HAL
open science

Méthode non-paramétrique des noyaux associés mixtes et applications

Francial Giscard Baudin Libengué Dobélé-Kpoka Libengue Dobele-Kpoka

To cite this version:

Francial Giscard Baudin Libengué Dobélé-Kpoka Libengue Dobele-Kpoka. Méthode non-paramétrique des noyaux associés mixtes et applications. Mathématiques générales [math.GM]. Université de Franche-Comté; Université Joseph Ki-Zerbo (Ouagadougou, Burkina Faso), 2013. Français. NNT : 2013BESA2007 . tel-01124288

HAL Id: tel-01124288

https://theses.hal.science/tel-01124288v1

Submitted on 6 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Doctorale Carnot-Pasteur (France)
&
École Doctorale des Sciences et Technologie (Burkina Faso)

THÈSE EN CO-TUTELLE

pour obtenir le grade de
Docteur de l'Université de Franche-Comté
et de l'Université de Ouagadougou
Spécialité : **Mathématiques et Applications**
Présentée par

Francial Giscard Baudin LIBENGUÉ DOBÉLÉ-KPOKA

Directeurs de Thèse :

Professeur Célestin C. KOKONENDJI (Université de Franche-Comté)
Professeur Blaise SOMÉ (Université de Ouagadougou)

Méthode non-paramétrique des noyaux associés mixtes et applications

Soutenue le **13 juin 2013 à Ouagadougou**

Après avis des Rapporteurs :

Pascal SARDA, Professeur à l'Université de Toulouse II-Le Mirail, France
Anne-Françoise YAO-LAFOURCADE, Professeure à l'Université Blaise Pascal, France

Devant le Jury composé, des Directeurs de thèse, des rapporteurs et :

Président :	Professeur Gane Samb LO,	Université Saint Louis, Sénégal
Examinateurs :	Professeur Clovis NITIEMA,	Université Ouaga II, Burkina Faso
	Professeur Ousséni SO,	Institut Des Sciences, Burkina Faso
	Professeur Longin SOME,	Université de Ouagadougou, Burkina Faso

Dédicaces

Je dédie cette thèse à :

Dieu le Père Tout Puissant qui ne cesse de faire pour moi des merveilles ;

Emmanuel LIBENGUE et Françoise KONAMO, le couple qui m'a mis au monde.

Je le dédie aussi à vous : Paméla DENEMANDJI, Patrick BAÏPO, Arlette LIBENGUE, Nadine LIBENGUE, Fridolin LIBENGUE, Junior LIBENGUE, Jenifer LIBENGUE, Cyinthia LIBENGUE, Jésus LIBENGUE, Divina LIBENGUE, Adoreinne LIBENGUE et Exaucée LIBENGUE, pour votre amour, votre soutien sans faille et tout ce vous avez pu m'apporter pour franchir les obstacles les plus difficiles.

Enfin, en mémoire de ma défunte et chère maman Elizabeth, et ma soeur Ella qui auraient été fières du fruit de tant d'efforts consentis.

"KANGA BÉ TI MO, GBOU MBÉTI A GOUÉ YONGORO".

Remerciements

Je suis très heureux d'exprimer ici tous mes remerciements, mes profondes gratitude et reconnaissance à mon directeur de thèse, le Professeur Célestin C. Kokonendji, sans qui ce travail n'aurait pu voir le jour. Merci de m'avoir proposé ce sujet de thèse aussi intéressant. Merci d'avoir pris le temps de m'enseigner avec patience les nécessaires bases de la statistique mathématique et de développer en moi l'aptitude d'analyse des projets de recherche. Ta rigueur scientifique, tes critiques pertinentes et ta présence attentive tout au long de ce parcours ont été essentiels à la réalisation de ce travail. J'espère que cette thèse est digne de ton attente et de ton investissement personnel.

Je remercie chaleureusement mon second directeur de thèse le Professeur Blaise Somé, Directeur de Laboratoire d'Analyse Numérique, d'Informatique et de Biomathématique (LANIBIO) de l'Université de Ouagadougou qui a guidé mes premiers pas en recherche depuis mon DEA et qui s'est fortement impliqué pour le financement de cette thèse en me recommandant personnellement à l'Agence Universitaire de la Francophonie (AUF). Merci d'avoir créé le LANIBIO qui nous pousse à rechercher le meilleure de nous même.

Je remercie cordialement le Professeur Pascal Sarda, de l'université de Toulouse II-Le Mirail et la Professeure Anne-Françoise Yao-Lafourcade, de l'Université Blaise Pascal de Clermont Ferrand, d'avoir accepté être rapporteurs externes de cette thèse. J'adresse également mes remerciements au Professeur Clovis Nitiéma de l'Université de Ougadougou qui a accepté être rapporteur interne de cette thèse.

Mes remerciements au Professeur Gane Samb Lo de l'Université Gaston Berger de Saint Louis au SENEGAL, qui a accepté présider ce jury, merci aussi à tous les autres membres du jury à savoir les Professeurs Longhin Somé et Ousséni So de l'Université de Ougadougou.

Mes profonds remerciements à l'Agence Universitaire de la Francophonie (AUF), pour avoir investi sur moi en finançant entièrement cette thèse. Ces remerciements vont aussi en l'endroit de ses agents en particulier Madame Fabar Sané, Madame Zoé Aubièrge Wangré, Madame Anne Laure Lejeune, Monsieur Benjamin Sia, Monsieur Ousmane Barra et Monsieur Youssouf Ouattara pour leurs services.

Je n'oublierai pas de remercier le Professeur Christian Maire, Directeur du Laboratoire de Mathématiques de Besançon (LMB) ainsi que son prédécesseur le Professeur Patrick Hild de m'avoir accueilli dans ce cadre idéal de travail. Mes remerciements vont aussi en l'endroit du Professeur Gilles Lancian, Directeur de l'École Doctorale Carnot Pasteur ainsi que tous les autres membres du LMB sans oublier Mme Catherine Pagani.

J'aimerais remercier tous les membres d'Équipe de Probabilités-Statistique du LMB en particulier les Maîtres de Conférences : Yacouba Boubacar Maïnassara, Davit Varron, Stéphane Chretien et Landy Rabhéhasaina pour leur collaboration.

Je remercie particulièrement Monsieur Nicolas Klutchnikoff, Maître de Conférences à l'ENSAI, pour m'avoir appris la théorie des minimax lors de mon séjour à Rennes. Merci aussi à toute l'équipe qui m'avait accueilli.

J'adresse mes remerciements à toute l'équipe du Décanat, à tous les Enseignants Chercheurs et en particuliers ceux du Département Mathématiques & Informatique de la Faculté des Sciences de l'Université de Bangui.

Mes remerciement s'adressent aussi à tous les membres du LANIBIO et en particulier au Pr. G. Kabré, Dr. O. Sawadogo, Dr. E. Gouba, Dr. V. Kana sans oublier Dr. Y. Paré, Dr. W. Mbwentchou, Dr. T. Sadou, Dr. M. Kéré et Mr K. Somé.

Je remercie aussi tous les doctorants du LMB et du LANIBIO, en particulier Cyril Moypemna, Matthieu Somé, Khoirin Nisa, Basad Al-sarray, Julien Grépat, Mohamed Gazibo et Étienne Ouédraogo. Je vous encourage à continuer jusqu'au bout.

Je remercie les Professeurs Sékou Traoré et Faustin Touadéra qui m'ont encouragé et orienter vers la recherche. Mes remerciements vont aussi à l'encontre de Pr. Jean L. Syssa Magalé, Pr. Joachim R. Sioké, Dr. Jean Mboliguipa, Dr. Ernest Mada, Dr. Jocelyn Gonessa et Dr. Doriano B. Pougaza pour les précieux conseils qu'ils n'ont cessé de me donner.

Que mes tantes Bernadette Konam, Simone Konam, Germaine Millier et Françoise Libengué ainsi que mes oncles Jean Aurelien Libengué, Dominique Konamo et Narcisse Millier soient aussi remerciés pour leurs soutiens infaillibles.

Mes sincères remerciement aux familles Grébongo, Moussapiti, Dénémandji, Douloumatchi, Bondéboli, Wambobo, Mokambo, Lho-Konam, Kpawilina, Légué, Adiallo, Yandoba, Adrisse, Bompangué et Mbito pour leurs soutiens infaillible et inconditionnel tout au long de mon parcours.

Merci à tous mes collaborateurs, je pense notamment à Frédéric Wango, Hugues et Rebecca Penda, Ange Wakamba, Crepin Bouté, Stevens Orondrou, Stève Djetel, Placide Douam, Arsène Solamosso, Dieudonné Mandéhou, Vianney Rama, Yvon Senguéla, Martinien Penda Rodrigue Rama et Mariam Saba.

Ben-Sthal Sakoma et Bienvenu Pakouzou, mes compagnons de misères, nous avons toujours partagé le même sort, vos soutien et conseil n' ont cessé de me réconforter. Que vous en soyez ainsi remerciés.

Merci à vous tous que je n'ai peut être pas nommément cité ici, pour les agréables occasions de partage que nous avons eues je pense à Mylène et Kevin Kokonendji, Daniel Bongo ainsi que Chérubin Magba.

Pour le meilleur, c'est surtout une profonde pensée pour mes quatre grands parents décédés, et qui auraient été fiers de moi.

Je n'oublierai pas la Communauté Centrafricaine du Burkina (CCAB), l'Union des Élèves, Étudiants et Stagiaires Centrafricains vivant au Burkina Faso (UEESCABF), les jeunes légionnaires de la Paroisse Saint Michel, la chorale Notre Dame de l'Unité et le groupe liturgique de Scholasticat Saint Camille.

Production scientifique

i. Parues

[**i.**1] Senga Kiessé, T., **Libengué, F.G.**, Zocchi, S.S. & Kokonendji, C.C. (2010). The R package for general discrete triangular distributions, http ://www.crn.r-project.org/web/packages/TRIANGG/index.html

[**i.**2] Kokonendji, C.C. & **Libengué, F.G.** (2011). Méthode des noyaux associés continus et estimation de densité. *Procedings du 43 ème Journée de Statistique de la SFdS*, 86-91, Tunis 23-27 mai 2011.

[**i.**3] Kokonendji, C.C., **Libengué, F.G.** & Varron, D. (2012). Covergence des estimateurs à noyaux associés de densité. *Journées de Statistique de la SFdS*, 5 pages, Bruxelles 21-25 mai 2012. http ://jds2012.ulb.ac.be/showabstract.php ?id=98

[**i.**4] **Libengué, F.G.**, Somé, S.M. & Kokonendji, C.C. (2013). Estimateur par noyaux a ssociés mixtes d'un modèle de mélange. *Les 45e Journées de Statistique de la SFdS, Toulouse 27 - 31 mai 2013*.

[**i.**5] Somé, S.M., **Libengué, F.G.** & Kokonendji, C.C. (2013). Estimation de densité par noyau bêta bivarié avec structure de corrélation. *Les 45e Journées de Statistique de la SFdS, Toulouse 27 - 31 mai 2013*.

ii. Soumises

[**ii.**1] Kokonendji, C.C. & **Libengué, F.G.** (2012). Asymptotic results for continuous associated kernel estimators for density functions. Soumis pour publication depuis le 19.11.2012 à *Annals of the Institute of Statistical Mathematics* (Ref. AISM-D-12-00087 ; Under Review).

[**ii.**2] Kokonendji, C.C. & **Libengué, F.G.** (2013). Non-classical associated kernels for non-standard density estimators. Soumis pour publication depuis le 07.01.2013 à *Statistical Methodology* (Ref.STAMET-D-13-00005 ; Under Review).

iii. En préparation

[**iii.**1] Klutchnikoff, N., Kokonendji, C.C. & **Libengué, F.G.** Minimax and Adaptive Properties of Associated Kernel Density.

[**iii.**3] Kokonendji, C.C., **Libengué, F.G.** & Somé, S.M. Mixed associated kernel estimators of a mixture model.

iv. Sans comité de lecture et séminaires universitaires

[**iv.**1] **Libengué, F.G.** (2011a). Introduction à la méthode des noyaux associés continus. *Deuxième Journée de Rencontre Dijon-Besançon en Probabilités et Statistique*. Dijon 2011.01.28.

[**iv.**2] **Libengué, F.G.** (2011b). Extended beta kernel estimators for bounded functions. *Séminaires de l'Equipe Probabilités et Statistique de l'Université de Franche-Comté*. Besançon 2011.02.14.

[**iv.**3] **Libengué, F.G.** (2011c). Estimation de densités par le noyau bêta étendu. *Séminaires du Laboratoire d'Analyse Numérique d'Informatique et de Biomathématique (LANIBIO) de l'Université de Ouagadougou*. Ouagadougou 2011.03.03.

[**iv.**4] **Libengué, F.G.** (2012a). Quelques résultats asymptotiques des estimateurs à noyaux associés de densité. *Séminaires de l'Equipe Probabilités et Statistique de l'Université de Franche-Comté*. Besançon 2012.02.20.

[**iv.**5] **Libengué, F.G.** (2012b). Convergences des estimateurs à noyaux associés de densité . *Séminaires des Doctorants de l'Université de Franche-Comté*. Besançon 2012.03.07.

[**iv.**6] **Libengué, F.G.**. (2012c). Convergences des estimateurs à noyaux associés de densité . *Séminaires du Laboratoire d'Analyse Numérique d'Informatique et de Biomathématique (LANIBIO) de l'Université de Ouagadougou*. Ouagadougou 2012.04.22

[**iv.**5] **Libengué, F.G.** (2013). Estimation par noyaux associés mixtes d'un modèle de mélange. *Séminaires de l'Equipe Probabilités et Statistique de l'Université de Franche-Comté*. Besançon 2013.02.11.

Résumé

Nous présentons dans cette thèse, l'approche non-paramétrique par noyaux associés mixtes, pour les densités à supports partiellement continus et discrets. Nous commençons par rappeler d'abord les notions essentielles d'estimation par noyaux continus (classiques) et noyaux associés discrets. Nous donnons la définition et les caractéristiques des estimateurs à noyaux continus (classiques) puis discrets. Nous rappelons aussi les différentes techniques de choix de paramètres de lissage et nous revisitons les problèmes de supports ainsi qu'une résolution des effets de bord dans le cas discret. Ensuite, nous détaillons la nouvelle méthode d'estimation de densités par les noyaux associés continus, lesquels englobent les noyaux continus (classiques). Nous définissons les noyaux associés continus et nous proposons la méthode mode-dispersion pour leur construction puis nous illustrons ceci sur les noyaux associés non-classiques de la littérature à savoir bêta et sa version étendue, gamma et son inverse, gaussien inverse et sa réciproque le noyau de Pareto ainsi que le noyau lognormal. Nous examinons par la suite les propriétés des estimateurs qui en sont issus plus précisément le biais, la variance et les erreurs quadratiques moyennes ponctuelles et intégrées. Puis, nous proposons un algorithme de réduction de biais que nous illustrons sur ces mêmes noyaux associés non-classiques. Des études par simulations sont faites sur trois types d'estimateurs à noyaux lognormaux. Par ailleurs, nous étudions les comportements asymptotiques des estimateurs de densité à noyaux associés continus. Nous montrons d'abord les consistances faibles et fortes ainsi que la normalité asymptotique ponctuelle. Ensuite nous présentons les résultats des consistances faibles et fortes globales en utilisant les normes uniformes et $\mathbb{L}^1$. Nous illustrons ceci sur trois types d'estimateurs à noyaux lognormaux. Par la suite, nous étudions les propriétés minimax des estimateurs à noyaux associés continus. Nous décrivons d'abord le modèle puis nous donnons les hypothèses techniques avec lesquelles nous travaillons. Nous présentons ensuite nos résultats minimax et enfin nous faisons une application sur les noyaux associés non-classiques bêta, gamma et lognormal. Enfin, nous combinons les noyaux associés continus et discrets pour définir les noyaux associés mixtes. De là, les outils d'unification d'analyses discrètes et continues sot utilisés, pour montrer les différentes propriétés des estimateurs à noyaux associés mixtes. Une application sur un modèle de mélange des lois normales et de Poisson tronquées est aussi donnée. Tout au long de ce travail, nous choisissons le paramètre de lissage uniquement avec la méthode de validation croisée par les moindres carrés.

Mots clés : Convergence, Densité mixte, Échelles de temps, Effet de bords, Estimation non-paramétrique par noyau, Modèle de mélange, noyau uni-modal, Paramètre de dispersion, Validation croisée.

Abstract

We present in this thesis, the non-parametric approach using mixed associated kernels for densities with supports being partially continuous and discrete.We first start by recalling the essential concepts of classical continuous and discrete kernel density estimators. We give the definition and characteristics of these estimators. We also recall the various technical for the choice of smoothing parameters and we revisit the problems of supports as well as a resolution of the edge effects in the discrete case. Then, we describe a new method of continuous associated kernels for estimating density with bounded support, which includes the classical continuous kernel method. We define the continuous associated kernels and we propose the mode-dispersion for their construction. Moreover, we illustrate this on the non-classical associated kernels of literature namely, beta and its extended version, gamma and its inverse, inverse Gaussian and its reciprocal, the Pareto kernel and the kernel lognormal. We subsequently examine the properties of the estimators which are derived, specifically, the bias, variance and the pointwise and integrated mean squared errors. Then, we propose an algorithm for reducing bias that we illustrate on these non-classical associated kernels. Some simulations studies are performed on three types of estimators lognormal kernels. Also, we study the asymptotic behavior of the continuous associated kernel estimators for density. We first show the pointwise weak and strong consistencies as well as the asymptotic normality. Then, we present the results of the global weak and strong consistencies using uniform and $\mathbb{L}^1$ norms. We illustrate this on three types of lognormal kernels estimators. Subsequently, we study the minimax properties of the continuous associated kernel estimators. We first describe the model and we give the technical assumptions with which we work. Then We present our results that we apply on some non-classical associated kernels more precisely beta, gamma and lognormal kernel estimators. Finally, we combine continuous and discrete associated kernels for defining the mixed associated kernels. Using the tools of the unification of discrete and continuous analysis, we show the different properties of the mixed associated kernel estimators. All through this work, we choose the smoothing parameter using the least squares cross-validation method.

Keywords : Asymmetric kernel, Boundary effect, Convergence, Cross-validation, Dispersion parameter, Mixed density, Mixture model, Nonparametric kernel estimation, Time-scales, Unimodal kernel.

