE(a, b, \hat{f}, f) = \iint E\{\hat{f}(y/x) - f(y/x)\}^2 m(x) dx dy. \quad (2.15)$$

L'erreur quadratique intégrée :

$$ISE(a, b, \hat{f}, f) = \iint \{\hat{f}(y/x) - f(y/x)\}^2 m(x) dx dy. \quad (2.16)$$

L'erreur quadratique intégrée asymptotique :

$$AISE = \frac{\Delta}{n} \sum_{j=1}^N \sum_{i=1}^n \left[\hat{f}(y'_j/X_i) - f(y'_j/X_i) \right]^2, \quad (2.17)$$

où $y' = \{y'_1, \dots, y'_N\}$ est un vecteur de points équidistants dans l'espace de Y et

$$\Delta = y'_{i+1} - y'_i.$$

2.4.1 Choix optimal

Les largeurs de fenêtres optimales peuvent être obtenues par la différentiation de (2.13) par rapport à a et b et en fixant les dérivés à zéro. En simplifiant ces dérivés, nous obtenons le système d'équations suivant :

$$\begin{cases} -\frac{c_1}{n} - \frac{c_2 b}{n} + 4c_3 a^5 b + 2c_5 a^3 b^3 = 0 \\ -\frac{c_1}{n} + 4c_4 a b^5 + 2c_5 a^3 b^3 = 0 \end{cases}$$

Hyndman et al. (1996) [8] ont montré que la solution du système est approximativement :

$$\begin{cases} a^* = c_1^{1/6} \left\{ 4 \left(\frac{c_3^5}{c_4} \right)^{1/4} + 2c_5 \left(\frac{c_3}{c_4} \right)^{3/4} \right\}^{-1/6} n^{-1/6}, \\ b^* = a^* \left(\frac{c_3}{c_4} \right)^{1/4} = c_1^{1/6} \left\{ 4 \left(\frac{c_4^5}{c_3} \right)^{1/4} + 2c_5 \left(\frac{c_4}{c_3} \right)^{3/4} \right\}^{-1/6} n^{-1/6}, \end{cases}$$

où c_i ($i = 1, \dots, 5$) sont définie précédemment. On remarque que a^* et b^* ne sont pas calculables car ils dépendent des fonctions inconnues $m(x)$ et $f(y/x)$.

2.4.2 La règle de référence

Cette méthode a été proposée dans Bashtannyk et Hyndman(2001) [2]. Elle consiste à supposer que la densité conditionnelle suit une loi normale et trouver les paramètres de lissage qui minimise le *MISE*. Cette technique robuste et fournie des résultats raisonnables, même pour des densités qui ne sont pas de distribution normale. En effet, les auteurs ont supposé que la densité conditionnelle de Y sachant que $X = x$ suit une loi normale de moyenne $r(x) = u + vx$ et d'écart type $\sigma(x) = p + qx$. D'où, $[Y/X = x] \overset{L}{\rightsquigarrow} N(u + vx, (p + qx)^2)$ et la densité conditionnelle est :

$$f(y/x) = \frac{1}{(p + qx)\sqrt{2\pi}} \exp \left\{ \frac{-1}{2(p + qx)^2} (y - u - vx)^2 \right\}. \quad (2.18)$$

Ils ont également supposé que la densité marginal $m(x)$ est connue, ils ont considéré deux situations, lorsque $m(x)$ est une loi Normale est lorsque $m(x)$ est une loi uniforme sur un intervalle.

Cas de loi uniforme

Pour trouver la règle de référence pour a et b quand $m(x)$ est une distribution uniforme sur $[a, b]$, on remplace la densité conditionnelle $f(y/x)$ (2.18) et la densité marginale dans les constants c_1, \dots, c_5 2.14, ensuite on fait l'intégration deux fois par rapport à x et y . On trouve ces constants en fonction de $p, q, R(K), a$ et b . Alors, sous la condition $v \neq 0$, la règle de référence est donnée comme suit : • **Cas $q \neq 0$:**

$$\begin{cases} a_U &= \left[\frac{2^{15/2} \sqrt{\pi} R^2(K) (a-b)^2 q}{3n\sigma_K^2 z w^{3/4} (\sqrt{w} + 2v^2 - 3q^2)} \right]^{1/6}, \\ b_U &= \frac{w^{1/4}}{\sqrt{2}} a_U, \end{cases}$$

où, $z = ((p + qa)^4 - (p - qb)^4) / (p + qa)^4 (p - qb)^4$, $w = 19q^4 + 4v^4 + 28q^2v^2$.

