

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieure et de la recherche scientifique

N° Attribué par la bibliothèque



Année univ.: 2018/2019

Estimation de la fonction de densité par les méthodes par noyaux associés

Mémoire présenté en vue de l'obtention du diplôme de

Master Académique

Université de Saïda - Dr Moulay Tahar

Discipline : MATHÉMATIQUES

Spécialité : Analyse Stochastiques, Statistique des Processus et Applications

par

Terras Hizia¹

Sous la direction de

Dr. F. Benziadi

Soutenu le 01/07/2019 devant le jury composé de

Mr. M. Laouni	Université Dr Tahar Moulay - Saïda	Président
Dr. F. Benziadi	Université Dr Tahar Moulay - Saïda	Encadreur
Dr. R. Rouane	Université Dr Tahar Moulay - Saïda	Examinatrice
Dr. S. Rahmani	Université Dr Tahar Moulay - Saïda	Examinatrice

1. e-mail : rattodja@gmail.com

Remerciement

Tout d'abord, nous tenons à remercier le "BON DIEU" le tout puissant de nous avoir accordé patience, courage et volonté afin de réaliser mener à terme ce modeste travail.

Je tiens à remercier tout particulièrement mon encadreur, *Dr. Fatima Benziadi* pour ses conseils, sa grande disponibilité et sa générosité. La pertinence de ses questions et de ses remarques ont toujours su me motiver et me diriger.

Je voudrais également remercier tous les membres de jury d'avoir accepté d'évaluer et d'examiner ce travail, merci pour toutes leurs remarques et critiques.

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

J'exprime mes remerciements à tous mes enseignants du département de Mathématiques qui m'ont initié aux valeurs authentiques, en signe d'un profond respect et d'un profond amour, ainsi que le personnel de l'administration. Je remercie tous les membres de laboratoire des Modèles Stochastique, Statistique et Application pour leur accueil et leur sympathie.

Mes sincères remerciements et ma gratitude vont aussi à Mr Djellouli Ghouti.

Merci à tous

Dédicace

Je dédie ce travail :

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

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

Je dédie ce travail à tout mes chères amis et spécialement Ikram, Fatima, Ines et Belkis

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

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

Table des matières

Introduction général	9
1 Estimation par noyaux associés unidimensionnelles	13
1.1 Estimation par noyaux classiques univariés	13
1.1.1 Résultats asymptotiques	14
1.2 Estimation par noyaux associés discrets univariés	15
1.2.1 Exemples de noyaux associés discrets	15
1.2.2 Résultats asymptotiques	17
1.2.3 Choix de fenêtre par la méthode de Validation croisée par les moindres carrés	22
1.3 Estimation à noyaux associés continus univariés	23
1.3.1 Constructions des noyaux associés continus par la méthode mode- dispersion	25
1.3.2 Estimateur à noyaux associés continus univariés	28
1.3.3 Résultats asymptotiques	29
1.3.4 Réduction de biais et noyaux associés modifiés	33
1.3.5 Choix du noyau	36
1.3.6 Choix de la fenêtre du lissage par la méthode de validation croisée par les moindres carrés	37
2 Estimation par noyaux associés multidimensionnels	39
2.1 Estimation par noyaux associés discrets multidimensionnels	39
2.1.1 Exemple(<i>Dirac Uniforme Discret</i>)	40
2.1.2 Estimateur à noyau associé discret multivarié	41
2.1.3 Propriétés asymptotiques	42

2.1.4	Choix de la matrice de fenêtre de lissage par la méthode de validation croisée	47
2.2	Estimation par noyaux associés continus multidimensionnels	47
2.2.1	Construction de noyaux multiples associés par la méthode de mode-dispersion	51
2.2.2	Résultats asymptotiques	52
2.2.3	Réduction du biais	56
3	Estimation par noyaux associés mixtes	61
3.1	Modèle	61
3.2	Définitions	64
3.3	Dérivabilité unifiée	65
3.4	Résultats asymptotiques	70
3.5	Choix de fenêtre par la méthode de validation croisée	73
3.6	Application numérique au modèle de mélange non-paramétrique	74
	Conclusion	79
	bibliographie	81

Introduction général

La statistique non-paramétrique a connu un développement considérable depuis une cinquantaine d'années. Un des plus vieux problèmes consisté à estimer la densité de probabilité $f(\cdot)$ et la fonction de répartition $F(\cdot)$ à partir d'un échantillon de variables aléatoires indépendantes et identiquement distribuées (i.i.d). L'estimation de la densité de probabilité à partir d'un ensemble fini d'observations est un problème fondamental qui a fait l'objet d'une très vaste littérature. Une bonne introduction à cette problématique peut être trouvée dans les livres de Silverman (1986)[33] et Tsybakov (2009)[34].

Deux types d'approches d'estimation de la densité de probabilité sont utilisées et largement discutées dans la littérature : l'approche paramétrique et l'approche non paramétrique. L'approche paramétrique a comme inconvénient principal de nécessiter une connaissance préalable du phénomène aléatoire considérée. L'approche non paramétrique estime la densité de probabilité directement à partir de l'information disponible sur l'ensemble d'observations. On dit que souvent dans cette approche les données parlent d'elles mêmes.

L'estimation de la densité par la méthode du noyau peut être vue comme une extension de la méthode d'estimation par histogramme, est la plus populaire parmi les multiples méthodes d'estimation non paramétriques de la densité. 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'interprète par la liberté de l'utilisateur dans le choix du noyau K . L'estimateur à noyau classique a été proposé initialement par Rosenblatt (1956)[26] pour estimer des densités de probabilité, par Parzen (1962)[24] pour estimer le mode d'une densité de probabilité et par Nadaraya (1964) [23] et Watson (1964) [39] pour estimer une fonction de régression à support non borné.

Dans ce cadre, le choix du noyau K est très peu influant. Les noyaux employés pour ce type de densité (ou de support) sont symétriques (dit aussi classiques). Nous citons comme exemple, le noyau rectangulaire, triangulaire, parabolique (Epanechnikov (1969)[9]), biweight et gaussien. Cependant, lorsqu'on veut estimer des densités à support borné au moins d'un côté, l'estimateur à noyau classique devient non consistant, à cause des effets du bord. Plusieurs solutions ont été proposées dans la littérature pour remédier à cette difficulté, citons data reflection de Schuster (1985)[32], boundary kernels de Müller (1991, 1993)[22] et empirical transformation de Marron et Ruppert (1994)[21]. La méthode d'estimation par les séries orthogonales de densités à supports bornés a été développée à partir des travaux de Cencov (1962)[7], et étudiée ensuite par plusieurs auteurs dont Kronmal et Tarter (1968)[19], Wahba (1981)[36], Bosq (2005)[5] et récemment Saadi et Adjabi (2009)[27]. Les méthodes d'interpolation par les fonctions splines Wahba (1975)[37]. Une autre méthode basée sur les développements en ondelettes a été proposée. Les estimateurs par ondelettes surpassent les estimateurs classiques dans la représentation des discontinuités et les oscillations locales. Les premiers résultats sur ces estimateurs sont donnés dans Walter (1992)[38] et Kerkycharian et Picard (1993)[15]. Le comportement asymptotique de ces estimateurs a été étudié par de nombreux auteurs tel que Tsybakov (2004)[35]. La solution la plus simple, qui est encore d'actualité est de remplacer le noyau symétrique par un noyau asymétrique, qui n'assigne pas un poids en dehors du support de la densité qu'on veut estimer. Cette idée est due à Chen (1999, 2000)[8], puis d'autres extensions ont été apportées par Jin et Kawczak (2003)[12], Scaillet (2004)[31], Senga Kiessé (2008)[30] et récemment, la notion du noyau K est généralisé en noyau associé $K_{x,h}$ par Kokonendji et Libengué (2011)[16] et Libengué (2013)[20]. En pratique, pour utiliser l'estimateur à noyau associé, il faut choisir le noyau associé $K_{x,h}$ et le paramètre de lissage h . Pour le noyau, le choix peut être adapté par le support de la fonction inconnue à estimer, et pourtant, en situation non-asymptotique (faible nombre de données), le côté arbitraire du choix du noyau peut avoir une incidence importante sur la qualité de l'estimation. En revanche, le paramètre de lissage est un facteur important et crucial dans l'estimation de la fonction densité par la méthode des noyaux associés. De petites ou de grandes valeurs de h peuvent conduire à une estimation sous ou sur-lissée. Deux catégories de méthodes classiques ont été proposées dans la littérature pour choisir le paramètre de lissage h . La première catégorie repose

sur la minimisation de l'erreur quadratique moyenne intégrée (MISE). Cette classe de méthodes est intéressante en théorie, mais sa difficulté majeure réside dans les applications. La seconde catégorie est de type validation croisée, elle est intéressante en pratique car elle se laisse guider seulement par les observations. L'inconvénient principal des deux catégories est qu'elles ont tendance à fournir des estimateurs sous ou sur-lissés lorsque les données sont de petite ou moyenne taille, ou encore lorsque on veut estimer des fonctions complexes.

Dans de nombreux domaines (médecine, économie, sociologie, environnement, finances, marketing) surviennent des phénomènes complexes à modéliser. Dans le cas univarié, ces derniers sont souvent décrits par une variable aléatoire réelle (v.a.r.) X à support connu $\mathbb{T} \subseteq \mathbb{R}$. Dans le cas multivarié, ils sont souvent décrits par des vecteurs aléatoires à valeurs réelles et à support connu dans $\mathbb{T}^d \subseteq \mathbb{R}^d$ avec $d \in \{1, 2, \dots\}$. Cet ensemble \mathbb{T} peut être constitué à la fois de parties continues (intervalles) et discrètes, on parlera alors d'échelles de temps ("time scales" en anglais). Cette notion a été introduite par Stefan Hilger dans sa thèse en 1988 pour unifier les analyses continues et discrètes, voir Hilger (1990) [11], Agarwal & Bohner (1999)[2], Bohner & Peterson (2001, 2003)[4], Sanyal (2008)[28] pour d'autres développements.

Dans ce travail, nous nous intéressons à l'étude des propriétés asymptotiques de l'estimateur à noyaux associés continus, discrets et mixtes. Ce mémoire est composé d'une introduction générale, de trois chapitres, d'une conclusion générale et d'une liste de références bibliographiques.

Dans le premier chapitre, nous présentons l'estimateur à noyau classique avec quelques résultats asymptotiques, nous introduisons la notion du noyau associé univarié (discret et continu), leur construction et l'estimateur à noyau associé de densité de probabilité en donnant ses propriétés de convergence. Le deuxième chapitre consacre à la méthode du noyau associé pour une estimation non-paramétrique de la fonction de la densité (fonction de masse) de probabilité multidimensionnelle avec les propriétés statistiques de l'estimateur (biais, variance, erreur quadratique moyenne intégrée). Une méthode classique pour le choix de la matrice des fenêtres sera également proposée. Enfin, le troisième Chapitre présente l'estimateur de densité mixte par les noyaux associés lesquels généralisent tous les noyaux étudiés précédemment. Il s'agit ici d'un mélange approprié des noyaux associés continus et discrets. Pour les passages entre les sous-ensembles successifs continus et discrets de $\mathbb{T} \subseteq \mathbb{R}$, on introduit de nou-

veaux outils d'analyse sur ces échelles de temps. Une application numérique concrète terminent ce travail.

Chapitre 1

Estimation par noyaux associés unidimensionnelles

1.1 Estimation par noyaux classiques univariés

On suppose que les X_i sont à valeurs réelles et que f est la densité par rapport à la mesure de Lebesgue sur $\mathbb{T} = \mathbb{R}$.

Définition 1.1.1. Une fonction \mathcal{K} de support \mathbb{S} est dite **noyau** si elle est une densité de probabilité symétrique (i.e $\mathcal{K}(-u) = \mathcal{K}(u)$), de moyenne $\mu_{\mathcal{K}}$ nulle ($\mu = \int_{\mathbb{S}} u\mathcal{K}(u)du = 0$), de variance $\sigma_{\mathcal{K}}^2$ finie ($\sigma_{\mathcal{K}}^2 = \int_{\mathbb{S}} u^2\mathcal{K}(u)du < +\infty$) et de carré intégrable ($\int_{\mathbb{S}} \mathcal{K}^2(u)du < +\infty$).

Précisons ici qu'en tant que densité de probabilité, le noyau \mathcal{K} est positif et de masse totale égale à 1 (i.e pour tout élément u de \mathbb{S} , $\mathcal{K}(u) > 0$ et $\int_{\mathbb{S}} \mathcal{K}(u)du = 1$).

Définition 1.1.2. Soit $h_n > 0$ la fenêtre de lissage et \mathcal{K} la fonction noyau vérifiant la définition (1.1.1), l'estimateur à noyau continu (classique) de f est défini en un point $x \in \mathbb{T}$ par :

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n \mathcal{K}\left(\frac{x - X_i}{h_n}\right). \quad (1.1)$$

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

1.1.1 Résultats asymptotiques

Théorème 1.1.1. (*la vitesse de la convergence presque complète*¹) Supposons que les hypothèses suivantes soient réalisées :

(H1) : $f(x) > 0$ et de classe C^k

(H2) : $\lim_{n \rightarrow \infty} h_n \rightarrow 0$, $n \rightarrow 0$ et $\lim_{n \rightarrow \infty} \left(\frac{nh_n}{\log(n)} \right) = 0$

(H3) : \mathcal{K} bornée et d'ordre k ²,

alors on a :

$$\widehat{f}_n(x) - f(x) = o(h^k) + o\left(\sqrt{\frac{\log(n)}{nh}}\right) \quad p.c \quad (1.2)$$

Théorème 1.1.2. (*Erreur quadratique moyenne*) Supposons que les hypothèses

(H1),(H2),(H3) et (H4) soient réalisées, où

(H4) : $\mathbb{E}(Y^2/X) < \infty$,

alors on a :

$$\begin{aligned} MSE(\widehat{f}_n(x)) &= \mathbb{E} \left[\{\widehat{f}_n(x) - f(x)\}^2 \right] \\ &= Bias^2\{\widehat{f}_n\} + Var\{\widehat{f}_n\} \\ &= \left(\frac{h^2}{2} f''(x) \mu(K) o(h^{2k}) \right)^2 + \frac{1}{2} \int_{-\infty}^{+\infty} \mathcal{K}^2(u) f(x - hu) du \\ &\quad - \frac{1}{n} \left(\int_{-\infty}^{+\infty} \mathcal{K}(u) f(x - hu) du \right) \\ &= o(h^{2k}) + o\left(\frac{1}{nh}\right). \end{aligned}$$

1. Soit $(u_n)_n$ une suite de v.a et u une v.a

- $\lim_{n \rightarrow \infty} u_n = u$ p.c ssi $\forall \epsilon > 0 \sum \mathbb{P}(|u_n - u| \geq \epsilon) < \infty$,
- on dit que $u_n = o(u)$ p.c ssi $\forall \epsilon > 0 \sum \mathbb{P}(|u_n| \geq \epsilon u) < \infty$.

2. un noyau K est dit d'ordre k ssi

$$\int_{\mathbb{S}} |u|^j K(u) du = \begin{cases} 0 & \text{si } j = 1, \dots, k-1 \\ \text{cst} < \infty & \text{si } j = k \end{cases}$$

1.2 Estimation par noyaux associés discrets univariés

Dans cette section, on présente l'essentiel des travaux de Aitchison & Aitken (1976)[3], Senga Kiessé (2008)[17] et Kokonendji & Senga Kiessé (2011)[29] sur les noyaux associés discrets. Désignons par $\mathbb{T} \subseteq \mathbb{Z}$ le support de la fonction de masse de probabilité (*f.m.p*) à estimer puis par $\mathbb{S}_{x,h} = \mathbb{S}_x \subseteq \mathbb{Z}$ celui du noyau associé.

Définition 1.2.1. *On appelle type de noyau discret toute fonction de masse de probabilité (*f.m.p*) K_θ , paramétrée par $\theta \in \Theta \subseteq \mathbb{R}^2$, de support $\mathbb{S}_\theta \subseteq \mathbb{Z}$ et de carré sommable.*

Donnons maintenant la définition améliorée du noyau associé discret.

Définition 1.2.2. *Soit $x \in \mathbb{T} \subseteq \mathbb{Z}$ et $h > 0$ avec \mathbb{T} le support de la fonction de masse de probabilité f à estimer. Une densité paramétrée $K_{x,h}$ de support $\mathbb{S}_{x,h} \subseteq \mathbb{Z}$ est appelée noyau associé discret lorsque les conditions suivantes sont vérifiées :*

$$x \in \mathbb{S}_{x,h} \tag{1.3}$$

$$\mathbb{E}(Z_{x,h}) = x + A(x, h) \tag{1.4}$$

$$\text{Var}(Z_{x,h}) = B(x, h) \tag{1.5}$$

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

Remarque 1.2.1. *Notons que, lorsque $B(x, h) \rightarrow 0$, le noyau associé discret $K_{x,h}$ est dit de "second ordre". Dans le cas contraire, il est dit du "premier ordre" (ou encore standard). Nous signalons aussi qu'il n'existe pas une méthode générale pour la construction des noyaux associés discrets. Il en résulte que celle-ci se fait au cas par cas. Toutefois, on peut remarquer que les noyaux associés discrets sont en général asymétriques et leurs supports peuvent ne pas dépendre de x et/ou h .*

1.2.1 Exemples de noyaux associés discrets

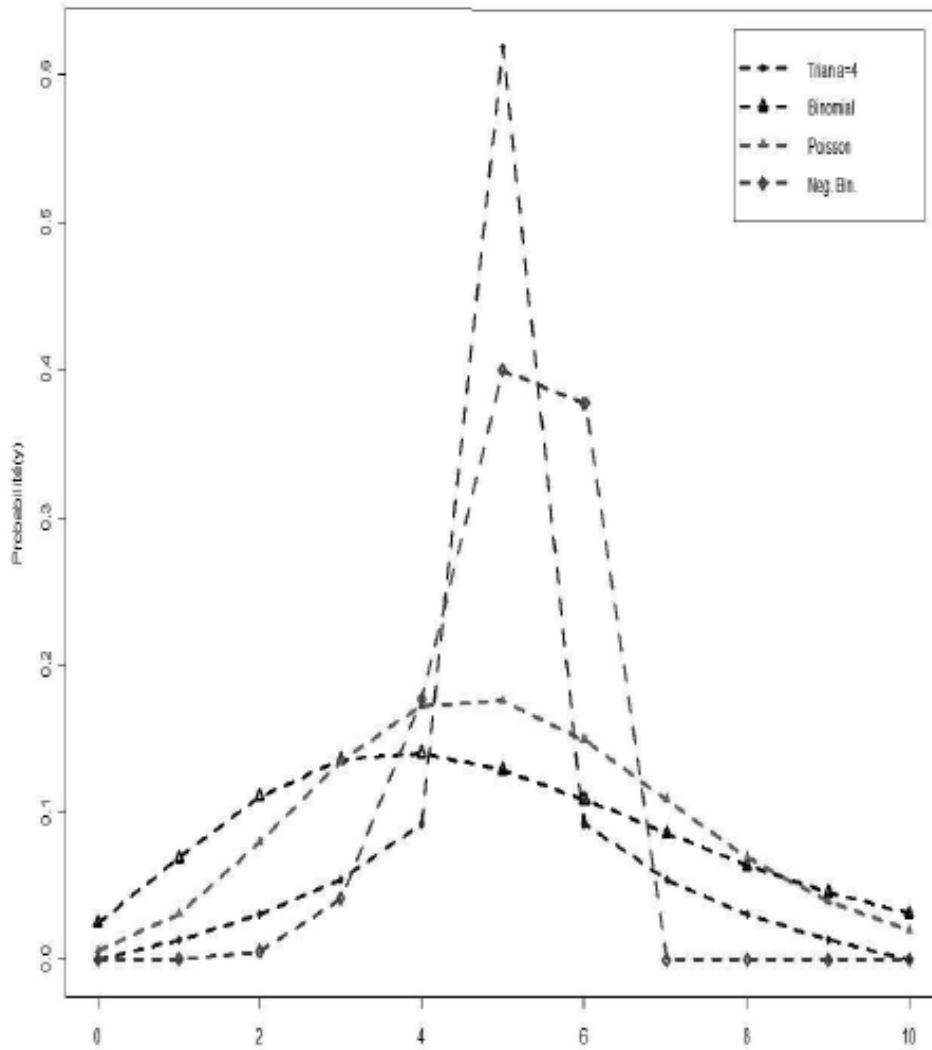


FIGURE 1.1 – Quelques noyaux associés discrets

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

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

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

Définition 1.2.3. Soit X_1, X_2, \dots, X_n , une suite de variables aléatoires discrètes (i.i.d) de fonction de masse de probabilité (fmp) inconnue f sur $\mathbb{T} \subseteq \mathbb{Z}$. L'estimateur à noyau associé discret de f est défini par :

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{j=1}^n K_{x,h}(X_j), \quad (1.6)$$

où $h > 0$ est le paramètre de lissage et $K_{x,h}$ est le noyau associé discret dépendant de x et h .

1.2.2 Résultats asymptotiques

Proposition 1.2.1. (Senga Kiessé (2008)[17]). Soit f la fmp à estimer sur $\mathbb{T} \subseteq \mathbb{Z}$ et \widehat{f}_n son estimateur à noyau associé discret en (1.6). Pour tout $x \in \mathbb{T}$ et $h > 0$, on a :

$$\begin{aligned} \mathbb{E}\{\widehat{f}_n(x)\} &= \mathbb{E}\{f(Z_{x,h})\}, \\ \widehat{f}_n(x) &\in [0, 1], \\ \sum_{x \in \mathbb{T}} \widehat{f}_n(x) &= C, \end{aligned}$$

où C est une constante strictement positive et finie.

Preuve : Pour tout $x \in \mathbb{T}$ on a :

$$\mathbb{E}\{\widehat{f}_n(x)\} = \sum_{y \in \mathbb{T} \cap \mathbb{S}_x} K_{x,h}(y) f(y) = \sum_{y \in \mathbb{T} \cap \mathbb{S}_x} f(y) \mathbb{P}(Z_{x,h} = y) = \mathbb{E}\{f(Z_{x,h})\},$$

car $K_{x,h}(y) f(y) = 0$ pour $y \notin \mathbb{T} \cap \mathbb{S}_x$ et la formule est ainsi montrée.

Ensuite, $\widehat{f}_n \in [0, 1]$ découle immédiatement de $K_{x,h}(X_i) \in [0, 1]$ pour tout X_i .