Table des matières

Introduction générale **19**

1 Estimateurs à noyaux : continus classiques et associés discrets **21**

1.1 Estimateurs à noyaux continus classiques 22

1.1.1 Caractéristiques 22

1.1.2 Exemples des noyaux continus classiques 27

1.1.3 Choix de fenêtre 29

1.1.4 Problèmes de supports 31

1.2 Estimateurs à noyaux associés discrets 32

1.2.1 Caractéristiques 32

1.2.2 Exemples de noyaux associés discrets 36

1.2.3 Choix de fenêtre 38

1.2.4 Une résolution de problèmes de bords 42

2 Méthode des noyaux associés continus **47**

2.1 Noyaux associés continus 48

2.1.1 Type de noyau et noyaux associés 48

2.1.2 Méthode mode-dispersion 50

2.1.3 Noyaux associés non-classiques : bêta et son extension, gamma et son inverse, gaussien inverse et sa réciproque, Pareto puis log-normal 52

2.2 Estimateurs à noyaux associés continus 64

2.2.1 Caractéristiques 64

2.2.2 Réduction de biais et noyaux associés modifiés 70

2.2.3 Illustration sur les noyaux associés non-classiques 76

2.2.4 Problèmes de choix de fenêtre 88

2.3 Conclusion 93

3 Convergence des estimateurs à noyaux associés continus **95**

3.1 Consistance et normalité asymptotique ponctuelles 96
3.1.1 Consistance faible et forte 96
3.1.2 Normalité asymptotique 97
3.2 Nouvelles convergences globales 98
3.2.1 Convergence uniforme 99
3.2.2 Convergence en norme L^1 101
3.3 Étude des estimateurs à noyaux lognormaux 104

4 Propriétés minimax des estimateurs à noyaux associés 111
4.1 Description du modèle 112
4.1.1 Risque minimax 112
4.1.2 Hypothèses techniques 114
4.2 Résultats 115
4.2.1 Contrôle du risque 115
4.2.2 Illustrations 119

5 Estimateurs à noyaux associés mixtes et applications 125
5.1 Calculs sur des échelles de temps 128
5.1.1 Définitions et notations 128
5.1.2 Dérivabilité unifiée 129
5.2 Définitions et propriétés 132
5.2.1 Noyaux associés mixtes 132
5.2.2 Estimateur de densités mixtes 134
5.3 Etudes de biais et variances 135
5.4 Application au modèle de mélange non-paramétrique 138

Conclusion et perspectives 143

Bibliographie 153

Annexes 153

A Programme sous R 155
A.1 Noyaux associés continus classiques 155
A.2 Noyaux associés continus non-classiques : bêta étendu 157
A.3 Estimateurs à noyaux associés continus 158
A.3.1 Programme général d'estimation 158

A.3.2 Calcul général de Biais, Variance et MSE 159
A.3.3 Cas particuliers 160
A.3.4 Choix de h par la validation croisée 163
A.4 Programme de simulation avec réplication : Cas de ISE 165
A.5 Estimateurs à noyaux associés mixtes 166

Liste des tableaux

1.1 Quelques noyaux continus classiques 27
1.2 Quelques noyaux associés discrets sur $\mathbb{T}_N \subseteq \mathbb{Z}$ 38

2.1 Quelques types de noyaux (pour la suite voir Tableau 2.2) 60
2.2 Quelques types de noyaux (suite du Tableau 2.1) 61
2.3 Noyaux associés bêta et sa version étendue 62
2.4 Quelques noyaux associés non-classiques 63
2.5 Quelques valeurs de $\Lambda(n, h, K) \simeq 1$ de la Proposition 2.2.2 pour $h = 0.05$ et $n = 1000$ 67
2.6 Quelques noyaux associés modifiés (suite voir Tableau 2.7) 86
2.7 Quelques noyaux associés modifiés (suite du Tableau 2.6) 87
2.8 Valeurs optimales de h obtenues aux bords par la validation croisée par les moindres carrés 90
2.9 Valeurs optimales de h obtenues à l'intérieur par la validation croisée par les moindres carrés 90
2.10 Biais et variances des trois estimateurs à noyaux lognormaux suivant le Modèle 1 91
2.11 Biais et variances des trois estimateurs à noyaux lognormaux suivant le Modèle 2 92
2.12 La moyenne des ISEs pour les trois estimateurs de densité à noyaux lognormaux 92

3.1 Tableau récapitulatif des différents r dans les Théorème 3.1.1, Théorème 3.2.2 et Théorème 3.2.4. 109

5.1 Classification des points de $\mathbb{T}$ 128
5.2 Biais et variances de l'estimateur à noyaux associés mixtes de la densité définie en (5.33) 141

Table des figures

1.1 Quelques noyaux classiques 28
1.2 Différentes formes du noyau gaussien en x (effet cible) 33
1.3 Différentes formes du noyau gaussien en x (effet h) 34
1.4 Quelques noyaux associés discrets 37
1.5 Différentes formes du noyau discret triangulaire général. 44

2.1 Type de noyau bêta étendu sur $[2, 5]$ 54
2.2 Noyau associé bêta étendu avec h fixé 55
2.3 Noyau associé bêta étendu avec x fixé 55
2.4 Comportements de quelques noyaux associés non-classiques de la droite réelle positive aux bords. 65
2.5 Comportements de quelques noyaux associés non-classiques de la droite réelle positive à l'intérieur. 66
2.6 Comportements des estimateurs noyaux associés non-classiques de la droite réelle positive. 71
2.7 Comportements des MSE aux bords dans l'estmation de la densité normale tronquée. 72
2.8 Comportements des MSE aux bords dans l'estmation de la densité exponentielle. 73

5.1 Un exemple de $\mathbb{T}$. 125
5.2 Exemple d'une densité mixte. 126

Introduction générale

Dans de nombreux domaines (p. ex. médecine, économie, sociologie, environnement, finances, markéting) 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 $\mathbb{T} \subseteq \mathbb{R}$. 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), Agarwal & Bohner (1999), Bohner & Peterson (2001, 2003), Sanyal (2008) pour d'autres développements.

On considère f la densité *mixte* et inconnue de la v.a.r. X à valeurs dans $\mathbb{T} \subseteq \mathbb{R}$. Puisqu'il n'est pas courant d'avoir de manière générale des modèles paramétriques sur de tels supports, cette thèse propose l'approche non-paramétrique pour estimer f par la méthode des noyaux dits *associés* afin de respecter les différentes structures topologiques de $\mathbb{T}$. Par exemple, si on étudie un processus de suivi d'un malade à trois phases connues dont la première correspond à une période d'hospitalisation où le patient doit subir des soins intensifs avec des prélèvements quasi-instantanés (i.e. continus), la deuxième représente une phase de soins externes où les prélèvements sont périodiques ou séquentiels dans le temps (i.e. discrets) et enfin une dernière phase de soins intensifs à nouveau jusqu'à la disparition ; alors le support de ce processus peut s'écrire

$$\mathbb{T} = [t_0, t_1] \cup \{t_2, t_3, ..., t_k\} \cup]t_k, +\infty[\subseteq \mathbb{R}, \tag{1}$$

où $t_j \in [-\infty, +\infty]$ avec $t_j \leqslant t_{j+1}$. Des cas usuels de (1) sont $\mathbb{R}$, $\mathbb{N}$, $\{1, 2, ..., k\}$, $[0, 1]$, $[0, +\infty[$ et $[0, 1[\cup \mathbb{N}^*$.

L'objet de cette thèse est d'unifier les estimateurs à noyaux depuis les travaux de Rosenblatt (1956) et Parzen (1962) sur $\mathbb{T} = \mathbb{R}$ jusqu'à la thèse de Tristan Senga Kiessé en 2008 sur $\mathbb{T} \subseteq \mathbb{N}$, en passant par les travaux de Chen (1999, 2000) pour $\mathbb{T} = [0, 1]$ et $\mathbb{T} = [0, +\infty[$ respectivement. Sans perte de généralité, on s'intéressera en particulier à de densités mixtes f par rapport à la mesure de Lebesgue pour les parties continues et la mesure de comptage pour les parties discrètes. Cette thèse est composée de cinq chapitres et une conclusion générale portant sur des discussions et perspectives.

Le Chapitre 1 est composé de deux parties bien distinctes. La première rappelle les propriétés essentielles des estimateurs à noyaux *continus classiques* sur $\mathbb{T} = \mathbb{R}$. On souligne la moindre importance du choix de ces noyaux contrairement à celui de fenêtres dont les diverses techniques sont brièvement présentées. Les problèmes de différentes structures de supports sont alors posés. La seconde partie récapitule la solution de "lissage discret" sur $\mathbb{T} \subseteq \mathbb{N}$ par des noyaux *associés discrets*. Différentes

propriétés fondamentales sont revisitées ; en particulier, on propose une résolution pratique de problème de bord par des noyaux triangulaires discrets.

Le Chapitre 2 est consacré à la méthode des noyaux *associés continus* sur un intervalle $\mathbb{T}$ de $\mathbb{R}$. On y introduit les deux notions du *type de noyau* ainsi que celles de noyau associé continu. Par la suite, la technique "mode-dispersion" est proposée pour construire des noyaux associés continus *non-classiques* puis illustrée sur de nombreux types de noyaux. On élabore les propriétés caractéristiques des estimateurs de densité correspondants dont un algorithme de réduction de leurs biais. Des exemples sont donnés, et les problèmes de choix efficaces de fenêtre sont également évoqués.

Dans le Chapitre 3, on traite de nouvelles convergences des estimateurs définis au chapitre précédent. On commence d'abord par montrer trois types de convergences ponctuelles en probabilité, presque-sûre et en loi. Sous des hypothèses simples de continuité sur des compacts relatives à la densité à estimer, on démontre ensuite des consistances globales en normes uniforme et L^1. Des études théoriques et par simulation sont menées sur les estimateurs de densité à différents noyaux lognormaux, gamma, gaussien inverse et sa réciproque.

Le Chapitre 4 fournit les propriétés minimax des estimateurs de densités à noyaux associés continus. On décrit d'abord le modèle en précisant les classes des densités et les hypothèses techniques avec lesquelles on travaille. Ensuite on présente les principaux résultats dont une application est faite sur les estimateurs à noyaux associés classiques d'une manière générale et les noyaux associés non-classiques dans les cas particuliers dans bêta, gamma et lognormal.

Enfin, le Chapitre 5 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 $\mathbb{T} \subseteq \mathbb{R}$ (1), on introduit de nouveaux outils d'analyse sur ces échelles de temps. Des simulations et quelques applications concrètes terminent ce travail.

Chapitre 1

Estimateurs à noyaux : continus classiques et associés discrets

L'estimation d'une densité de probabilité est similaire à sa construction à partir des données observées. La première approche est celle de l'estimation paramétrique. L'idée principale de cette approche est de supposer que la densité f à estimer appartient à une famille de distributions ayant un nombre fini de paramètres (nous citons essentiellement la méthode du maximum de vraisemblance et la méthode des moments). Son avantage réside dans la simplicité algorithmique de sa mise en œuvre. Par opposition, l'approche non-paramétrique ne fait aucune hypothèse a priori sur l'appartenance de f à une famille de lois connues. L'estimation ne concerne donc plus un paramètre dans cette famille de loi, mais directement la fonction elle-même (d'où le terme de non paramétrique).

La plus ancienne dans cette approche est la méthode de l'histogramme. Cette méthode est la plus naturelle et très répandue car elle est facilement implémentable. Cependant, l'estimateur de densité fourni par un histogramme ne peut pas être adapté à la situation assez courante où l'on dispose d'une information a priori sur la régularité de la densité à estimer. Plus précisément, si l'on sait par avance que la densité de l'échantillon observé est, par exemple, deux fois continûment différentiable, on aurait naturellement envie d'estimer cette densité par une fonction qui, elle aussi, est deux fois continûment différentiable. C'est ainsi que la méthode d'estimation par les noyaux (continus) fut introduite par Rosenblatt (1956) puis améliorée par Parzen (1962). Elle est basée sur la convolution de la loi empirique par une fonction noyau prise sous forme d'une densité de probabilité. Elle est la plus fréquente et la plus utilisée pour estimer les densités de probabilités dans ces dernières décennies. La plus récente dans cette approche est celle des noyaux associés discrets, introduite dans la thèse de Senga Kiessé (2008) puis améliorée par Kokonendji & Senga Kiessé (2011) et qui tient compte de la nature discrète des observations.

L'objet de ce chapitre est de rappeler les différentes techniques d'estimation par les noyaux continus (classiques) ainsi que les noyaux associés discrets dans le cas univarié sur le modèle de densité puisque celui-ci est simple et permet de tester les différentes innovations statistiques sans rentrer dans des calculs très fastidieux. Nous présentons dans un premier temps, les estimateurs à noyaux continus (classiques)

avec leurs propriétés de base puis les différentes méthodes de choix de fenêtres. Dans un second temps, nous décrivons succinctement les noyaux associés discrets et les propriétés élémentaires des estimateurs définis à partir de ces derniers. Un accent sera mis sur le choix du paramètre de lissage et nous terminons le chapitre par des illustrations.

Tout au long de ce chapitre, on suppose que les observations $X_1, \cdots, X_n$ sont des variables indépendantes de même loi (iid) de densité f.

1.1 Estimateurs à noyaux continus classiques

Nous présentons, dans cette section, la méthode d'estimation de densité par noyaux continus (classique). Pour simplifier, on suppose que les X_i sont à valeurs réelles et que f est la densité par rapport à la mesure de Lebesgue sur $\mathbb{T} = \mathbb{R}$.

1.1.1 Caractéristiques

Fixons la définition d'un noyau continu (classique).

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 ($\mu_{\mathcal{K}} := \int_{\mathbb{S}} u\mathcal{K}(u)du = 0$), de variance $\sigma^2_{\mathcal{K}}$ finie ($\sigma^2_{\mathcal{K}} := \int_{\mathbb{S}} u^2\mathcal{K}(u)du < +\infty$) 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 $\mathbb{S}$, $\mathcal{K}(u) \geq 0$ 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 :*

$$\widehat{f_n}(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h_n}\right). \tag{1.1}$$

De ce fait, on peut considérer l'estimation par noyau en un point donné $x \in \mathbb{T}$ comme la moyenne des valeurs des fonctions noyaux centrés sur chaque observation X_i de l'échantillon.

Dans l'expression (1.1), le paramètre de lissage h_n a une influence assez évidente. Il est largement admis (et partiellement démontré) que le choix de h_n est d'une importance capitale dans le processus. Lorsqu'il tend vers 0, la fonction noyau s'apparente à la fonction Dirac et plus on a d'informations (sur lissage) ; tandis que, lorsque h_n devient grand, plus on manque d'informations (sous lissage). Ainsi, l'obtention d'une estimation plus proche de la réalité implique donc un choix optimal de h_n et ceci est l'objet de la Section 1.1.3 du présent chapitre. Dans toute la suite, sauf mention expresse du contraire, nous remplacerons h_n par h.

Erreur Quadratique Moyenne

Lorsqu'on définit l'estimateur à noyau, il y a un certain nombre de critères qui permettent d'évaluer la similarité de l'estimateur $\widehat{f_n}$ par rapport à la vraie densité f à estimer. Parmi les nombreux critères proposés dans la littérature, figure l'Erreur Quadratique Moyenne ponctuelle en anglais "pointwise Mean Squared Error" (MSE). Ainsi, pour l'estimateur $\widehat{f_n}$ définie en (1.1) on a :

$$MSE(x) = \mathbb{E}\left[\left\{\widehat{f_n}(x) - f(x)\right\}^2\right], \quad x \in \mathbb{T}. \tag{1.2}$$

Il est intéressant de développer cette expression pour y faire apparaitre la variance et le biais ponctuels de l'estimateur $\widehat{f_n}$ notés $Var\left\{\widehat{f_n}(x)\right\}$ et $Biais\left\{\widehat{f_n}(x)\right\}$ définis respectivement en (1.3) et (1.4) pour tout point $x \in \mathbb{T}$ par :

$$Var\left\{\widehat{f_n}(x)\right\} = \mathbb{E}\left(\left[\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}\right]^2\right), \tag{1.3}$$

$$Biais\left\{\widehat{f_n}(x)\right\} = \mathbb{E}\left\{\widehat{f_n}(x)\right\} - f(x). \tag{1.4}$$

Ceci conduit à une nouvelle expression du $MSE(x)$ comme suit

$$\begin{aligned} MSE(x) &= \mathbb{E}\left[\left\{\widehat{f_n}(x)\right\}^2\right] - 2\mathbb{E}\left\{\widehat{f_n}(x)\right\} f(x) + \left[\mathbb{E}\left\{f(x)\right\}\right]^2, \\ &= \mathbb{E}\left[\left\{\widehat{f_n}(x)\right\}^2\right] - \left[\mathbb{E}\left\{\widehat{f_n}(x)\right\}\right]^2 + \left[\mathbb{E}\left\{\widehat{f_n}(x)\right\}\right]^2 - 2f(x)\mathbb{E}\left\{\widehat{f_n}(x)\right\} + f^2(x), \\ &= Var\left\{\widehat{f_n}(x)\right\} + Biais^2\left\{\widehat{f_n}(x)\right\}. \end{aligned} \tag{1.5}$$

Cette dernière expression montre bien le compromis pour la minimisation du MSE entre le biais (erreur systématique) et la variance (erreur aléatoire). A cause des termes $\mathbb{E}\left\{\widehat{f_n}(x)\right\}$ dans (1.3) et (1.4), on voit qu'une réduction de biais entraîne une augmentation de la variance et vice versa. Cependant, on peut constater que l'espérance et la variance de l'estimateur $\widehat{f_n}$ en un point x de $\mathbb{T}$ peuvent être exprimées en fonction du noyau classique K et de la densité à estimer f comme suit :

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \frac{1}{h}\mathbb{E}\left\{\mathcal{K}\left(\frac{x-t}{h}\right)\right\} = \frac{1}{h}\int_{\mathbb{T}} \mathcal{K}\left(\frac{x-t}{h}\right) f(t)dt \tag{1.6}$$

et

$$\begin{aligned} Var\left\{\widehat{f_n}(x)\right\} &= \frac{1}{nh^2} Var\left\{\sum_{i=1}^{n} \mathcal{K}\left(\frac{x-X_i}{h}\right)\right\} = \frac{1}{nh^2}\left[\mathbb{E}\left\{\mathcal{K}^2\left(\frac{x-X_1}{h}\right)\right\} - \mathbb{E}^2\left\{\mathcal{K}\left(\frac{x-X_1}{h}\right)\right\}\right], \\ &= \frac{1}{nh^2}\int_{\mathbb{T}} \mathcal{K}^2\left(\frac{x-t}{h}\right) f(t)dt - \frac{1}{nh^2}\mathbb{E}^2\left\{\mathcal{K}\left(\frac{x-X_1}{h}\right)\right\}. \end{aligned} \tag{1.7}$$

Il découle de (1.6) que le biais ne dépend pas directement de la taille de l'échantillon, mais plutôt du noyau. Par conséquent, l'obtention d'une estimation asymptotiquement sans biais implique un ajustement du noyau $\mathcal{K}$ et de paramètre de lissage h.

Une mesure, globale, de l'efficacité de l'estimateur $\widehat{f_n}$ est obtenue en intégrant le *MSE* sur tout le support $\mathbb{T}$ de f. Il s'agit de "l'Erreur Quadratique Moyenne Intégrée" en anglais "Means Integrated Squared Error" (MISE). En utilisant (1.5), elle s'écrit :

$$\begin{aligned} MISE(n,h,\mathcal{K},f) &= \int_{\mathbb{T}} MSE(x)dx \\ &= \int_{\mathbb{T}} Var\{\widehat{f_n}(x)\}dx + \int_{\mathbb{T}} Biais^2\{\widehat{f_n}(x)\}dx. \end{aligned} \tag{1.8}$$

Il est possible de reporter dans (1.8), les valeurs de la variance et du biais définies dans (1.3) et (1.4), mais cela va rendre les calculs lourds et conduit à des résultats rarement exploitables. D'où l'intérêt de faire quelques approximations et hypothèses supplémentaires pour garder les résultats généraux. Ainsi, en dépit des conditions vérifiées par le noyau classique dans la Définition 1.1.1, nous supposons dans la suite que la densité à estimer f admette des dérivées de tous ordres.