• **Cas $q = 0$:**

$$\begin{cases} a_U &= \left[\frac{4\sqrt{\pi} R^2(K) (a-b)^2 p^5}{3n\sigma_K^2 v^5} \right]^{1/6}, \\ b_U &= v a_U, \end{cases}$$

Cas de loi normale

Dans ce cas, ils ont supposé que $m(x)$ est normale avec une moyenne constante μ_h et une variance constante σ_h^2 . Ils supposent également que la densité (2.18) a une variance constante, $q = 0$. En refaisant les mêmes étapes que le cas uniforme, ils ont trouvé la règle de référence :

$$\begin{cases} a_N &= \left\{ \frac{16kR^2(K)p^5(288\pi^9\sigma_h^{58}\lambda^2(k))^{1/8}}{n\sigma_K^4 v^{5/2} \gamma^{3/4}(k) [\gamma^{1/2}(k) + v(18\pi\sigma_h^{10}\lambda^2(k))^{1/4}]} \right\}^{1/6}, \\ b_N &= \left\{ \frac{v^2 \gamma(k)}{3\sqrt{2}\pi\sigma_h^2 \lambda(k)} \right\}^{1/4} a_N, \end{cases}$$

où, $\lambda(k) = \int_{-\infty}^k \phi(t) dt$, $\phi(\cdot)$ est la densité normale standard et $\gamma(k) = \sqrt{2\pi}\sigma_h^3(3v^2\sigma_h^2 + 8p^2)\lambda(k) - 16k\sigma_h^2 p^2 e^{-k^2/2}$. La valeur k contrôle la taille de l'échantillon dans la direction

de x , les choix usuels de k sont 2 ou 3.

2.4.3 Validation croisée

Cette méthode a été traité dans le cas où a et b sont supposés égaux. Alors, dans ce cas le problème revient à trouver un seul paramètre de lissage noté h (*i.e* $a = b = h$).

L'estimateur de la densité conditionnelle sera définis par :

$$\hat{f}(y/x) = \frac{1}{h} \sum_{j=1}^n \omega_j(x) K \left(\frac{y - Y_j}{h} \right), \quad (2.19)$$

avec

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

K étant un noyau définie sur \mathbb{R} , et $h = h(n) \in \mathbb{R}_*^+$ est le paramètre de lissage.

Youndjé [21] a étudié le choix du paramètre de lissage avec la méthode de validation croisée, il a basé sur la minimisation de l'erreur quadratique intégrée pondérée définie par :

$$ISE(\hat{f}, f) = \iint \left\{ \hat{f}(y/x) - f(y/x) \right\}^2 W(x) W'(y) dx dy,$$

où W et W' sont des fonctions de poids positives. Ensuite, il a démontré que sous certaines hypothèses de régularité on a :

$$ISE(\hat{f}, f) = \frac{c_1}{nh^2} + c_2 h^4 + o \left(\frac{1}{nh^2} + h^4 \right), \quad p.s.$$

où c_1 et c_2 sont deux constants dépendent de la fonction de densité inconnue f , alors le paramètre de lissage qui minimise le ISE n'est pas calculable. Pour trouver un autre paramètre qui minimise le ISE , l'auteur a proposé de décomposer cette quantité de la manière suivante :

$$ISE(\hat{f}, f) = A + B - 2C,$$

où,

$$\begin{aligned} A &= \iint \hat{f}^2(y/x)W(\mathbf{x})W'(y)f(x)dx dy, \\ B &= \iint f^2(y/x)W(\mathbf{x})W'(y)f(x)dx dy, \\ C &= \iint \hat{f}(y/x)f(y/x)W(\mathbf{x})W'(y)f(x)dx dy. \end{aligned}$$

Puisque B est indépendant de h , choisir h minimisant ISE revient à choisir h minimisant $A - 2C$. Le fait que C s'écrit aussi :

$$C = E_{(X,Y)} \left[\hat{f}(y/x)f(y/x)W(\mathbf{x})W'(y) \right].$$

Youndjé [21] a approché A et C par :

