République Algérienne Démocratique et Populaire

Ministére 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

Zouaoui Rania

Titre :

Choix du paramètre de lissage par validation croisée dans l'estimation à noyau d'une densité conditionnelle

Membres du Comité d'Examen :

Pr.	**NECIR Abdelhakim**	UMKB	Prèsident.
Dr.	**CHERFAOUI Mouloud**	UMKB	Encadreur.
Dr.	**KHEIREDDINE Souraya**	UMKB	Examinatrice.

Juin 2021

Dèdicace

Je dédié ce travail avec amour, gratitude et grande appréciation et respect à mon soutien inconditionnel et mon refuge dans cette vie "mes chers parents" qui m'ont toujours encouragé.

A mes chers frères et mes sœurs ***Louiza*** *et* ***Dalel***

A toute ma famille chacun par son nom.

A tous mes amies

A tous ceux qui me sont chers.

Remeciements

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

Je remercie particulièrement mon encadreur Dr. **Cherfaoui Mouloud** pour sa disponibilité, son soutien et ses remarques précieuses qui m'ont aidé à bien présenter ce travail. Je suis très reconnaissant pour son aide et ses conseils. Sans son aide, tous cela n'auraient pas été possible alors encore une fois merci Dr. **Cherfaoui Mouloud.**

Je voudrais exprimer ma gratitude à tous les enseignants du département des Mathématiques de la Faculté des Sciences exactes et des science de la nature et de la vie de Biskra qui nous ont enseignés et soutenus dans la poursuite de nos études.

Je tiens à remercier toute ma famille,Ma chère cousin **Zouaoui Rania** et ma chère amie **Guedjoudj Imene** pour leurs soutien moral tout au long de la préparation de ce mémoire et mes condisciples de la promotion 2021.

Enfin, je remercie chaleureusement toutes les personnes qui m'ont aidé, et qui ont contribué de proche ou de loin à la réalisation de ce travail.

Table des matières

Dédicace i

Remerciements ii

Table des matières iii

Liste des figures v

Liste des tableaux vi

Introduction générale 1

1 Estimation à noyau d'une densité de probabilité univariée 4

Introduction 4

1.1 Définitions et critères d'erreur 4

1.2 Estimateur à noyaux classiques 6

1.2.1 Propriétés d'un estimateur à noyau : 6

1.2.2 Choix du noyau 9

1.2.3 Choix du paramètre de lissage 11

Conclusion 22

2 Estimation à noyau d'une densité conditionnelle 23

Introduction 23

2.1 Définition de l'estimateur 23
2.2 Propriétés asymptotiques de l'estimateur 25
2.2.1 Biais et variance 25
2.2.2 Convergence en norme L^1, pour x_0 fixé 26
2.2.3 Convergence en norme L^2 : 26
2.3 Choix du noyau et du paramètre de lissage 27
2.3.1 Choix optimal 28
2.3.2 La règle de référence 29
2.3.3 Validation croisée 30
2.4 Conclusion 33
Conclusion 33

3 Performances d'un estimateur à noyau d'une densité $f(y/x)$ 34
Introduction 34
3.1 Présentation des paramètres de l'application 34
3.2 Résultats et discussion 36
3.2.1 Première application : Variation de $UCV(h)$ 36
3.2.2 Deuxième application : Performances des estimateurs 38
Conclusion 39

Conclusion générale 40

Bibliographie 42

Table des figures

1.1 La forme des différents noyaux usuels. 10

1.2 Illustration des phénomènes sur-lissage ($h = 0.8$), sous-lissage ($h = 0.1$) et estimation idéal ($h^* = 0 : 3334$). 12

3.1 Présentation graphique de $f(y/x)$ donnée dans 3.2. 35

3.2 Variation de l'$UCV(h)$ en fonction de h. 37

3.3 Variation du paramètre de lissage optimal et du ISE moyenne en fonction de la taille de l'échantillon n. 38

Liste des tableaux

1.1 Noyaux usuels et leurs supports. 10

3.1 Les paramètres de lissage optimaux et les ISE moyennes associées aux estimateurs $\hat{f}_{ab}$ et $\hat{f}_h$. 38

Introduction générale

La théorie de l'estimation en général et en particulier l'estimation d'une densité est une des préoccupations majeures des statisticiens. Soit X une variable aléatoire de densité de probabilité inconnue f. Supposons que nous avons n observations $(x_1, x_2\ , ...x_n)$ provenant de X. Le problème consiste à trouver un estimateur pour la fonction f à partir de cet échantillon issu de X.

Dans la littérature, il existe deux types d'approches pour estimer la densité de probabilité : La première s'appelle l'approche paramétrique qui suppose l'existence d'un modèle connu avec des paramètres inconnus, Son objectif généralement est de connaître une fonction de valeur du paramètre. Le principal inconvénient de cette approche est qu'elle nécessite une connaissance préalable du phénomène aléatoire. La deuxième approche dite non paramétrique, propose de laisser parler les données, sans spécifier de forme sur cette fonction à estimer.

On s'intéresse dans le cadre de ce travail à l'approche non paramétrique. Il existe plusieurs méthodes non paramétriques pour l'estimation de la densité de probabilité, comme la méthode de l'histogramme, la méthode d'estimation par les séries orthogonales et la méthode du noyau. Rosenblatt (1956) [15], et Parzen (1962) [14] sont les premiers à proposer une classe d'estimateurs à noyau d'une densité univariée. Cette méthode est la plus utilisée vu la simplicité de sa forme, la qualité de l'estimation qu'elle assure, ses modes de convergence multiples et sa flexibilité qui s'interprète par la liberté de l'utilisateur dans le choix du noyau K et du paramètre de lissage h.

L'une des fonctions usuelles et importantes dans la statistique est bien que la fonction de densité conditionnelle. Les fonctions de densité conditionnelles sont un moyen utile d'afficher l'incertitude. Rosenblatt (1969) a étudié le problème d'estimation de la densité de Y à condition que $X = x$ ou X et Y soient des variables aléatoires univariées. Une correction de biais a été proposée par Hyndman, Bashtannyk et Grunwald (1996) [11]. Fan, Yao et Tong (1996) ont proposé un estimateur direct basé sur une estimation polynomiale locale. D'autre part, le paramètre de lissage est un facteur important et crucial dans l'estimation de la fonction de densité conditionnelle par la méthode de noyau. Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Il existe deux catégories de méthodes ont été proposées dans la littérature de pour choisir ce paramètre.

La première catégorie repose sur la minimisation de l'erreur quadratique moyenne intégrée ($MISE$), D'un point de vue pratique. La deuxième catégorie est de type validation croisée : validation croisée de la vraisemblance (LCV) proposée par Habbema et al. [7], validation croisée non biaisée (UCV), qui a été proposée par Rudemo [16] et Bowman [3], validation croisée biaisée (BCV) proposée par Scott et Terrell [18], ... Cette classe de méthodes elle est intéressante en pratique car elle se laisse guider seulement par les observations. Un des principaux intérêts de ces méthodes est leur caractère direct.

L'objectif du présent travail est de comprendre la méthode de l'UCV pour le choix du paramètre de lissage optimale dans le cadre d'estimation à noyaux d'une densité conditionnelle $f(y/x)$. Plus précisément, l'objectif est de décortiquer les différentes composantes de l'estimateur sans biais de L'ISE pondérée, associée à l'estimateur à noyau d'une densité conditionnelle, obtenue par la méthode de validation croisée.

Pour réaliser atteindre notre objectif nous avons organisé le présent mémoire comme suit : suivants :

1. Dans le premier chapitre nous allons présenter un bref rappel sur l'estimateur à noyau d'une densité $f(x)$ univariée où nous avons focalisé principalement sur la construction

des méthodes de validation croisée pour le choix du paramètre de lissage optimal.

2. Dans le deuxième chapitre, Après l'introduction de la notion de l'estimateur à noyau d'une densité conditionnelle $f(y/x)$, nous allons aborder le problème du choix du noyau et du paramètre de lissage où nous allons appuyer particulièrement sur le choix du paramètre de lissage par la méthode de validation croisée UCV.

3. Avant de conclure, dans le troisième chapitre nous allons présenter une application numérique (simulation) dont l'objectif est de mettre en évidence le mécanisme de la méthode de validation croisée sans biais (UCV) pour le choix du paramètre de lissage dans l'estimation à noyau d'une densité conditionnelle.

Chapitre 1

Estimation à noyau d'une densité de probabilité univariée

Introduction

Dans ce chapitre, la première section est consacrée à un rappel sur certains critères d'erreurs liées à l'estimation d'une densité de probabilité.Dans la deuxième section, après avoir décrit l'origine de l'estimation à noyau, nous avons énoncé ses propriétés, les noyaux classiques et les techniques de sélection du paramètre de lissage.