Enfin, en écrivant

$$C = \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{x \in \mathbb{T}} K_{x,h}(X_i) \right\}$$

pour tout $h > 0$ et en observant que :

$$\sum_{x \in \mathbb{T}} K_{x,h}(y) = \sum_{x \in \mathbb{T} \cap \mathbb{S}_x} K_{x,h}(y)$$

pour tout $y \in \mathbb{T} = \text{support}(X_i)$, il en résulte que :

- d'une part $C > 0$ car $K_{x,h}(x) > 0$ pour $y = x \in \mathbb{T} \cap \mathbb{S}_x$,
- d'autre part $C < +\infty$ car $0 \leq K_{x,h}(y) \leq 1$ pour tout $y \in \mathbb{T} \cap \mathbb{S}_x$ et $K_{x,h}(y) \rightarrow 0$ quand $x \rightarrow \infty$.

De manière pratique, quand nous calculons cette constante C , nous trouvons $C = 1$. ■

Le résultat suivant garantit que l'estimateur à noyau discret est asymptotiquement sans biais.

Proposition 1.2.2. (*Kokonendji & Senga Kiessé(2011)[29]*) Soit f la fmp à estimer sur $\mathbb{T} \subseteq \mathbb{Z}$ et \hat{f}_n son estimateur à noyau associé discret en (1.6). Pour tout $x \in \mathbb{T}$ et $h > 0$, si $h \rightarrow 0$ quand $n \rightarrow +\infty$ alors :

$$\mathbb{E}\{\hat{f}_n(x)\} = \sum_{y \in \mathbb{T}} f(y)K_{x,h}(y) \rightarrow f(x), \text{ quand } n \rightarrow +\infty.$$

Preuve : Puisque $K_{x,h}(y)f(y) = 0$ pour tout $y \notin \mathbb{S}_x \cap \mathbb{T}$, on suppose $\mathbb{T} \subseteq \mathbb{S}_x$ pour tout $x \in \mathbb{T}$. Pour tout $\eta > 0$ on note $\mathbb{S}_{x,\eta} = \{y \in \mathbb{S}_x : |y - x| < \eta\}$ et

$\bar{\mathbb{S}}_{x,\eta} = \{y \in \mathbb{S}_x : |y - x| \geq \eta\}$ son complémentaire. Sachant que $f(x) = f(x) \sum_{y \in \mathbb{S}_x} K_{x,h}(y)$,

nous exprimons :

$$\begin{aligned} |\mathbb{E}\{\hat{f}_n(x)\} - f(x)| &= \left| \sum_{y \in \mathbb{S}_x \subseteq \mathbb{T}} \{f(y) - f(x)\} K_{x,h}(y) \right| \\ &\leq \sum_{y \in \mathbb{S}_{x,\eta}} |f(y) - f(x)| K_{x,h}(y) + \sum_{y \in \bar{\mathbb{S}}_{x,\eta}} |f(y) - f(x)| K_{x,h}(y), \end{aligned}$$

où $\eta > 0$ est une constante. Puisque f est une fonction de masse de probabilité et par conséquent continue d'après la notion de continuité induite par la topologie de \mathbb{R} se traduit par :

$$\forall \epsilon > 0 \exists \eta > 0 : \forall y \in]x - \eta, x + \eta[\cap \mathbb{T} \Rightarrow |f(y) - f(x)| < \epsilon, \quad (1.7)$$

pour tout $\epsilon > 0$ il existe un $\eta = \eta(\epsilon) > 0$ pour lequel

$$\sum_{y \in \mathbb{S}_{x,h}} |f(y) - f(x)| K_{x,h}(y) \leq \epsilon. \quad (1.8)$$

Considérons la variable aléatoire $Z_{x,h}$. Pour $a > 0$ et pour tout $\epsilon > 0$, l'inégalité de Tchebychev-Markov s'écrit ici :

$$\mathbb{P}(|K_{x,h} - a| \geq \epsilon) \leq \frac{\mathbb{E}\{(Z_{x,h} - a)^2\}}{\epsilon^2}.$$

Ainsi, puisque f est une fonction de masse de probabilité et donc $f \leq 1$, nous pouvons majorer le second terme par :

$$\begin{aligned} \sum_{y \in \mathbb{S}_{x,\eta}} |f(y) - f(x)| K_{x,h}(y) &\leq 2 \sum_{y \in \mathbb{S}_{x,\eta}} K_{x,h}(y) = 2\mathbb{P}(|Z_{x,h} - x| > \eta) \\ &\leq \frac{2}{\eta^2} \mathbb{E}\{(Z_{x,h} - x)^2\} \\ &\leq \frac{2}{\eta^2} (\text{var}(Z_{x,h}) + \{\mathbb{E}(Z_{x,h}) - x\}^2). \end{aligned} \quad (1.9)$$

Finalement, sous le résultat (1.9) et la conditions 1.5 de la définition d'un noyau associé discret, les inégalités (1.8) et (1.9) permettent d'aboutir au résultat recherché. ■

Les résultats suivants concernent les faibles et fortes consistances ainsi que la normalité asymptotique des estimateurs à noyaux associés discrets.

Proposition 1.2.3. (*Abdous & Kokonendji (2009)[1]*) Soit f la fmp à estimer sur $\mathbb{S}_x \subseteq \mathbb{T}$ et \widehat{f}_n son estimateur à noyau associé discret en (1.6). Pour tout $x \in \mathbb{S}_x$ et sous les conditions (1.4) et (1.5) on a :

$$\mathbb{E}\{\widehat{f}_n(x) - f(x)\}^2 \longrightarrow 0, \text{ quand } n \longrightarrow +\infty.$$

Preuve : Sans perte de généralité, on suppose $\mathbb{S}_x \subseteq \mathbb{T}$ pour tout $x \in \mathbb{T}$. D'après la proposition (1.2.2) nous avons, pour tout $x \in \mathbb{T}$:

$$\text{biais}\{\widehat{f}_n(x)\} = \mathbb{E}\{\widehat{f}_n(x)\} - f(x) \longrightarrow 0 \text{ quand } n \longrightarrow +\infty \text{ et } h = h(n) \longrightarrow 0.$$

La variance ponctuelle peut être majorée successivement par :

$$\begin{aligned}
 nVar \left\{ \widehat{f}_n(x) \right\} &= Var \left\{ K_{x,h}(X_1) \right\} \\
 &\leq \mathbb{E} \left\{ K_{x,h}(X_1) \right\}^2 \\
 &\leq \sum_{y \in \mathbb{S}_x} f(y) \left\{ \mathbb{P} (Z_{x,h} = y) \right\}^2 \\
 &\leq \sum_{y \in \mathbb{S}_x} f(y) \mathbb{P} \{ Z_{x,h} = y \} \\
 &\leq 1.
 \end{aligned}$$

Les dernières inégalités s'obtiennent facilement grâce à $\mathbb{P} (Z_{x,h} = y) \leq 1$ et à $f(y) \leq 1$. De là, puisqu'on a $0 \leq var \left\{ \widehat{f}_n(x) \right\} \leq (1/n)$, alors $var \left\{ \widehat{f}_n(x) \right\}$ tend vers 0 pour tout $x \in \mathbb{N}$ lorsque $n \rightarrow +\infty$ et $h = h(n) \rightarrow 0$. Par conséquent, le résultat découle de la décomposition :

$$\mathbb{E} \left\{ \widehat{f}_n(x) - f(x) \right\}^2 = \text{biais}^2 \left\{ \widehat{f}_n(x) \right\} + Var \left\{ \widehat{f}_n(x) \right\}.$$

■

Proposition 1.2.4. (*Abdous & Kokonendji (2009)[1]*) Soit f la fmp à estimer sur $\mathbb{T} \subseteq \mathbb{Z}$ et \widehat{f}_n son estimateur à noyau associé discret en (1.6). Pour tout $x \in \mathbb{T}$ et sous les conditions (1.4) et (1.5) on a :

$$\widehat{f}_n(x) \xrightarrow{Ps} f(x), \text{ quand } n \rightarrow \infty,$$

où \xrightarrow{Ps} désigne la convergence presque sûre.

Corollaire 1.2.1. *Sous les hypothèses :*

$x \in \mathbb{S}_x$ et $\lim_{n \rightarrow \infty} \mathbb{E}(Z_{x,h}) = 0$, on a :

$$\mathbb{E} \left\{ \widehat{f}_n(x) - f(x) \right\}^2 = 0$$

Corollaire 1.2.2. Soient Y_1, Y_2, \dots, Y_n des variables aléatoires avec des secondes moments finis. S'il existe a et b des constants telles que $\mathbb{P} (Y_i \in [a, b]) = 1$ et $\epsilon > 0$ alors on a :

$$\mathbb{P} \left(\left| n^{-1} \sum_{i=1}^n Y_i \right| \geq \epsilon \right) \leq 2exp \left\{ - \frac{n\epsilon^2}{\epsilon(b-a) + 2Var(Y_1)} \right\}.$$

Preuve : Sous l'écriture

$$\widehat{f}_n(x) - f(x) = \left[\widehat{f}_n(x) - \mathbb{E} \left\{ \widehat{f}_n(x) \right\} \right] + \left[\mathbb{E} \left\{ \widehat{f}_n(x) \right\} - f(x) \right],$$

et en utilisant le corollaire (1.2.1) que réduit le résultat :

$$\mathbb{E} \left\{ \widehat{f}_n(x) \right\} - f(x) \xrightarrow{ps} 0 \text{ quand } n \longrightarrow \infty. \quad (1.10)$$

Pour utiliser le corollaire (1.2.2) posons $Y_i = K_{x,h}(X_i) - \mathbb{E} \{K_{x,h}(X_i)\}$ pour $i = 1, \dots, n$ Puis, on pose Z telle que

$$\lim_{n \rightarrow \infty} \mathbb{E}(Z) = x, \quad \lim_{n \rightarrow \infty} \text{Var}(Z) = 0,$$

où $Z = Z_{x,h}$ et sa fmp $K_{x,h}(\cdot)$, il est facile de vérifier que pour tout $x \in \mathbb{S}$ et $i = 1, \dots, n$

$$-1 \leq Y_i \leq 1 \quad \text{et} \quad \text{var}(Y_1) \leq \mathbb{E} \{K_{x,h}^2(X_1)\} \leq 1.$$

Par conséquent, selon le corollaire (1.2.2), pour tout $\epsilon > 0$, on a :

$$\begin{aligned} \mathbb{P} \left\{ \left| \widehat{f}_n(x) - \mathbb{E} \left\{ \widehat{f}_n(x) \right\} \right| \geq \epsilon \right\} &= \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n Y_i \right| \geq \epsilon \right) \\ &\leq 2 \exp \left(\frac{-n\epsilon^2}{2\epsilon+2} \right) \\ &\leq 2 \exp \left(\frac{-n\epsilon}{2} \right), \end{aligned}$$

par conséquent, le lemme de Borel-Cantelli conduit au résultat (3.2) puisque

$$\sum_{i \geq 1} \mathbb{P} \left\{ \left| \widehat{f}_n(x) - \mathbb{E} \left\{ \widehat{f}_n(x) \right\} \right| \geq \epsilon \right\} \leq 2 \sum_{i \geq 1} \exp \left(\frac{-n\epsilon}{2} \right) = \frac{2}{-1 + \exp(-\epsilon/2)} < \infty.$$

■

Proposition 1.2.5. (Abdous & Kokonendji (2009)[1]) Soit f la fmp à estimer sur $\mathbb{T} \subseteq \mathbb{Z}$ et \widehat{f}_n son estimateur à noyau associé discret définie en (1.6). Pour tout $x \in \mathbb{T}$ et sous les conditions (1.4) et (1.5) on a :

$$\frac{\widehat{f}_n(x) - \mathbb{E}\{\widehat{f}_n(x)\}}{(\text{Var}\{\widehat{f}_n(x)\})^{1/2}} \xrightarrow{\mathcal{L}} N(0, 1), \text{ quand } n \longrightarrow \infty, \quad (1.11)$$

où $\xrightarrow{\mathcal{L}}$ représente la convergence en loi.

Preuve : La déclaration (1.11) découle d'arguments assez classiques. En effet, puisque $\widehat{f}_n(x)$ est une somme des variables aléatoires i.i.d, une condition suffisante pour obtenir cette convergence est :

$$\lim_{n \rightarrow \infty} \frac{\sqrt{n} \mathbb{E} |K_{x,h}(X_1) - \mathbb{E} \{K_{x,h}(X_1)\}|^3}{[Var \{K_{x,h}(X_1)\}]^{3/2}} = 0. \quad (1.12)$$

Pour obtenir ce résultat, notez que le numérateur satisfait

$\mathbb{E} |K_{x,h}(X_1) - \mathbb{E} \{K_{x,h}(X_1)\}|^3 \leq 2^3$ tandis que le terme de variance est tel que

$$\lim_{n \rightarrow \infty} Var \{K_{x,h}(X_1)\} = f(x) \{1 - f(x)\}, \quad (1.13)$$

on a :

$$Var \{K_{x,h}(X_1)\} - f(x) \{1 - f(x)\}$$

$$= \underbrace{\left\{ \lim_{y \in \mathbb{S} \cap \mathbb{S}_x} f(y) [\mathbb{P}(y = Z)]^2 - f(x) \right\}}_{V_1} + \underbrace{\left\{ f(x)^2 - [\mathbb{E} \{\widehat{f}_n(x)\}]^2 \right\}}_{V_2},$$

pour V_1 est :

$$\begin{aligned} |V_1| &\leq \lim_{y \in \mathbb{S} \cap \mathbb{S}_x} |f(y) \mathbb{P}(y = Z) - f(x)| \mathbb{P}(y = Z) + f(x) \lim_{y \in \overline{\mathbb{S}} \cap \mathbb{S}_x} \mathbb{P}(y = Z) \\ &\leq f(x) [1 - \mathbb{P}(Z = x)] \\ &\quad + \lim_{y \in \mathbb{S} \cap \mathbb{S}_x \setminus \{x\}} |f(y) \mathbb{P}(y = Z) - f(x)| \mathbb{P}(y = Z) + f(x) \lim_{y \in \overline{\mathbb{S}} \cap \mathbb{S}_x} \mathbb{P}(y = Z) \\ &\leq 4[1 - \mathbb{P}(Z = x)] \longrightarrow 0 \text{ quand } n \longrightarrow \infty. \end{aligned}$$

Alors, la quantité V_2 vérifie que

$$|V_2| \leq 2|\mathbb{E} \{\widehat{f}_n(x)\} - f(x)| \longrightarrow 0 \text{ quand } n \longrightarrow \infty.$$

■

1.2.3 Choix de fenêtre par la méthode de Validation croisée par les moindres carrés

On propose la méthode de validation croisée par les moindres carrés pour le choix du paramètre de lissage dans le cas discret. Cette méthode est proposée dans les

travaux de Senga Kiessé (2008)[29] et Kokonendji & Senga Kiessé (2011)[17].

L'idée principale de la méthode adaptée ici est la même que la méthode de validation croisée. La seule différence est qu'ici, la fonction noyau $K_{x,h}$ dépend du point d'estimation x et de paramètre de lissage h . Elle a l'avantage de ne pas utiliser les approximations des dérivées de f .

On exprime le $MISE$ de l'estimateur défini en (1.6) par :

$$MISE(h) = \mathbb{E}\left\{\sum_{x \in \mathbb{T}_N} \widehat{f}_n^2(x)\right\} - 2\mathbb{E}\left\{\sum_{x \in \mathbb{T}_N} \widehat{f}_n(x)f(x)\right\} + \sum_{x \in \mathbb{T}_N} f_n^2(x), \quad (1.14)$$

avec le terme $\sum_{x \in \mathbb{T}_N} f_n^2(x)$ qui est non aléatoire et indépendant de h . Ensuite, on rem-

place le terme stochastique par son estimateur non biaisé $CV(h)$ défini par :

$$\begin{aligned} CV(h) &= \sum_{x \in \mathbb{T}_N} \widehat{f}_n^2(x) - \frac{2}{n} \sum_{x \in \mathbb{T}_N} \widehat{f}_{n,-i}^2(X_i) \\ &= \sum_{x \in \mathbb{T}} \left\{ \sum_{x \in \mathbb{T}} K_{x,h}(X_i) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1} \sum_{i \neq j} K_{X_i}(X_j), \end{aligned} \quad (1.15)$$

où $\widehat{f}_{n,-i}$ est calculé à partir de \widehat{f}_n sans l'observation X_i . Enfin, on obtient le paramètre de lissage optimal par

$$h_{uc} = \arg \min_{h>0} CV(h).$$

1.3 Estimation à noyaux associés continus univariés

Définition 1.3.1. *Un type de noyau continu K_θ , est une famille de densités de probabilité paramétrées par $\theta \in \Theta \subseteq \mathbb{R}^2$, de support un intervalle $\mathbb{S}_\theta \subseteq \mathbb{R}$ et de carré intégrable. Puisque Θ est de dimension deux alors on peut écrire $\theta = (\theta_1, \theta_2)$ où $\theta_i : \mathbb{R} \rightarrow \mathbb{R}$ avec $i \in \{1, 2\}$. Pour une commodité d'écriture, nous allons écrire θ comme fonction de deux paramètres réels a et b i.e $\theta = \theta(a, b)$. Pour la suite, nous désignons respectivement par $M(a, b)$ et $D(a, b)$ le mode M_θ et le paramètre de dispersion D_θ du type de noyau K_θ .*

Il est important d'observer minutieusement les relations liant le mode et la moyenne ainsi que les paramètres de dispersion autour de chacun.

Remarque 1.3.1. (i) *Le mode $M(a, b)$ d'un type de noyau $K_{\theta(a,b)}$ appartient toujours à son support $S_{\theta(a,b)}$.*

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

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

Définition 1.3.2. *On considère $x \in \mathbb{T} \subseteq \mathbb{R}$ et $h > 0$ avec \mathbb{T} le support de la densité f , à estimer. Une fdp paramétrée $K_{x,h}$ de support $\mathbb{S}_{x,h} \subseteq \mathbb{R}$ est appelée noyau associé continu lorsque les conditions suivantes sont satisfaites :*

$$x \in \mathbb{S}_{x,h}, \tag{1.16}$$

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

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

où $Z_{x,h}$ est une variable aléatoire de densité $K_{x,h}$ sur $\mathbb{S}_{x,h}$ et $A(x, h)$, $B(x, h)$ tendent vers 0 quand h tend aussi vers 0.

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

Proposition 1.3.1. *Soit \mathcal{K} un noyau classique. Pour un $x \in \mathbb{T} = \mathbb{R}$ donné et $h > 0$, alors le noyau associé classique de support \mathbb{S}_x ($\mathbb{S}_{x,h} = x - h \times$ les bornes du \mathbb{S}) et de plus :*

$$\mathbb{E}(Z_{x,h}) = x \text{ et } \text{Var}(Z_{x,h}) = h^2 \sigma_{\mathcal{K}}^2. \tag{1.19}$$

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

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

Preuve : Il est facile de voir que le noyau classique devient un cas particulier des noyaux associés continus $K_{x,h}$ qui dépendent en essence de x et h , à travers la relation :

$$K_{x,h}(\cdot) = \frac{1}{h} \mathcal{K} \left(\frac{x - \cdot}{h} \right). \tag{1.21}$$

A partir de (1.21) Pour un x fixé dans \mathbb{T} et pour tout t dans \mathbb{T} , il existe u dans \mathbb{S} tel que $u = (x - t)/h$ et donc $t = x - uh$. Puisque $t \in \mathbb{T}$, il vient de (1.16) que $\mathbb{S}_{x,h} = x - h\mathbb{S}$. Les deux derniers résultats s'obtiennent facilement en calculant les moyenne et variance de $K_{x,h}$, grâce à l'utilisation du changement de variables $u = (x - t)/h$.

■

1.3.1 Constructions des noyaux associés continus par la méthode mode-dispersion

Principe de construction

Etant donné un type de noyau $K_{\theta(a,b)}$ sur $\mathbb{S}_{\theta(a,b)}$, uni-modal de mode $M(a,b)$ et d'un paramètre de dispersion $D(a,b)$, la méthode *Mode-Dispersion* permet la construction d'une fonction $K_{\theta(x,h)}$, dépendant de x et h en résolvant en a et b le système :

$$\begin{cases} M(a,b) = x \\ D(a,b) = h. \end{cases} \quad (1.22)$$

Les solutions $a = a(x,h)$ et $b = b(x,h)$ du système (1.22) permettent d'exprimer $\theta(a,b)$ en fonction de x et h . On obtient alors $\theta(a,b) = \theta(a(x,h), b(x,h))$ que nous désignons simplement par $\theta(x,h)$. En remplaçant $\theta(a,b)$ par $\theta(x,h)$ puis a et b par leurs valeurs $a(x,h)$ et $b(x,h)$ dans les expressions des moyenne et variance du type de noyau $K_{\theta(a,b)}$, on aboutit à $K_{\theta(x,h)}$ avec ses caractéristiques données dans la définition (1.3.2).

Nous montrons dans la proposition suivante que la fonction noyau $K_{\theta(x,h)}$ issue de la méthode mode-dispersion est bien un noyau associé. En d'autres termes, nous montrons que $K_{\theta(x,h)}$ satisfait les différentes conditions de la définition (1.3.2).

Proposition 1.3.2. *Soit \mathbb{T} le support de la densité f , à estimer. Pour tout $x \in \mathbb{T}$ et $h > 0$, la fonction noyau $K_{\theta(x,h)}$ construite par la méthode mode-dispersion, de support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h), b(x,h))}$, vérifie :*

$$x \in \mathbb{S}_{\theta(x,h)}, \quad (1.23)$$

$$\mathbb{E}(Z_{\theta(x,h)}) - x = A_{\theta}(x,h), \quad (1.24)$$

$$\text{Var}(Z_{\theta(x,h)}) = B_{\theta}(x, h), \quad (1.25)$$

où $Z_{\theta(x,h)}$ est une variable aléatoire de densité $K_{\theta(x,h)}$ puis $A_{\theta}(x, h)$ et $B_{\theta}(x, h)$ tendent vers 0 quand h tend vers 0.