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{g}^{-i}(X_i, Y_i)}{\hat{h}^{-i}(X_i)} W(X_i)W'(Y_i) \quad \text{et} \quad \bar{C} = \frac{1}{n} \sum \int \left(\frac{\hat{g}^{-i}(X_i, Y_i)}{\hat{h}^{-i}(X_i)} \right)^2,$$

où,

$$\begin{aligned} \hat{g}^{-i}(x, y) &= \frac{1}{(n-1)h^2} \sum_{j \neq i}^n K\left(\frac{x-X_j}{h}\right) K\left(\frac{y-Y_j}{h}\right), \\ \hat{h}^{-i}(x) &= \frac{1}{(n-1)h} \sum_{j \neq i}^n K\left(\frac{x-X_j}{h}\right). \end{aligned}$$

Finalement, la règle de sélection du paramètre de lissage par la validation croisée est obtenue en choisissant le h minimisant le critère suivant :

$$CV(h) = \frac{1}{n} \sum_{i=1}^n \int \left(\frac{\hat{g}^{-i}(X_i, y)}{\hat{h}^{-i}(X_i)} \right)^2 W'(y) dy W(X_i) - \frac{2}{n} \sum_{i=1}^n \frac{\hat{g}^{-i}(X_i, Y_i)}{\hat{h}^{-i}(X_i)} W(X_i)W'(Y_i).$$

Conclusion

Dans le présent chapitre, nous avons mis, en premier lieu, en évidence la définition d'un estimateur à noyau d'une densité $f(y/x)$ lorsque les deux variable x et y sont univariées. Par la suite, notre intérêt s'est porté sur l'expression de l'estimateur $f(y/x)$ lorsque la variable explicative est multivariée ou fonctionnelle.

Chapitre 3

Exemple d'illustration numérique

Introduction

L'objectif du présent chapitre est de présenter un exemple numérique illustratif de la mise en oeuvre d'un estimateur à noyau d'une densité conditionnelle lorsque la variable explicative x est fonctionnelle. Nous focalisons principalement sur les performances de l'estimateur à noyau d'une densité de probabilité conditionnelle en fonction de la norme utiliser pour calculé la distance entre le point $x(t)$ et les observations $x_i(t)$, de la taille de l'échantillon simulé et la version de l'estimateur utilisé ($a = b = h$ ou $a \neq b$).

3.1 Présentation de l'application et ses paramètres

Pour reprendre à notre objectif nous avons implémenté un programme sous Matlab dont ses principales étapes sont comme suit :

1. Générer m échantillons (X, Y) de taille n d'une loi cible.
2. Estimer, respectivement, (a^*, b^*) et h^* qui minimise l'*ISE* moyenne.
3. Calculer \hat{f}_{ab} et \hat{f}_h et comparer leurs performances.

où \hat{f}_{ab} représente l'estimateur de f dans le cas a et b sont indépendants et qui est donné par :

$$\hat{f}_{ab}(y/x) = \frac{\sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{a}\right) K\left(\frac{y-Y_j}{b}\right)}{b \sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{a}\right)} \quad (3.1)$$

et f_h représente l'estimateur de f dans le cas $a = b = h$ et qui est donné par :

$$\hat{f}_h(y/x) = \frac{\sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{h}\right) K\left(\frac{y-Y_j}{h}\right)}{h \sum_{j=1}^n K\left(\frac{\|x-X_j\|_p}{h}\right)} \quad (3.2)$$

Pour l'exemple de simulation, nous prenons le modèle où la densité conditionnelle Y sachant $X = x$ suit une loi normale de moyenne $\|x\|_2$ et de variance 1 donnée par :

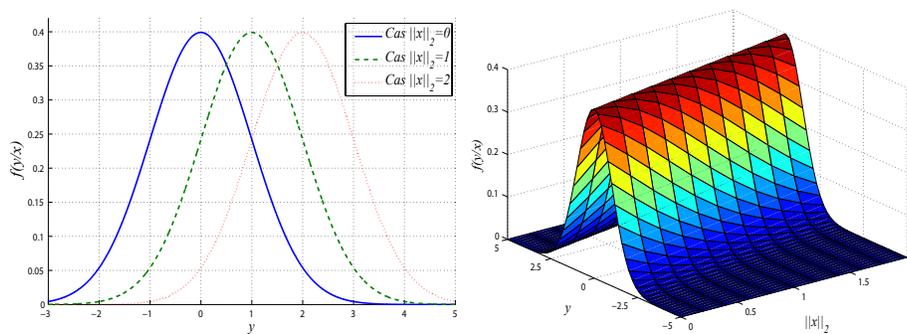
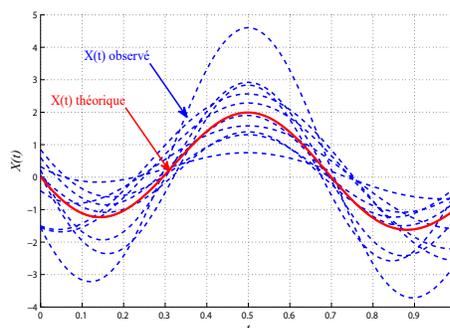
$$f(y/x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\|x\|_2)^2},$$