En reprenant l'expression de l'espérance de l'estimateur $\widehat{f_n}$ dans (1.6) et la variance dans (1.7) puis en posant $t = x - hu$, on a :

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \int_{\mathbb{T}} \mathcal{K}(u)\, f(x-hu)du \tag{1.9}$$

et

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{nh^2}\int_{\mathbb{T}} \mathcal{K}^2(u)\, f(x-hu)dt - \frac{1}{nh^2}\mathbb{E}^2\left\{\mathcal{K}\left(\frac{x-X_1}{h}\right)\right\}. \tag{1.10}$$

Le développement en séries de Taylor de $f(x-hu)$ au voisinage de x est alors

$$f(x-hu) = f(x) - huf'(x) + \frac{1}{2}h^2u^2f''(x) + o\left\{(hu)^2\right\}.$$

En l'injectant dans (1.9) et (1.10) on obtient alors :

$$\begin{aligned} \mathbb{E}\left\{\widehat{f_n}(x)\right\} &= \int_{\mathbb{T}} \mathcal{K}(u)\left\{f(x) - huf'(x) + \frac{1}{2}h^2u^2f''(x)\right\}dx + o\left\{(hu)^2\right\}, \\ &= f(x) + \frac{1}{2}h^2f''(x)\sigma^2_{\mathcal{K}} + o(h^2) \end{aligned} \tag{1.11}$$

et

$$\begin{aligned} Var\left\{\widehat{f_n}(x)\right\} &= \frac{1}{nh}\int_{\mathbb{T}} \mathcal{K}^2(u)\, f(x)du + R_n + o\left(\frac{1}{nh}\right), \\ &= \frac{1}{nh}f(x)\int_{\mathbb{T}} \mathcal{K}^2(u)\, du + o\left(\frac{1}{nh}\right), \end{aligned} \tag{1.12}$$

avec $R_n = (1/nh^2)\left(-huf'(x) + (1/2)h^2u^2f''(x) - \mathbb{E}^2\left[\mathcal{K}\left\{(x-X_1)/h\right\}\right]\right) \simeq o\left\{1/(nh)\right\}$. De ces deux derniers résultats (1.12) et (1.11), on peut déduire les formes approximées et asymptotiques du *MSE* et du *MISE* notées respectivement *AMSE* et *AMISE* par :

$$AMSE(x) = \frac{1}{nh}f(x)\int_{\mathbb{T}} \mathcal{K}^2(u)du + \frac{1}{2}h^4\sigma^4_{\mathcal{K}}f''^2(x) \tag{1.13}$$

et

$$AMISE(n,h,\mathcal{K},f) = \frac{1}{nh}\int_{\mathbb{T}} \mathcal{K}^2(u)du + \frac{1}{4}h^4\sigma_{\mathcal{K}}^4 \int_{\mathbb{T}} f''^2(x)dx. \tag{1.14}$$

Parmi toutes les qualités que peut avoir un estimateur (ou une suite d'estimateurs), on s'intéresse souvent à sa consistance, i.e., au fait qu'une suite d'estimateurs $\widehat{f_n}$ converge ou non vers f en une distance $d\left(\widehat{f_n}, f\right)$ donnée. Le paragraphe suivant nous donne quelques résultats de convergence des estimateurs à noyaux classiques de la littérature.

Convergence des estimateurs à noyaux classiques

Rappelons d'abord les hypothèses (sur le noyau et la fenêtre de lissage) par lesquelles ces résultats ont été établis.

$$\int_{\mathbb{T}} \mathcal{K}(u)du = 1, \sup_{u\in\mathbb{T}}|\mathcal{K}(u)| < +\infty \text{ et } \int_{\mathbb{T}} |\mathcal{K}(u)|\, du < +\infty. \tag{1.15}$$

$$h \to 0 \quad \text{quand} \quad n \to +\infty \tag{1.16}$$

$$nh \to +\infty \quad \text{quand} \quad n \to +\infty. \tag{1.17}$$

$$nh^2 \to +\infty \quad \text{quand} \quad n \to +\infty. \tag{1.18}$$

$$\frac{nh}{\log n} \to +\infty \quad \text{quand} \quad n \to +\infty. \tag{1.19}$$

Convergence en moyenne quadratique ponctuelle et intégrée

Ces résultats sont établis dans les travaux respectivement de Parzen (1962) et Tiago de Oliviera (1963). Le premier montre la convergence du *MSE* et le second donne celle du *MISE*.

Théorème 1.1.3 *(Parzen [1962]). Soit f, une densité continue sur $\mathbb{T} = \mathbb{R}$ et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ vérifiant (1.15). Si le paramètre h satisfait (1.16) et (1.17) alors*

$$MSE\left\{\widehat{f_n}(x)\right\} \xrightarrow{\mathbb{P}} 0, \quad \forall x \in \mathbb{T} = \mathbb{R},$$

où $\xrightarrow{\mathbb{P}}$ désigne la convergence en probabilité.

Théorème 1.1.4 *(Tiago de Oliviera [1963]). Soit f une densité de puissance p^{eme}-intégrable et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ satisfaisant (1.15). Si h vérifie (1.16) et (1.17) alors*

$$MISE\left(\widehat{f_n}\right) \xrightarrow{\mathbb{P}} 0.$$

Consistances faible et forte

Les trois résultats donnés ici sont issus des travaux de Parzen (1962), Nadaraya (1965), et Silverman (1986). Les deux premiers montrent la consistance faible tandis que le dernier propose quant à lui la consistance forte.

Théorème 1.1.5 *(Parzen [1962]). Soit f, la densité à estimer et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ satisfaisant (1.15). Si h vérifie (1.18) et de plus si la transformée de Fourier $\mathcal{TF}(z) := \int \exp(-izu)\mathcal{K}(u)du$ est absolument intégrable, alors*

$$\sup_{x\in\mathbb{T}=\mathbb{R}} \left|\widehat{f_n}(x) - f(x)\right| \xrightarrow{\mathbb{P}} 0.$$

Théorème 1.1.6 *(Nadaraya [1965]). Soit f, une densité uniformément continue et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ positif et à variations bornées. Pour tout h vérifiant (1.16) tel que $\sum_{k\geqslant 1} \exp(-\varepsilon n h^2) < +\infty$, alors*

$$\sup_{x\in\mathbb{T}=\mathbb{R}} \left|\widehat{f_n}(x) - f(x)\right| \xrightarrow{\mathbb{P}} 0.$$

Théorème 1.1.7 *(Silverman [1986]). Soit f, une densité uniformément continue et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ positif et à variations bornées. Pour tout h satisfaisant (1.16) et (1.19), alors*

$$\sup_{x\in\mathbb{T}=\mathbb{R}} \left|\widehat{f_n}(x) - f(x)\right| \xrightarrow{p.s} 0,$$

où $\xrightarrow{p.s}$ est la convergence presque-sûre.

Normalité asymptotique

Ce dernier résultat est tiré des travaux de Parzen (1962). Il montre que l'estimateur à noyau est asymptotiquement normal.

Théorème 1.1.8 *(Parzen [1962]). Soit f, une densité continue sur $\mathbb{T} = \mathbb{R}$ et $\widehat{f_n}$ son estimateur à noyau (classique) $\mathcal{K}$ vérifiant (1.15). Si h satisfait (1.16) et (1.17) alors*

$$\frac{\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}}{\sqrt{Var\left\{\widehat{f_n}(x)\right\}}} \xrightarrow{\mathcal{L}} N(0,1), \quad \forall x \in \mathbb{T} = \mathbb{R},$$

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

1.1.2 Exemples des noyaux continus classiques

Nous résumons dans le Tableau 1.1, quelques exemples de noyaux classiques de la littérature. Le lecteur peut éventuellement se référer à Scott (1977), Epanechnikov (1969), Tsybakov (2004), et Silverman (1986) pour plus de détails. Tous ces noyaux vérifient la Définition 1.1.1 et sont en général de support $[-1,1]$ ou $\mathbb{R}$. Nous avons donné aussi dans ce tableau la famille des noyaux définie par :

$$F_p(x) = \left\{2^{2p+1}B(p+1,p)\right\}^{-1}\left(1-x^2\right)^p, \tag{1.20}$$

où $B(p+1,p)$ est la fonction bêta d'Euler et p est un entier naturel.

Table 1.1 – Quelques noyaux continus classiques

Noyaux	Support	Densité	Efficacité
Epanechnikov ($p=1$)	$[-1,1]$	$\frac{3}{4}(1-x^2)$	1
Cosinus	$[-1,1]$	$\frac{\pi}{4}cos(\frac{\pi}{2}x)$	$0,999$
Biweight ($p=2$)	$[-1,1]$	$\frac{15}{16}(1-x^2)^2$	$0,994$
Triweight ($p=3$)	$[-1,1]$	$\frac{35}{32}(1-x^2)^3$	$0,987$
Triangulaire	$[-1,1]$	$1-\lvert x\rvert$	$0,986$
Gaussien	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)$	$0,946$
Unifiorme ($p=0$)	$[-1,1]$	$\frac{1}{2}$	$0,930$
Double Epanechnikov	$[-1,1]$	$3\lvert x\rvert(1-\lvert x\rvert)$	$0,816$
Double Exponentielle	$\mathbb{R}$	$\frac{1}{2}\exp\left\{-\frac{1}{2}\lvert x\rvert\right\}$	$0,759$
Famille des lois K_p	$[-1,1]$	$F_p(x)$: Voir (1.20)	–

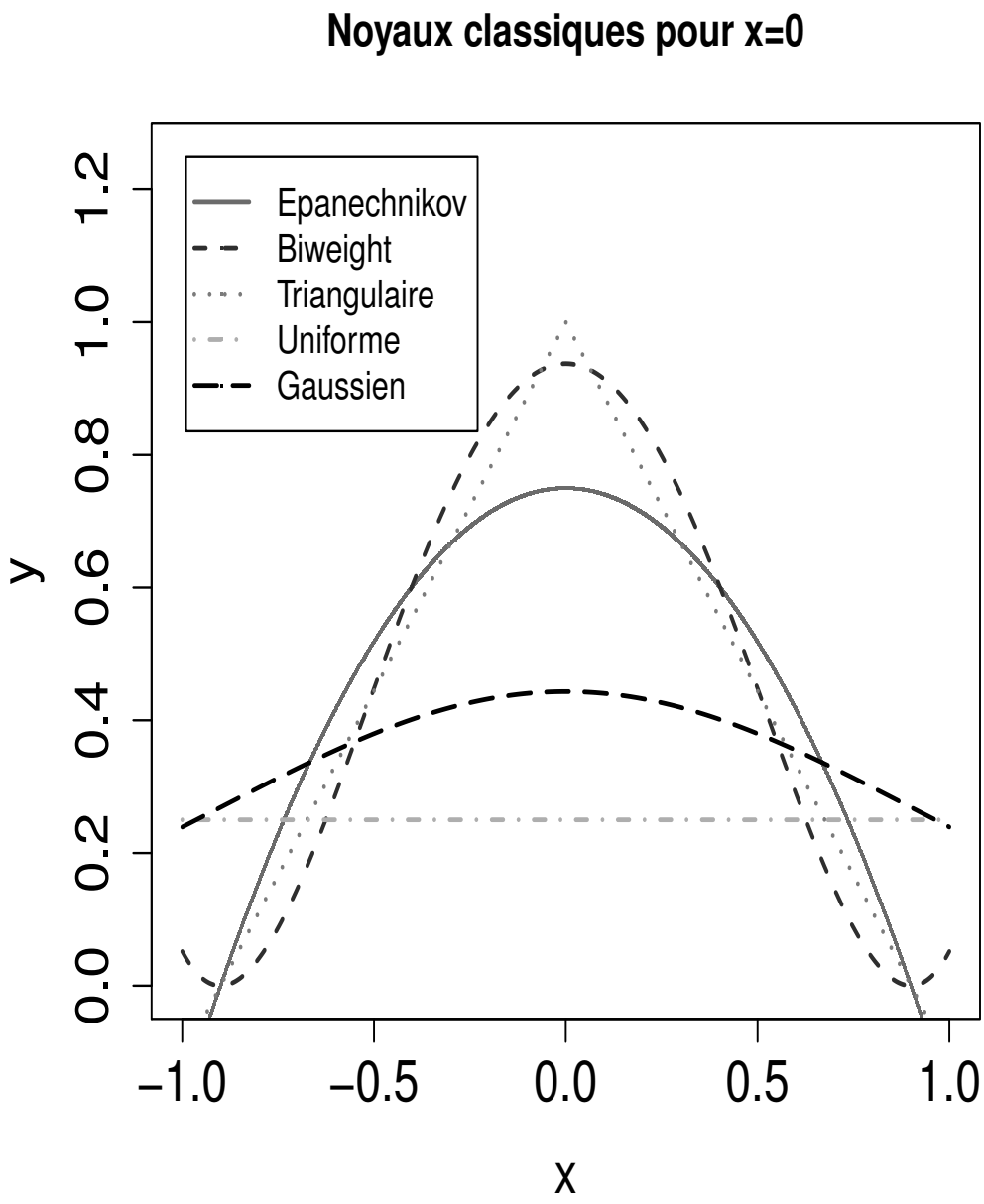

FIGURE 1.1 – Quelques noyaux classiques

1.1.3 Choix de fenêtre

Comme nous l'avons précédemment signalé, pour avoir une bonne estimation par la méthode des noyaux, il faut bien choisir le paramètre de lissage h puisque celui-ci a un rôle crucial dans le processus. Lorsque la fenêtre h est très petite, l'estimateur est très volatile et on parle de sous-lissage ("under-smoothing" en anglais). En revanche, lorsque h grandit, l'estimateur est alors de moins en moins influencé par les données. On parle d'un effet de sur-lissage ("over-smoothing" en anglais). En pratique, il est primordial de trouver la bonne dose de lissage qui permet d'éviter le sous-lissage et le sur-lissage. Les méthodes existantes pour le choix de h peuvent être classées en deux catégories.

La première catégorie est constitué des méthodes purement théoriques qui sont basées sur la minimisation de l'erreur quadratique moyenne intégrée (MISE). En effet, la valeur idéale théorique de h notée h_{id} s'obtient en minimisant le MISE asymptotique donné en (1.14). Ainsi, pour un échantillon de taille n donné et pour un noyau (classique) $\mathcal{K}$ fixé, cette valeur idéale de h est donnée par

$$h_{id} = \frac{1}{n^{1/5}} \left\{ \frac{\int_{\mathbb{T}} \mathcal{K}^2(t)dt}{\sigma_{\mathcal{K}}^4 \int_{\mathbb{T}} (f'')^2 (x)dx} \right\}^{1/5}. \tag{1.21}$$

Ce paramètre de lissage idéal h_{id} obtenu n'est pas directement utilisable puisqu'il dépend encore de la quantité inconnue $(f'')^2 (x)$.

La deuxième catégorie est celle dite des méthodes pratiques. Elle est intéressante puisqu'elle se laisse seulement guider par les observations. Elles ont été sujets des travaux de nombreux auteurs parmi lesquels nous pouvons citer Scott et al. (1977), Rudemo (1982), Stone (1984), Bowman (1984), Marron (1987), Berlinet & Devroye (1989), Park & Marron (1992), Sarda & Vieu (1991), Cuevas et al. (1994) ainsi que Yondjé et al. (1996a, 1996b).

Dans ce qui suit, nous allons décrire deux de ces méthodes pratiques à savoir méthode de ré-injection ("Plug-in" en anglais), la méthode de validation croisée par les moindres carrées ("Least Squares Cross Validation" en anglais).

Méthode Plug-in

Il s'agit ici d'estimer la quantité $\int_{\mathbb{R}} (f'')^2 (x)dx$ dans l'expression de h_{id} donnée en (1.21). Plusieurs approches ont été proposées dans la littérature mais nous en retenons deux. La première consiste à supposer que f appartient à une famille de lois paramétriques puis à estimer ces paramètres afin d'identifier $\int_{\mathbb{T}} (f'')^2 (x)dx$. L'autre est totalement non paramétrique et fait appel à un estimateur à noyau de $\int_{\mathbb{T}} (f'')^2 (x)dx$.

Pour la première approche, en supposant que f appartienne à la famille gaussienne centrée et de variance σ^2 on trouve :

$$\int_{\mathbb{T}} (f'')^2 (x)dx = \frac{3}{8\sqrt{\pi}} \sigma^{1/5} \approx 0.212\sigma^{1/5}.$$

La valeur optimale de h notée h_{opt} est obtenue en remplaçant σ dans l'expression de h_{id} par son estimateur $\widehat{\sigma} = \sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)}$ avec $\overline{X} = \sum_{i=1}^{n} X_i/n$. Ce qui conduit à :

$$h_{opt} = 1.06\left(\frac{\widehat{\sigma}}{n^{1/5}}\right).$$

Cette approche donne de bons résultats lorsque la population est réellement normalement distribuée. Toutefois, elle peut aussi donner une distribution trop lissée dans le cas d'une population multimodale ; voir Silverman (1986).

La seconde approche est celle de Sheather & Jones (1991) appelée communément plug-in à trois étapes. Elle consiste à remplacer $\int_{\mathbb{T}} f''^2(x)dx$ dans (1.14) par son estimateur

$$\widehat{R}_a = \int_{\mathbb{T}} \left(\widehat{f''}\right)^2(x)dx = \frac{1}{n^2a^5}\sum_{i,j=1}^{n} L^{(4)}\left(\frac{X_i - X_j}{a}\right),$$

où $L^{(4)}$ désigne la dérivée quatrième du noyau suffisamment lisse L et a est un nouveau paramètre de lissage appelé paramètre pilote. Cet estimateur est obtenu en remarquant que sous des conditions de régularités suffisantes

$$\int_{\mathbb{T}} (f'')^2(x)dx = \int_{\mathbb{R}} f^{(4)}(x)f(x)dx.$$

Le nouveau paramètre de lissage $\widehat{a}$ minimisant la quantité $\mathbb{E}\left[\left\{\widehat{T}_a - \int_{\mathbb{T}} (f'')^2(x)dx\right\}^2\right]$ admet la représentation asymptotique suivante :

$$\widehat{a} = \left\{\frac{2L^{(4)}(0)}{\sigma_L^2 \int_{\mathbb{T}} (f'')^2(x)dx}\right\}^{1/7} n^{-1/7}.$$

Une fois de plus, $\widehat{a}$ dépend de la quantité inconnue $\int_{\mathbb{R}} f''^2(x)dx$ qu'on estime de nouveau par

$$\widehat{R}_b = \int_{\mathbb{T}} \widehat{f''}^2(x)dx = \frac{1}{n^2b^7}\sum_{i,j=1}^{n} L^{(6)}\left(\frac{X_i - X_j}{b}\right),$$

avec $b = 0.912\widehat{\lambda}n^{-1/9}$ où $\widehat{\lambda}$ est l'estimateur de λ qui représente une mesure d'échelle de f (par exemple son écart inter-quartile).

La mise en œuvre de ces différentes approches de plug-in fait usage des fonctions nommées respectivement bw.nrd0 et bw.SJ disponible dans R Development Core Team (2012).

Méthode de la validation croisée par les moindres carrés

L'idée de base des méthodes de validation croisée consiste à trouver une fonction de score $CV(h)$ dont le calcul est plus simple que $MISE(h)$. Le paramètre de lissage h sélectionné ici n'est pas déterministe mais dépend plutôt des observations. Pour plus

de détails, on peut se référer à de nombreux auteurs tels que Bowman (1984), Marron (1987), Rudemo (1982), Stone (1984). L'approche la plus utilisée est celle de validation croisée par les moindres carrés. Elle consiste à minimiser une estimateur convenable de $ISE(h)$ défini par

$$ISE(h) = \int_{\mathbb{T}} \left\{\widehat{f_n}(x) - f(x)\right\}^2 dx = \int_{\mathbb{T}} \widehat{f_n}^2(x)dx - 2\int_{\mathbb{T}} \widehat{f_n}(x)f(x)dx + \int_{\mathbb{T}} f^2(x)dx.$$

Compte tenu du fait que la quantité $\int_{\mathbb{T}} f^2(x)dx$ ne dépend pas de h, on choisit le paramètre de lissage de façon à ce qu'il minimise un estimateur de $CV(h)$ définie par

$$CV(h) = \int_{\mathbb{T}} \left\{\widehat{f_n}(x)\right\}^2 dx - \frac{2}{n}\sum_{i=1}^{n} \widehat{f_{n,-i}}(X_i).$$

où $\widehat{f_{n,-i}}(X_i) = \dfrac{1}{(n-1)h}\sum_{j\neq i} K\left(\dfrac{x - X_j}{h}\right)$ est l'estimateur calculé à partir de l'échantillon privé de l'observation X_i. Et la valeur optimale de h notée h_{CV} est donnée par

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

1.1.4 Problèmes de supports

Malgré leur robustesse sur les modèles les plus généraux, les estimateurs à noyaux (classiques) connaissent quelques inconvénients parmi lesquels se situent les problèmes de support. Le premier cas est celui des effets de bord dans l'estimation des densités à support borné au moins d'un côté. Ceci est dû au fait que le noyau classique étant symétrique (voir la Définition 1.1.1) assigne des poids à l'extérieur du support de la densité à estimer lorsque le lissage se fait aux points de bord. Cela crée ainsi les biais dits de bord qui rendent l'estimateur non consistant. Plusieurs solutions à ce problème ont été proposées dans la littérature.