1.1 Définitions et critères d'erreur

Définition 1.1.1 *Soit X une variable aléatoire absolument continue de densité de probabilité $f(x)$.L'espérance mathématique de X est définie par :*

$$E[X] = \int x f(x) dx$$

Définition 1.1.2 *La variance d'une variable aléatoire X absolument continue est définie*

par :

$$V[X] = E\left[X^2\right] - [E[X]]^2$$

Définition 1.1.3 *La moyenne empirique d'un échantillon de variables aléatoires réelles ou vectorielles* $(x_1, x_2\ , ...x_n)$*est défini par la moyenne arithmétique des valeurs :*

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Définition 1.1.4 *Soit* $(x_1, x_2\ , ...x_n)$ *un n-échantillon de loi* $f(x)$ *sur* $\mathbb{R}$*. La fonction de répartition est définie par :*

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

Pour estimer la similitude entre l'estimateur $\hat{f}$ et la vrais densité f à estimer, les critères utilisés pour mesurer l'erreur sont généralement :

- **L'erreur quadratique moyenne (MSE)**

$$\begin{aligned} MSE(f(x), \hat{f}(x)) &= \mathbb{E}\left(\hat{f}(x) - f(x)\right)^2 \\ &= \left[f(x) - \mathbb{E}\left(\hat{f}(x)\right)\right]^2 + \mathbb{E}\left(\hat{f}^2(x)\right) - \left[\mathbb{E}\left(\hat{f}(x)\right)\right]^2 \\ &= \mathbb{V}ar(\hat{f}(x)) + Biais^2\hat{f}(x). \end{aligned} \tag{1.1}$$

- **L'erreur quadratique intégrée (ISE)**

$$\begin{aligned} ISE(f, \hat{f}) &= \int [f(x) - \hat{f}(x)]^2 dx \\ &= \int \left[f(x)^2 - 2f(x)\hat{f}(x) + \hat{f}^2(x)\right] dx. \end{aligned}$$

- **L'erreur quadratique moyenne intégrée (MISE)**

$$MISE\left(f,\hat{f}\right) = \int MSE\left(f(x),\hat{f}(x)\right)dx = \int \mathbb{E}\left(f(x)-\hat{f}(x)\right)^2 dx$$
$$= \int \left[Biais^2\left(\hat{f}(x)\right) + \mathbb{V}ar\left(\hat{f}(x)\right)\right]dx.$$

1.2 Estimateur à noyaux classiques

Rosenblatt (1956) [15], suivi de Parzen (1962) [14], ont proposé une classe d'estimateurs à noyau d'une densité de probabilité. L'idée de construction de cet estimateur consiste à évaluer la densité f au point x, en comptant le nombre d'observations tombées dans un certain voisinage de $x \in \mathbb{R}$.

Définition 1.2.1 *Soit (X_n)une suite de variables aléatoires indépendantes et de même loi, de densité de probabilité inconnue f.On veut estimer f à partir d'un échantillon $(x_1, x_2\ , ...x_n)$. On appelle estimateur à noyau de Parzen-Rosenblatt de $f : \mathbb{R} \to \mathbb{R}_+$ la fonctionnelle donnée par :*

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

où $h = h(n)$ est appelé paramètre de lissage et la fonction K est appelée noyau. Tel que h et K vérifient respectivement les conditions (1.2) et (1.3).

$$h(n) \to 0 \text{ et } nh \to \infty \text{ quand } n \to \infty \tag{1.2}$$

$$K(u) = K(-u);\quad \int_{\mathbb{R}} K(u)du = 1;\quad \int_{\mathbb{R}} uK(u)du = 0;\quad \int_{\mathbb{R}} u^2K(u)du = \sigma_K^2 < \infty. \tag{1.3}$$

1.2.1 Propriétés d'un estimateur à noyau :

Nous présentons dans cette partie les principales propriétés statistiques de l'estimateur $\hat{f}$:

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 [20]) :

Espérance de l'estimateur :

$$E\left[\hat{f}\left(x\right)\right] = f(x) + \frac{h^2}{2} f^{''}(x)\mu_2(K) + o(h^2)$$

Le biais de l'estimateur $\hat{f}\left(x\right)$ est

$$\begin{aligned} Biais\left(\hat{f}\left(x\right)\right) &= E\left(\hat{f}\left(x\right)\right) - f\left(x\right) \\ &= \frac{h^2}{2} f^{''}\left(x\right)\mu_2(K) + o(h^2) \end{aligned}$$

Variance de l'estimateur :

$$\begin{aligned} V\left(\hat{f}\left(x\right)\right) &= \frac{f\left(x\right)}{nh}\int_{-\infty}^{+\infty} K^2\left(y\right)dy - \frac{f^{'}\left(x\right)}{n}\int_{-\infty}^{+\infty} yK^2\left(y\right)dy - \frac{1}{n}\left(f\left(x\right) + biais\hat{f}\left(x\right)\right)^2 \\ &= \frac{f\left(x\right)R\left(K\right)}{nh} + 0\left(\frac{1}{nh}\right) \end{aligned}$$

où $R\left(K\right) = \int_{-\infty}^{+\infty} K^2(u)du$; $\mu_2(K) = \int_{\mathbb{R}} y^2K(y)dy$

MSE et MISE de l'estimateur :

L'erreur quadratique moyenne de l'estimateur :

$$MSE\left(\hat{f}\left(x\right)\right) = \frac{f\left(x\right)R\left(K\right)}{nh} + \frac{h^4}{4}\mu_2^2(K)f^{''^2}\left(x\right) + 0\left(\frac{1}{nh} + h^4\right)$$

L'erreur quadratique moyenne intégrée de l'estimateur :

$$MISE\left(\hat{f}\right) = \frac{R\left(K\right)}{nh} + \frac{h^4}{4}\mu_2^2(K)R\left(f^{''}\right) + 0\left(\frac{1}{nh} + h^4\right)$$

où $R(g) = \int_{\mathbb{R}} g^2(x)dx$.

Comportement asymptotique de l'estimateur à noyau

Parzen [14] 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 *Soit f une densité continue et $\hat{f}$ son estimateur. Si le noyau K vérifie :*

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

on a :

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

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

$$\lim_{n\to\infty} MSE\left(\hat{f}(x), f(x)\right) = 0$$

$$\hat{f}(x) \overset{loi}{\to} \mathcal{N}\left(E\left(\hat{f}(x)\right), V\left(\hat{f}(x)\right)\right)$$

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

Si de plus, f une densité continue, de puissance $p^{ème}$-intégrable :

$$\lim_{n\to\infty} MISE\left(\hat{f}, f\right) = 0$$

Le Théorème suivant donne le résultat élaboré par Nadaraya [12] concernant la convergence uniforme presque complète de l'estimateur à noyau classique.

Théorème 1.2.2 *Soit f une densité uniformément continue et son estimateur à noyau K positif et à variations bornées :*
Si $\lim_{n\to\infty} h(n) = 0$ *et* $\sum_{i=1}^{\infty} \exp(-\gamma n h^2) < \infty, \forall \gamma > 0$:

$$\sup_{x\in\mathbb{R}} \left| \hat{f}(x) - f(x) \right| \to 0 \text{ avec une probabilité 1}$$

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,*

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

alors,

$$\forall f \in \mathcal{F}, \ \lim_{n\longrightarrow\infty} \int |\hat{f}(x) - f(x)| dx = 0, \ \textit{Presque Complétement},$$

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

Vitesse de convergence

Wahba [22] 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.

1.2.2 Choix du noyau

Dans la littérature, il existe plusieurs fonctions qui jouent le rôle d'un noyau, la Table 1.1 résume les noyaux les plus usuels dont leurs formes sont illustrées dans la Figure 1.1.

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	$K(u) = \frac{\pi}{4}\cos(\frac{\pi u}{2})$	$\|u\| \leq 1$
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	$K_E(u) = \frac{3}{4\sqrt{5}}\left(1 - \frac{u^2}{5}\right)$	$\|u\| \leq \sqrt{5}$

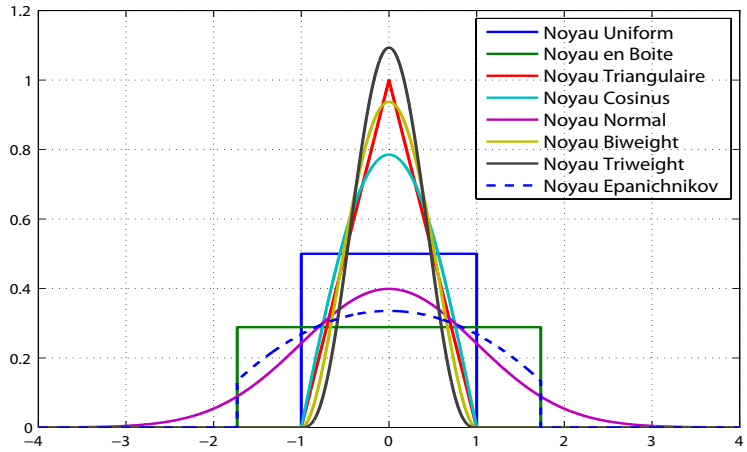

FIGURE 1.1 – La forme des différents noyaux usuels.

Nous pouvons considérer l'efficacité d'un noyau K (notée $eff(K)$) quelconque en comparant son $AMISE$ avec celui du noyau Epanechnikov K_e puisque ce dernier minimise le $AMISE$ lorsque h est choisi de façon optimale. Alors le critère d'efficacité peut être simplifié comme suit :

$$eff\,(K) = \frac{AMISE\,(K_e)}{AMISE\,(K)} \leq 1$$

Le tableau suivant donne quelques noyaux et leurs efficacités respectives :

Noyau	Efficacité
Epanchnikov	1.000
Quadratique	0.9939
Gaussien	0.9512
Triangulaire	0.9859
Uniforme	0.9295

A partir de ce dernier tableau,on constate que les valeurs d'efficacité des noyaux sont très proches, cela peut justifier que le choix du noyau dans l'estimation d'une densité n'est pas important.