et supposons que les variables explicatives X_i , sont issues d'un processus stochastique

$$X_i = X_i(t) = a_i \cos(2\pi t) + b_i \sin(3\pi t) + c_i(t - 0.45)(t - 0.75)e^{(-d_i t)}$$

pour $t \in [0, 1]$ avec $a_i \rightarrow N(-1, 1)$, $b_i \rightarrow N(-1, 1)$, $c_i \rightarrow U(1, 5)$ et $d_i \rightarrow U(1, 5)$; avec N et U designent respectivement la loi normale et la loi uniforme.

Un exemple de variation de $f(y/x)$ en fonction de $\|x\|_2$ est présenté dans la figure 3.1. La figure 3.2 à droite représente un exemple d'un échantillon de taille $n = 10$ issue de la variable X . La courbe rouge présentée dans figure 3.2 représente $E(X(t))$ théorique c'est-à-dire lorsque $a = b = E(a) = E(b) = -1$ et $c = d = E(c) = E(d) = 3$.


 FIGURE 3.1 – Exemples de la variation de la densité $f(y/x)$ en fonction de $\|x\|_2$.

 FIGURE 3.2 – Exemples d'un échantillon $x_i(t)$ de taille $n = 10$ issu de la variable X .

3.2 Résultats numérique et graphique

Pour l'application numérique nous considérons les deux versions de l'estimateur à noyau de $f(y/x)$ (voir les deux expressions (3.1) et (3.2)) qui sont conçues à l'aide du noyau gaussien et de la norme $p \in \{1, 2, \infty\}$.

Les résultats numériques qu'on a obtenus dans cette situation sur 100 ($m = 100$) échantillons de taille $n \in \{50, 100, 200, 500, 1000, 2000\}$ au point $x = E(X)$ (voir figure 3.2) sont résumés dans la Table 3.1 et sont présentés dans les Figures 3.3–3.4.

Norme	n	$a = b$		$a \neq b$	
		h^*	\overline{ISE}_h	(a^*, b^*)	$\overline{ISE}_{(a;b)}$
$\ \cdot\ _1$	50	0.8739	0.0187	(23.7655 ; 0.5509)	0.0064
	100	0.8707	0.0176	(32.6362 ; 0.4953)	0.0043
	200	0.8689	0.0166	(23.3421 ; 0.419)	0.0023
	500	0.8681	0.0162	(31.3126 ; 0.3695)	0.0014
	1000	0.8672	0.0162	(28.8502 ; 0.3176)	0.001
	2000	0.8665	0.0161	(35.6687 ; 0.2659)	0.0006
$\ \cdot\ _2$	50	0.8771	0.0176	(23.6759 ; 0.5784)	0.0062
	100	0.8703	0.0173	(21.494 ; 0.5151)	0.0039
	200	0.867	0.0173	(17.8311 ; 0.4541)	0.003
	500	0.8668	0.0165	(30.8016 ; 0.348)	0.0014
	1000	0.8667	0.016	(33.0062 ; 0.3195)	0.0008
	2000	0.8662	0.0156	(39.1335 ; 0.2431)	0.0005
$\ \cdot\ _\infty$	50	0.8817	0.0209	(28.8429 ; 0.5458)	0.0067
	100	0.8747	0.0177	(23.7255 ; 0.4911)	0.0041
	200	0.8674	0.0176	(26.9714 ; 0.4362)	0.0024
	500	0.8676	0.0165	(29.9083 ; 0.3765)	0.0016
	1000	0.8663	0.0164	(37.9307 ; 0.2872)	0.0009
	2000	0.8655	0.0164	(41.8221 ; 0.2384)	0.0006

TABLE 3.1: Résultats de simulation en fonction de la norme utilisée dans la définition de l'estimateur $\hat{f}(Y/x)$.