Preuve : Tout d'abord, par la méthode mode-dispersion on a :

$$\theta(x, h) = \theta(a(x, h), b(x, h)),$$

ce qui permet d'avoir $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$. Puisque $K_{\theta(a,b)}$ est uni-modal et de mode $M(a, b)$ appartenant à $\mathbb{S}_{\theta(a,b)}$ (grâce au point (i) de la remarque (1.3.1), par la méthode mode-dispersion on a le premier résultat (1.23) comme suit :

$$M(a, b) = x \in \mathbb{S}_{\theta(a,b)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}.$$

De plus, pour une variable aléatoire donnée $Z_{\theta(a,b)}$ associée au type noyau uni-modal $K_{\theta(a,b)}$, on peut écrire

$$\mathbb{E}(Z_{\theta(a,b)}) = M(a, b) + \epsilon(a, b),$$

où $\epsilon(a, b)$ est la différence entre le mode et la moyenne de $K_{\theta(a,b)}$. La méthode mode-dispersion conduit à $M(a, b) = x$ et $\epsilon(a, b) = \epsilon(a(x, h), b(x, h))$. On a par conséquent :

$$\mathbb{E}(Z_{\theta(x,h)}) - x = \epsilon(a(x, h), b(x, h)).$$

En prenant $A_{\theta}(x, h) = \epsilon(a(x, h), b(x, h))$ puis, en utilisant la définition du paramètre de dispersion autour du mode, on obtient le deuxième résultat (1.24). Enfin, comme $K_{\theta(a,b)}$ admet un moment de second ordre, alors la variance de $K_{\theta(x,h)}$ existe et est une fonction de x et h . On peut l'écrire comme suit :

$$\text{Var}(Z_{\theta(x,h)}) = B_{\theta(a(x,h),b(x,h))},$$

avec $B_{\theta(a(x,h),b(x,h))}$ tend vers zéro quand h tend vers zéro d'après le point (iii) de la remarque (1.3.1) On obtient le dernier résultat (1.25) en prenant :

$$B_{\theta}(x, h) = B_{\theta(a(x,h),b(x,h))}.$$

■

Donnons maintenant un exemple illustrant la construction des noyaux associés non-classique.

Exemple(Noyau bêta) : Le noyau associé bêta a été introduit par Chen (1999)[8], mais dans ses travaux, celui-ci n'a pas montré comment il l'a construit. Nous le construisons dans ce paragraphe par la méthode mode-dispersion. Nous signalons à l'avance que les résultats obtenus sont les mêmes que ceux de Chen (1999)[8]. Considérons le type de noyau bêta noté $BE_{\theta(a,b)}$ de paramètres $a > 1$ et $b > 0$ (a et b sont tous des paramètres de forme) et défini sur $\mathbb{S}_{BE} = [0, 1]$ par :

$$BE_{\theta(a,b)}(u) = \frac{1}{\mathcal{B}(a,b)} u^{a-1} (1-u)^{b-1} \mathbf{1}_{[0,1]}(u),$$

où $\mathcal{B}(a,b)$ est la fonction bêta. Son mode et son paramètre de dispersion sont respectivement

$$(a-1)/(a+b-2) \text{ et } 1/(a+b-2).$$

La résolution du système (1.22) donne pour tout $x \in [0, 1]$ et $h > 0$:

$$\theta(x, h) = \left(\frac{x}{h} + 1, \frac{1-x}{h} + 1 \right).$$

Ce qui permet d'écrire le noyau associé bêta $BE_{\theta(x,h)}$ comme suit :

$$BE_{\theta(x,h)}(u) = \frac{1}{\mathcal{B}\left(\frac{x}{h} + 1, \frac{1-x}{h} + 1\right)} u^{x/h} (1-u)^{(1-x)/h} \mathbf{1}_{[0,1]}(u).$$

Son support est

$$\mathbb{S}_{BE_{\theta(x,h)}} = [0, 1] = \mathbb{S}_{BE}.$$

En injectant les composantes $\theta(x, h)$

$$a(x, h) = (x/h) + 1 \text{ et } b(x, h) = \{(1-x)/h\} + 1,$$

dans les expressions des moyenne et variance de $BE_{\theta(a,b)}$ définies respectivement par $a/(a+b)$ et $ab/\{(a+b)^2(a+b+1)\}$, on trouve :

$$A_{\theta}(x, h) = \frac{h(1-2x)}{1+2h} \text{ et } B_{\theta}(x, h) = \frac{h\{x(1-x) + h + h^2\}}{(1+3h)(1+2h)^2}.$$

Remarque 1.3.2. *Remarquons cependant que certains types de noyaux ne satisfont pas le principe mode-dispersion, car il y en a qui ont des modes explicites et d'autres n'en ont pas. C'est notamment le cas du type de noyau Weibul, de paramètres $a > 1$ et $b > 0$ (respectivement paramètres de forme et d'échelle). Il est de moyenne et variance*

$$b\Gamma(1 + 1/a) \text{ et } b^2\{\Gamma(1 + 2/a) - (2/a) \log(2)\},$$

où $\Gamma(a)$ représente la fonction gamma. Il est défini sur $]0, +\infty[$ par :

$$W_{\theta(a,b)}(u) = \frac{a}{b^a} u^{a-1} \exp\left\{-\left(\frac{x}{b}\right)^a\right\} 1_{[0,+\infty[}(u).$$

Son mode est égal à $b(1 - 1/a)^{1/a}$ et son paramètre de dispersion est b . Le système (1.22) donne :

$$\begin{cases} (1 - 1/a)^{1/a} & = x/h \\ b & = h. \end{cases}$$

On constate par là que les solutions $a(x, h)$ et $b(x, h)$ du système précédent seront telles que $b(x, h) = h$ est indépendant de x mais $a(x, h)$ dépend fortement de x et h d'une manière non linéaire.

1.3.2 Estimateur à noyaux associés continus univariés

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

Définition 1.3.3. *Soit f la densité à estimer sur $\mathbb{T} \subseteq \mathbb{R}$, $h > 0$ et $K_{\theta(x,h)}$ un noyau associé continu. L'estimateur f_n de f issu de $K_{\theta(x,h)}$ est défini par :*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{\theta(x,h)}(X_i), x \in \mathbb{T} \subseteq \mathbb{R}. \quad (1.26)$$

Donnons maintenant les propriétés de cet estimateur.

1.3.3 Résultats asymptotiques

Proposition 1.3.3. *L'estimateur défini en (1.26) vérifie les propriétés suivantes :*

$$\mathbb{E}\{\widehat{f}_n(x)\} = \mathbb{E}\{f(Z_{\theta(x,h)})\}, \quad (1.27)$$

$$\widehat{f}_n(x) \geq 0, \quad (1.28)$$

$$\int_{\mathbb{T}} \widehat{f}_n(x) dx = \Lambda(n, h, K) \quad \text{n'est pas toujours égale 1,} \quad (1.29)$$

où $Z_{\theta(x,h)}$ est une variable aléatoire de loi $K_{\theta(x,h)}$.

La remarque suivante explicite la dernière propriété de cette proposition.

Remarque 1.3.3. *La troisième propriété de la proposition (1.3.3) montre que l'estimateur à noyau associé n'est pas forcément une densité de probabilité. Sa masse totale $\Lambda(n, h, K) = \Lambda$ est positive et finie i.e.*

$$\int_{\mathbb{T}} K_{\theta(x,h)}(u) dx < \infty,$$

pour tout $u \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}$. Cependant, on peut facilement vérifier que $\Lambda = 1$ dans le cas des noyaux classiques.

Preuve : La première propriété s'obtient à partir de la définition de $\mathbb{E}\{\widehat{f}_n(x)\}$ pour tout $x \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}$ et $Z_{\theta(x,h)}$ une variable aléatoire de loi $K_{\theta(x,h)}$, comme suit :

$$\mathbb{E}\{\widehat{f}_n(x)\} = \int_{\mathbb{S}_{\theta(x,h)} \cap \mathbb{T}} K_{\theta(x,h)}(t) f(t) dt = \int_{\mathbb{S}_{\theta(x,h)} \cap \mathbb{T}} f(t) K_{\theta(x,h)}(t) dt = \mathbb{E}\{f(Z_{\theta(x,h)})\}.$$

Pour la deuxième propriété, partant du fait que $K_{\theta(x,h)}$ est une densité de probabilité, on a $K_{\theta(x,h)}(X_j) \in [0, 1]$, pour tout $x \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}$, $X_j \in \mathbb{T}$ et $h > 0$. Puisque

les $(K_{\theta(x,h)}(X_j))_j$ sont aussi (i.i.d), alors $\frac{1}{n} \sum_{i=1}^n K_{\theta(x,h)}(X_j) \in [0, 1]$. Quant à la der-

nière propriété, elle se déduit de la remarque (1.3.3) dans le cas des noyaux associés classiques. ■

Nous présentons dans la proposition suivante les formes assez particulières des biais et variance de l'estimateur à noyaux associés continus.

Proposition 1.3.4. *Soit \widehat{f}_n l'estimateur à noyau associé (1.26) de $f \in \mathcal{C}^2(\mathbb{T})$. Pour tout $x \in \mathbb{T}$ et $h = h_n > 0$, alors :*

$$\text{Biais}\{\widehat{f}_n(x)\} = A_\theta(x, h)f'(x) + \frac{1}{2}\{A_\theta^2(x, h) + B_\theta(x, h)\}f''(x) + o(h^2). \quad (1.30)$$

De plus, si f est bornée sur \mathbb{T} alors

$$\text{Var}\{\widehat{f}_n(x)\} = \frac{1}{n}f(x)\|K_{\theta(x, h)}\|_2^2 + o\left(\frac{1}{nh^{r_2}}\right), \quad (1.31)$$

où $r_2 = r_2(K_{\theta(x, h)}) > 0$ est le plus grand réel tel que

$$\|K_{\theta(x, h)}\|_2^2 = \int_{\mathbb{S}_{\theta(x, h)} \cap \mathbb{T}} K_{\theta(x, h)}^2(u)du \leq c_2(x)h_n^{-r_2} \quad \text{et } 0 < c_2(x) < \infty.$$

Remarque 1.3.4. *Signalons que les estimateurs à noyaux associés continus sont aussi (comme dans le cas classique), asymptotiquement sans biais. Cependant, l'expression du biais donné dans cette proposition diffère du cas classique par la présence du terme $A_\theta(x, h)f'(x)$ qui est visiblement non négligeable. Ceci peut être interprété comme l'équivalent des effets de bords dans le cas classique.*

Preuve : Pour tout $x \in \mathbb{T}$ et $r > 0$, on obtient le biais de \widehat{f}_n en procédant de la manière suivante :

$$\begin{aligned} \text{Biais}(\widehat{f}_n(x)) &= \mathbb{E}(f(Z_{\theta(x, h)})) - f(x) \quad (i) \\ &= f\{\mathbb{E}(Z_{\theta(x, h)})\} + \frac{1}{2}\text{Var}(Z_{\theta(x, h)})f'\{\mathbb{E}(Z_{\theta(x, h)})\} - f(x) \\ &\quad + o(\mathbb{E}\{Z_{\theta(x, h)} - \mathbb{E}(Z_{\theta(x, h)})\}^2) \quad (ii) \\ &= f\{x + A_\theta(x, h)\} + \frac{1}{2}B_\theta(x, h)f'\{x + A_\theta(x, h)\} - f(x) \\ &\quad + o\{B_\theta(x, h)\} \quad (iii) \\ &= A_\theta(x, h)f'(x) + \frac{1}{2}\{A_\theta^2(x, h) + B_\theta(x, h)\}f''(x) + o(h^2). \quad (iv) \end{aligned}$$

L'égalité (i) vient de la proposition (1.3.3). Le (ii) est obtenu après développement de $f(Z_{\theta(x, h)})$ en séries de Taylor au voisinage de $\mathbb{E}(Z_{\theta(x, h)})$ à l'ordre deux. Pour le (iii), on remplace dans (ii) successivement, $f(Z_{\theta(x, h)})$ par son approximation de Taylor, puis $\mathbb{E}(Z_{\theta(x, h)})$ par son expression dans la définition (1.3.2).

Enfin, le (iv) s'obtient en approximant $f(x + A_\theta(x, h))$ par les séries de Taylor au

voisinage de x à l'ordre deux et en soustrayant $f(x)$ par la suite. En fait, le reste $o(h^2)$ vient de (1.20) et

$$o\left(\mathbb{E}\left\{Z_{\theta(x,h)} - \mathbb{E}(Z_{\theta(x,h)})\right\}^2\right) = \mathbb{E}\left(o_P\left\{Z_{\theta(x,h)} - \mathbb{E}(Z_{\theta(x,h)})\right\}^2\right),$$

où $o_P(\cdot)$ est le taux de convergence en probabilité. Concernant la variance on a :

$$\begin{aligned} \text{Var}\{\widehat{f}_n(x)\} &= \frac{1}{n}\mathbb{E}\left\{K_{\theta(x,h)}^2(X_1)\right\} - \frac{1}{n}\left[\mathbb{E}\{K_{\theta(x,h)}(X_1)\}\right]^2 \\ &= \frac{1}{n}\int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}}K_{\theta(x,h)}^2(u)f(u)du - \frac{1}{n}\left[\mathbb{E}\{K_{\theta(x,h)}(X_1)\}\right]^2 \\ &= I_1 - I_2. \end{aligned}$$

A partir de (1.27) et (1.30) le second terme se comporte de la manière suivante :

$$I_2 = (1/n)\left[\mathbb{E}\{K_{\theta(x,h)}(X_1)\}\right]^2 \simeq (1/n)f^2(x) \simeq o(1/n),$$

puisque f est bornée pour tout $x \in \mathbb{T}$. En utilisant le développement de f en séries de Taylor aux voisinages de x , le premier terme I_1 donne :

$$I_1 = \frac{1}{n}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}}K_{\theta(x,h)}^2(u)f(u)du = \frac{1}{n}f(x)\int_{\mathbb{S}_{x,h}\cap\mathbb{T}}K_{\theta(x,h)}^2(u)du + R(x,h),$$

avec

$$R(x,h) = \frac{1}{n}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}}K_{\theta(x,h)}^2(u)\left[(u-x)f'(x) + \frac{(u-x)^2}{2}f''(x) + o\{(u-x)^2\}\right]du.$$

Sous l'hypothèse que $\|K_{\theta(x,h)}\|_2^2 \leq c_2(x)h_n^{-r_2}$ nous déduisons successivement :

$$0 \leq R(x,h) \leq \frac{1}{nh^{r_2}}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}}c_2(x)\left\{(u-x)f'(x) + \frac{(u-x)^2}{2}f''(x)\right\}du \simeq o(n^{-1}h^{-r_2}).$$

■

Le théorème suivant nous présente les erreurs quadratiques moyennes ponctuelles et globales.

Théorème 1.3.1. (*Erreurs quadratiques ponctuelles et globales*) *Le MSE et le MISE de l'estimateur à noyaux associés continus sont comme dans le cas des noyaux classiques, définis respectivement par :*

$$MSE(x) = Var \left\{ \widehat{f}_n(x) \right\} + Biais^2 \left\{ \widehat{f}_n(x) \right\}, \quad (1.32)$$

et

$$MISE(\widehat{f}_{n,h,K,f}) = \int_{\mathbb{T}} Var \{ \widehat{f}_n(x) \} dx + \int_{\mathbb{T}} Biais^2 \{ \widehat{f}_n(x) \} dx. \quad (1.33)$$

En remplaçant la variance et le biais par leurs valeurs asymptotiques on obtient respectivement les expressions asymptotiques du MSE et de MISE données par :

$$MSE(x) = \left[A_{\theta}(x, h) f'(x) + \frac{1}{2} \{ A_{\theta}^2(x, h) + B_{\theta}(x, h) \} f''(x) \right]^2 + \frac{1}{n} f(x) \|K_{\theta(x,h)}\|_2^2, \quad (1.34)$$

et

$$MISE \left(\widehat{f}_{n,h,K,f} \right) = \int_{\mathbb{T}} \left(\left[A_{\theta}(x, h) f'(x) + \frac{1}{2} \{ A_{\theta}^2(x, h) + B_{\theta}(x, h) \} f''(x) \right]^2 \right) dx + \int_{\mathbb{T}} \frac{1}{n} \|K_{\theta(x,h)}\|_2^2 f(x) dx. \quad (1.35)$$

La proposition suivante donne le paramètre du lissage optimal.

Proposition 1.3.5. *Soit $f \in \mathcal{C}^2(\mathbb{T})$, la densité à estimer. Si les dérivées premières et secondes de f sont bornées alors la fenêtre optimale de lissage qui minimise le MISE est :*

$$h = C(x) n^{-1/(r_2+2)},$$

avec $r_2 = r_2(K_{\theta(x,h)})$ défini dans (1.31).

Preuve : La définition (1.3.2) et la proposition (1.3.1) permettent de dire qu'il existe deux constantes positives $c_1^*(x)$ et $c_1^{**}(x)$ telles que :

$$A_{\theta}(x, h) \leq c_1^*(x)h \text{ et } B_{\theta}(x, h) \leq c_1^{**}(x)h^2.$$

D'après la proposition (1.3.4), on a :

$$\text{Biais} \left\{ \widehat{f}_n(x) \right\} \leq hc_1(x) \text{ et } \text{Var} \left\{ \widehat{f}_n(x) \right\} \leq n^{-1}h^{-r_2}c_2(x),$$

avec $c_1(x) = \sup_{x \in \mathbb{T}} |c_1^*(x)f'(x) + c_1^{**}(x)f''(x)|$. De (1.32), on déduit que :

$$MSE(x) \leq h^2c_1^2(x) + n^{-1}h^{-r_2}c_2(x). \quad (1.36)$$

En intégrant (1.36), on obtient :

$$MISE\left(\widehat{f}_{n,h,K,f}\right) \leq h^2C_1(x) + n^{-1}h^{-r_2}C_2(x),$$

avec $C_1(x)$ et $C_2(x)$ les primitives respectivement de $c_1^2(x)$ et $c_2(x)$ sur \mathbb{T} . En prenant le second membre égal à 0 on trouve le résultat. ■

1.3.4 Réduction de biais et noyaux associés modifiés

Pour réduire le biais de $\widehat{f}_{n,h,K,f}$ défini en (1.30), on procède en deux étapes. La première étape consiste à définir les régions de bords et de l'intérieur et la seconde traite de la modification du noyau associé qui conduit à un biais réduit.

Algorithme de réduction du biais

Voici comment on peut réduire le biais ponctuel de $\widehat{f}_{n,h,K,f}$:

Première étape : On divise le support $\mathbb{T} = [t_1, t_2]$ en deux régions de longueur $\alpha(h) > 0$ avec $\alpha(h) \rightarrow 0$ quand $h \rightarrow 0$,

(i) *région intérieure* (La plus grande région pouvant contenir 95% des observations) notée $\mathbb{T}_{\alpha(h),0}$ et définie par l'intervalle :

$$\mathbb{T}_{\alpha(h),0} =]t_1 + \alpha(h), t_2 - \alpha(h)[,$$

(ii) *régions de bords* (pouvant être vide) représentées par les deux intervalles $\mathbb{T}_{\alpha(h),-1}$ et $\mathbb{T}_{\alpha(h),+1}$ respectivement définis par :

$$\mathbb{T}_{\alpha(h),-1} = [t_1, t_1 + \alpha(h)] \text{ (gauche),}$$

et

$$\mathbb{T}_{\alpha(h),+1} = [t_2 - \alpha(h), t_2] \text{ (droite).}$$

Les régions de bords constituent le complémentaire de la région intérieure et on note :

$$\mathbb{T}_{\alpha(h),0}^c = \mathbb{T}_{\alpha(h),-1} \cup \mathbb{T}_{\alpha(h),+1}.$$

Deuxième étape : On modifie le noyau associé $K_\theta(x, h)$ correspondant à $A_\theta(x, h)$ et $B_\theta(x, h)$ en une nouvelle fonction noyau $K_{\tilde{\theta}(x,h)}$ correspondant à

$$\tilde{A}(x, h) = (\tilde{A}_{-1}(x, h), \tilde{A}_0(x, h), \tilde{A}_{+1}(x, h)) \text{ et}$$

$$\tilde{B}(x, h) = (\tilde{B}_{-1}(x, h), \tilde{B}_0(x, h), \tilde{B}_{+1}(x, h)) \text{ de sorte que, pour un } h \text{ fixé,}$$

$$\begin{aligned} \tilde{\theta}(x, h) &= (\tilde{\theta}_{-1}(x, h), \tilde{\theta}_0(x, h), \tilde{\theta}_{+1}(x, h)) \\ &= \begin{cases} \tilde{\theta}_{-1}(x, h) & \text{si } x \in \mathbb{T}_{\alpha(h),-1} \\ \tilde{\theta}_0(x, h) : \tilde{A}_0(x, h) = 0 & \text{si } x \in \mathbb{T}_{\alpha(h),0} \\ \tilde{\theta}_{+1}(x, h), & x \in \mathbb{T}_{\alpha(h),+1} \end{cases} \end{aligned} \quad (1.37)$$

soit continue sur \mathbb{T} et constant sur $\mathbb{T}_{\alpha(h),0}^c$ voir Chen (2010)[8].

La proposition suivante montre que la fonction noyau modifié $K_{\tilde{\theta}(x,h)}$ de support $\mathbb{S}_{\tilde{\theta}(x,h)} = \mathbb{S}_{\theta(x,h)}$ est un noyau associé continu.

Proposition 1.3.6. *La fonction $K_{\tilde{\theta}(x,h)}$ issue de (1.37) est un noyau associé de même type $K = K_{\theta(x,h)}$.*

Preuve : Nous montrons ici que $K_{\tilde{\theta}(x,h)}$ satisfait les conditions de la définition (1.3.2). En partant du fait que $K_{\theta(x,h)}$ est un noyau associé puis en utilisant la première étape de l'algorithme de réduction de biais, on a pour tout $j \in J = \{-1, 0, +1\}$,

$$x \in \mathbb{T}_{(\alpha(h),j)} \implies x \in \mathbb{T} \implies x \in \mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\tilde{\theta}(x,h)}.$$

Les deux derniers résultats découlent de la proposition (1.3.2), c'est-à-dire que pour toute variable aléatoire $Z_{\tilde{\theta}(x,h)}$ de densité $K_{\tilde{\theta}(x,h)}$ on a :

$$\mathbb{E}(Z_{\tilde{\theta}(x,h)}) = x + A_{\tilde{\theta}(x,h)},$$

et

$$\text{Var}(Z_{\tilde{\theta}(x,h)}) = B_{\tilde{\theta}(x,h)},$$

avec $A_{\tilde{\theta}(x,h)}$ et $B_{\tilde{\theta}(x,h)}$ tendent vers 0 quand h tend vers 0. En prenant

$\tilde{A}(x, h) = A_{\tilde{\theta}(x,h)}$ et $\tilde{B}(x, h) = B_{\tilde{\theta}(x,h)}$ il vient que pour tout $j \in J$, $\tilde{A}_j(x, h) = A_{\tilde{\theta}_j(x,h)}$ et $\tilde{B}_j(x, h) = B_{\tilde{\theta}_j(x,h)}$. En particulier, à partir de (1.3.2) on obtient $A_{\tilde{\theta}_0(x,h)} = 0$.