1.2.3 Choix du paramètre de lissage

L'estimation de la fonction de probabilité par la méthode des noyaux est principalement conditionnée par le paramètre de lissage h.Ce paramètre est indispensable pour l'efficacité du lissage et la qualité de l'estimation .dans le sens où une petite perturbation de h est suffisante pour que les caractéristiques de $\hat{f}$ changent complétement (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.À l'opposé, lorsque h est trop grand, un sur-lissage risque de camoufler les particularités de la véritable fonction de densité.

L'exemple présenté dans la Figure 1.2, réalisé dans le cadre d'estimation d'une densité d'une loi normale, centrée réduite, à partir d'un échantillon de taille $n = 200$, est une illustration de l'influence du choix du paramètre de lissage sur les caractéristiques graphiques de l'estimateur en question.Les graphes des trois estimateurs présentés, mis en évidence le phénomène de sur-lissage dans le cas $h = 0.8$ (trop grand), le phénomène de sous-lissage dans le cas $h = 0.1$ (trop petit) et l'estimation idéale dans le cas $h^* = 0.3334$ (h^* est

l'optimal au sens du ISE).

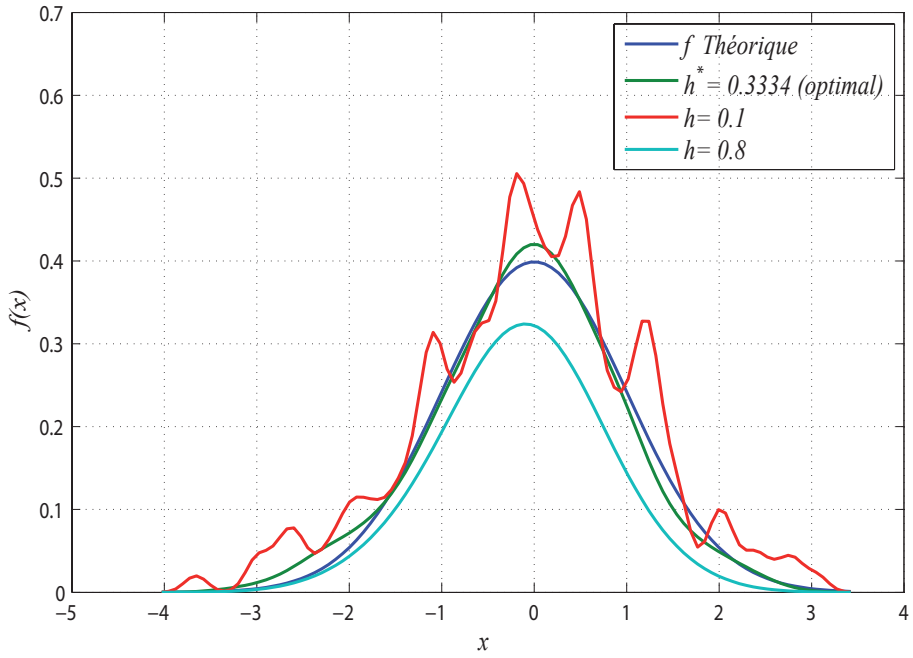

FIGURE 1.2 – Illustration des phénomènes sur-lissage ($h = 0.8$), sous-lissage ($h = 0.1$) et estimation idéal ($h^* = 0 : 3334$).

Plusieurs méthodes pour choisir ce paramètre ont été développées dans la littérature et qui sont regroupées en deux familles :

– Méthodes de plug-in (ré-injection),

– Méthodes de Cross-Validation (Validation-croisée).

Toutes ces méthodes nous fournissent un paramètre de lissage qui est optimale pour la distribution à estimer.Celles-ci diffèrent au niveau du choix du critère à optimiser, mais ces méthodes ont toujours des inconvénients, 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.

Méthodes plug-in (ré-injection)

Estimateur optimal : La décision d'un choix optimal pour le paramètre de lissage suppose la spécification d'un critère d'erreur qui puisse être optimisé.Dans ce sens, il est claire que l'optimalité n'est pas un concept absolu : elle est liée aux choix du critère qui peut faire intervenir à la fois la densité inconnue f et l'estimateur $\hat{f}$ (donc h et le noyau K). Supposons qu'on cherche à minimiser l'erreur Quadratique Intégrée Moyenne ($MISE$), c'est à dire :

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

Pour déterminer le paramètre de lissage $h^\star$ qui minimise $MISE$ nous avons besoin du résultat suivant :

Théorème 1.2.4 *(Scott [17])*
Si f a une dérivée seconde absolument continue, $f^{(2)} \in \mathbb{L}^2$, le noyau $K \in \mathbb{L}^2$ et une densité de probabilité continue, symétrique de variance $\sigma_K^2 > 0$, alors, sous les conditions :

$$h\left(n\right) \to 0 \text{ et } nh\left(n\right) \to \infty$$

$$MISE = \frac{h^4}{4}\sigma_K^4 \int \left(f^{''}\left(x\right)dx\right)^2 + \frac{\int K^2\left(x\right)dx}{nh} + 0\left(\frac{1}{n}+h^5\right) \tag{1.4}$$

où $\mathbb{L}^2$ est l'ensemble des fonctions f définies sur $\mathbb{R}$, telles que : $\int \left|f\left(x\right)\right|^2 dx < \infty$
L'erreur quadratique Intégrée Moyenne Asymptotique $AMISE = MISE - o(h^5 + \frac{1}{n})$ est :

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

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 :

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

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

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

où h^* peut aussi s'écrire :

$$h^* = \psi(K)\varphi(f)n^{-1/5}, \tag{1.7}$$

et

$$\psi(K) = \left[\frac{R(K)}{\sigma_K^4}\right]^{1/5}, \quad \varphi(f) = \left[\frac{1}{R(f'')}\right]^{1/5}, \quad \text{pour} \quad R(f'') \neq 0,$$

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 :

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

Outre sa nature asymptotique, la largeur de fenêtre optimale $h^\star$ dépend de la densité inconnue f à travers $R\left(f^{\shortparallel}\right)$.Cette largeur de fenêtre "idéale" (relativement au critère d'erreur retenu) n'est donc pas directement calculable. Une façon classique de remédier à ce dernier problème consiste à remplacer la quantité $R\left(f^{\shortparallel}\right)$ par un estimateur approprié ou une quantité bien définie.

Estimateur "Rule Of Thumb" (règle de référence) : L'estimateur "Rule Of Thumb "du paramètre de lissage, noté h_{rot}. Pour f comme étant la distribution normale centrée de moyenne 0 et de variance σ^2, on a alors :

$$R\left(f^{\shortparallel}\right) = \int \left(f^{\shortparallel}(x)\right)^2 dx = \frac{3}{8}\pi^{\frac{1}{2}}\sigma^{-5} \tag{1.8}$$

La valeur h_{rot} est obtenue en substituant la valeur obtenue dans la formule (1.7) :

$$\begin{aligned} h_{rot} &= (4\pi)^{-\frac{1}{10}} \left(\frac{3}{8}\pi^{-\frac{1}{2}}\sigma\right) n^{\frac{-1}{5}} \\ &= \left(\frac{4}{3}\right)^{\frac{1}{5}} \sigma n^{\frac{-1}{5}} \\ &= 1.06\sigma n^{\frac{-1}{5}} \end{aligned} \tag{1.9}$$

Il suffit donc d'estimer σ à partir des données et de substituer cet estimateur dans la formule (1.9).D'après Silverman (1986) [20], cette formule donnera de bon résultat si la population est réellement normalement distribuée mais celle-ci peut donner une distribution trop lissée si la population est plutôt multimodale.Dans ce cas, de meilleurs résultat peuvent être obtenus, si on utilise l'interquartile IQ définie par :

$$IQ = \frac{X_{\left(\frac{3n}{4}\right)} - X_{\left(\frac{n}{4}\right)}}{1.34}$$

où $X_{\left(\frac{3n}{4}\right)}$ et $X_{\left(\frac{n}{4}\right)}$ représente le troisième quartile et le premier quartile respectivement de l'échantillon observé.Si X suit une loi normale, l'interquartile est $IQ = 1.394$ alors h_{rot}

de l'équation (1.9) devient :

$$h_{rot} = 1.06IQn^{\frac{-1}{5}}$$

Cette dernière formule peut aussi donner une distribution trop lissée si la vraie densité est multimodale.Parfois cette dernière donne 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 dans la formule (1.9) où A est défini par $A = min(\sigma, IQ)$, donc la formule pour h_{rot} devient alors :

$$h_{rot} = 1.06An^{\frac{-1}{5}}$$

Cette expression elle est moins efficace dans le cas des distributions multimodales et asymétriques.