Schuster (1985), Silverman (1986) et Cline & Hart (1991) ont fait usage de la méthode dite des données reflétées ("reflection data method" en anglais). Cette méthode est spécialement conçue pour le cas des densités vérifiant $f^{(1)}(0) = 0$ et $f^{(1)}(t_1) = 0$, où $f^{(1)}$ désigne la dérivée première de f définie sur un intervalle du type $[0, t_1]$. Ensuite, Gasser & Müler (1979), Gasser et al. (1985), Jones (1993), Zhang & Karunamuni (2000) et Karunamuni & Albert (2005) ont suggéré l'utilisation de la méthode des noyaux de bord ("boundary kernels method" en anglais). Cette méthode est plus générale que celle des données reflétées dans le sens où elle peut s'adapter à n'importe quelle forme de la densité. Cependant, un inconvénient de cette méthode est que les estimations faites pourraient être négatives à proximité des extrémités, surtout quand $f(0) \simeq 0$ ou $f(t_1) \simeq 0$. Pour corriger cette déficience de la méthode des noyaux de bord, certains remèdes ont été proposés ; se référer à Jones (1993), Jones & Foster (1996), Gasser & Müler (1979), Karunamuni et al. (2006) ainsi que Karunamuni & Zhang (2008) pour un approfondissement.

Toujours dans l'optique de résoudre le problème des effets de bord, Cowling & Hall (1996) ont proposé la méthode des pseudo-données ("pseudo-data method" en anglais). L'idée de cette méthode consiste à générer des observations supplémentaires $X_{(i)}$ à l 'aide de ce qu'ils appellent les "trois points pour mieux régner", puis les combiner avec les observations d'origine X_i pour former un type d'estimateur à noyau. Dans le même ordre d'idée, Marron et Ruppert (1994) ont proposé la méthode de transformation ("transformation method" en anglais). Elle est constituée de trois étapes. Tout d'abord, une transformation g est choisie dans une famille paramétrique de telle sorte que la densité des variables $Y_i = g(X_i)$ ait une dérivée première qui est approximativement égale à 0 aux bords. Ensuite, un estimateur à noyau de la réflexion est appliquée à Y_i. Enfin, cet estimateur est converti par la formule de changement de variables pour obtenir une estimation de f.

Notons aussi que d'autres auteurs comme Bouezmarni et al. (2005) ainsi que Gustafson et al. (2007) ont abordé le même sujet en combinant parfois quelques-unes de ces méthodes.

Récemment, certains auteurs ont proposé, dans le cas des densités à support compact, l'utilisation de noyaux dont le support coïncide avec celui de la densité à estimer. Ceci a efficacement résolu le problème des effets de bord puisque les noyaux utilisés ici sont généralement asymétriques et peuvent changer de forme selon la position du point d'estimation. C'est notamment le cas de Chen (1999, 2000) avec les noyaux bêta et gamma pour estimer les densités à support respectivement $[0, 1]$ et $[0, +\infty[$ puis Scaillet (2004) avec les noyaux inverses gaussien et sa réciproque pour les densités à support $]0, +\infty[$.

Dans cette même optique, Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011) ont proposé l'utilisation des noyaux associés discrets pour le lissage des données catégorielles ou de dénombrement ; puisque jusqu'alors, certains auteurs comme Simonof (1996) ainsi que Simonof & Tutz (2000) dans leur tentative d'estimation des données discrètes font usage des noyaux continus. Nous présentons dans le prochain paragraphe la méthode d'estimation par noyaux associés discrets.

1.2 Estimateurs à noyaux associés discrets

Cette section présente l'essentiel des travaux de Aitchison & Aitken (1976), Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011) sur les noyaux associés discrets. Nous désignons par $\mathbb{T} := \mathbb{T}_N \subseteq \mathbb{N}$ le support de la fonction de masse de probabilité ($f.m.p.$) à estimer puis par $\mathbb{S}_{x,h} := \mathbb{S}_x \subseteq \mathbb{N}$ celui du noyau associé.

1.2.1 Caractéristiques

Nous présentons, successivement, les définitions améliorées des types de noyaux et noyaux associés discrets de Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011). Nous rappelons ensuite les estimateurs à noyaux associés discrets. Enfin, examinons quelques unes de leurs propriétés importantes pour la suite du présent travail.

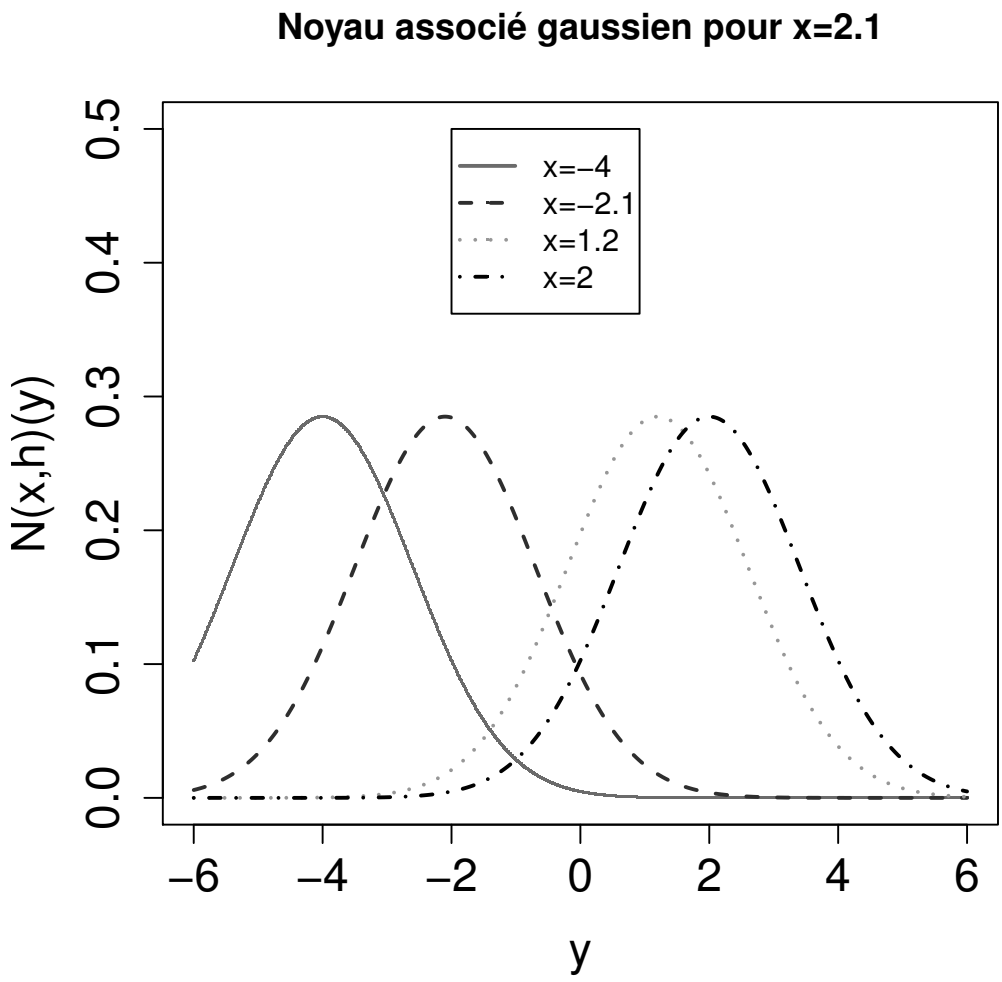

FIGURE 1.2 – Différentes formes du noyau gaussien en x (effet cible)

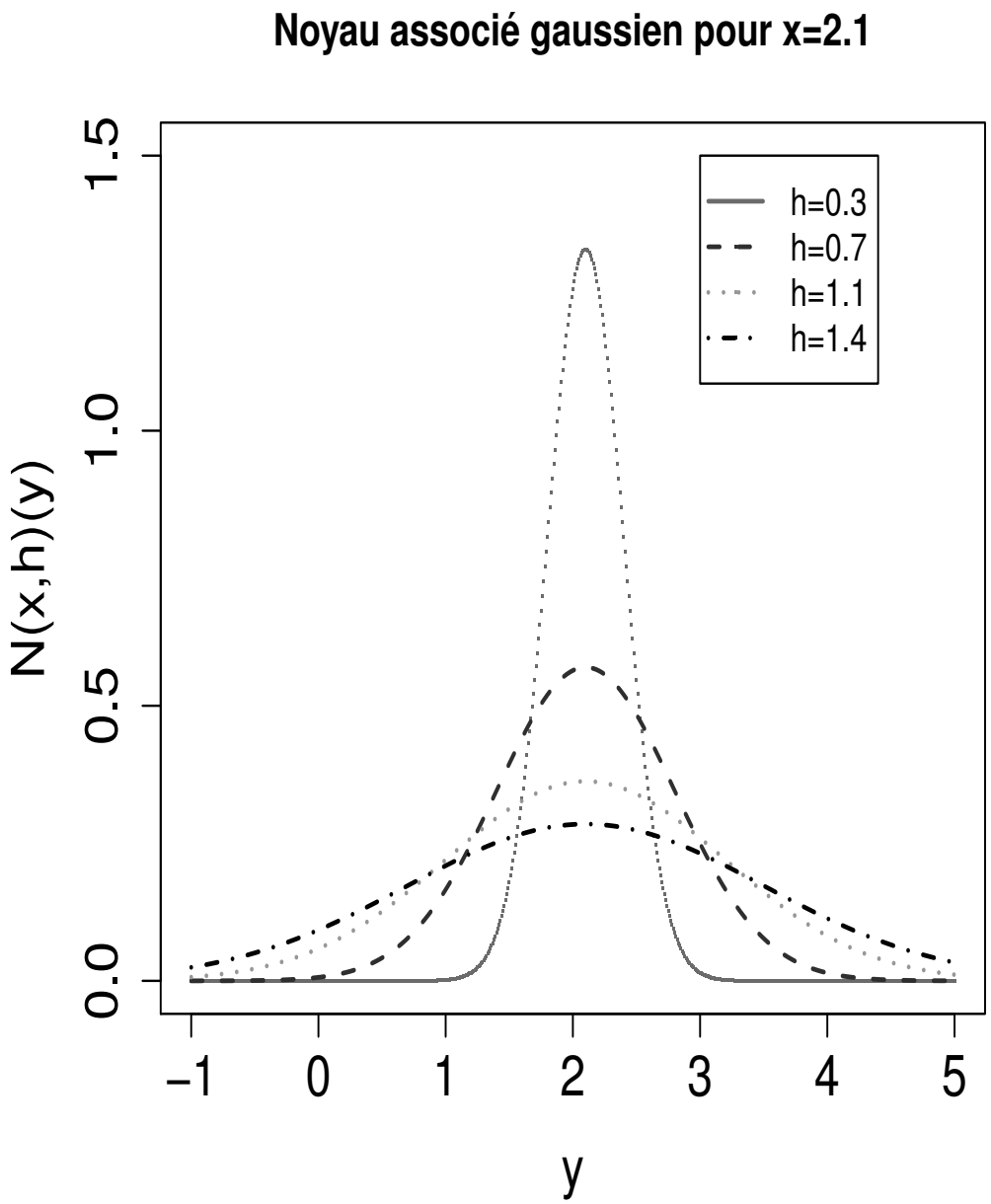

FIGURE 1.3 – Différentes formes du noyau gaussien en x (effet h)

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.*

Dans le reste de cette section, nous ne considérerons que les types de noyaux discrets uni-modaux pour la construction des noyaux associés discrets. Donnons maintenant la définition améliorée du noyau associé discret.

Définition 1.2.2 *Soit $x \in \mathbb{T}_N \subseteq \mathbb{Z}$ et $h > 0$ avec $\mathbb{T}_N$ 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 :*

$$x \in \mathbb{S}_{x,h} \tag{1.22}$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + A(x,h) \tag{1.23}$$

$$Var(\mathcal{Z}_{x,h}) = B(x,h) \tag{1.24}$$

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

Notons que, lorsque $B(x,h) \nrightarrow 0$, 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. La définition suivante présente l'estimateur à noyau associé discret.

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

$$\widehat{f_n}(x) = \frac{1}{n}\sum\nolimits_{i=1}^{n} K_{x,h}(X_i) \tag{1.25}$$

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

Nous rappelerons dans le prochain paragraphe quelques propriétés élémentaires de cet estimateur. Le lecteur peut se référer éventuellement à Senga Kiessé (2008), Abdous & Kokonendji (2009) et Kokonendji & Senga Kiessé (2011) pour les démonstrations et exemples détaillés. Dans le reste de ce paragraphe, $\mathcal{Z}_{x,h}$ désigne une variable aléatoire discrètes de loi $K_{x,h}$.

Proposition 1.2.4 *(Senga Kiessé [2008]). Soit f la fmp à estimer sur $\mathbb{T}_N \subseteq \mathbb{Z}$ et $\widehat{f_n}$ son estimateur à noyau associé discret en (1.25). Pour tout $x \in \mathbb{T}_N$ et $h > 0$, on a respectivement*

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \mathbb{E}\left\{f\left(\mathcal{Z}_{x,h}\right)\right\}$$

$$\widehat{f_n}(x) \in [0,1]$$

$$\sum\nolimits_{x\in\mathbb{T}_N} \widehat{f_n}(x) = C$$

où C est une constante strictement positive et finie.

La proposition suivante montre que les estimateurs à noyaux associés discrets sont asymptotiquement sans biais en chaque point $x \in \mathbb{T}_N$.

Proposition 1.2.5 *(Kokonendji & Senga Kiessé [2011]). Soit f la fmp à estimer sur $\mathbb{T}_N \subseteq \mathbb{Z}$ et $\widehat{f_n}$ son estimateur à noyau associé discret en (1.25). Pour tout $x \in \mathbb{T}_N$ et $h > 0$, si $h \to 0$ quand $n \to +\infty$ alors*

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \sum\nolimits_{y \in \mathbb{T}_N} f(y) K_{x,h}(y) \to f(x), \quad \textit{quand} \quad n \to +\infty.$$

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

Proposition 1.2.6 *(Abdous & Kokonendji [2009]). Soit f la fmp à estimer sur $\mathbb{T}_N \subseteq \mathbb{Z}$ et $\widehat{f_n}$ son estimateur à noyau associé discret en (1.25). Pour tout $x \in \mathbb{T}_N$ et sous les conditions (1.23) et (1.24) on a*

$$\mathbb{E}\left\{\widehat{f_n}(x) - f(x)\right\}^2 \xrightarrow{\mathbb{P}} 0, \quad \textit{quand} \quad n \to +\infty.$$

où $\xrightarrow{\mathbb{P}}$ est la convergence en probabilité.

Proposition 1.2.7 *(Abdous & Kokonendji [2009]). Soit f la fmp à estimer sur $\mathbb{T}_N \subseteq \mathbb{Z}$ et $\widehat{f_n}$ son estimateur à noyau associé discret en (1.25). Pour tout $x \in \mathbb{T}_N$ et sous les conditions (1.23) et (1.24) on a*

$$\widehat{f_n}(x) \xrightarrow{\mathbb{P}} f(x), \quad \textit{quand} \quad n \to +\infty.$$

où $\xrightarrow{P.S.}$ désigne la convergence presque sûre.

Proposition 1.2.8 *(Abdous & Kokonendji [2009]). Soit f la fmp à estimer sur $\mathbb{T}_N \subseteq \mathbb{Z}$ et $\widehat{f_n}$ son estimateur à noyau associé discret en (1.25). Pour tout $x \in \mathbb{T}_N$ et sous les conditions (1.23) et (1.24) on a*

$$\frac{\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}}{\left[Var\left\{\widehat{f_n}(x)\right\}\right]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \quad \textit{quand} \quad n \to +\infty.$$

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

1.2.2 Exemples de noyaux associés discrets

Nous présentons dans le tableau suivant quelques noyaux associés discrets de la littérature. Le lecteur trouvera les versions détaillées de ces noyaux ainsi que leurs applications dans les travaux de Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011). La constante $P(m,h)$ donnée dans le tableau (le dénominateur du noyau associé triangulaire) est la constante de normalisation. Elle est explicitement donnée par :

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

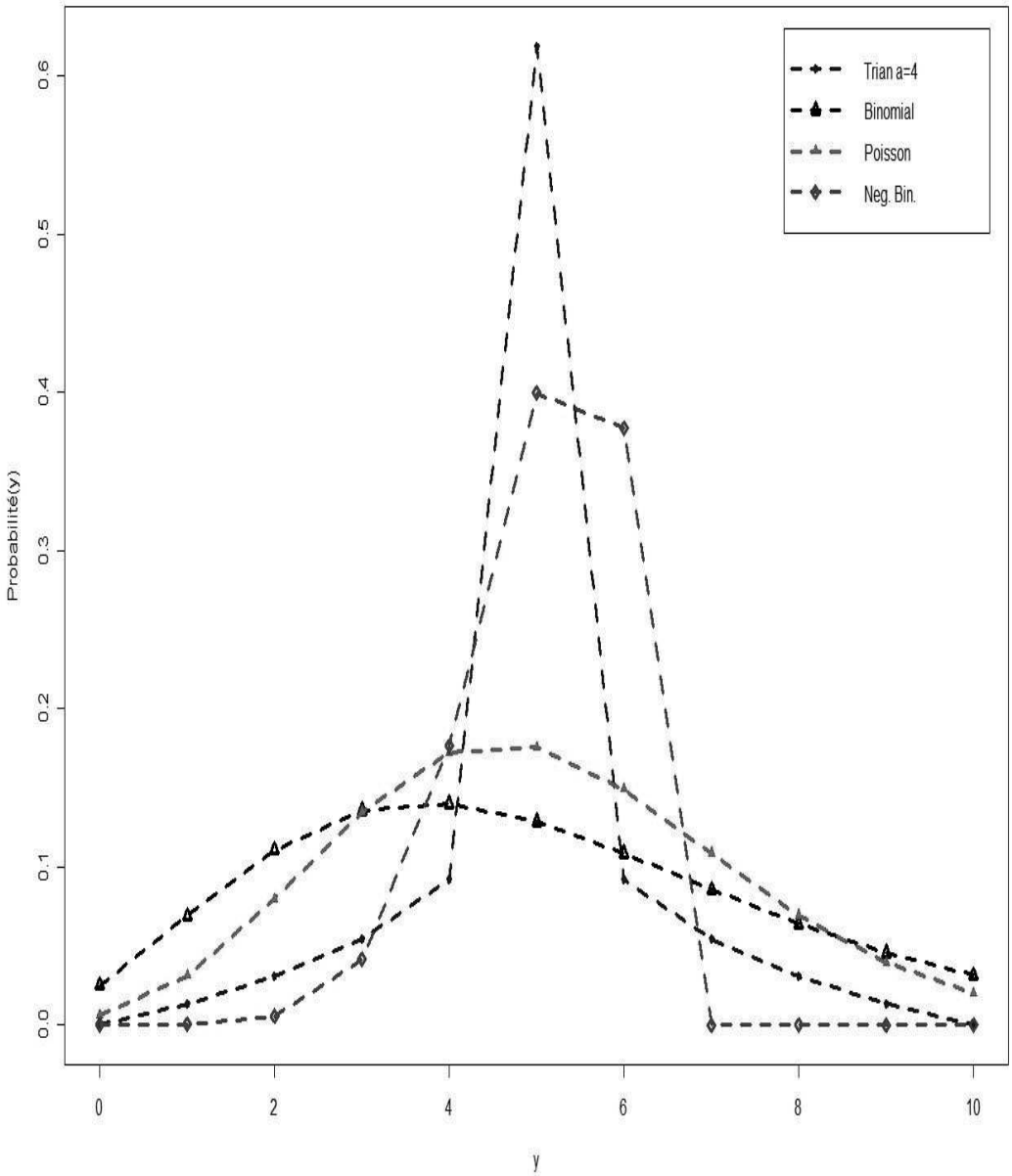

FIGURE 1.4 – Quelques noyaux associés discrets

TABLE 1.2 – Quelques noyaux associés discrets sur $\mathbb{T}_N \subseteq \mathbb{Z}$

Noyau associé	$K_{x,h}(u)$	$\mathbb{T}_N$
Aitchison & Aitken	$(1-h)\,\mathbb{I}_{\{u=x\}} + \dfrac{h}{c-1}\mathbb{I}_{\{u\neq x\}}$	$\{0,1,...,c-1\}$
Binomial	$\dfrac{(x+1)!}{u!(x+1-u)}\left(\dfrac{x+h}{x+1}\right)^u\left(\dfrac{1-h}{x+1}\right)^{x+1-u}$	$\{0,1,...,x+1\}$
Binomial Négatif	$\dfrac{(x+u)!}{u!x!}\left(\dfrac{x+h}{2x+1+h}\right)^u\left(\dfrac{x+1}{2x+1+h}\right)^{x+1}$	$\mathbb{N}$
Poisson	$\dfrac{(x+h)^u e^{-(x+h)}}{u!}$	$\mathbb{N}$
Triangulaire	$\dfrac{(m+1)^h - \lvert u-x\rvert^h}{P(m,h)}$	$\{0,\pm1,...,\pm k\}$
Wang & Van Ryzin	$(1-h)\,\mathbb{I}_{\{u=x\}} + \dfrac{1}{2}(1-h)\,h^{\lvert u-x\rvert}\mathbb{I}_{\{\lvert u-x\rvert\geqslant 1\}}$	$\mathbb{Z}$

1.2.3 Choix de fenêtre

Dans les travaux de Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011), ceux-ci ont proposé deux méthodes pour le choix du paramètre de lissage dans le cas discret. Il s'agit en fait des méthodes de validation croisée par les moindres carrés et l'excès de zéro que nous allons brièvement résumer dans les paragraphes suivants. Ensuite, nous présentons aussi les méthodes bayésiennes pour le choix de paramètres de lissage proposées par Zougab et al. (2012, 2013a, 2013b, 2013c).