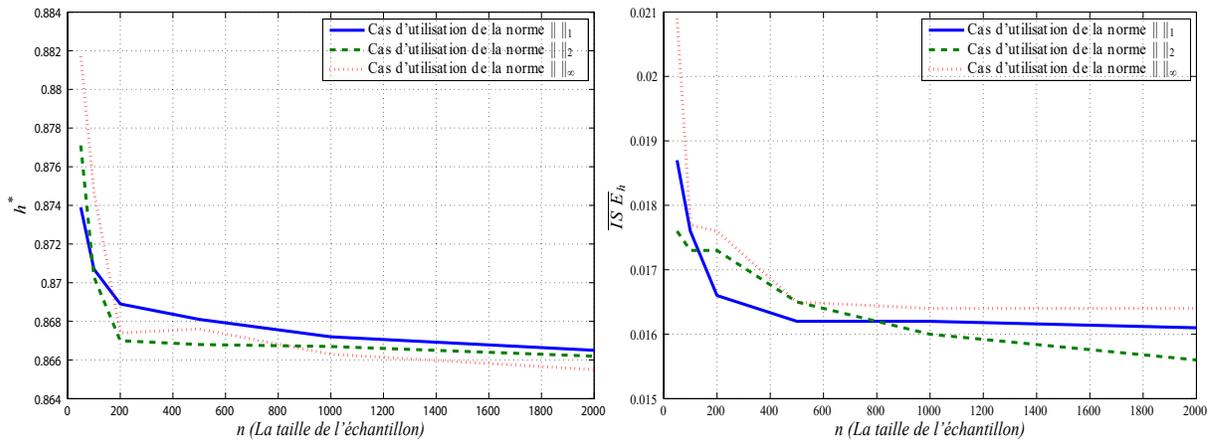


FIGURE 3.3 – Variation du h^* et du \overline{ISE}_{h^*} en fonction de la taille de l'échantillon.