Estimateur de Sheather et Jones : Sheather et Jones (1991) [19], recommandent l'utilisation de l'estimateur naturel $\hat{R}_a\left(f^{\shortparallel}\right)$, en faisant observer que le terme de biais $\frac{R(K^{\shortparallel})}{na^5}$ est positif et peut donc servir à annuler le terme de biais (négatif) de l'erreur quadratique moyenne entre $\hat{R}_a\left(f^{\shortparallel}\right)$ et $R\left(f^{\shortparallel}\right)$. Afin de faire disparaitre quelques effets indésirables du terme du biais.Sheather et Jones [19], choisissent d'estimer $R\left(f^{\shortparallel}\right)$ par :

$$S\left(a\right) = \frac{1}{n\left(n-1\right)a^5}\sum_{i=1}^{n}\sum_{j=1}^{n}L^{(4)}\left(\frac{x_i - x_j}{a}\right) \tag{1.10}$$

avec : $L^{(4)}$ c'est 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 :

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

Le paramètre de lissage optimal est estimé par :

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

avec $\int_{\mathbb{R}} u^2 K(u) du = \sigma_K^2$ qui peut être réécrite sous la forme :

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

avec :

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

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

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

$a = 0.920\hat{\lambda} n^{\frac{-1}{7}}$ et $b = 0.912\hat{\lambda} n^{-\frac{1}{9}}$ et $\hat{\lambda}$ (estimateur de λ) représente une mesure d'échelle de f

Méthodes de Validation Croisée :

Validation croisée de la vraisemblance (LCV) : Habbema, Hermans et Vandenbroek en 1974 [7] qui ont proposé cette méthode fondée sur un critère non asymptotique du maximum de vraisemblance, mais l'interprétation entre les données est purement heuristique.Pour f_h est l'estimateur de la densité f obtenu par la méthode du noyau. Rappelons que la vraisemblance pour l'échantillon $(x_i)_{(1 \leq i \leq n)}$ est défini par :

$$LCV(h) = \prod_{i=1}^{n} f_{h,i}(x_i)$$

où $f_{h,i}(x_i) = \frac{1}{(n-1)h} \sum_{\substack{i=1 \\ j \neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right), i = \overline{1,n}$ est l'estimateur de la densité construit à

partir de l'ensemble de points sauf au point x_i. Alors,

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

pour K un noyau gaussien on obtient :

$$LCV\left(h\right)=\prod_{i=1}^{n}\frac{1}{\left(n-1\right)h}\sum_{\substack{i=1\\j\neq i}}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x_i-x_j}{h}\right)^2\right) \tag{1.11}$$

Le paramètre de lissage h choisi par cette méthode est la valeur de h qui maximise $LCV\left(h\right)$

$$h_{lcv}^{*}=\arg\max_{h}\left(LCV\left(h\right)\right).$$

Validation croisée non biaisée (UCV) La méthode de validation croisée non biaisée a été proposée par Rudemo [16] en 1982 et Bowman [3] en 1984.Soit f_h l'estimateur de f. Le paramètre de lissage choisi pour la méthode de la validation croisée est la valeur de ce paramètre h qui minimise un estimateur du ISE.On peut choisir le paramètre de lissage de façon à ce qu'il minimise un estimateur de :

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

Puisque $\int_{\mathbb{R}}f^2(x)dx$ ne dépend pas du paramètre de lissage h.Tout d'abord, nous voulons

trouver une estimation pour $\int_{\mathbb{R}} f_h(x) f(x) dx$. Noter que :

$$\int_{\mathbb{R}} f_h(x) f(x) dx = E(f_h(x))$$

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

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

et le critère à optimiser est :

$$UCV(h) = \int_{\mathbb{R}} f_h^2(x)dx - \frac{2}{n}\sum_{i=1}^{n} f_{h,i}(x_i), \tag{1.12}$$

où $f_{h,i}(x_i) = \frac{1}{(n-1)h}\sum_{\substack{i=1\\ j\neq i}}^{n} K\left(\frac{x_i - x_j}{h}\right), i = \overline{1,n}$ est l'estimateur à noyau basé sur les $(n-1)$ observations différentes de x_i.

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 [21].

Théorème 1.2.5 *(Stone) :Soit h_{ucv} l'estimateur de h qui minimise $UCV(h)$ et f_h l'estimateur de f :*

$$\frac{\int (f(x) - f_{h_{ucv}}(x))^2 dx}{\inf_h \int_{\mathbb{R}} [f(x) - f_h(x)]^2 dx} \xrightarrow{p.s} 1$$

Le théorème suivant nous donne la moyenne et la variance du $UCV(h)$:

Théorème 1.2.6 *Pour h fixé et K noyau positif*

$$E[UCV(h)] = AMISE - R(f) + 0\left(h^4 + \frac{1}{nh}\right)$$

$$V[UCV(h)] = \frac{4}{n}\left[R\left(f^{\frac{3}{2}}\right) - R^2(f)\right] + 0\left(h^4 + \frac{1}{n^2 h}\right)$$

Nous donnons maintenant le théorème sur la normalité asymptotique.

Théorème 1.2.7 *Si K est un noyau positif satisfait aux conditions suivantes :*

1. *$f^{'''}$ est absolument continue, $f^{(4)}$ intégrable, $R(f^{(4)}\sqrt{f}) < \infty$ et $R(\sqrt{f^{(4)}}f) < \infty$;*
2. *$K \geq 0$ un noyau symétrique avec $\mu_2 > 0$, $K^{'}$est continue :*

$$h_{ucv} \overset{loi}{\to} \mathcal{N}\left(h^{\star}; \frac{2R(\rho)\,R(f)}{25n^2h^{\star 7}\mu_2^4 R\left(f^{''}\right)^2}\right)$$

avec $\rho(c) = cRK(w)K^{'}(w+c)dw - 2cK^{'}(c)$, $-2 \leq c \leq 2$ et $h^{\star} = Cn^{-\frac{1}{5}}$ .Le paramètre de lissage optimal défini dans (1.7)

L'écart type de $h_{ucv} - h^{\star}$ est défini par :

$$\sqrt{V\left[h_{ucv} - h^{\star}\right]} = \frac{\sqrt{2}C^{-\frac{7}{2}}}{5\mu_2^2 R\left(f^{''}\right)}\left[R(\rho)\,R(f)\right]^{\frac{1}{2}} n^{\frac{3}{10}}$$

Proposition 1.2.1 *Soit $(X_1, X_2, ..., X_n)$ un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f.En utilisant le noyau gaussien, on obtient :*

$$UCV(h) = \frac{1}{2n^2h\sqrt{\pi}}\left(n + 2\sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n}\exp\left(-\left(\frac{x_i - x_j}{2h}\right)^2\right)\right) - \frac{2}{\sqrt{2}\pi n(n-1)h}\sum_{i=1}^{n}\sum_{i\neq j,j=1}^{n}\exp\left(-\frac{(x_i - x_j)^2}{2h^2}\right).$$

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 [10]. Pour d'autres études voir Hall [9], Burman [4], Scott et Terrell [18].

Validation croisée biaisée (BCV) Le critère de validation croisée biaisée, a été introduit par Scott et Terrell (1987) [18].L'idée de cette méthode est de trouver la valeur de h qui minimise un estimateur du $AMISE$.On a :

$$AMISE = \frac{h^4}{4}\sigma_K^4 R(f'') + \frac{R(K)}{nh}. \tag{1.13}$$

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 (1.13), il est claire 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 [18] ont proposé la forme de l'estimateur de $AMISE$ à minimiser (critère de $BCV(h)$), qui se résume comme suit :

Proposition 1.2.2 *(Scott et Trell [18]) Soit $(X_1, X_2, ..., X_n)$ un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f.Pour un noyau K, on obtient :*

$$BCV(h) = \frac{R(K)}{nh} + h^4\frac{\mu_2^2(K)}{4n^2}\sum_i\sum_{j,j\neq i} K_h^{(2)}K_h^{(2)}(X_i - X_j). \tag{1.14}$$

Si on considéré que le noyau K est un noyau gaussien alors la Proposition 1.2.2 peut être réécrite sous la forme suivante :

Proposition 1.2.3 *Soit ($X_1, X_2, ..., X_n$) un n-échantillon i.i.d issu d'une variable aléatoire X de fonction de densité f.En choisissant le noyau gaussien, on obtient :*

$$BCV(h) = \frac{1}{2nh\sqrt{\pi}} + \frac{1}{64n^2h\sqrt{\pi}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left[\left(\frac{x_i - x_j}{h}\right)^4 - 12\left(\frac{x_i - x_j}{h}\right)^2 + 12\right]\times$$
$$\exp\left[-\frac{(x_i - x_j)^2}{4h^2}\right].$$

le paramètre de lissage h_{bcv} *qui minimise* $BCV(h)$:

$$h_{bcv} = \arg\min_{h} \left(BCV(h)\right)$$

La méthode de validation croisée biaisée a été introduite 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.Cette méthode nous fournis plusieurs minimums locaux pour la fonctionnelle cible à minimiser.

Conclusion

Dans ce chapitre, nous avons présenté un rappelle sur l'estimateur à noyau d'une densité de probabilité (sa définition, le choix de noyau, le choix du paramètre de lissage) où nous nous somme focalisés principalement sur le choix du paramètre de lissage par les méthodes de validation croisée.

Chapitre 2

Estimation à noyau d'une densité conditionnelle

Introduction

L'estimateur à noyau d'une densité est l'un des estimateurs les plus étudiés et les plus performants. Dans ce chapitre nous allons vous présenter l'estimateur à noyau de la densité conditionnelle En 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.... Ensuite, nous nous sommes concentrés sur le problème du choix du noyau et de paramètre de lissage, en particulier par la méthode UCV, dans ce cas.