■

Ainsi, à partir de la définition (1.3.2) puis en utilisant le noyau associé modifié $K_{\tilde{\theta}(x,h)}$, on définit l'estimateur à noyau associé modifié \tilde{f}_n de f par :

$$\tilde{f}_n(x) = \frac{1}{h} \sum_1^n K_{\tilde{\theta}(x,h)}(X_i). \quad (1.38)$$

La proposition suivante donne les expressions du biais et la variance de \hat{f}_n .

Proposition 1.3.7. *Soit $\mathbb{T} = \bigcup_{j \in J} \mathbb{T}_{\alpha(h),j}$ ($J = \{-1, 0, +1\}$), le support de la densité f à estimer, \tilde{f}_n et \hat{f}_n les estimateurs à noyaux associés continus de f donnés respectivement définis en (1.26) et (1.38). Pour tout $x \in \mathbb{T}_{\alpha(h),0}$ on a :*

$$\text{Biais} \left\{ \tilde{f}_n(x) \right\} = \frac{1}{2} B_{\tilde{\theta}_0}(x, h) f''(x) + o(h^2),$$

et

$$\text{Var} \left\{ \tilde{f}_n(x) \right\} \simeq \text{Var} \left\{ \hat{f}_n(x) \right\} \text{ quand } h \longrightarrow 0.$$

Preuve : On obtient le premier résultat en remplaçant dans (1.30) respectivement, \hat{f}_n , A_θ et B_θ par \tilde{f}_n , \tilde{A}_0 et \tilde{B}_0 . Pour le second résultat, il suffit de montrer que :

$$\|K_{\theta(x,h)}\|_2^2 \simeq \|K_{\tilde{\theta}(x,h)}\|_2^2 \text{ quand } h \longrightarrow 0.$$

En effet, puisque $K_{\theta(x,h)}$ et $K_{\tilde{\theta}(x,h)}$ sont des noyaux associés de même type K , $r_2 = r_2(K_{\theta(x,h)})$ et $\tilde{r}_2 = \tilde{r}_2(K_{\tilde{\theta}(x,h)})$, il existe un plus grand réel $r = r(K)$ tel que

$$\|K_{\theta(x,h)}\|_2^2 \leq c_2(x) h^{-r} \text{ et } \|K_{\tilde{\theta}(x,h)}\|_2^2 \leq \tilde{c}_2(x) h^{-r},$$

avec $0 < c_2(x), \tilde{c}_2(x) < \infty$. En prenant

$$c(x) = \sup\{c_2(x), \tilde{c}_2(x)\},$$

on a :

$$\|K_{\theta(x,h)}\|_2^2 \leq c(x)h^{-2r} \text{ et } \|K_{\tilde{\theta}(x,h)}\|_2^2 \leq c(x)h^{-2r}.$$

Puisque $c(x)/nh^{2r} = o(n^{-1}h^{-2r})$ alors :

$$\|K_{\tilde{\theta}(x,h)}\|_2^2 \simeq \|K_{\theta(x,h)}\|_2^2.$$

■

1.3.5 Choix du noyau

Nous proposons dans ce paragraphe, un critère pour le choix du noyau essentiellement dans la partie la plus importante du support \mathbb{T} en l'occurrence $\mathbb{T}_{(\alpha(h),0)}$. Ce critère est basé sur la comparaison des erreurs quadratiques moyennes intégrées asymptotiques. D'après la proposition précédente, nous pouvons écrire l'erreur quadratique moyenne ponctuelle de \tilde{f}_n sur $\mathbb{T}_{(\alpha(h),0)}$ notée MSE_0 , de la manière suivante :

$$MSE_0(x) = \frac{1}{4}B_{\frac{2}{\theta}}(x, h)\{f''(x)\}^2 + \frac{1}{n}\|K_{\tilde{\theta}(x,h)}\|_2^2 f(x),$$

par conséquent le $MISE_0$ est donné par :

$$MISE_0(\tilde{f}_{n,h,K,f}) = \int_{\mathbb{T}_{\alpha(h),0}} \left[\frac{1}{4}B_{\frac{2}{\theta}}(x, h)\{f''(x)\}^2 + \frac{1}{n}\|K_{\tilde{\theta}(x,h)}\|_2^2 f(x) \right] dx.$$

Ainsi, pour des noyaux associés de même support \mathbb{S} , le noyau optimal (le noyau qui estime mieux) noté $K_{\theta \text{ opt}}$ dans la région intérieure de \mathbb{T} (i.e. sur $\mathbb{T}_{\alpha(h),0}$) est celui qui a le plus petit $MISE_0$ pour tout $h > 0$. On écrit alors :

$$K_{\theta \text{ opt}} = \arg_{K_{\theta \text{ sur } \mathbb{S}}} \min MISE_0(\tilde{f}_{n,h,K}).$$

Remarque 1.3.5. *Lorsque $\mathbb{T} = \mathbb{R}$, ceci coïncide avec le cas classique et le noyau optimal est celui d'Epachnikov.*

1.3.6 Choix de la fenêtre du lissage par la méthode de validation croisée par les moindres carrés

Considérons $K_{\theta(x,h)}$ le noyau associé construit par la méthode mode-dispersion avec $x \in \mathbb{T}$ et $h > 0$. Le paramètre optimal h_{CV} de h est obtenu par :

$$h_{CV} = \arg_{h>0} \min CV(h),$$

où

$$\begin{aligned} CV(h) &= \int_{x \in \mathbb{T}} \left\{ \widehat{f}_n(x) \right\}^2 dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i) \\ &= \int_{x \in \mathbb{T}} \left\{ -\frac{1}{n} \sum_{i=1}^n K_{\theta_{x,h}}(X_i) \right\}^2 dx - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i), \end{aligned}$$

et $\widehat{f}_{n,-i}(u) = (n-1)^{-1} \sum_{j \neq i} K_{\theta(u,h)}(X_j)$ est calculé à partir de $\widehat{f}_n(u)$ sans l'observation

X_j .

Chapitre 2

Estimation par noyaux associés multidimensionnels

2.1 Estimation par noyaux associés discrets multidimensionnels

Définition 2.1.1. (Noyau associé discret multivarié) Soient $x \in \mathbb{T}_d \subseteq \mathbb{Z}^d$ et H une matrice des fenêtres, avec \mathbb{T}_d est le support de la fmp f à estimer. Une fmp $K_{x,H}(\cdot)$ de support $S_{x,H} \subseteq \mathbb{Z}^d$ est appelée noyau associé discret multivarié si

$$x \in S_{x,H}, \quad (2.1)$$

$$E(Z_{x,H}) = x + A(x, H), \quad (2.2)$$

$$Cov(Z_{x,H}) = B(x, H), \quad (2.3)$$

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

Définition 2.1.2. (Noyau associé discret produit) Soient $\mathbb{T}_1^{[j]}$ le support des marges univariés de f , $x = (x_1, x_2, \dots, x_d)^t$ est le vecteur cible h_j avec $j = 1, \dots, d$ sont les fenêtres de lissage univariés et $K_{x_j, h_j}^{[j]}$ est le j -ème noyau associé discret univarié de support S_{x_j, h_j} . "Le noyau associé discret multivarié produit" est défini comme suit :

$$K_{x,H}(\cdot) = \prod_{j=1}^d K_{x_j, h_j}^{[j]}(\cdot), \quad \forall x_j \in \mathbb{T}_1^{[j]} \subseteq \mathbb{Z}. \quad (2.4)$$

2.1.1 Exemple(*Dirac Uniforme Discret*)

Le noyau catégoriel multivarié Dirac Uniforme Discret introduit par Aitchison et Aitken (1976)[3] et Racine et Li (2004)[25] est donné comme suit :

$$DirDU_{x,H,c}(y) = \prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{y_j=x_j}} \left(\frac{h_j}{c_j - 1}\right)^{1 - \mathbb{1}_{y_j=x_j}}, \quad (2.5)$$

où $S_{x,c} = \times_{j=1}^d \{0, 1, \dots, c_j - 1\} = \mathbb{T}_d$, avec $\times_{j=1}^d$ est le produit Cartésien des ensembles $\{0, 1, \dots, c_j - 1\}$ tel que $c_j \in \{2, 3, \dots\} \forall j = 1, \dots, d$, $H = \text{Diag}_d(h_j)$ et $\mathbb{1}_A$ est la fonction indicatrice de A .

Ce noyau vérifie bien les équations (2.1) et (2.3). En effet, on a :

$$\begin{aligned} \mathbb{E}(Z_{x,H}) &= \sum_{y \in \mathbb{T}_d} (y_1, \dots, y_d) \left[\prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{y_j=x_j}} \left(\frac{h_j}{c_j - 1}\right)^{1 - \mathbb{1}_{y_j=x_j}} \right] \\ &= \left(\sum_{y_1=0}^{c_1-1} y_1 \left[\prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{y_j=x_j}} \left(\frac{h_j}{c_j - 1}\right)^{1 - \mathbb{1}_{y_j=x_j}} \right], \dots, \right. \\ &\quad \left. \sum_{y_d=0}^{c_d-1} y_d \left[\prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{y_j=x_j}} \left(\frac{h_j}{c_j - 1}\right)^{1 - \mathbb{1}_{y_j=x_j}} \right] \right) \\ &= \left(x_1 + h_1 \left(1 - \frac{x_1}{c_1 - 1} + \frac{h_1 c_1}{2}\right), \dots, x_d + h_d \left(1 - \frac{x_d}{c_d - 1} + \frac{h_d c_d}{2}\right) \right)^t \\ &= x + H \left(1 - \frac{x_1}{c_1 - 1} + \frac{h_1 c_1}{2}, \dots, 1 - \frac{x_d}{c_d - 1} + \frac{h_d c_d}{2} \right)^t \\ &= x + A(x, H), \end{aligned}$$

avec $a(x, H) \rightarrow 0$ quand $H \rightarrow 0_d$, et

$$\begin{aligned} Cov(Z_{x,H}) &= Cov \left(\prod_{j=1}^d Z_{x_j, h_j, c_j}^{[j]} \right) \\ &= \text{Diag}_d \left(\text{Var} \left(Z_{x_j, h_j, c_j}^{[j]} \right) \right)_j \\ &= H \text{Diag}_d \left(x_j^2 \frac{c_j^2 (1 - h_j) - c_j}{(c_j - 1)^2} - x_j \frac{c_j^2 (1 - h_j) - c_j}{c_j - 1} + \frac{c_j}{2} \left(\frac{2c_j - 1}{3} - \frac{h_j c_j}{2} \right) \right)_j \\ &= B(x, H), \end{aligned}$$

avec $B(x, H) \rightarrow 0_d$ lorsque $H \rightarrow 0_d$.

Remarque 2.1.1. *La version unidimensionnelle du noyau Dirac Uniforme Discret (2.5) peut s'écrire :*

$$\text{DirDU}_{x,h,c}(y) = (1-h)\mathbb{1}_{y=x} + \left(\frac{h}{c-1}\right)\mathbb{1}_{y \neq x}, \quad (2.6)$$

dont le support est $\mathbb{S}_{x,c} = \{0, 1, \dots, c-1\}$, avec fixé telque $c \in \{2, 3, \dots\}$ et $h \in [0, 1]$.

2.1.2 Estimateur à noyau associé discret multivarié

Définition 2.1.3. *Soient x_1, \dots, x_n des vecteurs aléatoires (i.i.d) de fmp multivariée commune inconnue f à estimer sur \mathbb{T}_d . L'estimateur à noyau associé discret multivarié \hat{f}_n de f est de la forme :*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x,H}(X_i), \quad x \in \mathbb{T}_d, \quad (2.7)$$

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

Définition 2.1.4. *(L'estimateur à noyau associé produit) Selon la définition (2.1.2), le produit des noyaux associés discrets univariés (2.4) conduit à une version utile de (2.7) que nous appelons "estimateur à noyau associé produit" :*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{x_j, h_j}^{[j]}(X_{ij}), \quad x_j \in \mathbb{T}_1^{[j]} \subseteq \mathbb{Z}, \quad (2.8)$$

où $\mathbb{T}_1^{[j]}$ est le support de la marge univariée de f pour $j = 1, \dots, d$, $x = (x_1, \dots, x_d)^t \in \times_{j=1}^d \mathbb{T}_1^{[j]}$, $X_i = (X_{i1}, \dots, X_{id})^t$ pour $i = 1, \dots, n$ et h_1, \dots, h_d sont les paramètres de lissage unidimensionnels. La fonction $K_{x_j, h_j}^{[j]}$ est le j^{eme} noyau associé discret univarié de support $S_{x_j, h_j} \subseteq \mathbb{Z}$.

2.1.3 Propriétés asymptotiques

Dans cette partie, nous examinons les différentes propriétés à distance finie et asymptotique de l'estimateur à noyaux associé produit donné dans (2.8), en utilisant les noyaux associés discrets du type Dirac Uniforme Discret.

Exemple(Cas du noyau Dirac Uniforme Discret) On considère les noyaux associés : Dirac Uniforme Discret du exemple, alors \widehat{f}_n est toujours une fmp, à savoir $C_n = 1$. En effet, dans le cas du noyau Dirac Uniforme Discret avec $\mathbb{T}_d = \times_{j=1}^d \{0, 1, \dots, c_j - 1\}$ on a :

1.

$$\begin{aligned} \sum_{x \in \mathbb{T}_d} \widehat{f}_n(x) &= \sum_{x \in \mathbb{T}_d} \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{X_{ij}=x_j}} \left(\frac{h_j}{c_j - 1} \right)^{1 - \mathbb{1}_{X_{ij}=x_j}} \\ &= \sum_{x \in \mathbb{T}_d} \prod_{j=1}^d (1 - h_j)^{\mathbb{1}_{X_{1j}=x_j}} \left(\frac{h_j}{c_j - 1} \right)^{1 - \mathbb{1}_{X_{1j}=x_j}} \\ &= \prod_{j=1}^d [(1 - h_j) + h_j] \\ &= 1. \end{aligned}$$

2. Biais

$$\mathbb{E} \left(\widehat{f}_n(x) \right) = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \text{Dir} DU_{x,H,c}(x_i) \right) = \mathbb{E} \left(f(Z_{x_1, h_1, c_1}^{[1]}, \dots, Z_{x_d, h_d, c_d}^{[d]}) \right),$$

où $Z_{x_j, h_j, c_j}^{[j]}$ sont des variables aléatoires de la loi d'Aitchison & Aitken, elles sont indépendantes de moyenne :

$$\mu_j = x_j + \dot{h}_j, \text{ où } \dot{h}_j = h_j \left(1 - x_j - \frac{x_j}{c_j - 1} + \frac{h_j c_j}{2} \right)$$

et de variance :

$$\sigma_j^2 = h_j \left(x_j^2 \frac{c_j^2 (1 - h_j) - c_j}{(c_j - 1)^2} - x_j \frac{c_j^2 (1 - h_j) - c_j}{c_j - 1} + \frac{c_j}{2} \left(\frac{2c_j - 1}{3} - \frac{h_j c_j}{2} \right) \right).$$

En utilisant le développement en séries de Taylor à l'ordre 2 et en remplaçant les dérivées partielles par les différences finies, on obtient :

$$\begin{aligned}
\mathbb{E}(\widehat{f}_n(x)) &= f(\mu_1, \mu_2, \dots, \mu_d) + \frac{1}{2} \sum_{j=1}^d \text{Var}(Z_{x_j, h_j, c_j}^{[j]}) f_{jj}^{(2)} + o\left(\sum_{j=1}^d h_j^2\right) \\
&= f(x_1, \dots, x_d) + \sum_{j=1}^d h_j f_j^{(1)} + \frac{1}{2} \sum_{j=1}^d \text{Var}(Z_{x_j, h_j, c_j}^{[j]}) f_{jj}^{(2)} + o\left(\sum_{j=1}^d h_j^2\right).
\end{aligned}$$

Donc, le biais de l'estimateur est donné comme suit :

$$\begin{aligned}
\text{Biais}(\widehat{f}_n(x)) &= \sum_{j=1}^d h_j \left(1 - x_j - \frac{x_j}{c_j - 1} + \frac{h_j c_j}{2}\right) f_j^{(1)} \\
&\quad + \frac{1}{2} \sum_{j=1}^d h_j \left(x_j^2 \frac{c_j^2 (1 - h_j) - c_j}{(c_j - 1)^2} - x_j \frac{c_j^2 (1 - h_j) - c_j}{c_j - 1}\right. \\
&\quad \left. + \frac{c_j}{2} \left(\frac{2c_j - 1}{3} - \frac{h_j c_j}{2}\right)\right) f_{jj}^{(2)} \\
&\quad + o\left(\sum_{j=1}^d h_j^2\right).
\end{aligned}$$

3. Variance

La variance de $\widehat{f}_n(x)$ se décompose autour du vecteur cible comme suit :

$$\begin{aligned}
\text{Var}(\widehat{f}_n(x)) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{x_j, h_j}^{[j]}(X_{ij})\right) \\
&= \frac{1}{n} \left[f(x) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = x_j)\right)^2 \right. \\
&\quad \left. + \sum_{y \in \mathbb{T}_x^d \setminus x} f(y) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = y_j)\right)^2 \right] \\
&\quad - \frac{1}{n} \left[\left(\prod_{j=1}^d f(x) \mathbb{P}(Z_{x_j, h_j}^{[j]} = x_j)\right)^2 \right. \\
&\quad \left. + \left(\sum_{z \in \mathbb{T}_x^d \setminus x} f(z) \prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = z_j)\right)^2 \right] \\
&= \frac{1}{n} f(x)(1 - f(x)) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = x_j)\right)^2 + R,
\end{aligned}$$

où

$$R = \frac{1}{n} \left[\sum_{y \in \mathbb{T}_x^d \setminus x} f(y) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = y_j) \right)^2 - \left(\sum_{z \in \mathbb{T}_x^d \setminus x} f(z) \prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j}^{[j]} = z_j) \right)^2 \right] \longrightarrow o\left(\frac{1}{n}\right).$$

Donc

$$\text{Var}(\widehat{f}_n(x)) = \frac{1}{n} f(x)(1 - f(x)) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j, c_j}^{[j]} = x_j) \right)^2 + o\left(\frac{1}{n}\right).$$

4. L'erreur quadratique moyenne est donnée par :

$$\begin{aligned} \text{MSE}(x) &= E[\{\widehat{f}_n(x) - f(x)\}^2] = \text{Var}\{\widehat{f}_n(x)\} + \text{Biais}^2\{f_n(x)\} \\ &= \frac{1}{n} f(x)(1 - f(x)) \left(\prod_{j=1}^d \mathbb{P}\{Z_{x_j, h_j, c_j}^{[j]} = x_j\} \right)^2 \\ &\quad + \left(\sum_{j=1}^d h'_j f_j^{(1)} + \frac{1}{2} \sum_{j=1}^d \text{Var}(Z_{x_j, h_j, c_j}^{[j]}) f_{jj}^{(2)} \right)^2. \end{aligned}$$

5. L'erreur quadratique moyenne intégrée est donnée par :

$$\begin{aligned} \text{MISE}(\widehat{f}_n) &= \sum_{x \in \mathbb{T}_d} \text{Var}(\widehat{f}_n(x)) + \sum_{x \in \mathbb{T}_d} \text{Biais}^2(\widehat{f}_n(x)) \\ &= \sum_{x \in \mathbb{T}_d} \frac{1}{n} f(x)(1 - f(x)) \left(\prod_{j=1}^d \mathbb{P}(Z_{x_j, h_j, c_j}^{[j]} = x_j) \right)^2 \\ &\quad + \sum_{x \in \mathbb{T}_d} \left(\sum_{j=1}^d h'_j f_j^{(1)} + \frac{1}{2} \sum_{j=1}^d \text{Var}(Z_{x_j, h_j, c_j}^{[j]}) f_{jj}^{(2)} \right)^2. \end{aligned}$$

Dans ce qui suit, nous donnons les propriétés asymptotiques de l'estimateur.

Proposition 2.1.1. *Soit $f : \mathbb{T}_d \rightarrow \mathbb{R}$ une fonction discrète multidimensionnelle et $x \in \mathbb{T}_d$, \widehat{f}_n est l'estimateur non-paramétrique de f défini dans (2.8) alors :*

$$\text{Biais}(\widehat{f}_n(x)) = \mathbb{E}(\widehat{f}_n(x)) - f(x) \rightarrow 0 \text{ quand } h_j \rightarrow 0, j = 1, \dots, d \text{ et } n \rightarrow \infty.$$

Preuve : En utilisant la notion de continuité de la fonction discrète, pour toute ϵ il existe η :

$$\forall \epsilon > 0, \exists \eta > 0 : \text{pour } \|y - x\| \cap \mathbb{T}^d \implies |f(y) - f(x)| < \eta. \quad (2.9)$$

Nous commençons par :

$$\begin{aligned} |\mathbb{E}(\widehat{f}_n(x)) - f(x)| &= \left| \sum_{y \in \mathbb{T}_x^d} \{f(y) - f(x)\} K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d) \right| \\ &\leq \sum_{y \in \mathbb{T}_{x\eta}^d} |f(y) - f(x)| K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d) \\ &\quad + \sum_{y \in \overline{\mathbb{T}}_{x\eta}^d} |f(y) - f(x)| K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d), \end{aligned}$$

où $\mathbb{T}_{x\eta}^d = \{y \in \mathbb{T}_x^d : \|y - x\| < \eta\}$ et $\overline{\mathbb{T}}_{x\eta}$ son complémentaire.