Validation croisée par les moindres carrés

L'idée principale de la méthode adaptée ici est la même que celle présentée dans la Section 1.1.3. La seule différence est qu'ici, la fonction noyau $K_{x,h}$ dépend intrinsèquement 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 (1.25) par

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

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

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

$$\begin{aligned} CV(h) &= \sum_{x\in\mathbb{T}_N} \widehat{f_n}^{\,2}(x) - \frac{2}{n}\sum\nolimits_{x\in\mathbb{T}_N} \widehat{f_{n,-i}}(X_i) \\ &= \sum_{x\in\mathbb{T}_N} \left\{ \sum_{x\in\mathbb{T}_N} K_{x,h}(X_i) \right\}^2 - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{X_i}(X_j), \end{aligned}$$

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

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

Excès de zéros

Cette méthode proposée par Kokonendji et al. (2007) repose sur la particularité des données de comptage (pour $\mathbb{T}_N = \mathbb{N}$) qui consiste à avoir un nombre important de zéros dans l'échantillon. Notons $h_0 = h_0(n, K)$ une fenêtre adaptée. Pour un noyau associé discret $K_{x,h}$, le choix de h_0 est fait de sorte que

$$\sum_{y\in\mathbb{T}_N} \mathbb{P}(\mathcal{Z}_{x,h_0} = 0) = n_0, \tag{1.26}$$

où n_0 est le nombre de zéros dans l'échatillon. L'équation (1.26) s'obtient à partir de

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \sum_{y\in\mathbb{T}_N} \mathbb{P}(\mathcal{Z}_{x,h} = y) f(y)$$

en prenant $y = 0$ et $f(0) = 1$ afin d'identifier le nombre de zéros théorique au nombre de zéros empiriques. Le paramètre de lissage h_0 obtenu par excès de zéros est proche de ceux obtenus par la minimisation du MISE et validation croisée. Le lecteur peut aussi consulter l'article de Marsh & Mukhopadhyay (1999) qui applique cette méthode sur le modèle poissonien.

Approche bayésienne globale

Cette approche a été adaptée à l'estimation d'une fmp aux noyaux asymétriques par Kuruwita et al. (2010) puis aux noyaux associés discrets par Zougab et al. (2013b). L'idée consiste à considérer d'abord une suite de variables aléatoires discrètes $X_1, X_2, \cdots, X_n$ de fmp f et de réalisations $\mathbf{x} = (x_1, x_2, \cdots, x_n)$. Puis, on détermine l'estimateur de la vraisemblance conditionnelle de données $\mathbf{x}$ sachant h définie par

$$L(x_1, x_2, \cdots, x_n; h) = \pi(x_1, x_2, \cdots, x_n|h) = \prod_{i=1}^{n} \widehat{f_n}(x_i),$$

où $\widehat{f_n}$ est l'estimateur défini en (1.25). Ensuite, on utilise la technique de la validation croisée pour estimer $f(x_i)$ à partir de l'ensemble des points sauf en x_i. Ce qui permet

d'avoir

$$L(x_1, x_2, \cdots, x_n; h) = \pi(x_1, x_2, \cdots, x_n|h) = \prod_{i=1}^{n} \frac{1}{n-1} \sum_{j=1, i\neq j}^{n} K_{x_i,h}(x_j). \tag{1.27}$$

Par la suite, on choisit la loi *a priori* de h notée $\pi(h)$ de sorte que les méthodes de Monte Carlo pour les Chaînes de Markov (MCMC) puissent fonctionner correctement en la prenant à constante près de la forme

$$\pi(h) \propto \frac{1}{1+h^2}. \tag{1.28}$$

Ce qui n'a pas été le cas pour certains auteurs comme Brewer (1998), qui ont proposé la loi gamma comme loi *a priori* pour h. Ensuite, on détermine par le théorème de Bayes la loi a posteriori de h définie par

$$\pi(h|x_1, x_2, \cdots, x_n) = \frac{\pi(x_1, x_2, \cdots, x_n|h)\pi(h)}{\pi(x_1, x_2, \cdots, x_n)} = \frac{\pi(h) \prod_{i=1}^{n} \widehat{f_n}(x_i)}{\pi(x_1, x_2, \cdots, x_n)}, \tag{1.29}$$

où $\pi(x_1, x_2, \cdots, x_n) = \int \pi(x_1, x_2, \cdots, x_n|h)dh$. Puisque le calcul ce cette intégrale entraîne des calculs fastidieux, alors on remplace la loi *a posteriori* par le produit de la loi *a priori* avec l'estimateur de la vraisemblance. Ce qui conduit à

$$\pi(h|x_1, x_2, \cdots, x_n) \propto \frac{1}{1+h^2} \prod_{i=1}^{n} \frac{1}{n-1} \sum_{j=1, i\neq j}^{n} K_{x_i,h}(x_j). \tag{1.30}$$

Enfin, on utilise les méthodes de MCMC pour estimer h. Plus précisément, on applique l'algorithme à marche aléatoire selon les quatre étapes suivantes :
Etape 1-) Initialisation de h^0

Etape 2-) Pour $t \in \{1, \cdots, N\}$:
a) Générer $\tilde{h} \sim \mathcal{N}(h^{(t)}, \tau)$
b) Calculer la probabilité d'acceptation $\rho = \min\left\{1, \frac{\pi(\tilde{h}|x_1, x_2, \cdots, x_n)}{\pi(h^{(t)}|x_1, x_2, \cdots, x_n)}\right\}$ puis considérer $h^{(t+1)} = \begin{cases} \tilde{h} & \text{si } u < \rho, \quad u \sim \mathcal{U}_{[0,1]} \\ h^{(t)} & \text{sinon} \end{cases}$

Etape 3-) $t = t + 1$ et aller à 2.

Etape 4-) Calculer l'estimateur de Bayes : $\widehat{h} = \frac{1}{N - N_0} \sum_{t=N_0+1}^{N} h^{(t)}$.

Approche bayésienne locale

L'objectif de cette approche est d'estimer le paramètre de lissage en chaque point x dans le support $\mathbb{T}_N$ d'une fmp f caractérisant une suite de variable aléatoire $X_1, X_2, \cdots, X_n$

de réalisation $x_1, x_2, \cdots, x_n$. Pour cela, on le considère comme paramètre du modèle f_h qu'on utilise pour construire un estimateur bayésien de h pour chacun des point x de $\mathbb{T}_N$. Soit $\pi(h)$ la loi *a priori* de h. Par la formule de Bayes, la loi *a posteriori* de h au point x prend la forme suivante

$$\pi(h|x) = \frac{f_h(x)\pi(x)}{\int f_h(x)\pi(h)dh}. \tag{1.31}$$

Comme le modèle f_h est inconnu, on le remplace par son estimateur à noyau associé discret $\widehat{f_h}$. Ainsi, la loi *a posteriori* de $\pi(h|x)$ devient

$$\widehat{\pi}(h|x, X_1, X_2, \cdots, X_n) = \frac{\widehat{f_h}(x)\pi(x)}{\int \widehat{f_h}(x)\pi(h)dh}. \tag{1.32}$$

L'utilisation de l'estimateur de Bayes sous la perte quadratique conduit au meilleur estimateur $\widehat{h}$ de h au point x par

$$\widehat{h}_n(x) = \int h\widehat{\pi}(h|x, X_1, X_2, \cdots, X_n)dh. \tag{1.33}$$

Lorsque la loi *a priori* (1.31), la densité *a posteriori* (1.32) ainsi que la moyenne *a posteriori* (1.33) ne s'obtiennent pas explicitement, on peut utiliser les MCMC.

Approche bayésienne adaptative

Soit $X_1, X_2, \cdots, X_n$ une suite de variables aléatoires d'observations $x_1, x_2, \cdots, x_n$ et de fmp inconnue f. L'approche que nous décrivons ici consiste dans un premier temps à l'estimateur à noyau associé adaptatif donné pour tout $x \in \mathbb{T}_N$ par

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h_i}(x_i) \tag{1.34}$$

où $K_{x,h}$ est le noyau associé et h_i est le paramètre de lissage adaptatif associé à chaque observation x_i. En utilisant la technique de validation croisée on estime $f(x_i)$ comme en (1.34) sur l'ensemble des observations sauf en x_i et on a

$$\widehat{f}_{-i}(x_i) = \widehat{f}(x_i|\{x_{-i}\}, h_i) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} K_{x,h_i}(x_j). \tag{1.35}$$

Soit $\pi(h_i)$ la loi *a priori* de h_i, à partir de la formule de Bayes, la loi *a posteriori* pour chaque h_i prend la forme suivante :

$$\widehat{\pi}(h_i|x_i, \{x_{-i}\}) = \frac{\widehat{f}(x_i|\{x_{-i}\}, h_i)\pi(h_i)}{\int \widehat{f}(x_i|\{x_{-i}\}, h_i)\pi(h_i)dh_i}. \tag{1.36}$$

Par l'estimateur de Bayes sous la perte quadratique, le meilleur estimateur pour h_i est la moyenne de $\widehat{\pi}(h_i|x_i, \{x_{-i}\})$ donnée par

$$\widetilde{h}_i = \mathbb{E}\left[h_i|x_i, \{x_{-i}\}\right] = \int h\widehat{\pi}(h_i|x_i, \{x_{-i}\})dh_i. \tag{1.37}$$

Les expressions (1.36) et (1.37) donnent dans certains cas des résultats explicites grâce à l'usage des priors conjugués ; le lecteur peut se référer à Brewer (2000) et Zougab et al. (2013a).

1.2.4 Une résolution de problèmes de bords

Cette section se propose de rappeler la résolution de problèmes de bords dans l'estimation d'une fmp par le noyau triangulaire général $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$, proposé par Kokonendji & Zocchi (2010). La fmp de $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ est notée $f(\cdot; m, a_1, a_2, h_1, h_2)$ où $m \in \mathbb{Z}$ est son mode, $a_1, a_2 \in \mathbb{N}$ sont respectivement ses bras gauche et droit et enfin $h_1, h_2 \in \mathbb{R}_+$ sont ses degrés à gauche et à droite. En posant respectivement

$$\begin{aligned} \mathbb{T}_{a_1,m} &= \{m-k\ ;\ k=0,1,\cdots,a_1\} \quad ; \quad \mathbb{T}^*_{a_1,m} = \mathbb{T}_{a_1,m} \setminus \{m\} \\ \mathbb{T}_{m,a_2} &= \{m+k\ ;\ k=0,1,2,\cdots,a_2\} \quad \text{et} \quad \mathbb{T}^*_{m,a_2} = \mathbb{T}_{m,a_2} \setminus \{m\}, \end{aligned}$$

on remarque facilement que

$$\mathbb{T}_{m,a_1,a_2} = \mathbb{T}^*_{a_1,m} \cup \mathbb{T}_{m,a_2} = \mathbb{T}_{a_1,m} \cup \mathbb{T}^*_{m,a_2} = \mathbb{T}^*_{a_1,m} \cup \{m\} \cup \mathbb{T}^*_{m,a_2}.$$

L'expression explicite de $f(\cdot; m, a_1, a_2, h_1, h_2)$ est donnée par

$$f(y; m, a_1, a_2, h_1, h_2) = \frac{1}{D(a_1, a_2, h_1, h_2)} \left\{ \left[1 - \left(\frac{m-y}{a_1+1} \right)^{h_1} \right] \mathbf{1}_{\mathbb{T}^*_{a_1,m}}(y) + \left[1 - \left(\frac{y-m}{a_2+1} \right)^{h_2} \right] \mathbf{1}_{\mathbb{T}_{m,a_2}}(y) \right\},$$

avec

$$D(a_1, a_2, h_1, h_2) = (a_1 + a_2 + 1) - (a_1+1)^{-h_1} \sum_{k=1}^{a_1} k^{h_1} - (a_2+1)^{-h_2} \sum_{k=1}^{a_2} k^{h_2}.$$

Il a été signalé dans Kokonendji & Zocchi (2010) que lorsque $h_1 = h_2 = h$, on a la loi triangulaire générale d'ordre h notée $\mathcal{DT}(m; a_1, a_2; h, h) = \mathcal{DT}(m; a_1, a_2; h)$. Les principales caractéristiques de cette loi sont données par le théorème suivant :

Théorème 1.2.9 *(Kokonendji & Zocchi [2010]). Pour une variable aléatoire discrète Y de loi $\mathcal{DT}(m; a_1, a_2; h_1, h_2)$ on a :*

$$\begin{aligned} \mathbb{E}(Y) &= m + A(a_1, a_2; h_1, h_2) \\ Var(Y) &= B(a_1, a_2; h_1, h_2) - [A(a_1, a_2; h_1, h_2)]^2 \end{aligned}$$

avec

$$\begin{aligned} A(a_1, a_2; h_1, h_2) = \frac{1}{D(a_1, a_2, h_1, h_2)} \Bigg[& \frac{a_2(a_2+1)}{2} - \frac{a_1(a_1+1)}{2} \\ & + \sum_{k=1}^{a_1} k \left(\frac{k}{a_1+1} \right)^{h_1} - \sum_{k=1}^{a_2} k \left(\frac{k}{a_2+1} \right)^{h_2} \Bigg] \end{aligned}$$

et

$$B(a_1,a_2;h_1,h_2)=\frac{1}{D(a_1,a_2,h_1,h_2)}\left[\frac{a_2\,(a_2+1)\,(2a_2+1)}{6}+\frac{a_1\,(a_1+1)\,(2a_1+1)}{6}\right.$$
$$\left.-\sum_{k=1}^{a_1}k^2\left(\frac{k}{a_1+1}\right)^{h_1}-\sum_{k=1}^{a_2}k^2\left(\frac{k}{a_2+1}\right)^{h_2}\right].$$

De plus, lorsque :

$$\begin{array}{llll}
h_1\to 0 \;\; et \;\; h_2\to 0, & alors & \mathcal{DT}(m;a_1,a_2;h_1,h_2) & tend\ vers \quad \mathcal{D}(m);\\
h_1\to +\infty \;\; et \;\; h_2\to +\infty, & alors & \mathcal{DT}(m;a_1,a_2;h_1,h_2) & tend\ vers \quad \mathcal{U}(\mathbb{T}_{m,a_1,a_2});\\
h_1\to 0 \;\; et \;\; h_2\to +\infty, & alors & \mathcal{DT}(m;a_1,a_2;h_1,h_2) & tend\ vers \quad \mathcal{U}(\mathbb{T}_{m,a_2});\\
h_1\to +\infty \;\; et \;\; h_2\to 0, & alors & \mathcal{DT}(m;a_1,a_2;h_1,h_2) & tend\ vers \quad \mathcal{U}(\mathbb{T}_{a_1,m}).
\end{array}$$

La Figure 1.5 illustre les différentes formes qu'on peut obtenir en modifiant les paramètres de $\mathcal{DT}(m;a_1,a_2;h_1,h_2)$.