2.1 Définition de l'estimateur

X et Y sont des variables aléatoires univariées. Si $g(x, y)$ est la densité joint de (X, Y) et $m(x)$ est la densité marginal de X.On définit la densité conditionnelle de $Y/(X = x)$ comme suit :

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

Hyndmanet al. (1996) [11] ont considéré la forme modifiée de l'estimateur de Rosenblatt pour définir l'estimateur de la densité conditionnelle, donnée par :

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

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

$$\hat{g}(x,y) = \frac{1}{nab}\sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right) K\left(\frac{y-Y_j}{b}\right), \tag{2.2}$$

$$\hat{m}(x) = \frac{1}{na}\sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right), \tag{2.3}$$

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$\{(X1,Y1),...,(Xn,Yn)\}$ un échantillon aléatoire du couple (X,Y) $i.i.d$ qui est à valeurs dans $\mathbb{R}$.

D'après les formules (2.2) et (2.3) l'estimateur de la densité conditionnelle s'écrit comme suit :

$$\hat{f}(y/x) = \frac{\frac{1}{nab}\sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right) K\left(\frac{y-Y_j}{b}\right)}{\frac{1}{na}\sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right)} = \frac{\sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right) K\left(\frac{y-Y_j}{b}\right)}{b \sum_{j=1}^{n} K\left(\frac{x-X_j}{a}\right)}. \tag{2.4}$$

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

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

On peut réécrire (2.4) comme suit :

$$\hat{f}\left(y/x\right)=\frac{1}{b}\sum_{j=1}^{n}w_j\left(x\right)K\left(\frac{y-Y_j}{b}\right)\ \text{avec}\ w_j\left(x\right)=\frac{K\left(\frac{x-X_j}{a}\right)}{\sum\limits_{j=1}^{n}K\left(\frac{x-X_j}{a}\right)}.$$

2.2 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).

2.2.1 Biais et variance

En se basant sur la dérivation et le développement de Taylor à l'ordre 2, Hyndman et al. [11] ont obtenu la forme asymptotique du biais et de la variance de l'estimateur.

Le biais de l'estimateur :

$$\begin{aligned}E\left[\hat{f}(y/x)\right]-f(y/x)\ &=\frac{a^2\sigma_K^2}{2}\left\{2\frac{m'(x)}{m(x)}\frac{\partial f(y/x)}{\partial x}+\frac{\partial^2 f(y/x)}{\partial x^2}+\frac{b^2}{a^2}\frac{\partial^2 f(y/x)}{\partial y^2}\right\}\\&\quad -o(a^4)+o\left(b^4\right)+o\left(a^2b^2\right)+o\left(\frac{1}{na}\right),\end{aligned}\tag{2.5}$$

La variance de l'estimateur :

$$V[\hat{f}(y/x)]=\frac{R(K)f(y/x)}{nabm(x)}\left[R(K)-bf(y/x)\right]+o\left(\frac{1}{n}\right)+o\left(\frac{b}{an}\right)+o\left(\frac{a}{bn}\right),\tag{2.6}$$

où $R(K)=\int K^2(u)du$, à condition que $a\to 0$ et $b\to 0$ lorsque $n\to\infty$. Youndjé [23] 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 $\mathcal{P}_{\alpha}^{\beta}$ possédant les propriétés suivantes :

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

2.2.2 Convergence en norme $\mathbf{L}^1$, 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 $y \longmapsto f(y/x_0)$ 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 [23] 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 complétement.

2.2.3 Convergence en norme L^2 :

Sachant que l'erreur quadratique moyenne (MSE) est la somme du carré du biais (2.5) avec la variance (2.6), alors l'erreur quadratique moyenne asymptotique ($AMSE$) est de la forme :

$$\begin{aligned} AMSE\left(\hat{f}(x), f(x)\right) = \ & \frac{a^4\sigma_K^4}{4}\left[2\frac{m'(x)}{m(x)}\frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} + \frac{b^2}{a^2}\frac{\partial^2 f(y/x)}{\partial y^2}\right]^2 \\ & + \frac{R(K)f(y/x)}{nabm(x)}\left[R(K) - bf(y/x)\right] + o\left(\frac{1}{n}\right) + o\left(\frac{b}{an}\right) \\ & + o\left(\frac{a}{bn}\right) + o(a^6) + o(b^6) + o(a^2b^4) + o(a^4b^2). \end{aligned} \tag{2.7}$$

On constate que cet estimateur est constitué à condition que $a \to 0$, $b \to 0$ et $nab \to \infty$ lorsque $n \to \infty$.

Comme d'autres problèmes de lissage, les petits paramètres de lissage donnent des petits biais et grandes variances, alors que les grands paramètres de lissage donnent des grands biais et petites variances. Les paramètres de lissage qui sont choisis pour minimiser (2.7) 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 (2.7) avec $m(x)$, sa forme est donnée comme suit :

$$MISE \approx \frac{c_1}{nab} - \frac{c_2}{na} + c_3 a^4 + c_4 b^4 + c_5 a^2 b^2, \tag{2.8}$$

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 :

$$\begin{cases} c_1 = \int R^2(K)dx, \\ c_2 = \int\int R(K) f^2(y/x)dydx, \\ c_3 = \int\int \frac{\sigma_K^4 m(x)}{4} \left\{ 2\frac{m'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} \right\}^2 dydx, \\ c_4 = \int\int \frac{\sigma_K^4 m(x)}{4} \left\{ \frac{\partial^2 f(y/x)}{\partial y^2} \right\}^2 dydx, \\ c_5 = \int\int \frac{\sigma_K^4 m(x)}{2} \left\{ 2\frac{m'(x)}{m(x)} \frac{\partial f(y/x)}{\partial x} + \frac{\partial^2 f(y/x)}{\partial x^2} \right\} \left\{ \frac{\partial^2 f(y/x)}{\partial y^2} \right\} dydx, \end{cases} \tag{2.9}$$

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

2.3 Choix du noyau et du paramètre de lissage

L'estimation de la fonction de densité probabilité univariée et unimodale par la méthode des noyaux est principalement conditionnée par le choix d'un noyau K et d'un paramètre de lissage h. Donc le noyau K peut être sélectionné parmi les fonctions rangées dans la table 1.1 et c'est ce que nous avons réalisé dans le premier chapitre.

On s'intéresse à étudier les procédures de sélection du paramètre de lissage h 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 :

$$MISE(a,b,\hat{f},f) = \iint E\{\hat{f}(y/x) - f(y/x)\}^2 m(x)dxdy. \tag{2.10}$$

L'erreur quadratique intégrée :

$$ISE(a,b,\hat{f},f) = \iint \{\hat{f}(y/x) - f(y/x)\}^2 m(x)dxdy. \tag{2.11}$$

L'erreur quadratique intégrée asymptotique :

$$AISE = \frac{\Delta}{n}\sum_{j=1}^{N}\sum_{i=1}^{n}\left[\hat{f}(y'_j/X_i) - f(y'_j/X_i)\right]^2, \tag{2.12}$$

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

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

2.3.1 Choix optimal

Les largeurs de fenêtres optimales peuvent être obtenues par la différentiation de (2.8) par rapport à a et b et en fixant les dérivés à zéro. En simplifiant ces dérivés, nous obtenons le système d'équations suivant :

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

Hyndman et al. (1996) [11] 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 $(i = 1, ..., 5)$ sont définie précédemment.On remarque que a^*et b^*ne sont pas calculables car ils dépendent des fonctions inconnues $m(x)$ et $f(y/x)$.

2.3.2 La règle de référence

Cette méthode 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. Bashtannyk et Hyndman (2001) [2] 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, $[Y/X = x] \overset{L}{\rightsquigarrow} N\left(u + vx, (p + qx)^2\right)$ et la densité conditionnelle est :

$$f(y/x) = \frac{1}{(p + qx)\sqrt{2\pi}} \exp\left\{\frac{-1}{2(p + qx)^2}(y - u - vx)^2\right\}. \tag{2.13}$$

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 $[\alpha, \beta]$, on remplace la densité conditionnelle $f(y/x)$ (2.13) et la densité marginale dans les constants $c_1, ..., c_5$ (2.9), ensuite on fait l'intégration deux fois par rapport à x et y. On trouve ces constants en fonction de p, q, $R(K)$, α et β. Alors, sous la condition $v \neq 0$, la

règle de référence est donnée comme suit :

- **Cas $q \neq 0$:**

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

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

- **Cas $q = 0$:**

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

Cas de loi normale

Si $m(x)$ est une loi normale avec une moyenne constante μ_h et une variance constante σ_h^2. Ils supposent également que la densité (2.13) a une variance constante, $q = 0$. En refaisant les mêmes étapes que le cas uniforme, ils ont trouvé la règle de référence :

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

avec, $\lambda(k) = \int_{-\infty}^{k} \phi(t)dt$, $\phi(.)$ 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.