D'après (2.9) nous avons :

$$\sum_{y \in \mathbb{T}_{x\eta}^d} |f(y) - f(x)| K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d) < \epsilon. \quad (2.10)$$

Puisque f est une fonction de masse de probabilité donc $f \leq 1$ nous pouvons majorer le second terme par :

$$\begin{aligned} \sum_{y \in \overline{\mathbb{T}}_{x\eta}^d} |f(y) - f(x)| K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d) &\leq 2 \sum_{y \in \overline{\mathbb{T}}_{x\eta}^d} K_{x_1, h_1}^1(y_1) K_{x_2, h_2}^2(y_2) \cdots K_{x_d, h_d}^{[d]}(y_d) \\ &= 2\mathbb{P}\{|Z_{x_1, h_1}^{[1]} - x_1| \geq \eta \cdot |Z_{x_d, h_d}^{[d]} - x_d| \geq \eta\} \\ &\leq 2 \left[\frac{\text{Var}(Z_{x_1, h_1}^{[1]})}{\eta^2} \cdots \frac{\text{Var}(Z_{x_d, h_d}^{[d]})}{\eta^2} \right], \end{aligned}$$

par conséquent,

$$\sum_{y \in \overline{\mathbb{T}}_{x\eta}^d} |f(y) - f(x)| K_{x_1, h_1}^{[1]}(y_1) \cdots K_{x_d, h_d}^{[d]}(y_d) \leq \frac{2}{\eta^2} \prod_{j=1}^d \text{Var}(Z_{x_j, h_j}^j). \quad (2.11)$$

Finalement, sous les conditions de la définition d'un noyau associé discret multivarié (définition (2.1.1)), les inégalités (2.10) et (2.11) permettent d'aboutir au résultat

recherché. ■

Proposition 2.1.2. *Soit $f : \mathbb{T}_d \rightarrow \mathbb{R}$ une fonction discrète multidimensionnelle et $x \in \mathbb{T}_d$, si \widehat{f}_n est l'estimateur non-paramétrique de f défini dans (2.8) alors*

$$MSE(x) = \mathbb{E}(\widehat{f}_n(x) - f(x))^2 \rightarrow 0 \text{ quand } h_j \rightarrow 0, j = 1, \dots, d \text{ et } n \rightarrow \infty.$$

Preuve : On a :

$$\mathbb{E} \left(\widehat{f}_n(x) - f(x) \right)^2 = \text{Var} \left(\widehat{f}_n(x) + \text{Biais}^2(\widehat{f}_n(x)) \right).$$

D'après la proposition (2.1.1), le deuxième terme est évident, la variance ponctuelle peut être majorée successivement par :

$$\begin{aligned} n \text{Var} \left(\widehat{f}_n(x) \right) &= \text{Var}(K_{x,H}(X_1)) \\ &\leq \mathbb{E}(K_{x_1, h_1}^{[1]} \cdots K_{x_d, h_d}^{[d]})^2 \\ &\leq \mathbb{E}(K_{x_1, h_1}^{[1]})^2 \cdots \mathbb{E}(K_{x_d, h_d}^{[d]})^2 \\ &\leq \sum_{y_1 \in \mathbb{T}_{x_1}} f(y_1) \mathbb{P}(Z_{x_1, h_1}^{[1]} = y_1)^2 \cdots \sum_{y_d \in \mathbb{T}_{x_d}} f(y_d) \mathbb{P}(Z_{x_d, h_d}^{[d]} = y_d)^2 \\ &\leq \sum_{y_1 \in \mathbb{T}_{x_1}} f(y_1) \mathbb{P}(Z_{x_1, h_1}^{[1]} = y_1) \cdots \sum_{y_d \in \mathbb{T}_{x_d}} f(y_d) \mathbb{P}(Z_{x_d, h_d}^{[d]} = y_d) \\ &\leq 1. \end{aligned}$$

Les dernières inégalités s'obtiennent facilement grâce à

$$\mathbb{P}(Z_{x_j, h_j}^j = y_j) \leq 1 \text{ et } f(y_j) \leq 1 \forall j,$$

puisqu'on a $0 \leq \text{Var}(\widehat{f}_n(x)) \leq (1/n)$, alors $\text{Var}(\widehat{f}_n(x))$ tend vers 0, d'où le résultat. ■

Proposition 2.1.3. *Soit $f : \mathbb{T}_d \rightarrow \mathbb{R}$ une fonction discrète multidimensionnelle et $x \in \mathbb{T}_d$, \widehat{f}_n est l'estimateur à noyau associé de f défini dans (2.8) alors :*

$$MISE(\widehat{f}_{n,H,K,f}) = \sum_{x \in \mathbb{T}_d} MSE(x) \rightarrow 0 \text{ quand } h_j \rightarrow 0, j = 1, \dots, d \text{ et } n \rightarrow \infty.$$

Preuve : La preuve découle de la proposition (2.1.2). ■

2.1.4 Choix de la matrice de fenêtre du lissage par la méthode de validation croisée

Les méthodes de validation croisée sont largement utilisées dans l'estimation par noyau des densités univariées et d'une manière restreinte dans le cas multivariée avec des noyaux symétriques. Nous proposons dans ce travail la méthode least square cross-validation (CV) avec les noyaux associés discrets multivariés. Le principe de cette méthode est de minimiser par rapport à H le critère *MISE*,

$$MISE = \sum_{x \in \mathbb{T}^d} [\hat{f}_n(x) - f(x)]^2.$$

La matrice des fenêtres optimale notée H_{CV} est donnée par :

$$H_{CV} = \arg \min_{\mathcal{H}} CV(H),$$

où \mathcal{H} est l'espace des matrices de lissage symétriques définies positives et

$$\begin{aligned} CV(H) &= \sum_{x \in \mathbb{T}_d} \hat{f}_n^2(x) - \frac{2}{n} \sum_{i=1}^n \hat{f}_{n,-i}(x_i) \\ &= \sum_{x \in \mathbb{T}_d} \left\{ \frac{1}{n} \sum_{i=1}^n K_{x,H}(x_i) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K_{x_i,H}(x_j), \end{aligned}$$

où $\hat{f}_{n,-i}$ est calculé à partir de \hat{f}_n sans l'observation x_i .

2.2 Estimation par noyaux associés continus multidimensionnels

Définition 2.2.1. Un type de noyau K_θ est un fdp paramétré avec le support $\mathbb{S}_\theta \subseteq \mathbb{R}^d$, $\theta \in \Theta \subseteq \mathbb{R}^d$, tel que K_θ est carré intégrable, uni-modal avec le mode $M \in \mathbb{R}^d$ et

en admettant D une matrice de dispersion de dimension $d \times d$ (ce qui est symétrique et défini positif) et $\theta = \theta(M, D)$ est vecteur de dimension d .

Notons par $D_{x_0} = \int_{\mathbb{R}^d} (x - x_0)(x - x_0)^t K_\theta(dx)$ la matrice de dispersion de K_θ autour du vecteur fixé x_0 .

Corollaire 2.2.1. Soient X_1, \dots, X_n des vecteurs aléatoires (i.i.d) avec une fonction de densité inconnue f sur \mathbb{T}_d , un sous-ensemble de \mathbb{R}^d ($d \geq 1$). Comme souvent observé en pratique, le sous-ensemble \mathbb{T}_d peut être non borné, partiellement ou totalement borné comme :

$$\mathbb{T}_d = \mathbb{R}^{d_\infty} \times [z, \infty[^{d_z} \times [u, w]^{d_{uw}} \quad (2.12)$$

avec u, w et z des réels telque $u < w$, d_∞ , d_z et d_{uw} sont des valeurs positives dans $\{0, 1, \dots, d\}$ tels que $d = d_\infty + d_z + d_{uw}$.

Définition 2.2.2. Soit \mathbb{T}_d est le support de la fdp à estimer avec $\mathbb{T}_d \subseteq \mathbb{R}^d$ et H la matrice des fenêtres. Pour un vecteur cible $x \in \mathbb{T}_d$ on considère un vecteur aléatoire $Z_{x,H}$ du fdp $K_{x,H}(\cdot)$ paramétrée de support $\mathbb{S}_{x,H} \subseteq \mathbb{R}^d$. La fonction $K_{x,H}(\cdot)$ est appelée "noyau multivarié associé ou (général)" si les conditions suivantes sont satisfaites :

$$x \in \mathbb{S}_{x,H}, \quad (2.13)$$

$$\mathbb{E}(Z_{x,H}) = x + A(x, H), \quad (2.14)$$

$$\text{Cov}(Z_{x,H}) = B(x, H), \quad (2.15)$$

où $A(x, H) = (a_1(x, H), \dots, a_d(x, H))^t$ et $B(x, H) = (b_{ij}(x, H))_{i,j=1,\dots,d}$ tendent respectivement vers le vecteur nul et la matrice nulle quand $H \rightarrow 0$.

La proposition suivante transforme la fonction noyau continu symétrique \mathcal{K} en noyau associé classique

Proposition 2.2.1. Soit $\mathbb{R}^d = \mathbb{T}_d$ le support de la densité à estimer. Soit \mathcal{K} un noyau classique du support $\mathbb{S}_d \subseteq \mathbb{R}^d$ de moyen $\mu = 0$ et de matrice de variance-covariance Σ . Pour un vecteur cible $x \in \mathbb{R}^d$ et H la matrice de fenêtre de lissage, le noyau classique multidimensionnel est transformé en noyau associé classique :

(i)

$$\mathcal{K}_{x,H}(\cdot) = \frac{1}{\det H} \mathcal{K} \{ H^{-1}(x - \cdot) \}, \quad (2.16)$$

sur

$$\mathbb{S}_{x,H} = x - H\mathbb{S}_d,$$

avec $\mathbb{E}(Z_{x,H}) = x$ (i.e $A(x, H) = 0$) et $\text{Cov}(Z_{x,H}) = H\Sigma H$,

(ii)

$$\mathcal{K}_{x,H}(\cdot) = \frac{1}{(\det H)^{1/2}} \mathcal{K} \{ H^{-1/2}(x - \cdot) \},$$

sur

$$\mathbb{S}_{x,H} = x - H^{1/2}\mathbb{S}_d,$$

avec $\mathbb{E}(Z_{x,H}) = x$ (i.e $A(x, H) = 0$) et $\text{Cov}(Z_{x,H}) = H^{1/2}\Sigma H^{1/2}$.

Preuve : Nous ne prouvons que (i) parce que (ii) est similaire à (i). À partir de (2.20) et la formule suivante :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_j), \quad \forall x \in \mathbb{T}_d = \mathbb{R}^d, \quad (2.17)$$

avec $K_H(y) = (\frac{1}{\det H})K(H^{-1}y)$, on en déduit facilement l'expression (2.16). À partir de (2.16) et la définition (2.2.2), pour un $x \in \mathbb{T}_d = \mathbb{R}^d$ fixé et pour tout $t \in \mathbb{T}_d = \mathbb{R}^d$, il existe $u \in \mathbb{S}_d$ tel que $u = H^{-1}(x - t)$ et donc $t = x - Hu$. Cela implique, à partir de (2.13), que $\mathbb{S}_{x,H} = x - H\mathbb{S}_d$. Les deux derniers résultats sont simplement dérivés du calcul de la matrice de covariance et du vecteur moyen de $Z_{x,H}$ en effectuant la substitution précédente $u = H^{-1}(x - t)$. ■

Dans la proposition suivante, nous soulignons que tous les noyaux multiples associés sont des noyaux associés à plusieurs variables (produit) sans structure de corrélation dans la matrice de fenêtre du lissage.

Proposition 2.2.2. Soit $\times_{j=1}^d \mathbb{T}_1^{[j]} = \mathbb{T}_d$ est le support de densité f à estimer avec $\mathbb{T}_1^{[j]} \subseteq \mathbb{R}$ les supports des marges univariées de f . Soit $x = (x_1, \dots, x_d)^t \in * \times_{j=1}^d \mathbb{T}_1^{[j]}$

et $H = \text{Diag}(h_{11}, \dots, h_{dd})$ avec $h_{jj} > 0$. Soit $K_{Z_j}^{[j]h_{jj}}$ un noyau associé univarié (voir définition (2.2.2) pour $d = 1$) avec sa variable aléatoire correspondante $Z_{x_j}^{[j]h_{jj}}$ de support $\mathbb{S}_{x, h_{jj}} \subseteq \mathbb{R}$, pour tout $j = 1, \dots, d$. Alors le noyau associé multiple est également un noyau associé à plusieurs variables (produit)

$$K_{x, H}(\cdot) = \prod_{j=1}^d K_{x_j}^{[j]h_{jj}}(\cdot), \quad (2.18)$$

sur $\mathbb{S}_{x, H} = \times_{j=1}^d \mathbb{S}_{x, h_{jj}}$ avec $\mathbb{E}(Z_{x, H}) = (x_1 + a_1(x_1, h_{11}), \dots, x_d + a_d(x_d, h_{dd}))^T$ et $\text{Cov}(Z_{x, H}) = \text{Diag}(b_{jj}(x_j, h_{jj}))_{j=1, \dots, d}$. En d'autres termes, les variables aléatoires $Z_{x_j}^{[j]h_{jj}}$ sont des composantes indépendantes du vecteur aléatoire $Z_{x, H}$.

Preuve : De (2.20) et la formule :

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d K_{x_j, h_{jj}}^{[j]}(X_{ij}), \quad \forall x_j \in \mathbb{T}_1^{[j]} \subseteq \mathbb{R}, \quad (2.19)$$

on en déduit facilement l'expression (2.18). Les autres résultats sont obtenus directement en calculant le vecteur moyen et la matrice de covariance de $(Z_{x, H} = Z_{x_1, h_{11}}^{[1]}, \dots, Z_{x_d, h_{dd}}^{[d]})^t$ qui est le vecteur aléatoire du fdp (2.18). ■

Définition 2.2.3. L'estimateur de noyau associé produit \widehat{f}_n de f est défini par :

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x, H}(X_i), \quad \forall x \in \mathbb{T}_d \subseteq \mathbb{R}^d, \quad (2.20)$$

où H est une matrice de dimension $d \times d$ (i.e. symétrique et définie positive) telle que $H \equiv H_n \rightarrow 0_d$ (la matrice null de dimension $d \times d$) quand $n \rightarrow \infty$, et $K_{x, H}(\cdot)$ est dit noyau associé à plusieurs variables.

2.2.1 Construction de noyaux multiples associés par la méthode de mode-dispersion

La construction de noyaux associés généraux (multiples) à partir d'un type de noyau K_θ sur \mathbb{S}_θ avec $\theta = \theta(M, D)$ par la méthode de mode-dispersion multivariée nécessite la résolution du système des équations suivant : $(M, D) = (x, H)$, la solution est de la forme :

$$(\theta(M, D))^t = (x, \text{vech}H)^t, \quad (2.21)$$

telque vech la symétrie de H est défini comme :

$$\text{vech}(x, H) = (x, \text{vech}H),$$

où

$$\text{vech} H = (h_{11}, \dots, h_{1d}, h_{22}, \dots, h_{2d}, \dots, h_{(d-1)(d-1)}, h_{(d-1)d}, h_{dd})^t, \quad (2.22)$$

est un vecteur de dimension $[d(d+1)/2]$ obtenu en superposant les colonnes du triangle inférieur de H , voir E.G, Henderson & Searle (1979)[10], avec $\theta = \theta(M, D)$ est de dimension d où $x \in \mathbb{T}_d$ est un vecteur cible et $H = (h_{ij})_{i,j=1,\dots,d}$ est la matrice de fenêtre de lissage.

La solution de (2.21) s'elle existe, est un $d \times 1$ vecteur noté $\theta(x, H) = \theta(M(x, H), D(x, H))$.

La solution de (2.21) dépend de la flexibilité du type de noyau K_θ et conduit au noyau associé correspondant noté $K_{\theta(x,H)}$.

La proposition suivante montre que le noyau associé construit est un noyau associé multiple :

Proposition 2.2.3. *Le noyau associé $K_{\theta(x,H)}(\cdot)$ est obtenu de (2.21), ayant le support $\mathbb{S}_{\theta(x,H)}$, vérifié :*

$$x \in \mathbb{S}_{\theta(x,H)}, \quad (2.23)$$

$$\mathbb{E}(Z_{\theta(x,H)}) - x = A_\theta(x, H), \quad (2.24)$$

$$\text{Cov}(Z_{\theta(x,H)}) = B_\theta(x, H), \quad (2.25)$$

où $Z_{\theta(x,H)}$ est un vecteur aléatoire dont la fdp $K_{\theta(x,H)}$ et les deux paramètres

$$A_\theta(x, H) = (a_{\theta 1}(x, H), \dots, a_{\theta d}(x, H))^t \text{ et } B_\theta(x, H) = (b_{\theta ij}(x, H))_{i,j=1,\dots,d},$$

tendent respectivement au vecteur nul 0 et à la matrice nulle 0_d quand H tend vers 0_d .

Preuve : La méthode de mode-dispersion multivariée (2.21) implique $\theta(x, H) = \theta(M(x, H), D(x, H))$ ce qui conduit à $\mathbb{S}_{\theta(x, H)} = \mathbb{S}_{\theta(M(x, H), D(x, H))}$. Puisque K_θ est uni-modal du mode $M \in \mathbb{S}_\theta$ (car le vecteur mode M de K_θ appartient toujours à \mathbb{S}_θ), on a évidemment (2.23), alors on vérifie (2.24), car M est identifié à x dans (2.21). Soit $Z_{\theta(M, D)}$ le vecteur aléatoire dont la fdp $K_{\theta(M, D)}$ uni-modal. Puisque $K_\theta(M) \geq K_\theta(\mu)$ où μ est le vecteur moyen de K_θ , nous pouvons écrire :

$$\mathbb{E}(Z_{\theta(M, D)}) = M + \tau(M, D),$$

où $\tau(M, D)$ est la différence entre le vecteur moyen $\mathbb{E}(Z_{\theta(M, D)})$ et le mode M de $Z_{\theta(M, D)}$. Ainsi, à partir de la méthode mode-dispersion (2.21), nous avons $M = x$ et $\tau(M, D) = \tau(M(x, H), D(x, H))$, en prenant $A_\theta(x, H) = \tau(M(x, H), (x, H))$, Comme la condition suivante :

Si D_M tend vers la matrice nulle, alors D_μ il tend aussi à la matrice nulle.

Alors on a le deuxième résultat (2.24), et donc (2.14) est vérifié. De plus, puisque $K_{\theta(M, D)}$ admet un moment de second ordre, la matrice de covariance de $K_{\theta(M, D)}$ existe et peut être écrite sous la forme $\text{Cov}(Z_{\theta(M, D)}) = B_\theta(M, D)$, en résolvant (2.21) puis en prenant

$$B_\theta(x, H) = B_\theta(M(x, H), D(x, H)),$$

le dernier résultat (2.25) est valable dans le sens de (2.15) en utilisant à nouveau la même condition. ■

Définition 2.2.4. *L'estimateur \hat{f}_n de f à noyau associé construit dans la proposition (2.2.3), i.e. $K_{\theta(x, H)}(\cdot) \equiv K_{x, H}(\cdot)$ est donné par :*

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{\theta(x, H)}(X_i), \quad \forall x \in \mathbb{T}_d. \quad (2.26)$$

2.2.2 Résultats asymptotiques

Proposition 2.2.4. *Pour tout $x \in \mathbb{T}_d$,*

$$\mathbb{E}\{\hat{f}_n(x)\} = \mathbb{E}\{f(Z_{\theta(x, H)})\}, \quad (2.27)$$

$$\begin{aligned}\widehat{f}_n(x) &\geq 0, \\ \int_{\mathbb{T}_d} \widehat{f}_n(x) dx &= \Lambda,\end{aligned}\tag{2.28}$$

où la masse totale $\Lambda = \Lambda(n, H, K_\theta)$ est un réel positif et est égal une constante finie si $\int_{\mathbb{T}_d} K_{\theta(x,H)}(t) dx < \infty$ pour tout $t \in \mathbb{T}_d$.

Preuve : Le premier résultat (2.27) est directement obtenu à partir de (2.26) comme suit :

$$\mathbb{E}\{\widehat{f}_n(x)\} = \int_{S_{\theta(x,H)} \cap \mathbb{T}_d} K_{\theta(x,H)}(t) f(t) dt = \mathbb{E}\{f(Z_{\theta(x,H)})\}.$$

De plus, les estimateurs $\widehat{f}_n(x) \geq 0$ et la masse totale $\Lambda > 0$ découlent immédiatement du fait que $K_{\theta(x,H)}(\cdot)$ est un fdp. Enfin, les valeurs de X_j appartenant à l'ensemble \mathbb{T}_d , on a (2.28) quand :

$$\int_{\mathbb{T}_d} \widehat{f}_n(x) dx = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{T}_d} K_{\theta(x,H)} dx < \infty.$$

Proposition 2.2.5. Soit $x \in \mathbb{T}_d$ une cible et $H \equiv H_n \rightarrow 0_d$ quand $n \rightarrow \infty$ la matrice de fenêtre de lissage. supposons que f dans la classe $\mathcal{C}^2(\mathbb{T}_d)$, alors :

$$\begin{aligned}\text{Biais}\{\widehat{f}_n(x)\} &= A_\theta^t(x, H) \nabla f(x) + \frac{1}{2} \text{trace} [\{A_\theta(x, H) A_\theta^t(x, H) + B_\theta(x, H)\} \nabla^2 f(x)] \\ &\quad + o\{\text{trace}(H^2)\}.\end{aligned}\tag{2.29}$$

De plus, si f est borné sur \mathbb{T}_d alors il existe un plus grand nombre réel positif $r_2 = r_2(K_\theta)$ tel que :

$$\|K_{\theta(x,H)}\|_2^2 \leq c_2(x) (\det H)^{-r_2}, \quad 0 \leq c_2(x) \leq \infty,$$

et

$$\text{Var}\{\widehat{f}_n(x)\} = \frac{1}{n} \|K_{\theta(x,H)}\|_2^2 f(x) + o(n^{-1} (\det H)^{-r_2}),\tag{2.30}$$

avec

$$\|K_{\theta(x,H)}\|_2^2 = \int_{S_{\theta(x,H)}} K_{\theta(x,H)}^2(u) du$$

Preuve : En utilisant successivement (2.27) et le developpement de Taylor de f au voisinage de $\mathbb{E}(Z_{\theta(x,H)})$:

$$\begin{aligned}
 \text{Biais}\{\widehat{f}_n(x)\} &= \mathbb{E}\{f(Z_{\theta(x,H)})\} - f(x) \\
 &= f(\mathbb{E}(Z_{\theta(x,H)})) + \frac{1}{2}\mathbb{E}[\text{trace}\{(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t \\
 &\quad \nabla^2 f(\mathbb{E}(Z_{\theta(x,H)}))(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))\}] - f(x) \\
 &\quad + o[\mathbb{E}\{(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))\}] \\
 &= f(\mathbb{E}(Z_{\theta(x,H)})) - f(x) + \frac{1}{2}\text{trace}[\nabla^2 f(\mathbb{E}(Z_{\theta(x,H)})) \\
 &\quad \mathbb{E}\{(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))\}] \\
 &\quad + o[\text{trace}(\mathbb{E}\{(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t\})] \\
 &= f(\mathbb{E}(Z_{\theta(x,H)})) - f(x) + \frac{1}{2}\text{trace}[\nabla^2 f(\mathbb{E}(Z_{\theta(x,H)})) \text{Cov}(Z_{\theta(x,H)})] \\
 &\quad + o[\text{trace}\{\text{Cov}(Z_{\theta(x,H)})\}] \\
 &= f(x + A_{\theta}(x, H)) - f(x) + \frac{1}{2}\text{trace}[\nabla^2 f(x + A_{\theta}(x, H))B_{\theta}(x, H)] \\
 &\quad + o\{\text{trace}(B_{\theta}(x, H))\} \\
 &= A_{\theta}^t(x, H)\nabla f(x) + \frac{1}{2}\text{trace}[\nabla^2 f(x)\{A_{\theta}(x, H)A_{\theta}^t(x, H) + B_{\theta}(x, H)\}] \\
 &\quad + o\{\text{trace}\{H^2\}\}.
 \end{aligned}$$

En fait, le reste $o\{\text{trace}(H^2)\}$ provient de la $\text{trace}(B_{\theta}(x, H)) = o(\text{trace } H^2)$ déduite de la proposition (2.2.1) des noyaux associés classiques et

$$\begin{aligned}
 o(\mathbb{E}\{(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))\}) &= \\
 \mathbb{E}\{o_p(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))^t(Z_{\theta(x,H)} - \mathbb{E}(Z_{\theta(x,H)}))\}, &
 \end{aligned}$$

où $o_p(\cdot)$ est la vitesse de la convergence en proba.