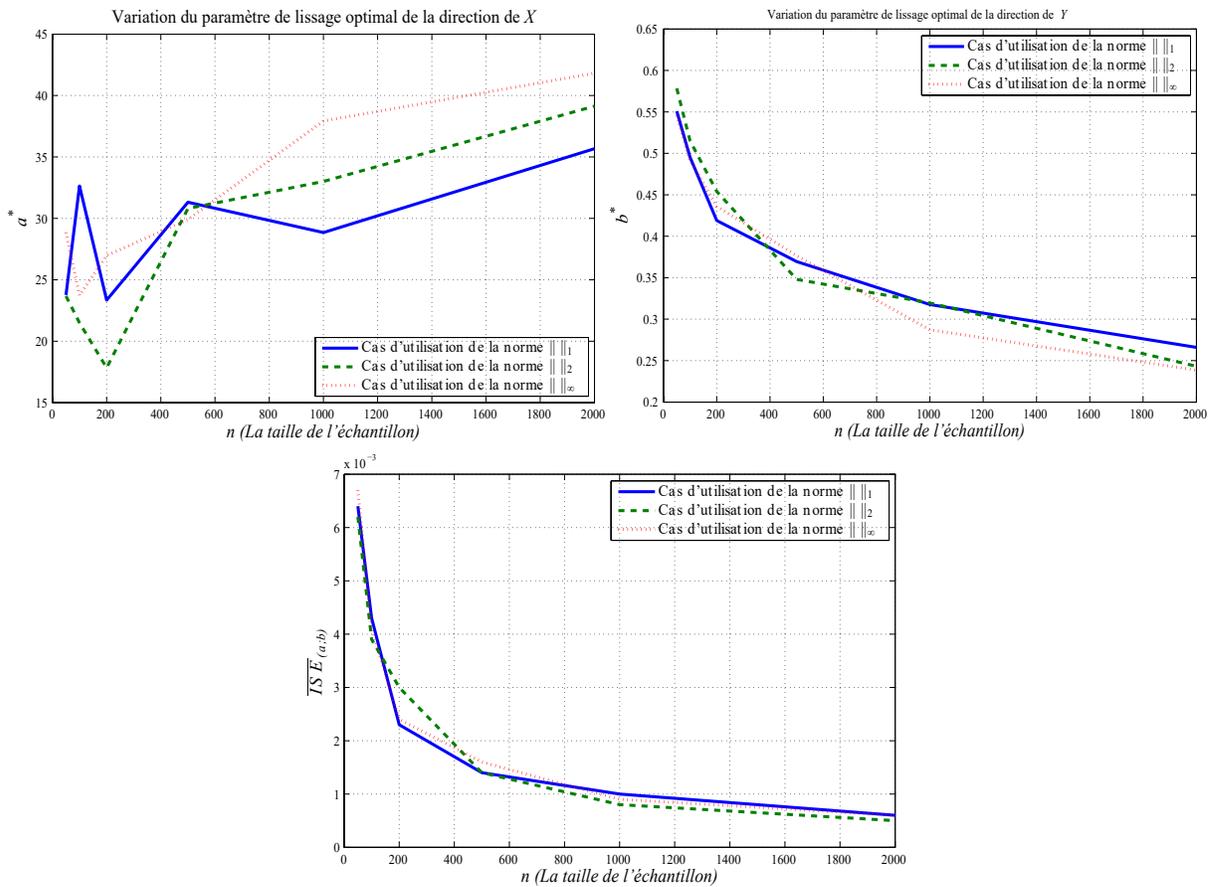


FIGURE 3.4 – Variation du (a^*, b^*) et du $\overline{ISE}_{(a^*, b^*)}$ en fonction de la taille de l'échantillon.

3.3 Discussion des résultats

En tenant compte des résultats numérique et graphiques obtenus dans la section précédente, on constate que :

- Dans toutes les situations considérées, les paramètres de lissage optimal décroissent en n et à mesure que la taille de l'échantillon augmente ce qui coïncide avec la propriété fondamentale (condition) du paramètre de lissage suivante : $\lim_{n \rightarrow \infty} h(n) = 0$.
- Indépendamment de la norme utilisée les estimateurs $\hat{f}_{ab}(y/E(X))$ et $\hat{f}_h(y/E(X))$ convergent vers $f(y/E(X))$ en L_2 (ISE moyenne) et ceci peut être justifié par la décroissance (convergence vers zéro) du ISE moyenne, associée aux estimateurs en question, en fonction de la taille de l'échantillon n .
- L'estimateur $\hat{f}_{ab}(y/E(X))$ est plus performant, au sens du ISE moyenne, que l'estimateur $\hat{f}_h(y/E(X))$ et ceci indépendamment de la taille de l'échantillon et la norme utilisée pour la conception de ces deux estimateurs.
- Les trois normes utilisées, nous fournis pratiquement des estimateurs de même performance (ISE moyenne). Mais, on constate qu'il y a une légère préférence :
 - Pour la norme $\|\cdot\|_2$ lorsque la taille de l'échantillon est très petite.
 - Pour la norme $\|\cdot\|_1$ lorsque la taille de l'échantillon est moyenne.
 - Pour la norme $\|\cdot\|_2$ lorsque la taille de l'échantillon est grande.

Dans le deuxième chapitre, on a signalé que, à cause de la positivité de la distance entre x et x_i ($\|x - x_i\|_p \geq 0$), le noyau doit être défini sur un support positif lorsque la variable explicative est fonctionnelle. Nos résultats montrent que même pour un noyau symétrique on peut avoir des résultats raisonnables.

Conclusion

Dans ce chapitre à travers d'une application numérique, basée sur la simulation, nous avons mis en relief la notion de l'estimateur à noyau d'une densité conditionnelle lorsque

la variable explicative est une variable aléatoire fonctionnelle.

Les résultats numériquement obtenus dans cette étude indiquent que l'estimateur d'une densité conditionnelle lorsque la variable explicative est une variable aléatoire fonctionnelle converge au sens du *ISE* moyenne vers la densité cible et ceci quel que soit la norme utilisée pour calculer la distance entre les variables explicatives fonctionnelle et quel que soit l'hypothèse imposée sur les deux paramètres de lissages respectivement de la direction de X et de la direction de Y ($a = b = h$ ou $a \neq b$).

Conclusion générale

Dans ce mémoire nous avons considéré l'estimation à noyau d'une fonction de densité conditionnelle de Y sachant que $X = x$ où Y est univariée et x univariée, multivariée ou fonctionnelle. La fonction de densité conditionnelle a un intérêt dans différents domaines statistiques, et elle peut être considérée comme une généralisation à la fois de la densité de probabilité classique et de la régression.