2.3.3 Validation croisée

Fan and Yim (2004) [6]and Hall, Racine and Li (2004) [8]ont proposé une méthode de validation croisée appropriée pour les estimateurs de densité conditionnels non paramétriques. Rappelons que, si a et b deux paramètres de lissage et K étant un noyau défini

sur $\mathbb{R}$ alors l'estimateur de la densité conditionnelle sera définie par :

$$\hat{f}(y/x) = \frac{\hat{g}(y,x)}{\hat{m}(x)} = \frac{\sum_{i=1}^{n} K_a(x - X_i) K_b(y - Y_i)}{\sum_{i=1}^{n} K_a(x - X_i)}, \tag{2.14}$$

où $a = c_1 n^{-\frac{1}{6}}$, $b = c_2 n^{-\frac{1}{6}}$ avec $c_1 > 0,\ c_2 > 0$, $K_a(u) = \frac{1}{a} K\left(\frac{u}{a}\right)$ et $K_b(u) = \frac{1}{b} K\left(\frac{u}{b}\right)$ $\hat{g}(y,x) = \frac{1}{n} \sum_{i=1}^{n} K_a(x - X_i) K_b(y - Y_i)$ et $\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} K_a(x - X_i)$ sont les estimateurs à noyau respectivement de $g(y,x)$ et $m(x)$.

Le choix du paramètre de lissage avec la méthode de validation croisée, il est basé sur la minimisation de l'estimateur l'erreur quadratique intégrée pondérée définie par :

$$ISE(\hat{f}, f) = \iint \left\{\hat{f}(y/x) - f(y/x)\right\}^2 W(x) W'(y) dxdy, \tag{2.15}$$

où W et W' sont des fonctions de poids positives.

Dans la pratique les deux fonctions de pondération W et W' doivent être fixées prélablement au choix de l'utilisateur. Mais généralement, ces deux fonctions sont fixées comme suit :

$$W(x) = f(x) \quad \text{et} \quad W'(y) = 1.$$

Ainsi, l'expression (2.15) sera réécrite comme suite :

$$\begin{aligned} ISE\left(\hat{f}, f\right) &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\hat{f}(y/x) - f(y/x)\right)^2 f(x)\, dydx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(y/x)^2 f(x) dxdy - 2 \int_{\mathbb{R}} \int_{\mathbb{R}} \hat{f}(y/x) f(y/x) f(x)\, dydx + \int_{\mathbb{R}} \int_{\mathbb{R}} f(y/x)^2 f(x) dxdy \\ &= A - 2B + C \end{aligned}$$

Puisque C est indépendant des deux paramètre de décision a et b, ce qui fait que la

constante C pourrait être ignorée.La fonction de validation croisée proposée par Hansen (2004) est définie par :

$$\hat{A} - 2\hat{B}$$

La méthodologie de validation croisée (leave-one-out-cross-validation) est appliquée pour estimer $\hat{A}$ et $\hat{B}$. Le principe de base de la validation croisée (leave-one-out-cross-validation) est de construire une estimation du $A - 2B$ à partir des données, puis minimisez cette estimation sur les paramètres de lissage pour obtenir des valeurs de ces paramètres. L'estimation de la densité conditionnelle au point y_i/x_i par validation croisée consiste à prendre en compte tous les points des données sauf l'observation (y_i, x_i) qui sera exclue de l'échantillons. Sois la notation $\hat{f}_{-i}(y_i/x_i)$ qui désigne l'estimateur $\hat{f}(y_i/x_i)$ avec l'observation i omise. C'est-à-dire :

$$\hat{f}_{-i}(y_i/x_i) = \frac{\sum_{j\neq i}^{n} K_a(x_i - x_j) K_b(y_i - y_j)}{\sum_{j\neq i}^{n} K_a(x_i - y_j)}$$

Les estimateurs de validation croisée de A et B sont :

$$\begin{aligned}
\hat{A} &= \frac{1}{n}\sum_{i=1}^{n} \int_{\mathbb{R}} \hat{f}_{-i}(y/x_i)^2 \, dy \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\frac{\sum_{j\neq i}^{n}\sum_{k\neq i} K_a(x_i - x_j) K_a(x_i - x_k) \int_{\mathbb{R}} K_b(y - y_j) K_b(y - y_k) \, dy}{\sum_{j\neq i}^{n} K_a(x_i - x_j)} \right]
\end{aligned}$$

$$\begin{aligned}
\hat{B} &= \frac{1}{n}\sum_{i=1}^{n} \hat{f}_{-i}(y_i/x_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\frac{\sum_{j\neq i}^{n} K_a(x_i - x_j) K_b(y_i - y_j)}{\sum_{j\neq i}^{n} K_a(x_i - x_j)} \right].
\end{aligned}$$

Finalement, la régle de sélection du paramètre de lissage par la validation croisée est :

$$\hat{A} - 2\hat{B} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}^{n}\sum_{k\neq i} K_a\left(x_i - x_j\right) K_a\left(x_i - x_k\right) \int_{\mathbb{R}} K_b\left(y - y_j\right) K_b\left(y - y_k\right) dy}{\sum_{j\neq i}^{n} K_a\left(x_i - x_j\right)}\right] - \frac{2}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}^{n} K_a\left(x_i - x_j\right) K_b\left(y_i - y_j\right)}{\sum_{j\neq i}^{n} K_a\left(x_i - x_j\right)}\right] \tag{2.16}$$

Remarque 2.3.1 *Si on fixe le noyau à un noyau gaussien ($K(u) = \phi(u)$) alors l'expression de $\hat{A}$ peut être simplifier (sans l'integrale) comme suit :*

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\sum_{j\neq i}^{n}\sum_{k\neq i} K_a\left(x_i - x_j\right) K_a\left(x_i - x_k\right) K_{\sqrt{2b}}\left(y_j - y_k\right)}{\sum_{j\neq i}^{n} K_a\left(x_i - x_j\right)}\right].$$

2.4 Conclusion

Dans ce chapitre nous avons introduit la notion de l'estimateur à noyau d'une densité conditionnelle $f(y/x)$. Nous avons constaté que sa construction se base sur le principe de l'estimation d'une densité classique qu'on a exposé dans le Chapitre 1.

Pour le choix des paramètres de l'estimateur en question, nous avons remarqué que les noyaux utilisés sont les même que dans le cas univariée et les procédures de sélection du paramètre de lissage sont construites avec la même démarche que le cas d'estimation d'une densité univariée.

A ce niveau, une question s'impose. Les méthodes de validation croisée pour le choix du paramètre de lissage dans l'estimation à noyau d'une densité conditionnelle ont-elles les mêmes avantages et inconvénients que dans le cas d'estimation d'une densité univariée ? La réponse à cette question, via une étude de simulation, pour le cas de la méthode UCV fera l'objet du prochain chapitre.

Chapitre 3

Performances d'un estimateur à noyau d'une densité $f(y/x)$

Introduction

Dans ce chapitre notre objectif est d'analyser, à base des échantillons simulés, les performances de méthodes de validation croisée UCV pour le choix du paramètre de lissage lors de l'estimation d'une densité de probabilité conditionnelle par la méthode du noyau, où nous allons considérer deux situations, à savoir : les paramètres de lissage a et b définis respectivement dans la direction de x et de y sont indépendant et lorsque l'hypothèse d'égalité de ces deux paramètres de lissage est imposée, c'est-à-dire $a = b = h$. Nous nous somme intéressé principalement à la comparaison de l'ISE moyenne des deux estimateurs en question ainsi qu'au problème des minimum locaux qui figure dans le cas d'estimation d'une densité univariée.

3.1 Présentation des paramètres de l'application

Dans ce qui suit, afin de ne pas confondre les deux estimateurs nous allons utiliser les notations suivantes : $\hat{f}_{ab}$ dans le cas a et b sont indépendants et $\hat{f}_h$ dans le cas $a = b = h$.

Notre objectif de vérifier dans un premier lieu l'effet de la taille de l'échantillon sur le comportement de la fonctionnelle $UCV(h)$ afin de tirer des conclusions sur la variabilité de cette fonctionnelle et l'existence d'éventuels minimum locaux. Dans un second lieu notre objectif et de comparer les performances des estimateurs $\hat{f}_{ab}$ et $\hat{f}_h$ lorsque les paramètres de lissage leurs associés sont sélectionnés à l'aide de l'UCV.

Comme exemple de modèle de simulation, nous avons considéré un modèle similaire au premier exemple traité dans [2]. En effet, nous avons considéré le modèle suivant :

$$Y = 1 + 5X + \varepsilon, \tag{3.1}$$

où X et ε sont deux variables aléatoires indépendantes issues d'une loi normale de paramètres $(0, 1)$ et d'une loi normale de paramètres $(0, 3)$, respectivement. Dans cet exemple, il est facile de démontrer que la densité de la variable aléatoire Y sachant X est définie comme suite :

$$f(y/x) = \frac{1}{3}\phi\left(\frac{y - 1 - 5x}{3}\right), \tag{3.2}$$

avec, $\phi(.)$ est la densité d'une distribution normale centrée réduite. L'allure graphique de la distribution conditionnelle $f(y/x)$ est donné dans la figure 3.1.