En ce qui concerne la variance (2.30), on a d'abord :

$$\begin{aligned}
 \text{Var}\{\widehat{f}_n(x)\} &= \frac{1}{n}\mathbb{E}\{K_{\theta(x,H)}^2(X_1)\} - \frac{1}{n}[\mathbb{E}\{K_{\theta(x,H)}(X_1)\}]^2 \\
 &= \frac{1}{n} \int_{\mathbf{S}_{\theta(x,H)} \cap \mathbb{T}_d} K_{\theta(x,H)}^2(u)f(u)du - \frac{1}{n}[\mathbb{E}\{K_{\theta(x,H)}(X_1)\}]^2 \\
 &= I_1 - I_2.
 \end{aligned}$$

De (2.27) et (2.29), on a le résultat suivant :

$$I_2 = (1/n)[\mathbb{E}\{K_{\theta(x,H)}(X_1)\}]^2 \simeq (1/n)f^2(x) \simeq o(1/n).$$

Puisque f est borné pour tout $x \in \mathbb{T}_d$. En utilisant le développement de Taylor de f au voisinage x , le premier terme :

$$I_1 = \frac{1}{n} \int_{\mathbb{S}_{\theta(x,H)} \cap \mathbb{T}_d} K_{\theta(x,H)}^2(u) f(u) du,$$

devient

$$I_1 = \frac{1}{n} f(x) \int_{\mathbb{S}_{\theta(x,H)} \cap \mathbb{T}_d} K_{\theta(x,H)}^2(u) du + R(x, H),$$

avec

$$\begin{aligned} R(x, H) = & \frac{1}{n} \int_{\mathbb{S}_{\theta(x,H)} \cap \mathbb{T}_d} K_{\theta(x,H)}^2(u) \left((u-x)^t \nabla f(x) + \frac{1}{2} (u-x)^t \nabla^2 f(x) (u-x) \right. \\ & \left. + o\{(u-x)^t (u-x)\} \right) du. \end{aligned}$$

Un argument similaire de Chen (1999, Lemma)[8] avec f borné sur \mathbb{T}_d montre l'existence de r_2 et la condition $\|K_{\theta(x,H)}\|_2^2 \leq c_2(x)(\det H)^{-r_2}$ mène successivement à

$$0 \leq R(x, H) \leq \frac{1}{n(\det H)^{r_2}} \int_{\mathbb{S}_{\theta(x,H)} \cap \mathbb{T}_d} c_2(x) \left((u-x)^t \nabla f(x) + \frac{1}{2} (u-x)^t \nabla^2 f(x) (u-x) \right) du \simeq o\{n^{-1}(\det H)^{-r_2}\}.$$

■

Nous rappelons maintenant les mesures naturelles permettant d'évaluer la similarité de l'estimateur à noyau multivarié associé \widehat{f}_n selon le vrai fdp f à estimer.

Théorème 2.2.1. (*Erreurs quadratiques ponctuelles et globales*)

L'erreur quadratique moyenne (MSE) est exprimée par :

$$\begin{aligned} MSE(x) &= \text{Biais}^2\{\widehat{f}_n(x)\} + \text{Var}\{\widehat{f}_n(x)\}, \\ &= \left\{ A_{\theta}^t(x, H) \nabla f(x) + \frac{1}{2} \text{trace} \left[\{ A_{\theta}(x, H) A_{\theta}^t(x, H) + B_{\theta}(x, H) \} \nabla^2 f(x) \right] \right. \\ &\quad \left. + o\{\text{trace}(H^2)\} \right\}^2 + \frac{1}{n} \|K_{\theta(x,H)}\|_2^2 f(x) + o(n^{-1}(\det H)^{-r_2}). \end{aligned}$$

La forme intégrée de MSE sur \mathbb{T}_d est :

$$\begin{aligned}
 MISE(\hat{f}_n) &= \int_{\mathbb{T}_d} MSE(x) dx \\
 &= \int_{\mathbb{T}_d} \left(\left[A_\theta^t(x, H) \nabla f(x) + \frac{1}{2} \text{trace}\{ (A_\theta(x, H) A_\theta^t(x, H) + B_\theta(x, H)) \right. \right. \\
 &\quad \left. \left. \nabla^2 f(x) \right] \right]^2 + \frac{1}{n} \|K_{\theta(x, H)}\|_2^2 f(x) \right) dx
 \end{aligned} \tag{2.31}$$

Remarque 2.2.1. *Le choix de la fenêtre du lissage est similaire que le cas discret multidimensionnel*

2.2.3 Réduction du biais

Nous proposons ci-dessous un algorithme pour éliminer le terme de gradient dans la plus grande région de \mathbb{T}_d .

L'algorithme de réduction de biais : L'algorithme de réduction de biais a deux étapes, la première étape consiste à définir les régions internes et externes. La seconde consiste à définir le noyau associé modifié qui conduit à un biais réduit.

– **Première étape :** Partitionner \mathbb{T}_d en deux régions d'ordre

$\alpha(H) = (\alpha_1(H), \dots, \alpha_d(H))^t$ qui est un vecteur de dimension d avec $\alpha_1(H), \dots, \alpha_d(H) \in \mathbb{R}$, où $\alpha(H)$ tend vers le vecteur nul 0 quand H tend vers la matrice nulle 0_d :

a) La région intérieure est la plus grande dans l'intérieur de \mathbb{T}_d afin de contenir au moins 95% des observations. Elle est indiquée par $\mathbb{T}_d^{\alpha(H), I}$.

b) Les régions limites représentant le complémentaire de $\mathbb{T}_d^{\alpha(H), I}$ dans \mathbb{T}_d , et il est noté $\mathbb{T}_d^{\alpha(H), B}$ qui pourrait être vide rappelons que $\mathbb{T}_d = \mathbb{T}_d^{\alpha(H), I} \cup \mathbb{T}_d^{\alpha(H), B}$ et $\mathbb{T}_d^{\alpha(H), I} \cap \mathbb{T}_d^{\alpha(H), B} = \emptyset$.

Puisque $\mathbb{T}_d \subseteq \mathbb{R}^d$ pourrait avoir chacun de ses d composants convexes sous la forme d'intervalles non bornés, partiellement ou totalement bornés, comme dans (2.12), il n'ya qu'un seul $\mathbb{T}_d^{\alpha(H), I}$ tel que

$$N_d = 1^{d_\infty} 2^{d_z} 3^{d_{uw}} - 1 \tag{2.32}$$

sous-régions limites de $\mathbb{T}_d^{\alpha(H), B}$. Une illustration est fournie pour $d = 2$ avec

$\mathbb{T}_2 = [0, 1] \times [0, 1]$, par conséquent, le nombre de limites des sous-régions (2.32) sont $N_2 = 3^2 - 1 = 8$.

- **Deuxième étape :** Changer le noyau associé général $K_{\theta(x,H)}$ dans sa version modifiée $K_{\tilde{\theta}(x,H)}$, cela conduit à remplacer le couple $(A_{\theta}(x, H), B_{\theta}(x, H))$ dans $(A_{\tilde{\theta}}(x, H), B_{\tilde{\theta}}(x, H))$ avec

$$A_{\tilde{\theta}}(x, H) = A_{\tilde{\theta}_B}(x, H)\mathbb{1}_{\mathbb{T}_d^{\alpha(H),B}}(x),$$

parce que $A_{\tilde{\theta}_I}(x, H)\mathbb{1}_{\mathbb{T}_d^{\alpha(H),I}}(x) = 0$ à l'intérieur et

$$B_{\tilde{\theta}}(x, H) = B_{\tilde{\theta}_I}(x, H)\mathbb{1}_{\mathbb{T}_d^{\alpha(H),I}}(x) + B_{\tilde{\theta}_B}(x, H)\mathbb{1}_{\mathbb{T}_d^{\alpha(H),B}}(x).$$

Ce noyau associé modifié est tel que, pour toute matrice de fenêtre fixe H ,

$$\tilde{\theta}(x, H) = \begin{cases} \tilde{\theta}_I(x, H) : A_{\tilde{\theta}_I}(x, H) = 0 & \text{si } x \in \mathbb{T}_d^{\alpha(H),I} \\ \tilde{\theta}_B(x, H), & \text{si } x \in \mathbb{T}_d^{\alpha(H),B} \end{cases} \quad (2.33)$$

doit être continu sur \mathbb{T}_d et constant sur $\mathbb{T}_d^{\alpha(H),B}$.

La proposition suivante montre que la fonction noyau construite est un noyau associé général.

Proposition 2.2.6. *La fonction $K_{\tilde{\theta}(x,H)}$ sur son support $\mathbb{S}_{\tilde{\theta}(x,H)} = \mathbb{S}_{\theta(x,H)}$ et obtenu à partir de (2.33) est également un noyau associé général.*

Preuve : Puisque $K_{\theta(x,H)}$ est un noyau associé pour tous les

$x \in \mathbb{T}_d = \mathbb{T}_d^{\alpha(H),I} \cup \mathbb{T}_d^{\alpha(H),B}$, on obtient la première condition (2.13) de la définition (2.2.2) car on a $\mathbb{S}_{\theta(x,H)} = \mathbb{S}_{\tilde{\theta}(x,H)}$.

Selon la proposition (2.2.3), il en résulte que, pour une variable aléatoire donnée $Z_{\tilde{\theta}(x,H)}$ dont la fdp $K_{\tilde{\theta}(x,H)}$, nous obtenons les deux dernières conditions (2.14) et (2.15) de la définition (2.2.2) comme respectivement

$$\mathbb{E}(Z_{\tilde{\theta}(x,H)}) = x + A_{\tilde{\theta}}(x, H) \text{ et } \text{Cov}(Z_{\tilde{\theta}(x,H)}) = B_{\tilde{\theta}}(x, H).$$

Les deux quantités $A_{\tilde{\theta}}(x, H)$ et $B_{\tilde{\theta}}(x, H)$ tendent respectivement vers le vecteur nul 0 et la matrice nulle 0_d comme H tend vers 0_d . En particulier, à partir de (2.33) nous

avons facilement $A_{\tilde{\theta}_I}(x, H) = 0$ à l'intérieur. ■

Remarque 2.2.2. *L'estimateur de noyau associé modifié \tilde{f}_n utilisant $K_{\tilde{\theta}(x,H)}$ est alors définis par :*

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{j=1}^n K_{\tilde{\theta}(x,H)}(X_j). \quad (2.34)$$

Le résultat suivant donne uniquement à l'intérieur $\mathbb{T}_d^{\alpha(H),I}$ de \mathbb{T}_d les expressions du biais et de la variance de \hat{f}_n . Bien entendu, les expressions correspondantes dans les régions limites $\mathbb{T}_d^{\alpha(H),B}$ sont fastidieuses à écrire en ce qui concerne les situations limites N_d ((2.32) de (2.34)).

Proposition 2.2.7. *Soient \hat{f}_n et \tilde{f}_n les estimateurs multivariés associés du noyau de f défini dans (2.26) et (2.34) respectivement. Alors pour $x \in \mathbb{T}_d^{\alpha(H),I}$ comme dans (2.33) :*

$$\text{Biais}\{\tilde{f}_n(x)\} = \frac{1}{2} \text{trace}(B_{\tilde{\theta}_I}(x, H) \nabla^2 f(x)) + o\{\text{trace}(H^2)\}, \quad (2.35)$$

et

$$\text{Var}\{\tilde{f}_n(x)\} \simeq \text{Var}\{\hat{f}_n(x)\} \text{ quand } n \rightarrow \infty.$$

Preuve : On obtient le premier résultat (2.35) en remplaçant dans (2.29) \hat{f}_n par \tilde{f}_n et B_θ, A_θ par $A_{\tilde{\theta}_I}$ et $B_{\tilde{\theta}_I}$ respectivement. Pour le dernier résultat, considérant (2.30) et il suffit de montrer que

$$\|K_{\theta(x,H)}\|_2^2 \simeq \|K_{\tilde{\theta}(x,H)}\|_2^2 \text{ quand } H \rightarrow 0_d.$$

Puisque $K_{\theta(x,H)}$ et $K_{\tilde{\theta}(x,H)}$ sont des noyaux associés multidimensionnelles du même type K_θ avec $r_2 = r_2(K_\theta)$ et $\tilde{r}_2 = \tilde{r}_2(K_{\tilde{\theta}})$, alors il existe un plus grand nombre réel positif commun

$$r_2^* = r_2^*(K_\theta),$$

tel que

$$\|K_{\theta(x,H)}\|_2^2 \leq c_2(x)(\det H)^{-r_2^*} \text{ et } \|K_{\tilde{\theta}(x,H)}\|_2^2 \leq \tilde{c}_2(x)(\det H)^{-r_2^*},$$

avec

$$0 < c_2(x), \tilde{c}_2(x) < \infty.$$

Prend $c(x) = \sup\{c_2(x), \tilde{c}_2(x)\}$, nous avons $\|K_{\theta(x,H)}\|_2^2 \leq c(x)(\det H)^{-2r_2^*}$ et $\|K_{\tilde{\theta}(x,H)}\|_2^2 \leq c(x)(\det H)^{-2r_2^*}$. Puisque $c(x)/n(\det H)^{2r_2^*} = o(n^{-1}(\det H)^{-2r_2^*})$ alors $\|K_{\tilde{\theta}(x,H)}\|_2^2 \simeq \|K_{\theta(x,H)}\|_2^2$.

Ainsi, nous définissons l'expression asymptotique du *MISE* de \hat{f}_n à l'intérieur $\mathbb{T}_d^{\alpha(H),I}$ comme suit : ■

$$MISE_{\tilde{\theta}_I}(\tilde{f}_n) = \int_{\mathbb{T}_d^{\alpha(H),I}} \left(\left\{ \frac{1}{2} \text{trace}(B_{\tilde{\theta}_I}(x, H) \nabla^2 f(x)) \right\}^2 + \frac{1}{n} \|K_{\theta(x,H)}\|_2^2 f(x) \right) dx.$$

Chapitre 3

Estimation par noyaux associés mixtes

3.1 Modèle

Nous nous intéressons à l'estimation non-paramétrique de densité gouvernant les variables aléatoires réelles à support connu $\mathbb{T} \subseteq \mathbb{R}$ constitué à la fois des intervalles et des ensembles discrets deux à deux disjoints. Par exemple, si on étudie un processus (fonction) de suivi d'un malade à p phases au cours desquelles le patient subit alternativement des soins intensifs avec des prélèvements quasi-instantanés (i.e. continus) correspondant à une période d'hospitalisation, et des soins externes où les prélèvements sont périodiques ou séquentiels dans le temps (i.e. discrets). Ce type de fonctions est généralement un modèle de mélange fini et s'écrit sous la forme :

$$f(x) = \sum_{j=1}^p \beta_j f_j(x) 1_{\mathbb{T}_j}(x), \quad (3.1)$$

où les f_j sont des fonctions de densité ou masse de probabilité (f.d.m.p.) et les constantes β_j sont les proportions du mélange avec $\sum_j \beta_j = 1$ (supposées connues dans ce travail pour simplifier) sur chaque composante \mathbb{T}_j partitions de \mathbb{T} (La Figure (3.1) présente une représentation graphique de f pour $p = 2$). Ce type de support est généralement appelé *échelle de temps* ("time-scales" en anglais). Elle est constituée de réunions des connexes continus et discrets deux à deux disjoints qu'on peut écrire

de la manière suivante :

$$\mathbb{T} = \left(\bigcup_J \mathbb{T}_{I_j} \right) \cup \left(\bigcup_{J'} \mathbb{T}_{N_{j'}} \right), \quad (3.2)$$

où les \mathbb{T}_{I_j} sont des intervalles et les $\mathbb{T}_{N_{j'}}$ sont des ensembles discrets au plus dénombrables. Sans perte de généralité, nous désignons par \mathbb{T}_I un intervalle de \mathbb{T} et par \mathbb{T}_N un sous ensemble discret de \mathbb{T} au plus dénombrable. Nous rappelons au lecteur que les échelles de temps sont des fermés de \mathbb{R} . Pour simplifier, nous réécrivons \mathbb{T} comme suit :

$$\mathbb{T} = \mathbb{T}_I \cup \mathbb{T}_N = [t_0, t_1] \cup \{t_2, t_3, \dots\}. \quad (3.3)$$

Sur les échelles de temps, plusieurs outils d'analyse ont été proposés pour travailler à la fois sur les intervalles et les ensembles discrets. Signalons que, au passage les échelles de temps sont des espaces métriques complets. Elles sont dotées des notions de distance, d'ouverts qui débouchent sur les concepts des limites et continuités que nous n'allons pas détailler dans ce travail. Sinon, en ce qui concerne la continuité, une fonction $f : \mathbb{T}_j \rightarrow \mathbb{R}$, est dite continue sur \mathbb{T}_j lorsqu'on a :

$$\forall \epsilon > 0, \exists \gamma_{x_j} : s \in \gamma_{x_j} \cap \mathbb{T}_j \implies |f(s) - f(x_j)| < \epsilon,$$

où γ_{x_j} est un voisinage de x_j selon la topologie induite sur \mathbb{T}_j .

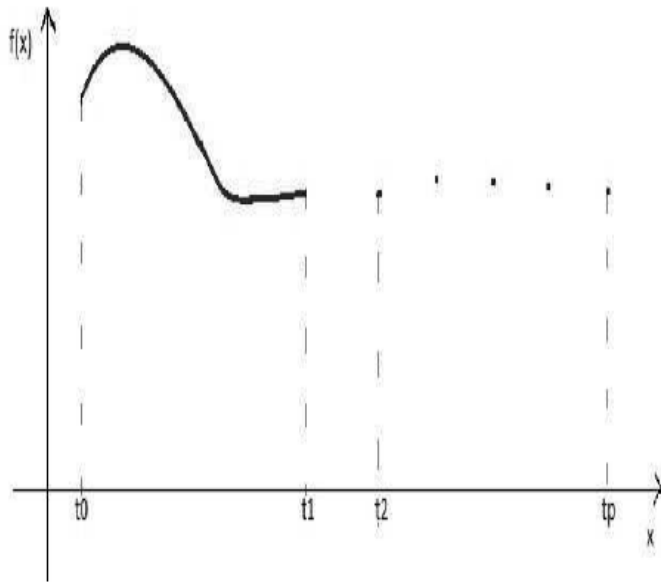


FIGURE 3.1 – Exemple d'une densité mixte

Ce chapitre se propose d'en faire usage dans le but d'estimer par les méthodes non-paramétriques les fonctions de type (3.1) sur le support \mathbb{T} défini en (3.3) tout en respectant la structure topologique de ce dernier. La fonction définie en (3.1) est une (f.d.m.p) dite *fonction mixte univariée*. La mixité est due ici au fait que la densité f à estimer est partiellement continue et discrète. La méthode appropriée pour ce type d'estimation est celle des noyaux associés puisque ces derniers sont construits dans l'esprit du strict respect de la nature topologique du support de f , cependant, nous devrions rester attentifs aux différents changements de structures de supports. Il faut noter que la force des noyaux associés pour ce type d'estimation réside dans leur capacité à dépendre intrinsèquement du point d'estimation x et de la fenêtre de lissage h interprétée comme paramètre de dispersion et qui joue le même rôle tant dans le cas discret que continu voir Jørgensen (1997)[13] et Jørgensen et Kokonendji (2011, 2013)[14]. Enfin, l'une dernière des raisons est leur flexibilité dans l'utilisation de l'analyse unifiant le discret et le continu.

3.2 Définitions

Nous donnons ici la définition des opérateurs de saut proposé par Hilger (1988)[11] lesquels jouent un rôle très important lors de changement de structure en passant d'un intervalle à un ensemble discret. Ensuite, nous faisons une classification de certains points particuliers que nous auront à utiliser.

Définition 3.2.1. (Hilger (1988)[11]) Soit \mathbb{T} une échelle de temps.

– (i) On appelle opérateur de saut en avant ("forward jump operator" en anglais) l'application :

$$\sigma : \mathbb{T} \longrightarrow \mathbb{T}, x \mapsto \sigma(x) = \inf\{s \in \mathbb{T}, s > x\}, \forall x \in \mathbb{T}. \quad (3.4)$$

– (ii) De même l'opérateur de saut en arrière ("backward jump operator" en anglais) est défini par l'application :

$$\rho : \mathbb{T} \rightarrow \mathbb{T}, x \mapsto \rho(x) = \sup\{s \in \mathbb{T}, s < x\}, \forall x \in \mathbb{T}.$$

Ces deux opérateurs nous permettent de procéder à la classification des points de \mathbb{T} comme dans le tableau (3.2). Nous signalons au passage que cette nomenclature est spécifique à ce travail.