En premier lieu, nous avons présenté la méthode du noyau pour l'estimation d'une densité de probabilité univariée (principe, propriétés, noyaux usuels, les procédures de sélection du paramètre de lissage,...).

En deuxième lieu, nous avons intéressé à l'estimation de la densité conditionnelle par la méthode du noyau. Après avoir défini l'estimateur de la densité conditionnelle dans le cas x et y sont des variables aléatoires univariées, nous avons présenté les expressions adéquates de ce dernier estimateur dans deux autres situations, à savoir : le cas où la variable explicative x est multivariée (vectorielle) et le cas où la variable explicative x est une fonction aléatoire. Nous avons abordé également le problème du choix du noyau et du paramètre de lissage dans ce cas. Nous sommes focalisés principalement sur le choix du paramètre de lissage le fait que le choix du noyau reste le même que de la densité classique, où les paramètres de lissage a et b sont considérés indépendants ($a \neq b$) et le cas où a et b sont supposé les mêmes ($a = b = h$).

Enfin, nous avons présenté une application numérique basée sur des échantillons simulés lorsque la variable explicative est fonctionnelle. Le but de l'application est d'une part de

mettre en évidence la mise en oeuvre de l'estimateur à noyau d'une densité conditionnelle lorsque la variable explicative est fonctionnelle d'autre part, d'analyser les performances de cet estimateur en fonction de la taille de l'échantillon, de l'hypothèse imposée sur les paramètres de lissage ($a \neq b$ ou $a = b$) et la norme utilisée pour sa conception.

Les résultats des simulations obtenus sur des échantillons de différentes tailles en utilisant le noyau normale et la norme $\|\cdot\|_p$, avec $p \in \{1, 2, \infty\}$ pour la construction de l'estimateur mettent en relief le problème du choix du paramètre de lissage, le problème du choix de la norme et de la convergence en ISE moyenne des différentes versions de l'estimateur à noyau de $f(y/x)$ lorsque la variable x est fonctionnelle.

Parmi les perspectives de ce travail, nous pouvons dégager plusieurs axes intéressants, tant sur le plan théorique que pratique :

- Considérer d'autres techniques d'estimation de sélection du paramètre de lissage.
- Considérer différentes noyaux ($K(x) \neq K(y)$).
- Considérer des études des cas réels (données réelles).
- Réaliser une étude similaire à celle exposée dans le présent mémoire sur d'autres champs d'application de la densité conditionnelle (régression, taux de hasard,...).

Bibliographie

- [1] Abou-Jaoude, S. (1976) *Sur une condition nécessaire et suffisante de L^1 convergence presque complète de l'estimateur de la partition fixe pour une densité*. C. R. Acad. Sci. Paris **283** : 1107 – 1110.
- [2] Bashtannyk, D. M., Hyndman, R. J. (2001) *Bandwidth selection for kernel conditional density estimation*. Computational statistics and data analysis **36** : 279 – 298.
- [3] Bowman, A. W. (1984) *An alternative method of cross-validation for the smoothing density estimates*. Biometrika **71** : 553 – 560.
- [4] Burman, P. (1985) *A Data Dependent Approach to Density Estimation*. of Zeitschrift Für Wahrscheinlichke its theorie and Verwandte Gebiete **69** : 609 – 628.
- [5] Devroye, L. (1983) *The Equivalence of Weak, Strong and Complete Convergence L^1 for Kernel Density Estimates*. The Annals of Statistics **11** : 896 – 904.
- [6] Hall, P. (1982) *Cross-validation in density estimation*. Biometrika **69** : 383 – 390.
- [7] Hall, P., Marron, S. J. (1991) *Local minima in cross-validation function*. Journal of the royal statistical society **90** : 149 – 173.
- [8] Hyndman, R. J., Bashtannyk, D.J., Grunwald, G. K. (1996) *Estimating and visualizing conditional densities*. Journal of Computational and Graphical Statistics **5** : 315 – 336.
- [9] Nadaraya, E. (1965) *On nonparametric estimation density function and regression*. Theory Probab. P.P.L **10** : 186 – 190.