Considérons maintenant $X_1,\cdots,X_n$ une suite de variables aléatoires discrètes iid de fmp inconnue f et de support $\mathbb{T}_N=\mathbb{N}$ ou $\mathbb{T}_N=\{0,1,\cdots,N\}$ avec N un entier naturel non nul. L'estimateur à noyau discret triangulaire général d'ordre h de f est défini par :

$$\widehat{f_n}(x)=\frac{1}{n}\sum_{i=1}^{n}K_{x,h;a_1,a_2}\left(X_i\right),\quad x\in\mathbb{T}_N, \tag{1.38}$$

où $(a_1,a_2)\in\mathbb{N}^2$ sont les deux bras, $h=h(n)>0$ est le paramètre de lissage vérifiant $\lim\limits_{n\to\infty}h(n)=0$, et $K_{x,h}(\cdot)\equiv f(\cdot;x,a_1,a_2,h)$ qui satisfait la Définition 1.2.3. Soit $\mathcal{Z}_{x,a_1,a_2,h}$ une variable aléatoire de loi $\mathcal{DT}(m;a_1,a_2;h)$, les biais et variances ponctuels de cet estimateur sont respectivement donnés par

$$\begin{aligned}
Bias\left[\widehat{f_n}(x)\right]&=f\left\{\mathbb{E}(\mathcal{Z}_{x,a_1,a_2,h})\right\}-f(x)+\frac{1}{2}Var\left(\mathcal{Z}_{x,a_1,a_2,h}\right)f^{(2)}(x)+o(h)\\
&=A(a_1,a_2;h)f^{(1)}(x)+\frac{1}{2}\left\{B(a_1,a_2;h)-\left[A(a_1,a_2;h)\right]^2\right\}f^{(2)}(x)+o(h),
\end{aligned}\tag{1.39}$$

$$\begin{aligned}
Var\left[\widetilde{f_n}(x)\right]&=\frac{1}{n}\,f(x)\left[1-f(x)\right]\left[\Pr\left(\mathcal{Z}_{x,a_1,a_2,h}=x\right)\right]^2+R_n\left(x;a_1,a_2,h\right)\\
&=\frac{1}{n\left[D(a_1,a_2,h)\right]^2}\,f(x)\left[1-f(x)\right]+R_n\left(x;a_1,a_2,h\right)
\end{aligned}$$

où $f^{(k)}$ est la différence finie d'ordre $k\in\{1,2\}$ et $R_n\left(x;a_1,a_2,h\right)\to 0$ quand $n\to\infty$; voir Kokonendji & Zocchi (2010) et Kokonendji & Senga Kiessé (2011) pour plus de détails. Notons cependant que lorsque $\bigcup\limits_{x\in\mathbb{T}}\mathbb{S}_{x,h}$ est différent de $\mathbb{T}$ (i.e. $\bigcup\limits_{x\in\mathbb{T}}\mathbb{S}_{x,h}\supsetneq\mathbb{T}$) alors on a les effets de bords. Pour corriger ce défaut, on procède comme suit :

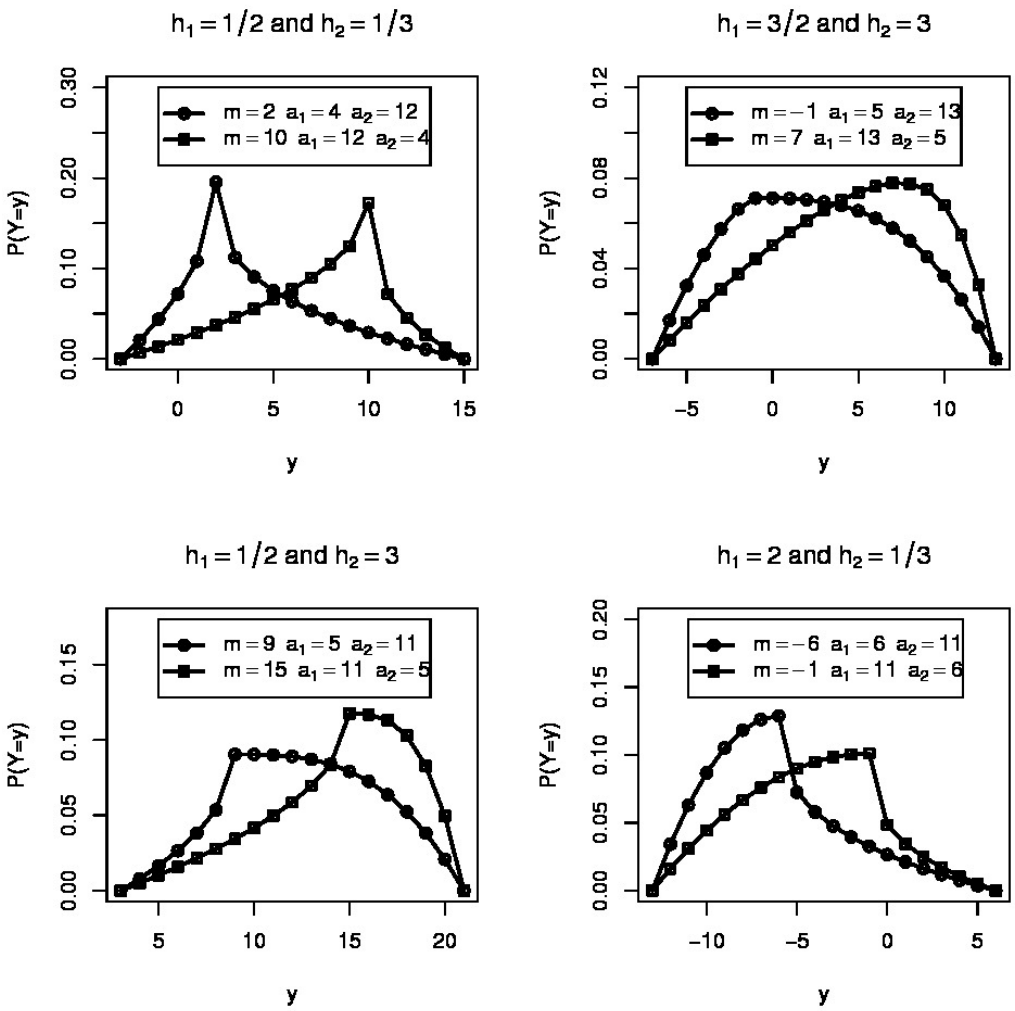

FIGURE 1.5 – Différentes formes du noyau discret triangulaire général.

Cas où $\mathbb{T} = \mathbb{N}$ et $a_1 \neq 0$

Ici, l'estimateur à noyau discret triangulaire général produit les effets de bords à gauche de $\mathbb{N}$ car la réunion $\bigcup_{x\in\mathbb{N}} \mathbb{T}_{x,a_1,a_2} = \{-a_1, \cdots, -1\} \cup \mathbb{N}$ contient d'une manière stricte le support $\mathbb{N}$ de f. Ainsi, on considère un nouveau bras gauche dit modifié et noté a_0 obtenu à partir de a_1 et défini de sorte que, pour $x \in \mathbb{N}$,

$$a_0 = \begin{cases} x & \text{si} \quad x \in \{0, 1, \cdots, a_1 - 1\} \\ a_1 & \text{si} \quad x \in \{a_1, a_1 + 1, \cdots, N, \cdots\}. \end{cases} \tag{1.40}$$

Cas où $\mathbb{T} = \{0, 1, \cdots, N\}$

Dans ce cas, l'estimateur à noyau discret triangulaire général produit les effets de bords à droite du support $\{0, 1, \cdots, N\}$ puisque

$$\bigcup_{x \in \{0,1,\cdots,N\}} \mathbb{T}_{x,a_1,a_2} = \{-a_1, \cdots, -1\} \cup \{0, 1, \cdots, N\} \cup \{N+1, \cdots, N+a_2\}.$$

En appliquant (1.40) à droite de $\{0, 1, \cdots, N\}$ qui est un voisinage de $x = N$, alors le nouveau bras modifié a_N obtenu à partir de a_2 est défini de sorte que, pour $x \in \{0, 1, \cdots, N\}$,

$$a_N = \begin{cases} a_2 & \text{si} \quad x \in \{0, 1, \cdots, N - a_2\} \\ N - x & \text{si} \quad x \in \{N - a_2 + 1, \cdots, N-1, N\}. \end{cases} \tag{1.41}$$

La combinaison de (1.40) et (1.41) fournit un noyau associé modifié et asymétrique qui est plus approprié pour l'estimation de toute pmf compact (ou fonctions discrètes).

Chapitre 2

Méthode des noyaux associés continus

Dans le chapitre précédent, nous avons vu que pour une suite de variables aléatoires $X_1, \cdots, X_n$ indépendantes et identiquement distribuées (iid) de fonction densité de probabilité (fdp) f inconnue de support $\mathbb{T}$, un sous ensemble de $\mathbb{R}^d$ (avec $d = 1$). l'estimateur $\widehat{f_n}$ de la fdp f est défini par

$$\widehat{f_n}(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathcal{K}\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{T} = \mathbb{R}. \tag{2.1}$$

La fonction noyau $\mathcal{K}(\cdot)$ en (2.1) vérifie la Définition 1.1.1 et est indépendante de la cible x et la fenêtre de lissage h. Le lecteur peut, en plus du chapitre précédent, se référer aux travaux de Rosenblatt (1956), Parzen (1962), Sylverman (1986), Devroye (1987), Scott (1992) et Tsybakov (2004) afin d'approfondir ces notions. Cette fonction noyau $\mathcal{K}(\cdot)$ avait été imaginée pour estimer des densités f à support non borné $\mathbb{T} = \mathbb{R}$. Depuis des travaux de Chen (1999, 2000) sur les noyaux bêta et gamma, respectivement pour estimer les densité à support $\mathbb{T} = [0, 1]$ et $\mathbb{T} = [0, \infty)$, de Scaillet (2004) sur les noyaux inverse gaussien et sa réciproque estimant les densités à support $\mathbb{T} = [0, \infty)$ ainsi que les travaux de Kokonendji & Senga Kiessé (2011) pour le cas discret (i.e. $\mathbb{T} \subseteq \mathbb{Z}$), est née une nouvelle famille d'estimateurs que nous appelons *estimateurs à noyaux associés* où la fonction noyau est paramétrée par le point d'estimation x et le paramètre de lissage h. Afin d'harmoniser l'écriture comme dans Kokonendji & Senga Kiessé (2011), on peut écrire les estimateurs à noyaux associés continus $\widehat{f_n}$ de la fdp f comme suit

$$\widehat{f_n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T} \subseteq \mathbb{R}. \tag{2.2}$$

Il est facile de voir que le noyau (classique) $\mathcal{K}(\cdot)$ en (2.1) devient un cas particulier des noyaux associés continus $K_{x,h}$ qui dépendent intrinsèquement de x et h, à travers la relation :

$$K_{x,h}(\cdot) = \frac{1}{h}\mathcal{K}\left(\frac{x - \cdot}{h}\right). \tag{2.3}$$

De nombreux auteurs ont raffiné les propriétés de ces estimateurs à noyaux associés dans des cas particuliers. C'est notamment le cas de Bouezmarni & Rolin (2003) qui ont travaillé sur la consistance des estimateurs à noyaux bêta. Kokonendji et al. (2009) ont

aussi traité le cas d'estimation semi-paramétrique avec les noyaux associés discrets. Bouezmarni & Roumbots (2010) ont présenté le cas multivarié essentiellement pour les données bornées. Kokonendji & Zocchi (2010) ont proposé (comme nous l'avons présenté dans le dernier paragraphe du chapitre précédent) une solution aux problèmes des effets de bords dans le cas discret en utilisant le noyau triangulaire général. Senga Kiesse & Rivoire (2010) ont pour leur part appliqué la régression semi-paramétrique sur des données réelles. Balakrishnan et al. (2011) ont étudié le cas de mélange du noyau Birnbaum-Sanders avec d'autres. Klutchnikoff & Bertin (2011) ont donné les propriétés minimax des estimateurs à noyau bêta. Enfin, Zougab et al. (2012a, 2012b, 2012c) ont traité le cas du choix de paramètres de lissage discret avec les méthodes bayésiennes.

Ce chapitre, se propose d'homogénéiser la théorie concernant les estimateurs à noyaux associés continus en fournissant une technique systématique de construction sans effets de bords sur n'importe quel type de support $\mathbb{T}$ de la densité f, à estimer. Il s'organise de la manière suivante : nous présentons d'abord une définition générale des noyaux associés continus incluant les cas classiques. Ensuite, nous donnons le principe de construction à partir de n'importe quelle fdp paramétrée et nous illustrons ceci sur les noyaux non-classiques de la littérature. Par la suite, nous utilisons ces noyaux associés pour estimer des densités. Différentes propriétés ponctuelles sont présentées, en particulier la convergence au sens du risque quadratique moyen et intégré asymptotique ("Asymptotic Mean Integrated Squared Error" (AMISE) en anglais) et l'algorithme de réduction de biais. Puis, des illustrations avec les types de noyaux tels que Pareto, lognormal, bêta et sa version étendue, gamma et son inverse ainsi que l'inverse gaussien et sa réciproque, seront données. Enfin nous présentons les résultats des études par simulation sur trois différents types d'estimateurs à noyaux lognormaux. Dans tout ce chapitre, nous désignons par $\mathbb{T} = \mathbb{T}_I$ les intervalles de $\mathbb{R}$ *a fortiori* bornés au moins d'un côté.

2.1 Noyaux associés continus

Nous détaillons dans cette section la méthode de construction des noyaux associés continus en commençant par les notions de type de noyaux.

2.1.1 Type de noyau et noyaux associés

Fixons d'abord la définition du type de noyau continu.

Définition 2.1.1 *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 $\theta_i : \mathbb{R} \to \mathbb{R}$ 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 2.1.2 *(i) Le mode $M(a,b)$ d'un type de noyau $K_{\theta(a,b)}$ appartient toujours à son support $\mathbb{S}_{\theta(a,b)}$.*

(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.

Signalons aussi qu'il existe plusieurs manières de définir le concept de "paramètre de dispersion" qui est souvent lié à celui de la "mesure de dispersion". Certains auteurs déduisent le paramètre de dispersion après avoir exprimé la variance de la loi en fonction de son espérance comme pour les lois appartenant à la famille exponentielle naturelle. D'autres le définissent comme le paramètre d'échelle ou encore le paramètre d'intensité de la loi (ce qui n'est autre que l'inverse de son paramètre d'échelle). Ce sont ces deux dernières considérations que nous donnons à ce paramètre dans la plupart de nos résultats. Nous conseillons au lecteur de se référer aux travaux de Jørgensen (1997), Jørgensen & Lauritzen (2000), Jørgensen et al. (2010) et Jørgensen & Kokonendji (2011) pour un approfondissement de cette notion. Donnons maintenant la définition précise d'un noyau associé continu.

Définition 2.1.3 *On considère $x \in \mathbb{T}_I \subseteq \mathbb{R}$ et $h > 0$ avec $\mathbb{T}_I$ 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 :*

$$x \in \mathbb{S}_{x,h}, \tag{2.4}$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + A(x,h), \tag{2.5}$$

$$Var(\mathcal{Z}_{x,h}) = B(x,h), \tag{2.6}$$

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

Observons les aspects cachés de cette définition.

Remarque 2.1.4

1. *Les conditions (2.5) et (2.6) montrent implicitement que la construction du noyau associé n'est pas unique voir [Proposition 2.2.6, Section 2.2.2].*
2. *Le support $\mathbb{S}_{x,h}$ du noyau associé $K_{x,h}$ n'est pas nécessairement symétrique par rapport à 0 ou x comme dans le cas classique. Il peut ne pas dépendre de x et/ou h (e.g. le support $\mathbb{S}_{x,h}$ du noyau d'Epanechnikov dépend de x et h. Toutefois, celui du noyau de Pareto dépend seulement de x, par contre, les supports des noyaux bêta, gamma et autres sont indépendants de x et h).*
3. *La condition (2.4) peut être remplacée par $\cup_{x\in\mathbb{T}_I}\mathbb{S}_{x,h} \supseteq \mathbb{T}_I$. Elle sous-entend que le choix du noyau associé doit se faire en tenant compte du support $\mathbb{T}_I$ de la densité f, de sorte que $\mathbb{S}_{x,h} = \mathbb{T}_I$ afin d'éviter les problèmes classiques des effets de bords.*
4. *La condition (2.5) montre que les noyaux associés sont de plus en plus concentrés sur le point d'estimation x lorsque h tend vers 0. Ceci met en évidence la flexibilité de ces derniers à changer de forme suivant la position du point d'estimation.*

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 2.1.5 *Soit $\mathcal{K}$ un noyau classique (Voir Définition 1.1.1) de support $\mathbb{S}$, de moyenne $\mu_{\mathcal{K}} = 0$ et de variance $\sigma^2_{\mathcal{K}} < \infty$. Pour un $x \in \mathbb{T} = \mathbb{R}$ donné et $h > 0$, alors le noyau associé classique est défini par (2.3) et son support est $\mathbb{S}_{x,h} = x - h\mathbb{S}$ et de plus*

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x \text{ et } Var(\mathcal{Z}_{x,h}) = h^2\sigma^2_{\mathcal{K}}. \tag{2.7}$$

En d'autres termes, (2.7) montre que les caractéristiques A et B du noyau associé classique $K_{x,h}$ sont :

$$A(x,h) = 0 \text{ et } B(x,h) = O(h^2). \tag{2.8}$$

Démonstration. A partir de (2.3), 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 (2.4) 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 $u = (x-t)/h$.■

Nous allons maintenant montrer la technique de construction des noyaux associés non-classiques à partir de n'importe quel type de noyau.

2.1.2 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 :

$$\begin{cases} M(a,b) = x \\ D(a,b) = h. \end{cases} \tag{2.9}$$

Les solutions $a = a(x,h)$ et $b = b(x,h)$ du système (2.9) 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 2.1.3. Nous présenterons des cas pratiques dans la Section 2.1.3.

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 2.1.3.

Proposition 2.1.6 *Soit $\mathbb{T}_I$ le support de la densité f, à estimer. Pour tout $x \in \mathbb{T}_I$ 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*

$$x \in \mathbb{S}_{\theta(x,h)}, \tag{2.10}$$
$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = A_\theta(x,h), \tag{2.11}$$
$$Var(\mathcal{Z}_{\theta(x,h)}) = B_\theta(x,h), \tag{2.12}$$

où $\mathcal{Z}_{\theta(x,h)}$ *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.

Démonstration. Tout d'abord, par la méthode mode-dispersion on a $\theta(x,h) = \theta\left(a(x,h), b(x,h)\right)$, 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 unimodal et de mode $M(a,b)$ appartenant à $\mathbb{S}_{\theta(a,b)}$ (grâce au point (i) de la Remarque 2.1.2), par la méthode mode-dispersion on a le premier résultat (2.10) 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 $\mathcal{Z}_{\theta(a,b)}$ associée au type noyau uni-modal $K_{\theta(a,b)}$, on peut écrire $\mathbb{E}(\mathcal{Z}_{\theta(a,b)}) = M(a,b) + \varepsilon(a,b)$, où $\varepsilon(a,b)$ 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 $\varepsilon(a,b) = \varepsilon(a(x,h), b(x,h))$. On a par conséquent

$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = \varepsilon(a(x,h), b(x,h)).$$

En prenant $A_\theta(x,h) = \varepsilon(a(x,h), b(x,h))$ puis, en utilisant la définition du paramètre de dispersion autour du mode, on obtient le deuxième résultat (2.11). 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 :

$$Var(\mathcal{Z}_{\theta(x,h)}) = B_{\theta(a(x,h),b(x,h))},$$

avec $B_{\theta(a(x,h),b(x,h))}$ qui tend vers zéro quand h tend vers zéro d'après le point (iii) de la Remarque 2.1.2. On obtient le dernier résultat (2.12) en prenant $B_\theta(x,h) = B_{\theta(a(x,h),b(x,h))}$.■

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 $b\Gamma(1 + 1/a)$ et $b^2\left\{\Gamma(1 + 2/a) - (2/a)\log(2)\right\}$, où $\Gamma(a)$ représente la fonction gamma. Il est défini sur $]0, +\infty[$ par :

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

Son mode est égal à $b(1 - 1/a)^{1/a}$ et son paramètre de dispersion est b. Le système (2.9) donne

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

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.

Un autre cas est celui du type de noyau Birnbaum-Saunders (proposé dans Birnbaum & Saunders, 1969) de paramètres $a > 0$ et $b > 0$ et défini par :

$$BS_{\theta(a,b)}(u) = \frac{\mathbf{1}_{[0,+\infty)}(u)}{2ab\sqrt{2\pi}}\left\{\left(\frac{b}{u}\right)^{1/2} + \left(\frac{b}{u}\right)^{3/2}\right\}\exp\left(\frac{-1}{2a^2}\left[\frac{u}{b} + \frac{b}{u} - 2\right]\right).$$

Ses moyenne et variance sont respectivement $b(1 + a^2/2)$ et $(ab)^2(1 + a^2/2)$. Le principe mode-dispersion conduit à

$$\begin{cases} \arg\max\limits_{u>0} BS_{\theta(a,h)}(u) = x \\ \qquad\qquad\qquad\qquad b = h. \end{cases}$$

Il y a lieu de remarquer ici que, le mode $M(a,h) = \arg\max\limits_{u>0} BS_{\theta(a,h)}(u)$ n'a pas une forme explicite. On l'obtient en résolvant une équation non linéaire en a. Balakrishnan et al. (2011) ont calculé quelques valeurs modales en faisant varier a de 0.5 à 5.0.

Donnons maintenant quelques exemples illustrant la construction des noyaux associés non-classique.