Type de points	Formulation
Continu à gauche (càg)	$\rho(t) = t$
Continu à droite (càd)	$\sigma(t) = t$
Discret à gauche (dàg)	$\rho(t) < t$
Discret à droite (dàd)	$\sigma(t) > t$

Tableau (3.2) -Classification des points de \mathbb{T}

Bien évidemment, lorsqu'un point t est à la fois continu à droite et gauche (i.e. càg et càd) alors on a $\rho(t) = t = \sigma(t)$ et on dit que t est un point continu. De même, un point est dit discret lorsqu'il est discret à droite et à gauche (i.e. dàd et dàg) ce qui se formule par $\rho(t) < t < \sigma(t)$.

Nous présentons dans la définition suivante la nouvelle notion de l'ordre de discrétisation des points sur la partie \mathbb{T}_N de \mathbb{T} .

Définition 3.2.2. On considère $x \in \mathbb{T}_N$ et $k \in \mathbb{N}$.

- On dit que x est un point discret d'ordre k , s'il est situé à k -pas symétriques des bords gauche et droit de \mathbb{T}_N .
- Lorsqu'il est situé à k_1 -pas de bord droit (respectivement k_2 -pas de bord gauche) de \mathbb{T}_N , alors il est dit discret d'ordre k_1 à droite (respectivement discret d'ordre k_2 à gauche).

Nous résumons dans le prochain paragraphe les différentes techniques de dérivabilité.

3.3 Dérivabilité unifiée

Définition 3.3.1. Soit $f : \mathbb{T} \rightarrow \mathbb{R}$, \mathbb{T}^κ et \mathbb{T}_κ des sous ensembles de \mathbb{T} , respectivement semi ouvert à droite et à gauche.

- On dit que f est Δ -dérivable au point x si :

$$\forall \epsilon > 0, \exists \eta > 0 : \forall s \in \gamma_\eta(x) =]x - \eta, x + \eta[\cap \mathbb{T},$$

ce qui implique :

$$|[f\{\sigma(x)\} - f(s)] - f^\Delta(x)\{\sigma(x) - s\}| < \epsilon |\sigma(x) - s|, \forall x \in \mathbb{T}^\kappa.$$

- De même, on dit que f est ∇ -dérivable en x si :

$$\forall \epsilon > 0, \exists \eta' > 0 : \forall s \in \gamma_{\eta'}(x) =]x - \eta', x + \eta'[\cap \mathbb{T},$$

ce qui implique :

$$|[f\{\rho(x)\} - f(s)] - f^\nabla(x)\{\rho(x) - s\}| < \epsilon |\rho(x) - s|, \forall x \in \mathbb{T}_\kappa.$$

On note f^Δ et f^∇ les fonctions Δ -dérivée et ∇ -dérivée de f .

Cette définition s'écrit aussi de la manière suivante :

$$f^\Delta(x) = \frac{f\{\sigma(x)\} - f(s)}{\sigma(x) - s} = \lim_{\substack{x > s \\ x \rightarrow s}} \frac{f(x) - f(s)}{x - s}, \quad (3.5)$$

et

$$f^\nabla(x) = \frac{f\{\rho(x)\} - f(s)}{\rho(x) - s} = \lim_{\substack{x < s \\ x \rightarrow s}} \frac{f(x) - f(s)}{x - s}. \quad (3.6)$$

Lorsque $\mathbb{T} = \mathbb{T}_I$ alors (3.5) et (3.6) représentent respectivement les notions usuelles de dérivabilité à droite et à gauche de x . Cependant, lorsque $\mathbb{T} = \mathbb{T}_N$ alors (3.5) et (3.6) sont similaires aux notions des différences finies décentrées à droite et à gauche de x .

Ainsi, nous introduisons la nouvelle notion unifiant celles de dérivée première et de différences finies de premier ordre de f en un point x de \mathbb{T} dans la définition suivante.

Définition 3.3.2. Soient $f : \mathbb{T} \rightarrow \mathbb{R}$ et $x \in \mathbb{T}_\kappa^\kappa$ (où \mathbb{T}_κ^κ est un ouvert de \mathbb{T}). On dit que f est dérivable en x si et seulement si elle est à la fois Δ -dérivable et ∇ -dérivable en x et sa fonction dérivée en x est de la forme :

$$f^{(1)}(x) = \frac{1}{\sigma(x) - \rho(x)} \left[\{\sigma(x) - x\} f^{\Delta^1}(x) + \{x - \rho(x)\} f^{\nabla^1}(x) \right], \quad \forall x \in \mathbb{T}. \quad (3.7)$$

Remarque 3.3.1. On vérifie facilement que cette définition coïncide avec la dérivabilité classique en x lorsque $\mathbb{T} = \mathbb{T}_I$ et elle est similaire aux notions des différences finies centrées en x lorsque $\mathbb{T} = \mathbb{T}_N$. Ce qui nous permet d'établir la formule de Taylor de la manière suivante :

$$f(x) = \sum_j^k \frac{(x-s)^j}{j!} f^{(j)}(s) + o\{(x-s)^k\}, \quad (3.8)$$

avec

$$f^{(j)}(x) = \frac{1}{\sigma(x) - \rho(x)} [\{\sigma(x) - x\} f^{\Delta^j}(x) + \{x - \rho(x)\} f^{\nabla^j}(x)], \quad (3.9)$$

où $f^{\Delta^j}(x)$ et $f^{\nabla^j}(x)$ sont respectivement Δ -dérivée et ∇ -dérivée de f d'ordre j . Dans (3.9), la quantité $f^{(j)}$ unifie les notions classiques de la dérivée j^{eme} sur \mathbb{T}_I et de différences finies ordre j sur \mathbb{T}_N . Précisons que pour les points discrets d'ordres k_1 à droite et k_2 à gauche, (3.9) s'écrit de la forme :

$$f^{(k_1, k_2)}(x) = \frac{1}{\sigma(x) - \rho(x)} \left[\{\sigma(x) - x\} f^{\Delta^{k_1}}(x) + \{x - \rho(x)\} f^{\nabla^{k_2}}(x) \right], \quad \forall k_1, k_2. \quad (3.10)$$

Notons qu'en général $f^{(j)}(\cdot) = f^{(j,j)}(\cdot)$ et en particulier pour $x \in \mathbb{T}_N = \mathbb{N}$, on a :

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x+1) + 2f(x) - 2f(x-1) + f(x-2)\}/2 & \text{si } x \in \mathbb{N}^* \setminus \{1\} \\ \{f(3) - 2f(2) + 2f(1) - f(0)\}/2 & \text{si } x = 1 \\ \{f(2) - 2f(1) + f(0)\}/2 & \text{si } x = 0. \end{cases} \quad (3.11)$$

Ce qui est nettement différent du résultat de Kokonendji & Senga Kiessé (2011)[17].

Pour ces auteurs, la quantité (3.11) est exprimée sous la forme :

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\}/4 & \text{si } x \in \mathbb{N} \setminus \{0, 1\} \\ \{f(3) - 3f(1) + 2f(0)\}/2 & \text{si } x = 1 \\ \{f(2) - 2f(1) + f(0)\}/2 & \text{si } x = 0. \end{cases} \quad (3.12)$$

■

Les échelles de temps sont munies de Δ mesure et ∇ -mesure qui tiennent compte de leurs structures. Ces deux mesures sont similaires à la mesures de Lebesgue lorsque \mathbb{T} est un intervalle puis à une mesure de dénombrement si \mathbb{T}_N est un ensemble discret. La définition suivante introduit l'outil d'unification de calcul intégral sur les échelles de temps.

Définition 3.3.3. Une fonction $F : \mathbb{T} \rightarrow \mathbb{R}$ est appelée Δ -primitive de $f : \mathbb{T} \rightarrow \mathbb{R}$ si $F^\Delta(x) = f(x)$ pour tout $x \in \mathbb{T}^\kappa$. Dans ce cas, on définit l'intégrale de f par

$$\int_a^x f(t) \Delta t = F(x) - F(a),$$

avec Δt la Δ mesure sur \mathbb{T} .

On définit de même la ∇ -primitive de f en utilisant l'une des relations :

$$F^\Delta(x) = F^\nabla\{\sigma(x)\} \quad (3.13)$$

$$F^\nabla(x) = F^\Delta\{\rho(x)\}. \quad (3.14)$$

Ainsi, on a le théorème suivant :

Théorème 3.3.1. (Hilger, (1988)[11]) Soient a et b des points de \mathbb{T} tels que $a < b$ et f une fonction Δ -mesurable sur T . L'intégrale de f sur \mathbb{T} satisfait les propriétés suivantes :

$$\int_a^b f(t)\Delta t = \begin{cases} \int_a^b f(t)dt & \text{si } \mathbb{T} = \mathbb{T}_I \\ \sum_a^b f(t) & \text{si } \mathbb{T} = \mathbb{T}_N. \end{cases} \quad (3.15)$$

Preuve : Pour la démonstration, on peut se référer à Hilger (1988)[11] ou Bohner & Peterson (2001, pages 26-34)[4]. On obtient les même propriétés en ce qui concerne les fonctions ∇ -mesurables en utilisant l'une des relations (3.13) et (3.14).

Nous présentons dans la suite suivante les estimateurs à noyaux associés mixtes.

Définition 3.3.4. Un type de noyau mixte K_θ de support \mathbb{S}_θ est une combinaison convexe des types de noyaux continus et discrets $K_\theta^{[j]}$ de support $\mathbb{S}_\theta^{[j]}$. On écrit alors

$$K_\theta(u) = \sum_j \beta_j K_\theta^{[j]}(u) \mathbb{1}_{\mathbb{S}_\theta^{[j]}}(u),$$

avec $\beta_j > 0$, tels que $\sum_j \beta_j = 1$.

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

$$\mathbb{S}_\theta = \bigcup_j \mathbb{S}_\theta^{[j]}. \quad (3.16)$$

Sans perte de généralité, nous désignerons par \mathbb{S}_{θ_1} le support d'un type de noyau continu et par \mathbb{S}_{θ_N} celui d'un type de noyau discret. Par conséquent, le support \mathbb{S}_θ de K_θ sera la réunion de \mathbb{S}_{θ_1} et \mathbb{S}_{θ_N} .

Nous allons maintenant définir le noyau associé mixte.

Définition 3.3.5. Soit $\mathbb{T} = \cup_j \mathbb{T}_j$ le support de la densité mixte f à estimer. Un noyau associé mixte $K_{\theta(x,h)}$ est une densité de probabilité paramétrée par le point d'estimation x et le paramètre de lissage h , et constitué à la fois de noyaux associés

continus et discrets. Il s'écrit :

$$K_{\theta(x,h)}(\cdot) = \sum_{j=1}^p \beta_j K_{x,h}^{[j]}(\cdot) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(\cdot), \quad (3.17)$$

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

$$\mathbb{S}_{\theta(x,h)} = \bigcup_{j=1}^p \mathbb{S}_{x,h}^{[j]}, \quad (3.18)$$

avec $\mathbb{S}_{x,h}^{[j]}$ supports de $K_{x,h}^{[j]}(\cdot) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(\cdot)$.

On vérifie facilement que $K_{\theta(x,h)}(\cdot)$ en (3.17) satisfait les conditions des noyaux associés (2.13)-(2.22) que l'on peut récrire de la forme :

$$x \in \mathbb{S}_{\theta(x,h)}, \quad (3.19)$$

$$\mathbb{E}(Z_{x,h,\beta}) = x + A_{\theta}(x, h), \quad (3.20)$$

$$\text{Var}(Z_{x,h,\beta}) = B_{\theta}(x, h), \quad (3.21)$$

où $A_{\theta}(x, h) = \sum_{j=1}^p A_j(x, h) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x)$ et $B_{\theta}(x, h) = \sum_{j=1}^p B_j(x, h) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x)$

tendent vers 0 lorsque h tend 0 et $Z_{\theta(x,h)}$ est une variable aléatoire de loi $K_{\theta(x,h)}$.

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

Définition 3.3.6. Soient X_1, X_2, \dots, X_n une suite de variables aléatoires i.i.d. à densité mixte inconnue f en (3.1) et de support $\mathbb{T} = \cup_j \mathbb{T}_j$. L'estimateur à noyau associé mixte $K_{\theta(x,h)}$ \hat{f}_n de f est défini par :

$$\hat{f}_n(x) = \sum_{j=1}^p \frac{\beta_j}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]} \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(X_i), \quad (3.22)$$

où n_j le nombre d'observations tombant dans \mathbb{T}_j et les $K_{x,h}^{[j]}(\cdot) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(\cdot)$ sont des noyaux associés de support $\mathbb{S}_{x,h}^{[j]}$.

Cet estimateur peut être écrit encore sous la forme :

$$\widehat{f}(x) = \sum_{j=1}^p \beta_j \widehat{f}_j(x) \mathbb{1}_{\mathbb{T}_j}(x), \quad (3.23)$$

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

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

3.4 Résultats asymptotiques

Proposition 3.4.1. Soit \widehat{f}_n estimateur à noyaux associés mixtes $K_{\theta(x,h)}$ de f en (3.1).

Pour toute variable aléatoire $Z_{\theta(x,h)}$ de loi $K_{\theta(x,h)}$, \widehat{f}_n vérifie les propriétés suivantes :

$$\mathbb{E}\{\widehat{f}(x)\} = \mathbb{E}\{f(Z_{\theta(x,h)})\}, \quad (3.24)$$

$$\int_{\mathbb{T}} \widehat{f}_n(x) \Delta x = \Lambda(n, h, K_{\theta}) \text{ n'est pas toujours égale } 1. \quad (3.25)$$

Preuve : Le premier résultat se démontre de la manière suivante. Supposons que le support \mathbb{T} est muni de la Δ -mesure et considérons $Z_{\theta(x,h)}$ une variable aléatoire de loi $K_{\theta(x,h)}$. Alors on a :

$$\begin{aligned} \mathbb{E}\{\widehat{f}(x)\} &= \int_{\mathbb{T}} K_{\theta(x,h)}(u) f(u) \Delta u \\ &= \int_{\mathbb{T}} f(u) K_{\theta(x,h)}(u) \Delta u \\ &= \mathbb{E}\{f(Z_{\theta(x,h)})\}. \end{aligned}$$

Pour le second résultat, on a :

$$\begin{aligned} \int_{\mathbb{T}} \widehat{f}(x) \Delta x &= \frac{1}{n_j} \sum_{i=1}^{n_j} \int_{\mathbb{T}} K_{\theta(x,h)}(u) \Delta x \\ &= \int_{\mathbb{T}} K_{\theta(x,h)}(u) \Delta x. \end{aligned} \quad (3.26)$$

Puisque le noyau associé mixte $K_{x,h}(\cdot)$ est une densité de probabilité par rapport à la variable aléatoire mais il ne l'est pas (nécessairement) par rapport à x . Aussi, il hérite la flexibilité des noyaux associés issus desquels il provient. Le noyau associé mixte $K_{\theta(x,h)}(u)$ n'est pas une densité de probabilité par rapport à la variable x . Par conséquent, l'intégrale $\int_{\mathbb{T}} K_{\theta(x,h)}(u) \Delta x$ est une constante positive dépendant simultanément de l'échantillon, du paramètre de dispersion et du type de noyau associé mixte. Nous notons cette constante $\Lambda(n, h, K_{\theta})$ et elle n'est pas toujours égale à 1. ■

On peut se référer à la proposition (1.3.3) de ce travail pour plus de précision. Signalons aussi que l'on peut retrouver le même résultat en utilisant la ∇ -mesure.

Nous présentons dans ces propositions, le biais et la variance de l'estimateur à noyau associé mixte. Puis, nous procédons au calcul des erreurs quadratiques moyennes asymptotiques respectivement ponctuelles et globales.

Proposition 3.4.2. *Soit $f \in \mathcal{C}^2(\mathbb{T})$ une densité mixte du type défini en (3.1) et \widehat{f}_n son estimateur à noyau associé mixte en (3.22). Pour tout x dans \mathbb{T} et $h = h_n > 0$, on a*

$$\text{Biais}\{\widehat{f}(x)\} = A_{\theta}(x, h)f^{(1)}(x) + \frac{1}{2}\{A_{\theta}^*(x, h) + B_{\theta}(x, h)\}f^{(2)}(x) + o(h^2), \quad (3.27)$$

avec

$$A_{\theta}(x, h) = \sum_j \beta_j A_j(x, h), B_{\theta}(x, h) = \sum_j \beta_j B_j(x, h) \text{ et } A_{\theta}^*(x, h) = \sum_j \beta_j A_j^*(x, h).$$

Si en plus, f est bornée sur \mathbb{T} alors

$$\text{Var}\{\widehat{f}_n(x)\} = \sum_{j=1}^p \frac{\beta_j^2}{n_j} f_j(x) \|K_{x,h}^{[j]}\|_2^2 \mathbf{1}_{\mathbb{T}_j}(x) + o\left(\frac{1}{nh^{r_2}}\right), \quad (3.28)$$

où $r_2 = r_2(K_{\theta}) > 0$ est le plus grand réel tel que $\|K_{\theta}\|_2^2 \leq c_2(x)h_n^{-r_2}$ et $0 < c_2(x) < \infty$.

Preuve : Pour le premier résultat, considérons $x \in \mathbb{T}$. En utilisant (3.22), on a :

$$\begin{aligned} \text{Biais}\{\widehat{f}(x)\} &= \mathbb{E} \left\{ \sum_{j=1}^p \frac{\beta_j}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(X_i) \right\} - \sum_{j=1}^p \beta_j f_j(x) \mathbb{1}_{\mathbb{T}_j}(x) \\ &= \sum_{j=1}^p \beta_j \left\{ \mathbb{E} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(X_i) \right] - f_j(x) \mathbb{1}_{\mathbb{T}_j}(x) \right\}. \end{aligned}$$

Puisque les supports $\mathbb{S}_{x,h}^{[j]}$ et \mathbb{T}_j sont égaux, alors on a :

$$\begin{aligned} \text{Biais}\{\widehat{f}(x)\} &= \sum_{j=1}^p \beta_j \left\{ \mathbb{E} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x)(X_i) \right] - f_j(x) \right\} \mathbb{1}_{\mathbb{T}_j}(x) \\ &= \sum_{j=1}^p \beta_j \left[\mathbb{E}\{\widehat{f}_j(x)\} - f_j(x) \right] \mathbb{1}_{\mathbb{T}_j}(x) \\ &= \sum_{j=1}^p \beta_j A_j(x, h) f_j^{(1)}(x) \mathbb{1}_{\mathbb{T}_j}(x) + \frac{1}{2} \sum_{j=1}^p \beta_j \{A_j^2(x, h) \\ &\quad + B_j(x, h)\} f_j^{(2)}(x) \mathbb{1}_{\mathbb{T}_j}(x) + o(h^2) \\ &= A_\theta(x, h) f^{(1)}(x) + \frac{1}{2} \{A_\theta^*(x, h) + B_\theta(x, h)\} f^{(2)}(x) + o(h^2), \end{aligned}$$

où les $f^{(j)}$ pour $j \in \{1, 2\}$ sont définies dans (3.9) puis

$$A_\theta(x, h) = \sum_j \beta_j A_j(x, h), \quad B_\theta(x, h) = \sum_j \beta_j B_j(x, h) \text{ et } A_\theta^*(x, h) = \sum_j \beta_j A_j^2(x, h).$$

Montrons maintenant le second résultat. Partant de la définition de \widehat{f}_n on a :

$$\begin{aligned} \text{Var} \left\{ \widehat{f}_n(x) \right\} &= \text{Var} \left\{ \sum_{j=1}^p \frac{\beta_j}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(X_i) \right\} \\ &= \sum_{j=1}^p \beta_j^2 \text{Var} \left\{ \frac{1}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbb{1}_{\mathbb{T}_j}(X_i) \right\} \mathbb{1}_{\mathbb{T}_j}(x) \\ &= \sum_{j=1}^p \beta_j^2 \text{Var} \{ \widehat{f}_j(x) \} \mathbb{1}_{\mathbb{T}_j}(x). \end{aligned}$$

En remplaçant $\text{Var}\{\widehat{f}_j(x)\}$ par sa valeur donnée dans la formule suivante

$$\text{Var}\{\widehat{f}_j(x)\} = \frac{1}{n_j} f_j(x) \|K_{x,h}^{[j]}\|_2^2 \mathbb{1}_{\mathbb{T}_j}(x) + o\left(\frac{1}{nh^2}\right), \quad (3.29)$$

on obtient le résultat. ■

Nous voulons attirer l'attention du lecteur ici sur ce résultat. Comme démontré dans la proposition (1.3.4), l'obtention de la variance de \widehat{f}_j fait usage du développement en série de Taylor. Ici, la formule utilisée est celle donnée en (3.8). Aussi, la norme $\|K_{x,h}^{[j]}\|_2^2$ se calcule en utilisant l'intégrale définie en (3.15).