- [10] Park, B. U., Marron, S. J. (1990) *Comparison of data-driven bandwidth selectors*. Journal of the American Statistical Association **85** : 66 – 72.
- [11] Parzen, E. (1962) *On estimation of a probability density function and mode*. Ann. Math. Statist. **33** : 1065 – 1076.
- [12] Rosenblatt, M. (1956) *Remarks in some nonparametric estimates of a density function*. Ann. Math. Statist. **27** : 832 – 837.
- [13] Rosenblatt, M. (1969) *Conditional probability density and regression estimators*. Dans le livre : *Multivariate Analysis II* de P.R. Shnaiah. 25 – 31.
- [14] Rudemo, M. (1982) *Empirical choice of histogram and kernel density estimators*. Scandinavian Journal of Statistics. **9** : 65 – 78.
- [15] Scott, D. W. (1985) *Averaged shift histograms : effective nonparametric density estimators in several dimensions*. The Annals of Statistics **13** : 1024 – 1040.
- [16] Scott, D. W., Terrell, G. R. (1987) *Biased and unbiased cross-validation in density estimation*. Journal of the American Statistical Association **82** : 1131 – 1146.
- [17] Sheather, S. J., Jones, M. C. (1991) *A reliable data-based bandwidth selection method for kernel density estimation*. J. Roy. Statist. Soc. **B 53** : 683 – 690.
- [18] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- [19] Stone, C. (1984) *An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates*. The Annals of Statistics **12** : 1285 – 1297.
- [20] Wahba, G. (1975) *Optimal properties of variable knot, kernel and orthogonal series methods for density estimation*. Ann. Stat. **3** : 15 – 29.
- [21] Youndjé, E. (2011) *Contribution à l'estimation non-paramétrique par la méthode du noyau*. Mémoire d'Habilitation à Diriger des Recherches.

- [22] Zougab, N. (2007) *étude comparative des méthodes de sélection du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau*. Thèse de Magister en Mathématiques Appliquées, Université de Béjaïa.

Annexe A : Abréviations et Notations

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

$\hat{f}(\cdot)$:	Estimateur à noyau de la densité f .
$\hat{f}(y/x)$:	L'estimateur de la densité conditionnelle .
h	:	Paramètre de lissage.
(a, b)	:	Paramètres de lissage respectivement dans la direction de X et de Y .
h^*	:	Paramètre de lissage optimal.
$K(\cdot)$:	fonction Noyau.
$E(X)$:	L'espérance de la variable aléatoire X .
$var(X)$:	Variance de la variable aléatoire X .
$Biais(\hat{f}(\cdot))$:	Biais de l'estimateur
ISE	:	L'erreur quadratique intégrée.
$MISE$:	L'erreur quadratique moyenne intégrée.
$AMISE$:	L'erreur quadratique moyenne intégrée asymptotique.
$AMSE$:	L'erreur quadratique moyenne asymptotique.
UCV	:	Validation croisée non biaisée.
BCV	:	Validation croisée biaisée.
$\ \cdot\ _p$:	La p -distance.

Résumé

L'objectif de ce travail est de comprendre l'estimation à noyau de la densité conditionnelle d'une variable réelle Y conditionnée par une variable aléatoire X fonctionnelle ou multivariée.

L'étude de simulation réalisée, sur des échantillons de différentes tailles issus d'un modèle où la variable explicative est fonctionnelle, indique que l'utilisation des normes $\|\cdot\|_p$ ($p \in \{1, 2, \infty\}$), pour la construction de l'estimateur, nous fournis des estimateurs pratiquement de même performances au sens du ISE moyenne et ceci indépendamment de l'hypothèse imposée sur les paramètres de lissage de la direction de X et de la direction de Y ($a = b$ ou $a \neq b$). **Mots clés :** densité conditionnelle, estimateur à noyau, variable fonctionnelle, simulation, paramètre de lissage.

Abstract

The objective of this work is to understand the kernel estimation of the conditional density of a real variable Y conditioned by a functional or multivariate random variable X .

The simulation study carried out on different sizes-samples from a model where the explanatory variable is functional indicates that the use of the standards $\|\cdot\|_p$ ($p \in \{1, 2, \infty\}$), for the construction of the estimator, provides us, estimators practically having the same performance, within the meaning of the mean *ISE*, and this independently of the assumption imposed on the smoothing parameters of the directions of X and of Y ($a = b$ or $a \neq b$).

Keywords : conditional density, kernel estimator, functional variable, simulation, smoothing parameter.