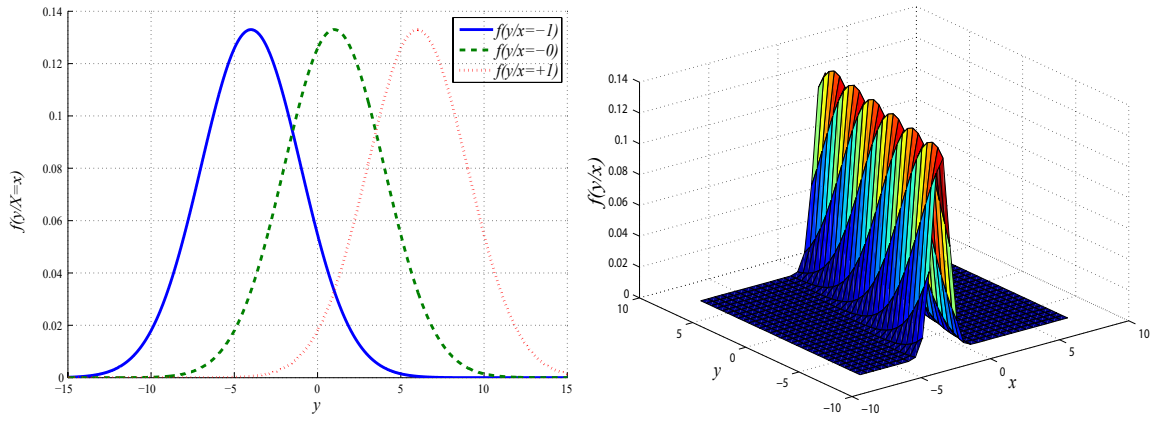

FIGURE 3.1 – Présentation graphique de $f(y/x)$ donnée dans 3.2.

3.2 Résultats et discussion

Afin d'atteindre notre objectif nous avons fixé les paramètres de l'étude de simulation comme suit :

Application 1 : Variation de $UCV(h)$:

- Le noyau gaussien pour la construction de $\hat{f}_h(y/x)$ et $\hat{f}_{ab}(y/x)$.
- La taille de l'échantillon $n \in \{10,\ 20,\ 100\}$
- Nous avons tracer $UCV(h)$ sur l'intervalle $[0.1, 1]$ avec un pas de 0.01, c'est-à-dire nous faisons varier h de 0.1 à 1 avec un pas 0.01.

Application 2 : Performances de $\hat{f}_h(y/x)$ et $\hat{f}_{ab}(y/x)$:

- Le noyau gaussien pour la construction de $\hat{f}_h(y/x)$ et $\hat{f}_{ab}(y/x)$.
- Le nombre d'échantillons $m = 50$.
- La taille de l'échantillon $n \in \{20,\ 50,\ 100,\ 150,\ 200\}$

3.2.1 Première application : Variation de $UCV(h)$

Les résultats numériques, fournis par le programme que nous avons implémenté sous Matlab, sont présentés dans la Figure 3.2.

Les résultats (Figure 3.2) obtenus dans cette première application montrent que :

– Dans le cadre d'estimation à noyau d'une densité conditionnelle l'utilisation de la méthode UCV pour la sélection du paramètre de lissage peut nous fournir des paramètres de lissage optimaux locaux et qui peut influer la qualité de l'estimateur conçu dans ce cas (phénomène de sur-lissage ou sous-lissage).

– Certes le phénomène des optimums locaux, dans la méthode UCV, peut dépendre de plusieurs paramètres tel le type de la densité cible (uni-modale, multimodale,) et son support et du noyau utilisé pour la construction de l'estimateur car ces paramètres interviennent d'une manière directe ou indirecte dans l'expression de $UCV(h)$. Mais

le paramètre essentiel qui contrôle le phénomène en question est bien que la taille de l'échantillon. En effet, on peut passer d'une grande fréquence d'existence de ce phénomène, dans une certaine situation, à une fréquence nulle simplement en augmentant la taille de l'échantillon. A titre d'exemple, sur des échantillons de taille $n = 10$ on constate qu'il existe plusieurs optimum locaux (voir Figure 3.2 (A)), pour que leurs nombre se réduit considérablement pour des échantillons de taille $n = 20$ (voir Figure 3.2 (B)) et pour que leurs nombre s'annule pour $n = 100$ (voir Figure 3.2 (C)).

– Le paramètre de lissage optimal sélectionné par la méthode UCV dans le cadre d'estimation à noyau d'une densité conditionnelle est extrêmement variable d'un échantillon à un autre malgré que si ces échantillons sont issu de la même distribution et ayant la même taille (Voir Figure 3.2 (D)).

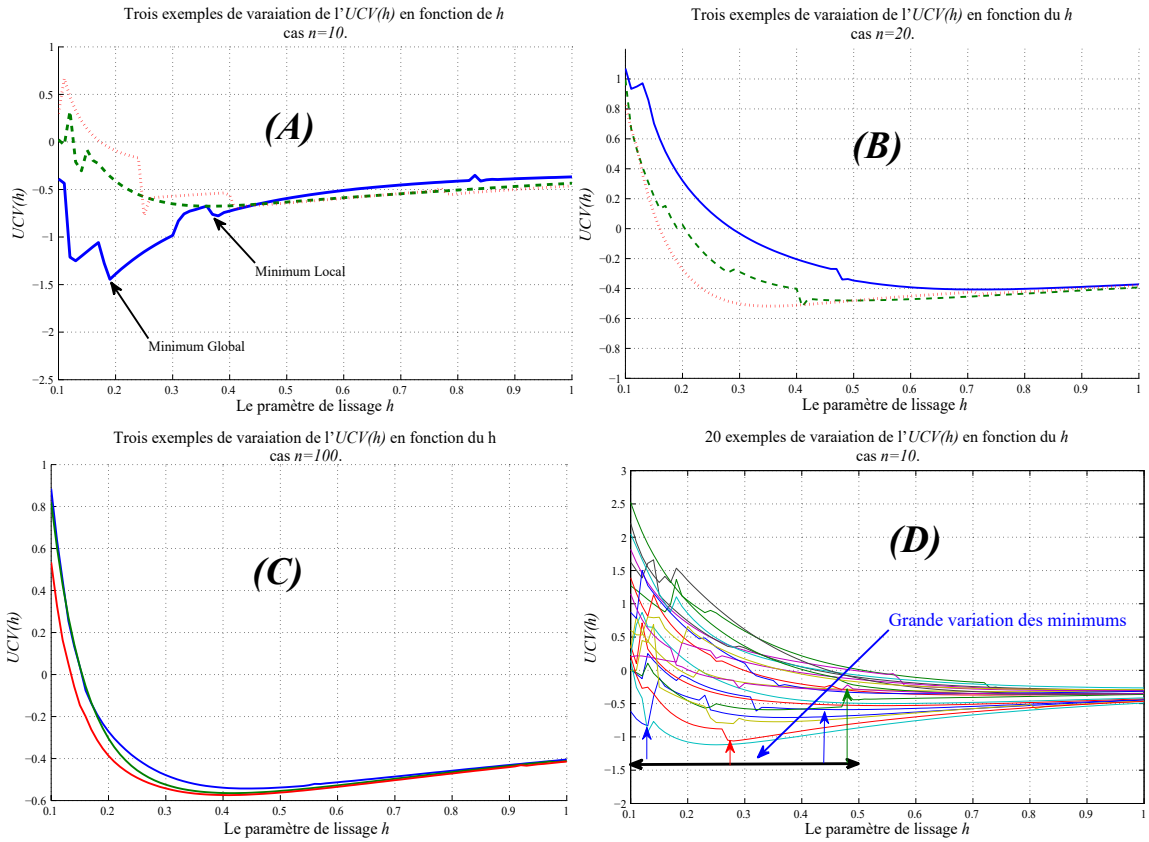

FIGURE 3.2 – Variation de l'$UCV(h)$ en fonction de h.

3.2.2 Deuxième application : Performances des estimateurs

Les résultats numériques, fournis par le programme que nous avons implémenté sous Matlab, sont résumés (rangés) dans la Table 3.1 et présentés dans la Figure 3.3.

	$a = b$		$a \neq b$	
n	h^*	$\overline{ISE}_h$	(a^*, b^*)	$\overline{ISE}_{(a;b)}$
20	0.6246	0.4654	(1.1339 0.5495)	0.3444
50	0.5452	0.0987	(0.9031 0.4607)	0.0315
100	0.4889	0.0566	(0.8994 0.3987)	0.0167
150	0.4614	0.0646	(0.7834 0.3417)	0.0126
200	0.4452	0.0361	(0.6761 0.3016)	0.0082

TABLE 3.1: Les paramètres de lissage optimaux et les ISE moyennes associées aux estimateurs $\hat{f}_{ab}$ et $\hat{f}_h$.

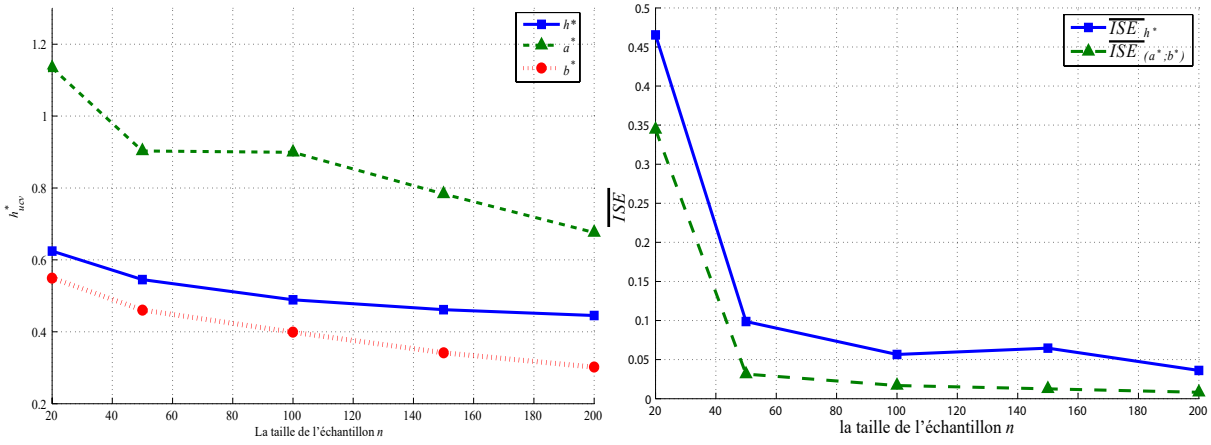

FIGURE 3.3 – Variation du paramètre de lissage optimal et du ISE moyenne en fonction de la taille de l'échantillon n.