2.1.3 Noyaux associés non-classiques : bêta et son extension, gamma et son inverse, gaussien inverse et sa réciproque, Pareto puis log-normal

Nous utilisons dans cette section, la méthode mode-dispersion pour la construction des différents noyaux associés continus non-classiques. Pour chacun des cas, nous présentons d'abord le type de noyau $K_{\theta(a,b)}$ avec ses paramètre de départ, y compris son mode $M(a,b)$ et son paramètre de dispersion $D(a,b)$. Ensuite, nous appliquons la méthode mode-dispersion pour obtenir la nouvelle paramétrisation $\theta(a(x,h), b(x,h)) = \theta(x,h)$ qui permet l'obtention de $K_{\theta(x,h)}$. Enfin, nous rappelons les expressions des moyennes et variances de $K_{\theta(a,b)}$ à partir desquels nous déduisons les caractéristiques $A_\theta(x,h)$ et $B_\theta(x,h)$.

Noyau bêta

Le noyau associé bêta a été introduit par Chen (1999), 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).

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 :

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

où $\mathscr{B}(a,b)$ est la fonction bêta. Son mode et son paramètre de dispersion sont respectivement $(a-1)/(a+b-2)$ et $1/(a+b-2)$. La résolution du système (2.9) donne

$$\theta(x,h) = \left(\frac{x}{h}+1, \frac{1-x}{h}+1\right) \text{ pour tout } x \in [0,1] \text{ et } h > 0.$$

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

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

Son support est

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

En injectant les composantes $a(x,h) = (x/h)+1$ et $b(x,h) = \{(1-x)/h\}+1$ de $\theta(x,h)$, 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

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

Noyau bêta étendu

Le noyau bêta étendu que nous introduisons ici est la généralisation du précédent. Il est jusque-là inconnu en statistique non paramétrique mais il est souvent utilisé en recherche opérationnelle (voir e.g. Grubbs, 1962). Il a deux paramètres de forme $a > 1$, et $b > 1$ et est défini sur $\mathbb{S}_{EB} = [t_0, t_1]$, $(t_0 < t_1 \leq \infty)$ par :

$$EB_{\theta(a,b;t_0,t_1)}(u) = \frac{1}{\mathscr{B}(a,b)(t_1-t_0)^{a+b-1}}(u-t_0)^{a-1}(t_1-u)^{b-1}\mathbf{1}_{[t_0,t_1]}(u).$$

Puisque son mode est $\{(a-1)t_1+(b-1)t_0\}/(a+b-2)$ et son paramètre de dispersion $1/(a+b-2)$, la résolution du système (2.9) conduit à :

$$\theta(x,h;t_0,t_1) = \left(\frac{x-t_0}{(t_1-t_0)h}+1, \frac{t_1-x}{(t_1-t_0)h}+1\right) \text{ pour tout } x \in [t_0,t_1] \text{ et } h > 0.$$

Le noyau associé bêta étendu construit est alors

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

avec $\Delta = \mathscr{B}\left([\{x-t_0\}/\{(t_1-t_0)h\}]+1, [\{t_1-x\}/\{(t_1-t_0)h\}]+1\right)$. Il est de support

$$\mathbb{S}_{EB_{\theta(x,h,t_0,t_1)}} = [t_0,t_1] = \mathbb{S}_{EB}.$$

En remplaçant a et b dans les expressions des moyenne et variance de $BE_{\theta(a,b;t_0,t_1)}$ (i.e dans $t_0+(t_1-t_0)a/(a+b)$ et $ab(t_1-t_0)^2/\{(a+b)^2(a+b+1)\}$), par les solutions $a(x,h;t_0,t_1)$

et $b(x, h; t_0, t_1)$ du système (2.9), on trouve :

$$A_\theta(x, h; t_0, t_1) = \frac{h\{(t_0 + t_1) - 2x\}}{1 + 2h} \tag{2.13}$$

$$B_\theta(x, h; t_0, t_1) = \frac{h\{x - t_0 + h(t_1 - t_0)\}\{(t_1 - x) + h(t_1 - t_0)\}}{(1 + 2h)^2(1 + 3h)}. \tag{2.14}$$

Les figures suivantes montrent la flexibilité du noyau associé bêta étendu. Plus précisément, nous présentons dans la Figure 2.1 le type de noyau bêta étendu pour $x \in [2, 5]$ puis, dans les deux derniers, nous donnons les différentes formes de ce noyau associé respectivement sous l'effet cible x et l'effet paramètre h.

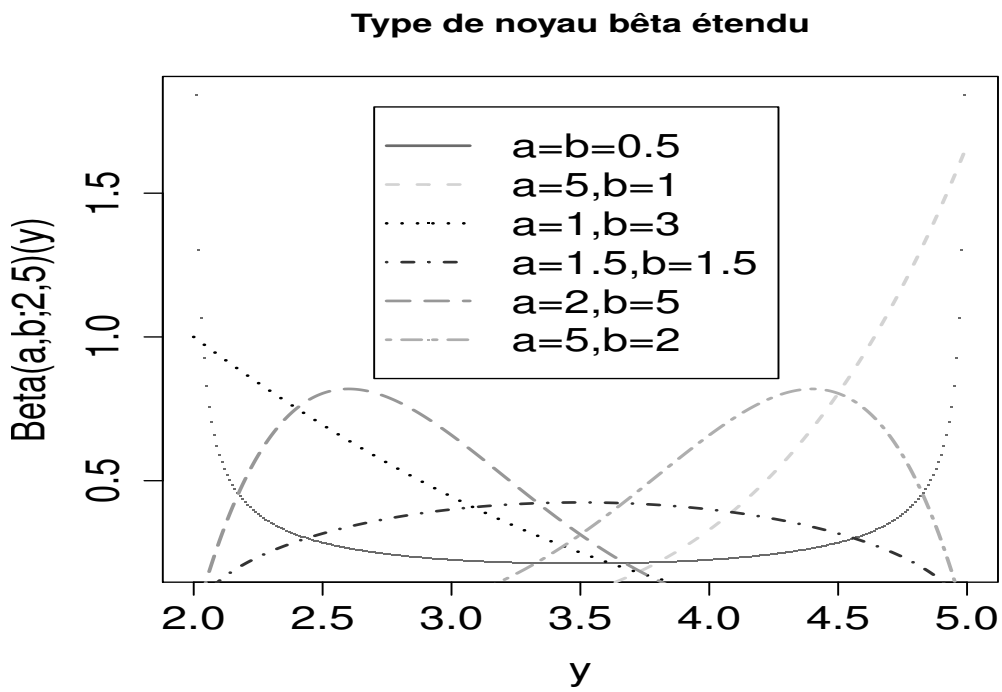

FIGURE 2.1 – Type de noyau bêta étendu sur $[2, 5]$

Noyau gamma

Ce noyau associé a été pour la première fois proposé par Chen (2000) dans le but d'estimer les densités sur $[0, +\infty[$ sans pourtant montrer comment il l'a construit. Nous proposons ici sa construction par la méthode mode-dispersion. Une fois de plus nous signalons d'avance que nos résultats coïncident avec ceux de Chen (2000).

Sans plus tarder notons $GA_{\theta(a,b)}$, le type de noyau gamma de paramètres $a > 1$ et $b > 0$ (avec a paramètre de forme et b celui d'échelle) et défini sur $\mathbb{S}_{GA} = [0, +\infty[$ par :

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

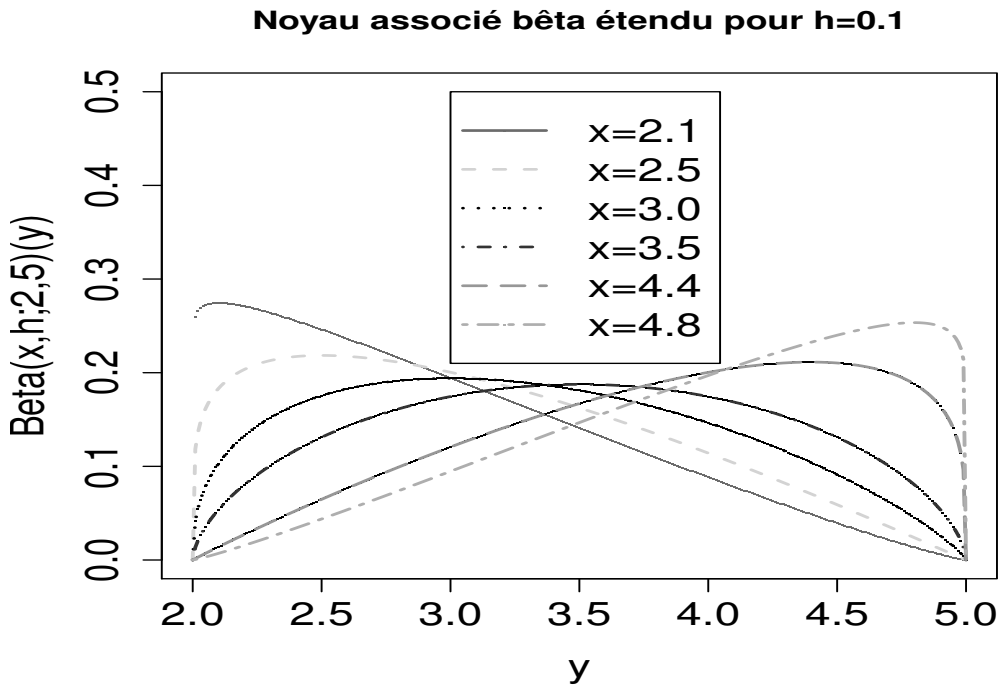

FIGURE 2.2 – Noyau associé bêta étendu avec h fixé

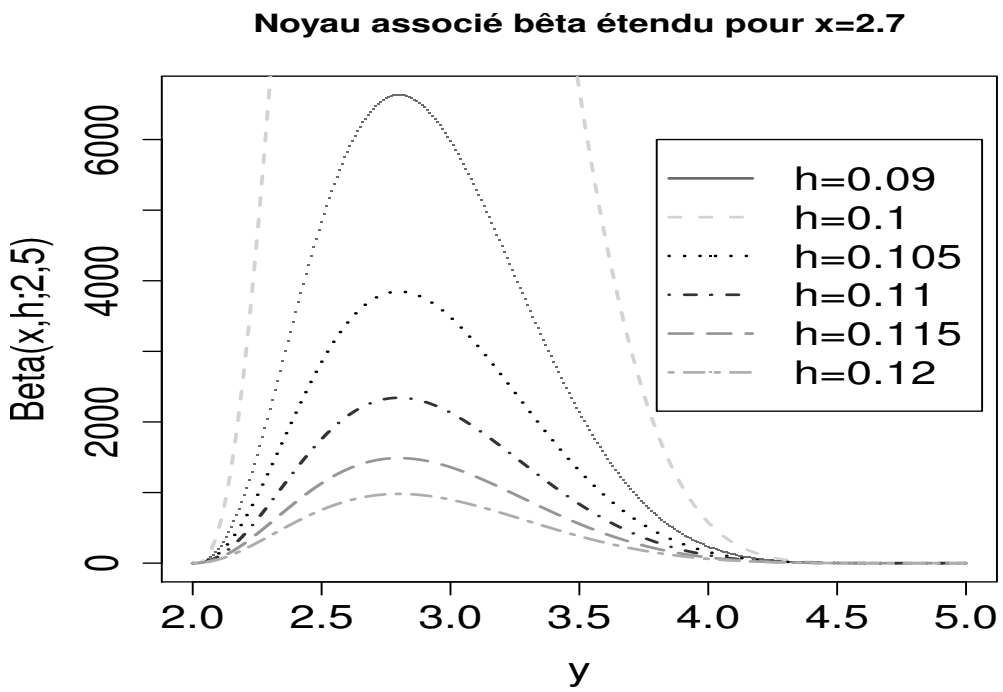

FIGURE 2.3 – Noyau associé bêta étendu avec x fixé

Son mode et son paramètre de dispersion étant respectivement $(a-1)b$ et b, en utilisant la méthode mode-dispersion on a :

$$\theta(x,h) = \left(\frac{x}{h}+1, h\right) \text{ pour tout } x > 0 \text{ et } h > 0.$$

Ainsi, le noyau associé gamma construit est :

$$GA_{\theta(x,h)}(u) = \frac{u^{x/h}}{\Gamma\left(\frac{x}{h}+1\right)h^{(x/h)+1}} \exp\left(-\frac{u}{h}\right)\mathbf{1}_{[0,+\infty[}(u).$$

Il est défini sur

$$\mathbb{S}_{GA_{\theta(x,h)}} = [0,+\infty[= \mathbb{S}_{GA}.$$

En écrivant la moyenne et la variance de $GA_{\theta(a,b)}$ en fonction des composantes de $\theta(x,h)$ on obtient :

$$A_\theta(x,h) = h \quad \text{et} \quad B_\theta(x,h) = h(x+h).$$

Noyau gamma inverse

Nous introduisons maintenant le type de noyau gamma inverse, respectivement de paramètres de forme et d'échelle $a > 0$ et $b > 0$, et défini sur $\mathbb{S}_{GI} =]0,+\infty[$ par :

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

Partant de son mode et son paramètre de dispersion définis respectivement par $b/(a+1)$ et $1/b$, on utilise le principe mode-dispersion pour obtenir :

$$\theta(x,h) = \left(\frac{1}{xh}-1, \frac{1}{h}\right) \text{ pour tout } h > 0 \text{ et } x \in]0, 1/h[.$$

Ce qui conduit au noyau associé gamma inverse noté $GI_{\theta(x,h)}$ et défini par

$$GI_{\theta(x,h)}(u) = \frac{h^{1-1/(xh)}}{\Gamma\left(\frac{1}{xh}-1\right)} u^{-1/(xh)} \exp\left(-\frac{1}{hu}\right)\mathbf{1}_{]0,+\infty[}(u).$$

Son support est

$$\mathbb{S}_{GI_{\theta(x,h)}} =]0,+\infty[= \mathbb{S}_{GI}.$$

De plus, en remplaçant a et b dans les expressions de moyenne et variance de $GI_{\theta(a,b)}$ (i.e $b/(a-1)$ et $b^2/\{(a-1)^2(a-2)\}$) par les composantes de $\theta(x,h)$, on trouve

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

Notons ici que pour tout $h > 0$ fixé, $\theta(x,h)$ n'est défini que sur $]0, 1/h[$. Ceci montre que le noyau associé inverse gamma n'est pas approprié pour l'estimation des densités sur la toute partie positive de la droite réelle.

Noyau gaussien inverse

Le noyau associé inverse gaussien ainsi que sa version réciproque ont été introduits par Scaillet (2004) sans aucune précision sur leur construction. Le lecteur peut aussi consulter l'ouvrage de Seshadri (1993) pour d'autres propriétés de ce type de noyau. Ici, nous utilisons la méthode mode-dispersion pour construire le noyau associé gaussien inverse. Nous précisons ici que la version proposée par Scaillet (2004) est celle modifiée du noyau associée que nous construisons ici [voir section 2.2.3] pour plus de précision.

Désignons par $IG_{\theta(a,b)}$ le type de noyau inverse gaussien de paramètres $a > 0$ et $b > 0$ (avec a paramètre de forme et b celui d'échelle). Il est défini sur $\mathbb{S}_{IG} =]0, +\infty[$ par

$$IG_{\theta(a,b)}(u) = \frac{\sqrt{b}}{\sqrt{2\pi t^3}} \exp\left\{-\frac{b}{2a}\left(\frac{u}{a} - 2 + \frac{a}{u}\right)\right\} \mathbf{1}_{\mathbb{R}_+^*}(u).$$

A travers son mode $a\left\{(1 + 9a^2b^{-2}/4)^{1/2} - 3ab^{-1}/2\right\}$ et son paramètre de dispersion $1/b$, on obtient d'après la méthode mode-dispersion

$$\theta(x,h) = \left(\frac{x}{\sqrt{1-3xh}}, \frac{1}{h}\right) \text{ pour tout } h > 0 \text{ et } x \in]0, 1/3h[.$$

Ainsi, le noyau associé gaussien inverse construit est,

$$IG_{\theta(x,h)}(u) = \frac{1}{h\sqrt{2\pi u^3}} \exp\left\{-\frac{\xi(x,h)}{2xh}\left(\frac{u\xi(x,h)}{x} - 2 + \frac{x}{u\xi(x,h)}\right)\right\} \mathbf{1}_{\mathbb{R}_+^*}(u),$$

avec $\xi(x,h) = (1 - 3xh)^{1/2}$. Son support est

$$\mathbb{S}_{IG_{\theta(x,h)}} =]0, +\infty[= \mathbb{S}_{IG}.$$

Sa moyenne et sa variance étant respectivement a et a^3/b, en y injectant les valeurs issues du système (5), on obtient

$$A_\theta(x,h) = x\left\{\frac{1}{\sqrt{1-3xh}} - 1\right\} \quad \text{et} \quad B_\theta(x,h) = \frac{x^3 h}{(1+3xh)^{3/2}}.$$

Il est important de noter que ce noyau associé n'est pas approprié pour estimer les densités définies sur la partie positive de la droite réelle, parce que pour une valeur de h fixée, $\theta(x,h)$ n'est pas défini en dehors de $]0, (3h)^{-1}[$. Nous illustrons ce disfonctionnement par la Figure 5.1

Noyau gaussien inverse réciproque

Considérons maintenant le type de noyau inverse gaussien réciproque noté $RG_{\theta(a,b)}$ avec $a > 0$ et $b > 0$ respectivement ses paramètres de forme et d'échelle. Il est défini sur $\mathbb{S}_{RG} =]0, +\infty[$ par :

$$RG_{\theta(a,b)}(u) = \frac{\sqrt{b}}{\sqrt{2\pi u}} \exp\left\{-\frac{b}{2a}\left(au - 2 + \frac{1}{au}\right)\right\} \mathbf{1}_{\mathbb{R}_+^*}(u).$$

A partir de son mode $(1/a)\left\{(1+a^2b^{-2}/4)^{1/2}-ab^{-1}/2\right\}$ et de son paramètre de dispersion $1/b$, on résout le système (5), puis on trouve

$$\theta(x,h)=\left(\frac{1}{\sqrt{x^2+xh}},\frac{1}{h}\right) \text{ pour tout } x\in]0,+\infty[\text{ et } h>0.$$

Ce qui conduit au noyau associé gaussien inverse réciproque défini par :

$$RG_{\theta(x,h)}(u)=\frac{1}{\sqrt{2\pi hu}}\exp\left\{-\frac{\zeta(x,h)}{2h}\left(\frac{u}{\zeta(x,h)}-2+\frac{\zeta(x,h)}{u}\right)\right\}\mathbf{1}_{\mathbb{R}_+^*}(u),$$

avec $\zeta(x,h)=(x^2+xh)^{1/2}$. Il est de support

$$\mathbb{S}_{RG_{\theta(x,h)}}=]0,+\infty[=\mathbb{S}_{RG}.$$

En écrivant sa moyenne $1/a+1/b$ puis sa variance $1/(ab)+2/b^2$ en fonction des composantes de $\theta(x,h)$ définies précédemment, on trouve

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

Noyau de Pareto

Nous introduisons ce noyau qu'on utilise dans les cas des valeurs extrêmes. Il est de paramètres $a>0$ et $b>0$ respectivement paramètres de forme et d'échelle. Il a pour support $\mathbb{S}_{PA_{\theta(a,b)}}=[a,+\infty[$ et est défini par :

$$PA_{\theta(a,b)}(u)=\frac{ba^b}{u^{b+1}}\mathbf{1}_{[a,+\infty[}(u).$$

Son mode et son paramètre de dispersion sont respectivement a et $1/b$. En résolvant le système (2.9), on obtient :

$$\theta(x,h)=\left(x,\frac{1}{h}\right) \text{ pour tout } x\in]0,+\infty[\text{ et } h>0.$$

On obtient finalement le noyau associé de Pareto défini par :

$$PA_{\theta(x,h)}(u)=\frac{1}{hu}\left(\frac{x}{u}\right)^{1/h}\mathbf{1}_{[x,+\infty[}(u).$$

Il est de support

$$\mathbb{S}_{PA_{\theta(x,h)}}=[x,\infty[\neq\mathbb{S}_{PA_{\theta(a,b)}}=[a,+\infty[,$$

où a est le mode du type de noyau $PA_{\theta(a,b)}$. En exprimant sa moyenne $ab/(b-1)$ pour $b>1$ et sa variance $a^2b/\{(b-1)^2(b-2)\}$ pour $b>2$ en fonction des composantes de $\theta(x,h)$, on obtient alors

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

Noyau log-normal

Nous proposons dans ce paragraphe l'utilisation du noyau associé lognormal, de moyenne $a \in \mathbb{R}$ et d'écart type $b > 0$, que nous notons $LN_{\theta(a,b)}$. Il est défini sur $\mathbb{S}_{LN} = [0, +\infty[$ par :

$$LN_{\theta(a,b)}(u) = \frac{1}{ub\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log(u) - a}{b}\right)^2\right\} \mathbf{1}_{[0,+\infty[}(u).$$

En utilisant son mode $\exp(a - b^2)$ et de son paramètre de dispersion b, on résout le système(2.9) pour avoir :

$$\theta(x, h) = \left(\log(x) + h^2, h\right) \text{ pour tout } x \in]0, +\infty[\text{ et } h > 0.$$

Ce qui conduit au noyau associé lognormal

$$LN_{\theta(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{1}{h}\log\left(\frac{u}{x}\right) - h\right)^2\right\} \mathbf{1}_{[0,+\infty[}(u)$$

défini sur

$$\mathbb{S}_{LN_{\theta(x,h)}} = [0, +\infty[= \mathbb{S}_{LN}.$$

Puis, en exprimant sa moyenne $e^{a+b^2/2}$ ainsi que sa variance $\left(e^{b^2} - 1\right)e^{2a+b^2}$ en fonction des solutions issues du système(2.9), on a :

$$A_\theta(x, h) = x\left(e^{(3h^2)/2} - 1\right) \quad \text{et} \quad B_\theta(x, h) = x^2 e^{(3h^2)}\left(e^{h^2} - 1\right).$$

Les constructions faites nous ont permis de savoir qu'en général le support du type de noyau et celui du noyau associé construit par le principe mode-dispersion coïncident sauf dans le cas des noyaux associés gamma inverse et gaussien inverse où le support du type de noyau et celui des noyaux associés diffèrent. Ceci est issue des contraintes liées à leurs paramétrisations $x < 1/(3h)$ pour le noyau associé gaussien inverse puis $x < 1/h$ dans le cas de gamma inverse.