Théorème 3.1. (*Risque quadratique moyen asymptotique*) En combinant (3.27) et (3.28), alors l'erreur quadratique moyenne ponctuelle et intégrée donner par :

$$MSE(x) = A_\theta(x, h)f^{(1)}(x) + \frac{1}{2}\{A_\theta^*(x, h) + B_\theta(x, h)\}f^{(2)}(x) + \sum_{j=1}^p \frac{\beta_j^2}{n_j} f_j(x) \|K_{x,j}^{[j]}\|_2^2 \mathbb{1}_{\mathbb{T}_j(x)}.$$

En intégrant par rapport à la Δ -mesure sur \mathbb{T} on trouve :

$$\begin{aligned} MISE(x) &= \int_{\mathbb{T}} [A_\theta(x, h)f^{(1)}(x) + \frac{1}{2}\{A_\theta^*(x, h) + B_\theta(x, h)\}f^{(2)}(x)] \Delta x \\ &+ \sum_{j=1}^p \frac{\beta_j^2}{n_j} \int_{\mathbb{T}} f_j(x) \|K_{x,j}^{[j]}\|_2^2 \mathbb{1}_{\mathbb{T}_j(x)} \Delta x. \end{aligned} \quad (3.30)$$

3.5 Choix de fenêtre par la méthode de validation croisée

Nous envisageons ici le choix de fenêtre de lissage par la méthode de validation croisée. Nous précisons ici que ce choix peut se faire de deux manières. La première consiste à choisir h globalement en utilisant directement le noyau associé mixte $K_{\theta(x,h)}$ dans la validation croisée en prenant en compte l'ensemble des observations. La deuxième manière consiste quant à elle à procéder au choix local de h sur chaque composante \mathbb{T}_j de \mathbb{T} . Ici, on fait intervenir uniquement les observations tombant dans \mathbb{T}_j . Dans tous les cas, le statisticien est appelé à doubler de vigilance lorsqu'on se trouve aux points limites i.e. au passage d'un intervalle \mathbb{T}_I à un ensemble \mathbb{T}_N discret en tenant compte des informations du passé. Pour ce travail, nous proposons un choix global par la méthode de validation croisée par les moindres carrés. L'approche de validation croisée par les moindres carrés nous amène à considérer le noyau associé

mixte $K_{\theta(x,h)}$ pour tout $x \in \mathbb{T} = \bigcup \mathbb{T}_j$ et $h > 0$. Le paramètre optimal h_{CV} de h est obtenu par :

$$h_{CV} = \arg \min_{h>0} CV(h),$$

où

$$\begin{aligned} CV(h) &= \int_{\mathbb{T}} \{\widehat{f}_n(x)\}^2 \Delta x - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i) \\ &= \int_{\mathbb{T}} \left\{ \frac{1}{n} \sum_{i=1}^n K_{\theta(x,h)}(X_i) \right\}^2 \Delta x - \frac{2}{n} \sum_{i=1}^n \widehat{f}_{n,-i}(X_i) \\ &= \int_{\cup_j \mathbb{T}} \left\{ \sum_{j=1}^p \frac{\beta_j}{n_j} \sum_{k=1}^{n_j} K_{\theta(x,h)}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbb{1}_{\mathbb{T}_j}(X_k) \right\}^2 \Delta x \\ &\quad - \frac{2}{n} \sum_{i=1}^n \left\{ \sum_{j=1}^p \frac{\beta_j}{n_j - 1} \sum_{i \neq k=1}^{n_j} K_{\theta(x,h)}^{[j]}(X_i) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}} \mathbb{1}_{\mathbb{T}_j}(X_k) \right\} \\ &= \sum_{j=1}^p \int_{\mathbb{T}} \left\{ \sum_{j=1}^p \beta_j \widehat{f}_j \mathbb{1}_{\mathbb{T}_j}(x) \right\}^2 \Delta x - \frac{2}{n} \sum_{i=1}^n \left\{ \sum_{j=1}^p \widehat{f}_{j,-i}(X_i) \right\}, \end{aligned}$$

où $\widehat{f}_{j,-i}(u) = (n_j - 1)^{-1} \sum_{i \neq k=1}^{n_j} K_{u,h}^{[j]}(X_k) \mathbb{1}_{\mathbb{S}_{x,h}^{[j]}}(u) \mathbb{1}_{\mathbb{T}_j}(X_k)$ est calculé à partir de $\widehat{f}_j(u)$

sans l'observation X_j .

3.6 Application numérique au modèle de mélange non-paramétrique

Nous présentons dans cette section, une première application de cette méthode dans le cadre d'estimation non-paramétrique d'un modèle de mélange mixte (i.e. partiellement discret et continu).

Considérons un modèle de mélange f (qui est bien-sur une densité mixte de probabilité) défini par :

$$f(x) = 0.65 f_1(x) \mathbb{1}_{[2,5]}(x) + 0.35 f_2(x) \mathbb{1}_{\{6,7,\dots,20\}}(x), \quad (3.31)$$

où f_1 et f_2 sont respectivement des fonctions de densité et de masse de probabilité inconnues de supports respectifs $\mathbb{T}_1 = [2, 5]$ et $\mathbb{T}_2 = \{6, 7, \dots, 20\}$.

Compte tenu du manque d'informations générales sur sa forme et son appartenance à une quelque famille de lois paramétriques, l'estimation de f définie en (3.31) ne nous

a pas laissé un autre choix que de procéder par l'approche non-paramétrique et plus précisément par la méthode des noyaux associés. Ainsi, nous proposons l'utilisation d'un noyau associé mixte constitué d'un mélange disproportionné des noyaux bêta étendu et triangulaire général discret.

Rappelons que le noyau bêta étendu construit par la méthode mode-dispersion est défini par :

$$EB_{x,h,t_0,t_1}(u) = \frac{(u-t_0)^{(x-t_0)/\{(t_1-t_0)h\}}(t_1-u)^{(t_1-x)/\{(t_1-t_0)h\}}}{(t_1-t_0)^{1+h^{-1}}\Delta} \mathbb{1}_{[t_0,t_1]}(u),$$

avec

$$\Delta = \mathcal{B}(\left[\frac{\{x-t_0\}}{\{(t_1-t_0)h\}}\right] + 1, \left[\frac{\{t_1-x\}}{\{(t_1-t_0)h\}}\right] + 1).$$

Il a pour support $[t_0, t_1]$. En posant $t_0 = 2$ et $t_1 = 5$ on a le bêta étendu sur $[2,5]$ défini par :

$$EB_{x,h,2,5}(u) = \frac{(u-2)^{(x-2)/3h}(5-u)^{(5-x)/3h}}{3^{1+1/h}\mathcal{B}\left(\left[\frac{\{x-2\}}{3h}\right] + 1, \left[\frac{\{5-x\}}{3h}\right] + 1\right)} \mathbb{1}_{[2,5]}(u).$$

Il a pour caractéristiques

$$A_1(x, h, 2, 5) = \frac{2h(4-x)}{1+2h}, \tag{3.32}$$

$$B_1(x, h, 2, 5) = \frac{-x^2 + 7x + 9h + 9h^2 - 10}{(1+2h)^2(1+3h)}. \tag{3.33}$$

L'estimateur de f_1 par le noyau bêta étendu a respectivement pour biais :

$$\begin{aligned} \text{Biais}\{\widehat{f}_1(x)\} &= \left\{ \frac{2h(4-x)}{1+2h} \right\} f_1^{(1)}(x) + \frac{1}{2} \left\{ \frac{2h(4-x)}{1+2h} \right\}^2 f_1^{(2)}(x) \\ &+ \left\{ \frac{-x^2 + 7x + 9h + 9h^2 - 10}{2(1+2h)^2(1+3h)} \right\} f_1^{(2)}(x) + o(h^2), \end{aligned}$$

et sa variance est donnée par :

$$\text{Var}\{\widehat{f}_1(x)\} = \begin{cases} \frac{\Gamma(2c+1)}{2^{2c+1}\Gamma^2(c+1)} \frac{1}{n} f(x) & \text{si } \frac{x-5}{3h} \text{ et } \frac{5-x}{3h} \longrightarrow c \\ \frac{3(x-2)^{-1/2}(5-x)^{-1/2}}{2\sqrt{\pi}nh^{1/2}} f(x) & \text{si } \frac{x-5}{3h} \text{ et } \frac{5-x}{3h} \longrightarrow +\infty. \end{cases}$$

De même, le noyau triangulaire général d'ordre h et de bras a_1 et a_2 est défini par :

$$\mathcal{DT}(u, x, a_1, a_2, h) = \frac{1}{D(a_1, a_2, h)} \left\{ \left[1 - \left(\frac{x-u}{a_1+1} \right)^h \right] \mathbb{1}_{\mathbb{T}_{a_1, x}^*}(u) + \left[1 - \left(\frac{u-x}{a_2+1} \right)^h \right] \mathbb{1}_{\mathbb{T}_{x, a_2}}(u) \right\},$$

avec

$$D(a_1, a_2, h) = (a_1 + a_2 + 1) - (a_1 + 1)^{-h} \sum_{k=1}^{a_1} k^h - (a_2 + 1)^{-h} \sum_{k=1}^{a_2} k^h.$$

En l'adaptant au support de f_2 (i.e. en posant $a_1 = 6$ et $a_2 = 20$) on a :

$$\mathcal{DT}(u, x, 6, 20, h) = \frac{1}{D(6, 20, h)} \left\{ \left[1 - \left(\frac{x-u}{7} \right)^h \right] \mathbb{1}_{\mathbb{T}_{6, x}^*}(u) + \left[1 - \left(\frac{u-x}{21} \right)^h \right] \mathbb{1}_{\mathbb{T}_{x, 20}}(u) \right\},$$

avec

$$D(6, 20, h) = 27 - 7^h \left(\sum_{k=1}^6 k^h - 3^h \sum_{k=1}^{20} k^h \right)$$

Ses caractéristiques sont données par

$$A_2(x, h) = A(6, 20, h) = \frac{1}{D(6, 20, h)} \left\{ 189 - 7^h \left(\sum_{k=1}^6 k^{h+1} - 3^h \sum_{k=1}^{20} k^{h+1} \right) \right\}, \quad (3.34)$$

$$B_2(x, h) = B(6, 20, h) = \frac{1}{D(6, 20, h)} \left\{ 2961 - 7^h \left(\sum_{k=1}^6 k^{h+2} - 3^h \sum_{k=1}^{20} k^{h+2} \right) \right\}. \quad (3.35)$$

L'estimateur \widehat{f}_2 de la fonction de masse de probabilité f_2 à noyau triangulaire général discret a pour biais :

$$\text{Biais}\{\widehat{f}_2(x)\} = A_2(x, h) f_2^{(1)}(x) + \frac{1}{2} \left\{ -A_2^2(x, h) + B_2(x, h) \right\} f_2^{(2)} + o(h^2),$$

et sa variance est donnée par :

$$\text{Var} \left\{ \widehat{f}_2(x) \right\} = \frac{1}{n_2 \{D(6, 20, h)\}^2} f_2(x) [1 - f(x)] + R_n(x, 6, 20, h),$$

avec $R_n(x, 6, 20, h) \rightarrow 0$ quand $n \rightarrow +\infty$ (se référer à Kokonendji& Zocchi, 2010[18]). Enfin, le noyau associé mixte $K_{\theta(x,h)}$ constitué des noyaux associés bêta étendu et triangulaire général est donné par :

$$K_{\theta(x,h)}(\cdot) = 0.65EB_{x,h,2,5}(\cdot)\mathbb{1}_{[2,5]}(x)\mathbb{1}_{[2,5]}(\cdot) + 0.35\mathcal{DT}_{x,6,20,h}(\cdot)\mathbb{1}_{\{6,7,\dots,20\}}(x)\mathbb{1}_{\{6,7,\dots,20\}}(\cdot). \quad (3.36)$$

Ses caractéristiques sont données par :

$$A_{\theta}(x, h) = 0.65A_1(x, h)\mathbb{1}_{[2,5]}(x) + 0.35A_2(x, h)\mathbb{1}_{\{6,7,\dots,20\}}(x), \quad (3.37)$$

$$B_{\theta}(x, h) = 0.65B_1(x, h)\mathbb{1}_{[2,5]}(x) + 0.35B_2(x, h)\mathbb{1}_{\{6,7,\dots,20\}}(x), \quad (3.38)$$

où les quantités $A_1(x, h)$, $B_1(x, h)$, $A_2(x, h)$ et $B_2(x, h)$ sont respectivement donnés en (3.32), (3.33), (3.34) et (3.35).

L'estimateur \widehat{f}_n de la densité mixte f définie en (3.31) à noyau associé mixte $K_{\theta(x,h)}$ en (3.36) est donné par :

$$\widehat{f}(x) = \frac{0.65}{n_1} \sum_{i=1}^{n_1} EB_{x,h,2,5}(X_i)\mathbb{1}_{[2,5]}(x)\mathbb{1}_{[2,5]}(X_i) + \frac{0.35}{n_2} \sum_{i=1}^{n_2} \mathcal{DT}_{x,6,20,h}\mathbb{1}_{6,7,\dots,20}(x)\mathbb{1}_{\{6,7,\dots,20\}}(X_i). \quad (3.39)$$

Son biais est défini par :

$$Biais\{\widehat{f}(x)\} = A_{\theta}(x, h)f^{(1)}(x) + \frac{1}{2}\{A_{\theta}^*(x, h) + B_{\theta}(x, h)\}f^{(2)} + o(h^2), \quad (3.40)$$

où les quantités A_{θ} et B_{θ} sont définies respectivement en (3.37) et (3.38), et la quantité $A_{\theta}^*(x, h)$ est donnée par :

$$A_{\theta}^*(x, h) = 0.65A_1^2(x, h)\mathbb{1}_{[2,5]}(x) + 0.35A_1^2(x, h)\mathbb{1}_{\{6,7,\dots,20\}}(x).$$

De plus $f^{(1)}$ et $f^{(2)}$ sont calculées à partir de (3.9).

La variance de \widehat{f} est donnée par :

$$\begin{aligned} Var\{\widehat{f}(x)\} &= \frac{0.65^2}{n_1} f_1(x) \|EB_{x,h,2,5}\|_2^2 \mathbb{1}_{[2,5]}(x) + \frac{0.35^2}{n_2} f_2(x) \|\mathcal{DT}_{x,6,20,h}\|_2^2 \\ &\quad \mathbb{1}_{\{6,7,\dots,20\}}(x) + o\left(\frac{1}{nh^{r_2}}\right) \\ &= \frac{1.2675(x-2)^{-1/2}(5-x)^{-1/2}}{2\sqrt{\pi}n_1h^{1/2}} f_1(x)\mathbb{1}_{[2,5]}(x) + \frac{0.1225}{n_2\{D(6,20,h)\}^2} \\ &\quad f_2(x)[1-f_2(x)]\mathbb{1}_{\{6,7,\dots,20\}}(x) + o\left(\frac{1}{nh^{1/2}}\right). \end{aligned} \quad (3.41)$$

Ici $r_2 = 1/2$.

Nous présentons dans le tableau (3.6), les différentes valeurs de biais, variance et MSE de f . Ces résultats montrent que la variance de l'estimateur à noyau associé mixte pour le modèle de mélange diminue en général lorsque la taille de l'échantillon augmente. Cependant, on note que son biais augmente aux bords de la partie continue du support et diminue aux bords de la partie discrète quand la taille de l'échantillon croît. Ceci peut s'expliquer d'une part par le fait que nous avons travaillé avec un mélange utilisant la version du noyau associé bêta étendu construit par la méthode mode-dispersion. D'autre part il peut être dû au choix du paramètre de lissage h . Nous signalons ici que notre choix de h est fait localement par composante de \mathbb{T} en utilisant la méthode de validation croisée par les moindres carrés. Il est très intéressant de comparer ce résultat en utilisant une valeur de h choisie globalement.

Taille	Cible	Biais	Variance	MSE
100	2.25	-0.4459 e-4	0.1016e-4	0.1036e-4
	4.50	-0.4152 e-4	0.1126e-4	0.1281e-4
	7.00	0.9213e-3	0.3538 e-4	0.1150e-3
	19.00	0.9477 e-3	0.7963 e-4	0.1694 e-3
500	2.25	-0.3241 e-3	0.1621e-5	0.1036e-6
	4.50	-0.3081 e-3	0.1012e-5	0.1106e-6
	7.00	0.7210e-2	0.0261e-5	0.5224 e-6
	19.00	0.7001e-2	0.0198 e-5	0.4921 e-6
1000	2.25	-0.3014 e-3	0.1315 e-5	0.1406 e-7
	4.50	-0.3009 e-3	0.1059 e-5	0.1149 e-7
	7.00	0.7012e-2	0.0197 e-5	0.4936 e-6
	19.00	0.6982e-2	0.0131e-5	0.4887 e-6

Tableau (1.11) -Biais et variances de l'estimateur à noyaux associés mixtes de la densité définie en (3.31).

Conclusion

Nous avons étudié dans ce travail une méthode non-paramétrique d'estimation de densités.

Tout d'abord, nous avons présenté l'état de l'art sur l'estimateurs à noyaux continus (classiques) de Rosenblatt (1956)[26] et Parzen (1962)[24] de la densité univarié continue et ses propriétés statistiques, et des estimateurs à noyaux associés discrets de Konondji & Senga Kiessé (2011) et continus. Nous avons commencé par une définition pour les noyaux continus (classiques) qui jusqu'alors se déduit de celle de l'estimateur puis nous avons aussi amélioré la définition des noyaux associés discrets et continus. Nous avons introduit la forme de l'estimateur à noyau associé discret/continu multivarié. Les propriétés statistiques de ces estimateur et la méthodes classiques pour le choix de la matrice de lissage ont été étudiées. Par la suite, nous avons défini les estimateurs à noyaux associés continus puis nous avons montré qu'ils sont sans effet de bord mais possèdent un biais plus grand (en nombre de termes) que les estimateurs à noyaux classiques. Ce problème peut être interprété comme l'équivalent des effets de bord dans le cas classique. Pour y remédier, nous avons développé un algorithme de réduction de biais qui consiste d'abord partitionner le \mathbb{T} de la densité à estimer en deux régions (régions de bord et celle de l'intérieur représentant la plus grande partie) puis à modifier le noyau associé de départ de sorte que le biais de l'estimateur issu de ce noyau soit de même ordre que celui de l'estimateur à noyau continu classique dans la région intérieure. Nous avons par la suite appliqué cet algorithme dans un exemple sur l'estimateur à noyau associé non-classique.

Par ailleurs, nous avons étudié les consistances ponctuelles puis globales des estimateurs à noyaux associés continus. Ces études nous ont permis de savoir que le comportement de ces estimateurs dépend d'une part du point d'estimation x , et autre part d'une certaine puissance du paramètre de lissage h . Enfin, nous avons combiné

les noyaux associés discrets de Kokonendji & Senga Kiessé (2011)[17] et les noyaux associés continus nouvellement présentés, pour définir les noyaux associés mixtes. Nous avons ensuite utilisé les outils d'analyse sur les échelles de temps pour montrer les propriétés des estimateurs à noyaux associés mixtes.

Bibliographie

- [1] Abdous, B. & Kokonendji, C.C. (2009). Consistency and asymptotic normality for discrete associated-kernel estimator. *African Diaspora Journal of Mathematics* 8 (2), 63-70.
- [2] Agarwal, R.P. & Bohner, M. (1999). Basic calculus on time scales and some of its applications, *Results in Mathematics* 35, 3-22.
- [3] Aitchison, J. & Aitken, C.G.G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413-420.
- [4] Bohner, M. & Peterson, A. (2001). *Dynamic Equations on Time Scales*, Birkhera Boston Inc., Boston.
- [5] Bosq, D. Estimation suroptimale de la densité par projection. *Canadian Journal of Statistics*, 33(1) : 21-37, (2005).
- [6] Chacon, J.E., Duong, T. & Wand, M.P. (2011). Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica*, 21, 807-840.
- [7] Cencov, N.N. Evaluation of an unknown distribution density from observations. *Soviet Mathematics*, 3 : 1559-1562, (1962).
- [8] Chen, S.X. (1999) Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* 31, 131-145.
- [9] Epanechnikov, V.A. (1969). Nonparametric estimation of a multivariate probability density, *Theory of Probability and Its Applications* 14, 153-158.
- [10] Henderson, H.V. & Searle, S.R. (1979). Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics, *Canadian Journal of Statistics* 7, 65-81.
- [11] Hilger, S. (1990). Analysis on measure chains a unified approach to continuous and discrete calculus, *Results in Mathematics* 18, 18-56.

-
- [12] Jin, X. and Kawczak, J. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, 4 : 103-124, (2003).
- [13] Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.
- [14] Jørgensen, B. & Kokonendji, C.C. (2011). Dispersion models for geometric sums, *Brazilian Journal of Probability and Statistics* 25, 263-293.
- [15] Kerkyacharian, G. and Picard, D. Density estimation by kernel and wavelets methods : optimality of besov spaces. *Statistics and Probability Letters*, 18(4) : 327-336, (1993).
- [16] Kokonendji, C.C. and Libengué, F. G. Méthode des noyaux associés continus et estimation de densité. *Journées de Statistique de la SFdS*, 6 pages, Tunis, (2011)
- [17] Kokonendji, C.C. & Senga Kiessé, T. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, 8, 497-516.
- [18] Kokonendji, C.C. & Zocchi, S.S. (2010). Extensions of discrete triangular distribution and boundary Biases in kernel estimation for discrete functions. *Statistics and Probability Letters* 80, 1655-1662.
- [19] Kronmal, R. and Tarter, M. The estimation of probability densities and cumulatives by fourier series methods. *Journal of the American Statistical Association*, pages 925-952, (1968).
- [20] Libengué, F.G. (2013). Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications. Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France & Université de Ouagadougou, Burkina Faso, June 2013, LMB no. 14334, Besançon, URL <http://hal.in2p3.fr/tel-01124288/document>.
- [21] Marron, J.S. and Ruppert, D. Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 653-671, (1994).
- [22] Müller, H.G. & Stadtmüller, U. (1999). Multivariate boundary kernels and a continuous least squares principle, *Journal of the Royal Statistical Society B* 61, 439-458.

-
- [23] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Applications* 9, 141-142.
- [24] Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33, 1065-1076.
- [25] Racine, J.S. and Li, Q. Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119 : 99-130, (2004).
- [26] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* 27, 832-837.
- [27] Saadi, N. and Adjabi, S. On the estimation of the probability density by trigonometric series. *Communications in Statistics Theory and Methods*, 38(19) : 3583-3595, (2009).
- [28] Sanyal, S. (2008), *Stochastic Dynamic Equations*, Ph.D. Thesis Manuscript to Missouri University of Sciences and Technology.
- [29] Senga Kiessé, T. (2009). *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat en Statistique, Université de Pau. [http : //tel.archives-ouvertes.fr/tel-00372180/fr/](http://tel.archives-ouvertes.fr/tel-00372180/fr/).
- [30] Senga Kiessé, T. (2008). *Approche Non-Paramétrique par Noyaux Associés Discrets des Données de Dénombrement*. Ph.D. Thesis Manuscript (in French), Université de Pau. [http : //tel.archives-ouvertes.fr/tel-00372180/fr/](http://tel.archives-ouvertes.fr/tel-00372180/fr/).
- [31] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics* 16, 217-226.
- [32] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics Theory and Methods* 40, 1123-1136.
- [33] Silverman, B.W. *Density estimation*. Chapman and Hall, London, (1986).
- [34] Tsybakov, A.B. *Introduction to Nonparametric Estimation*. Springer, New York, (2009).
- [35] Tsybakov, A.B. (2004). *Introduction à l'Estimation Non Paramétrique*. Springer, Paris.
- [36] Wahba, G. Data-based optimal smoothing of orthogonal series density estimates. *The Annals of Statistics*, pages 146-156, (1981).

- [37] Wahba, G. Optimal convergence properties of variable knot, kernel, and orthogonal series methods for density estimation. *The Annals of Statistics*, pages 15-29, 1975.
- [38] Walter, G.G. Approximation of the delta function by wavelets. *Journal of Approximation Theory*, 71(3) : 329-343, (1992).
- [39] Watson, G.S. (1964). Smooth regression analysis, *Sankhya Series A* 26, 359-372.