Les résultats numérique et graphique obtenus indiquent que :

- Dans toutes les situations considérées, les paramètres de lissage optimal décroissent en fur et à mesure que la taille de l'échantillon augmente ce qui coïncide avec la propriété fondamentale (condition) du paramètre de lissage suivante : $\lim_{n\to\infty} h(n) = 0$.
- En fur et à mesure que la taille de l'échantillon augmente les ISE moyennes, associés au deux estimateurs $\hat{f}_{(a,b)}(y/x)$ et $\hat{f}_h(y/x)$, décroît et tend vers zéro, ce qui signifier que les deux estimateur en question convergent vers la vraie densité $f(y/x)$ en norme L_2 (ISE moyenne) lorsque la taille de l'échantillon tend vers l'infinie ($n \to \infty$).
- L'estimateur $\hat{f}_{ab}(y/x)$ et plus performant, au sens du ISE moyenne, que l'estimateur $\hat{f}_h(y/x)$ et ceci pour toutes les tailles d'échantillons considérées.

Conclusion

Dans ce chapitre à travers d'une application numérique, basée sur la simulation, nous avons mis en relief d'une part les inconvénients de la méthode UCV pour choix de paramètre de lissage dans l'estimation à noyau d'une densité conditionnelle qui sont les mêmes que ceux de l'UCV appliqué au cas classique.

D'autre part, les résultats obtenus indiquent que lorsque on sélectionne le paramètre de lissage de l'estimateur à noyau $\hat{f}(y/x)$ à l'aide de la méthode UCV, il est préférable d'imposer l'hypothèse d'indépendance des deux paramètres de lissage a et b, respectivement dans la direction de X et Y ($a \neq b$) plutôt que l'hypothèse de leurs égalité ($a = b = h$) et ceci afin d'obtenir un estimateur de $f(y/x)$ plus performant au sens du ISE moyenne.

Conclusion générale

La méthode du noyau est un outil très efficace pour estimer une densité de probabilité en générale et une densité conditionnelle en particulier. Par ailleurs, le choix du paramètre de lissage est d'une ultime importance pour la qualité du lissage dans l'estimation de la fonction de densité par la méthode du noyau.
Dans ce travail, nous nous sommes intéressées au problème du paramètre de lissage dans l'estimation à noyau d'une densité conditionnelle qui est d'une grande importance dans la pratique. Plus précisément, nous nous sommes intéressées à la méthode de validation croisée non biais (UCV) pour le choix du paramètre de lissage optimal dans le cadre d'estimation à noyau d'une densité conditionnelle.

Dans le premier chapitre, nous avons présenté la méthode du noyau de Parzen-Rosenblatt d'une densité de probabilité univariée où nous avons mis en évidence sa forme, ses propriétés locales et globales, le problème du choix de noyau et enfin, les différentes techniques de sélection du paramètre de lissage notamment, la famille des Méthodes de sélection basées sur la validation croisée.

Dans le deuxième chapitre, nous avons introduit l'estimateur à noyau d'une densité conditionnelle. Nous avons rappelé la définition de cet estimateur (sa forme), ses propriétés (le biais, la variance, la convergence en normes L^1 et L^2). Par la suite, le problème de choix du noyau ainsi que du paramètre de lissage ont été abordé. En effet, après une brève remarque sur le choix du noyau qui reste le même pour la fonction de densité simple, nous avons exposé les différentes méthodes de sélection du paramètre de lissage (choix optimal,

la règle de référence, validation croisée) mais nous somme focalisé particulièrement sur la méthode de UCV.

Enfin, dans le troisième chapitre, nous avons présenté une étude de simulation qui porte sur l'analyse des performances de la méthode de l'UCV pour la sélection du paramètre de lissage (a, b) (où a et b sont respectivement les paramètres de lissage dans la direction de X et Y) dans l'estimation à noyau d'une densité conditionnelle $f(y/x)$.

Les résultats des simulations obtenus sur plusieurs échantillons de différentes tailles en utilisant le noyau normal montrent d'une part que les inconvénients de la méthode UCV dans l'estimation à noyau d'une densité univariée restent valables dans le cas d'estimation d'une densité conditionnelle. D'autre part, si on opte pour la méthode UCV pour la sélection du paramètre de lissage dans l'estimation à noyau d'une densité conditionnelle, afin d'obtenir un estimateur plus performant au sens du ISE il est préférable d'imposé l'hypothèse d'indépendance des deux paramètres de lissage a et b que l'hypothèse de leurs égalité.

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

- Réaliser une simulation extensive toute en considérant d'autres noyaux et d'autres lois.
- Revoir le présent travail lorsque nous considérons d'autres méthodes de validation croisée (BCV, la LCV).
- Revoir le même travail afin évaluer la fréquence de l'existence du phénomène des minimum locaux.

Bibliographie

[1] Abou-Jaoude, S. (1976) *Sur une condition nécessaire et suffisante de L^1 convergence presque compléte 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 Date 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] Fan, J. and Yim, T.H. (2004) *A cross-validation method for estimating conditional densities*, Biometrika, 91(**4**) : 819–834.

[7] Habbema, J. D. F., Hermans, J., and Van den Broek, K. (1974), *A Stepwise Discriminant Analysis Program Using Density Estimation* in Compstat 1974, ed. G. Bruckmann, Physica Verlag, Vienna, pp. 101 − −110.

[8] Hall, P., Racine, J. and Li, Q., (2004) *Cross-validation and the estimation of conditional probability densities.* Journal of the American Statistical Association. 99 (**468**) : 1015–1026.

[9] Hall, P. (1982) *Cross-validation in density estimation.* Biometrika **69** : 383 − 390.

[10] Hall, P., Marron, S. J. (1991) *Local minima in cross-validation function.* Journal of the royal statistical society **90** : 149 – 173.

[11] Hyndman, R. J., Bashtannyk, D.J., Grunwald, G. K. (1996) *Estimating and visualizing conditional densities.* Journal of Computational and Graphical Statistics **5** : 315 – 336.

[12] Nadaraya, E. (1965) *On nonparametric estimation density function and regression.* Theory Probab. P.P.L **10** : 186 – 190.

[13] Park, B. U., Marron, S. J. (1990) *Comparison of data-driven bandwidth selectors.* Journal of the American Statistical Association **85** : 66 – 72.

[14] Parzen, E. (1962) *On estimation of a probability density function and mode.* Ann. Math. Statist. **33** : 1065 – 1076.

[15] Rosenblatt, M. (1956) *Remarks in some nonparametric estimates of a density function.* Ann. Math. Statist. **27** : 832 – 837.

[16] Rudemo, M. (1982) *Empirical choice of histogram and kernel density estimators.* Scandinavian Journal of Statistics. **9** : 65 – 78.

[17] Scott, D. W. (1985) *Averaged shift histograms : effective nonparametric density estimators in several dimensions.* The Annals of Statistics **13** : 1024 – 1040.

[18] Scott, D. W., Terrell, G. R. (1987) *Biased and unbiased cross-validation in density estimation.* Journal of the American Statistical Association **82** : 1131 – 1146.

[19] 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.

[20] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

[21] Stone, C. (1984) *An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates.* The Annals of Statistics **12** : 1285 – 1297.

[22] Wahba, G. (1975) *Optimal properties of variable knot, kernel and orthogonal series methods for density estimation*. Ann. Stat. **3** : 15 − 29.

[23] Youndjé, E. (2011) *Contribution à l'estimation non-paramétrique par la méthode du noyau*. Mémoire d'Habilitation à Diriger des Recherches.

[24] 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*. Thése de Magister en Mathématiques Appliquées, Université de Béjaïa.

Résumé

L'objectif du présent travail est d'illustrer, à base des échantillons simulés, les performances de la méthode validation croisée non biaisée, UCV, pour le choix du paramètre de lissage dans l'estimation d'une densité de probabilité conditionnelle par la méthode du noyau.
L'application numérique réalisée, sur des échantillons de différentes tailles, montre que la méthode UCV présente deux problèmes majeurs. D'une part, que la méthode 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 notamment dans le cas de petits échantillons.
Mots clés : Densité conditionnelle, Estimation à noyau, Paramètre de lissage, UCV, optimum local.

Abstract

This work aims to illustrate, based on simulated samples, the performances of the unbiased cross-validation method, UCV, for the choice of the smoothing parameter in the kernel estimation of a conditional probability density.
The numerical application, on different samples-sizes, shows that the UCV method presents two major problems. On the one hand, that the method can be extremely variable from one sample to another. On the other hand, the functional to be minimized often tends to have several minimums, especially in the case of small samples.
Key words : Conditional density, Kernel estimation, Smoothing parameter, UCV, local optimum.