Nous signalons aussi qu'il existe d'autres méthodes de construction qu'utilisent certains auteurs mais cela reste à leur discrétion. Nous pouvons pour cela citer les travaux de Jin & Kawczak (2003) sur les noyaux Birnbaum-Sanders et lognormal.

Nous résumons, dans les tableaux suivants, les différents résultats obtenus dans cette section.

Table 2.1 – Quelques types de noyaux (pour la suite voir Tableau 2.2)

	Bêta étendu	Bêta	Gamma	Gamma inverse
$\mathbb{S}_{\theta(a,b)}$	$[t_1, t_2], \quad 0 \leq t_1 < t_2$	$[0,1]$	$[0,\infty)$	$(0,\infty)$
$K_{\theta(a,b)}$	$\dfrac{(u-t_1)^{a-1}(t_2-u)^{b-1}}{\mathscr{B}(a,b)(t_2-t_1)^{a+b-1}}$	$\dfrac{u^{a-1}(1-u)^{b-1}}{\mathscr{B}(a,b)}$	$\dfrac{b^{-a}u^{a-1}\exp\left(-\dfrac{u}{b}\right)}{\Gamma(a)}$	$\dfrac{b^{a}u^{-(a+1)}\exp\left(-\dfrac{b}{u}\right)}{\Gamma(a)}$
$M(a,b)$	$\dfrac{(a-1)t_2+(b-1)t_1}{a+b-2}$	$\dfrac{a-1}{a+b-2}$	$(a-1)b$	$\dfrac{b}{a+1}$
$D(a,b)$	$(a+b-2)^{-1}$	$(a+b-2)^{-1}$	b	b^{-1}
$\mathbb{E}\left(\mathcal{Z}_{\theta(a,b)}\right)$	$t_1+\dfrac{a(t_2-t_1)}{a+b}$	$\dfrac{a}{a+b}$	ab	$\dfrac{b}{a-1}$
$Var\left(\mathcal{Z}_{\theta(a,b)}\right)$	$\dfrac{ab(t_2-t_1)^2}{(a+b)^2(a+b+1)}$	$\dfrac{ab}{(a+b)^2(a+b+1)}$	ab^2	$\dfrac{b^2}{(a-1)^2(a-2)}$

Table 2.2 – Quelques types de noyaux (suite du Tableau 2.1)

	Gaussien inverse	Gaussien inverse réciproque	Pareto	Lognormal
$\mathbb{S}_{\theta(a,b)}$	$(0,\infty)$	$(0,\infty)$	$[a,\infty)$	$(0,\infty)$
$K_{\theta(a,b)}$	$\dfrac{\sqrt{b}\exp\left\{-\frac{b}{2a}\left(\frac{u}{a}-2+\frac{a}{u}\right)\right\}}{\sqrt{2\pi t^3}}$	$\dfrac{\sqrt{b}\exp\left\{-\frac{b}{2a}\left(au-2+\frac{1}{au}\right)\right\}}{\sqrt{2\pi u}}$	$\dfrac{ba^b}{u^{b+1}}$	$\dfrac{1}{ub\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2}\left(\dfrac{\log(u)-a}{b}\right)^2\right\}$
$M(a,b)$	$a\left\{\left(1+\dfrac{9a^2}{4b^2}\right)^{1/2}-\dfrac{3a}{2b}\right\}$	$\dfrac{1}{a}\left\{\left(1+\dfrac{a^2}{4b^2}\right)^{1/2}-\dfrac{a}{2b^1}\right\}$	a	$\exp(a-b^2)$
$D(a,b)$	b^{-1}	b^{-1}	b^{-1}	b
$\mathbb{E}\left(\mathcal{Z}_{\theta(a,b)}\right)$	a	$a^{-1}+b^{-1}$	$\dfrac{ab}{b-1}$	$e^{a+b^2/2}$
$Var\left(\mathcal{Z}_{\theta(a,b)}\right)$	a^3/b	$1/(ab)+2/b^2$	$\dfrac{a^2b}{(b-1)^2(b-2)}$	$\left(e^{b^2}-1\right)e^{2a+b^2}$

Table 2.3 – Noyaux associés bêta et sa version étendue

$K_{\theta(x,h)}$	Bêta étendu	Bêta
$\mathbb{S}_{\theta(x,h)}$	$[t_1, t_2], \quad 0 \leq t_1 < t_2$	$[0, 1]$
$\theta(x,h)$	$\left(\frac{x-t_1}{(t_2-t_1)h}+1, \frac{t_2-x}{(t_2-t_1)h}+1\right)$	$\left(\frac{x}{h}+1, \frac{1-x}{h}+1\right)$
$A_\theta(x,h)$	$\frac{h\{(t_1+t_2)-2x\}}{1+2h}$	$\frac{h(1-2x)}{1+2h}$
$B_\theta(x,h)$	$\frac{h\{x-t_1+h(t_2-t_1)\}\{(t_2-x)+h(t_2-t_1)\}}{(1+2h)^2(1+3h)}$	$\frac{h\{x(1-x)+h+h^2\}}{(1+3h)(1+2h)^2}$

Table 2.4 – Quelques noyaux associés non-classiques

$K_{\theta(x,h)}$	Gamma	Gamma inv.	Gaussien inv.	Gaussien inv. Rec.	Pareto	Lognormal		
$\mathbb{S}_{\theta(x,h)}$	$(0,\infty)$	$(0,\infty)$	$(0,\infty)$	$(0,\infty)$	$[x,\infty)$	$(0,\infty)$		
$\theta(x,h)$	$\left(\frac{x}{h}+1,h\right)$	$\left(\frac{1}{xh}-1,\frac{1}{h}\right)$	$\left(\frac{x}{\sqrt{1-3xh}},\frac{1}{h}\right)$	$\left(\frac{1}{\sqrt{x^2+xh}},\frac{1}{h}\right)$	$\left(x,\frac{1}{h}\right)$	$(\log(x)+h^2,h)$		
$A_\theta(x,h)$	h	$\frac{2x^2h}{1-2xh}$	$x\left\{\frac{1}{\sqrt{1-3xh}}-1\right\}$	$\sqrt{x^2+xh}-x+h$	$\frac{xh}{1-h}$	$x\left(e^{(3h^2)/2}-1\right)$		
$B_\theta(x,h)$	$h(x+h)$	$\frac{x^3h(1-3xh)^{-1}}{(1-2xh)^2}$	$\frac{x^3h}{(1+3xh)^{3/2}}$	$h\{\sqrt{x^2+xh}+2h\}$	$\frac{(xh)^2(1-2h)^{-1}}{(1-h)^2}$	$x^2e^{(3h^2)}\left(e^{h^2}-1\right)$		
$r_2(x)$	$1/2$	$1/2$	$1/2$	$1/2$	$1/2$	$1/2$	1	1

Maintenant que nous savons nettement construire les noyaux associés $K_{\theta(x,h)}$, nous sommes en mesure de définir avec précision l'estimateur à noyau associé continu pour une densité f inconnue sur un compact $\mathbb{T}_I$. C'est l'objet de la section suivante.

2.2 Estimateurs à noyaux associés continus

Cette section se propose d'étudier finement les estimateurs à noyaux associés continus non-classiques. Elle s'organise de la manière suivante. Nous proposons d'abord la définition des estimateurs à noyaux associés puis nous examinons point par point leurs propriétés comparativement à celles des estimateurs présentés dans le chapitre précédent. En commençant par les biais, variance et risque quadratique. Ensuite, nous proposons une technique de réduction de biais (lorsque celui-ci n'est pas de même ordre que dans le cas classique). Enfin, nous illustrons avec les noyaux associés précédemment construits.

2.2.1 Caractéristiques

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

Définition 2.2.1 *Soit f la densité à estimer sur $\mathbb{T}_I \subseteq \mathbb{R}$, $h > 0$ et $K_{\theta(x,h)}$ un noyau associé continu. L'estimateur $\widehat{f_n}$ de f issu de $K_{\theta(x,h)}$ est défini par :*

$$\widehat{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\theta(x,h)}(X_i). \tag{2.15}$$

Cet estimateur est similaire à celui défini en (2.2) sauf qu'ici $K_{\theta(x,h)}$ est construit par la méthode mode-dispersion. Donnons maintenant les premières propriétés de cet estimateur.

Proposition 2.2.2 *L'estimateur défini en (2.15) vérifie les propriétés suivantes*

$$\mathbb{E}\left\{\widehat{f_n}(x)\right\} = \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\}, \tag{2.16}$$

$$\widehat{f_n}(x) \geqslant 0, \tag{2.17}$$

$$\int_{\mathbb{T}_I} \widehat{f_n}(x)dx =: \Lambda(n,h,K) \text{ n'est pas toujours égale } 1, \tag{2.18}$$

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

La remarque suivante explicite la dernière propriété de cette Proposition.

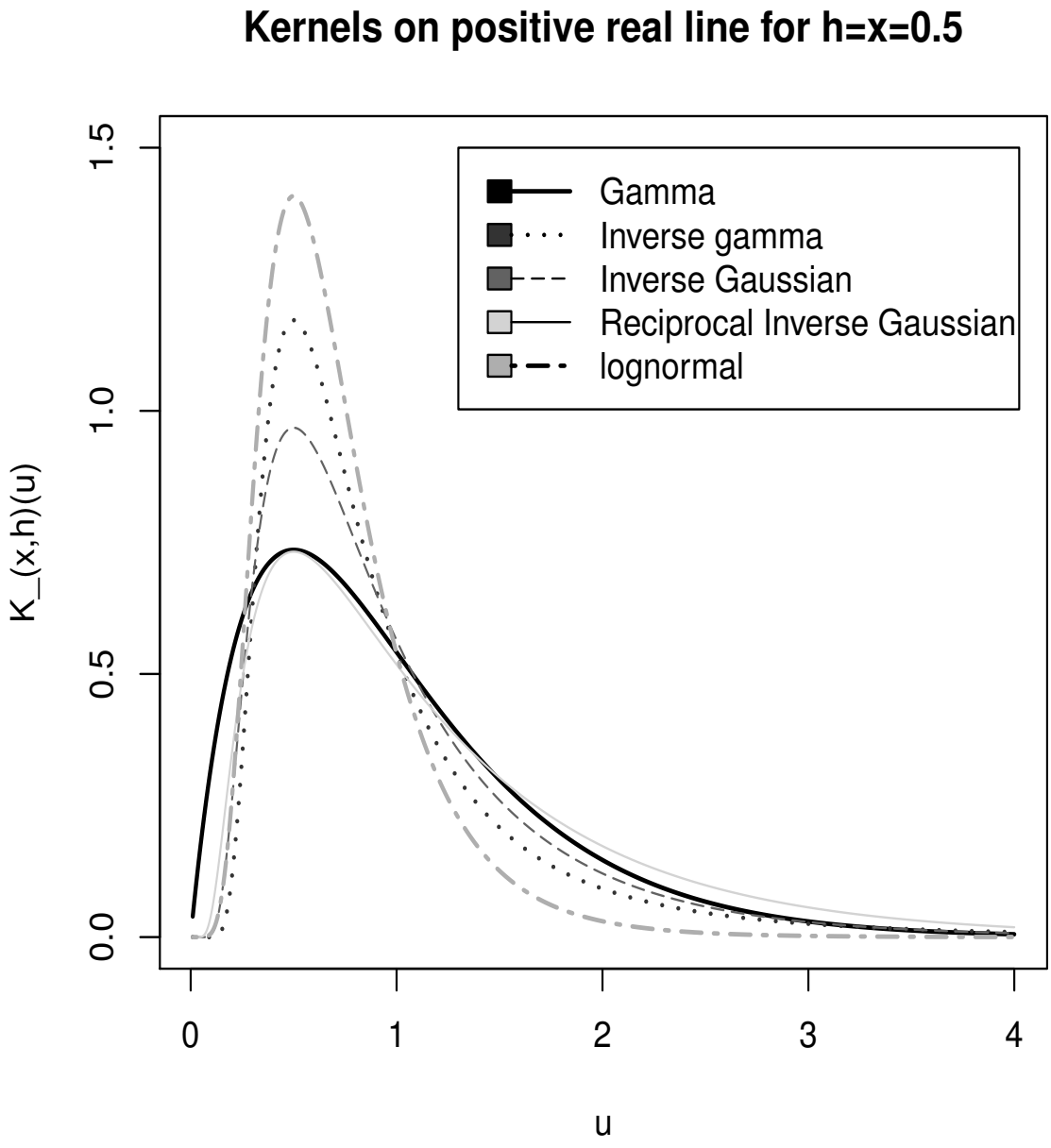

FIGURE 2.4 – Comportements de quelques noyaux associés non-classiques de la droite réelle positive aux bords.

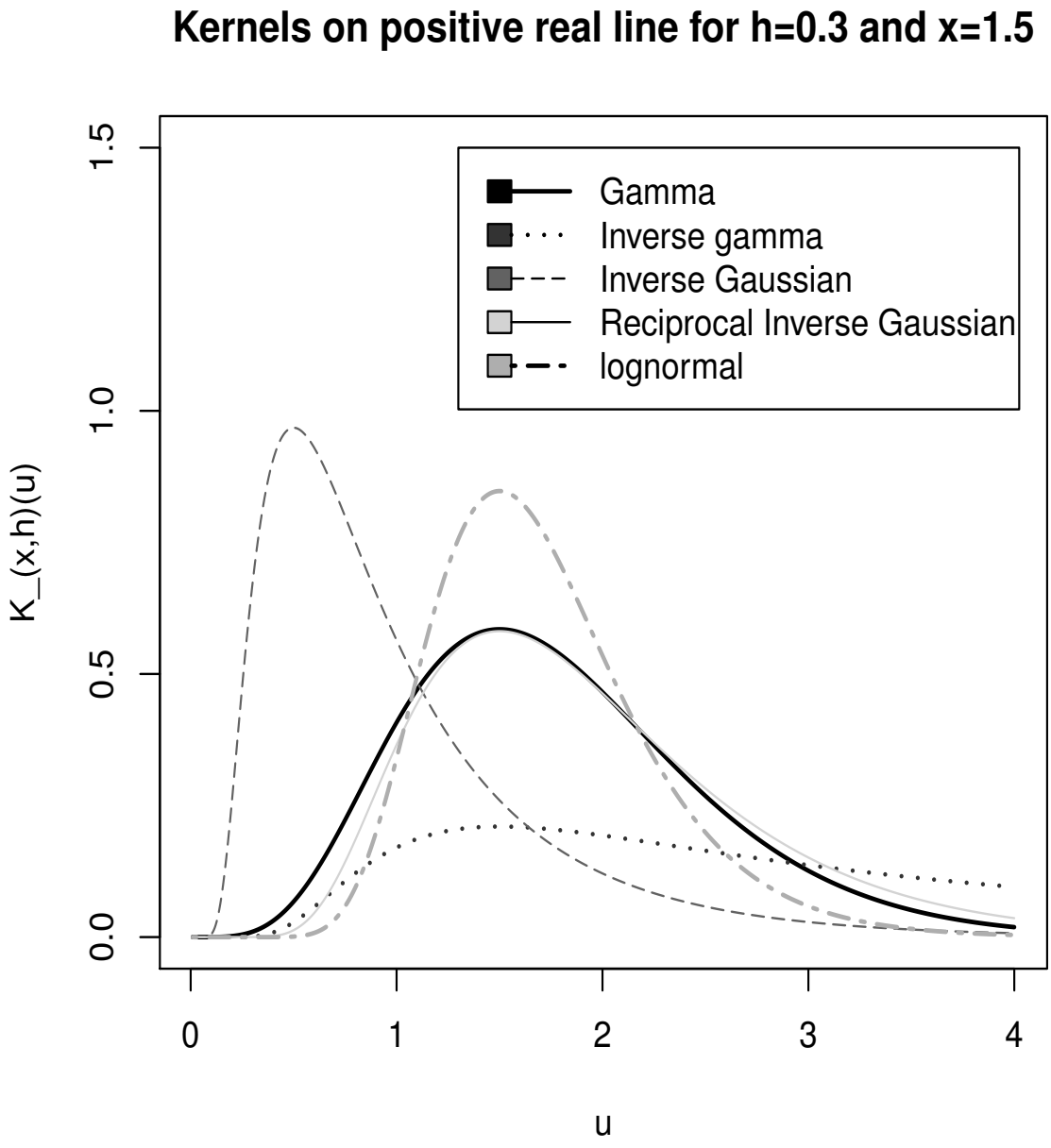

FIGURE 2.5 – Comportements de quelques noyaux associés non-classiques de la droite réelle positive à l'intérieur.

Remarque 2.2.3 *La troisième propriété de la Proposition 2.2.2 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.*

$$\int_{\mathbb{T}_I} K_{\theta(x,h)}(u)dx < \infty \text{ pour tout } u \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}_I.$$

Cependant, on peut facilement vérifier que $\Lambda = 1$ dans le cas des noyaux classiques.

Démonstration de la Proposition 2.2.2. La première propriété s'obtient à partir de la définition de $\mathbb{E}\left\{\hat{f}_n(x)\right\}$ pour tout $x \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}_I$ et $\mathcal{Z}_{\theta(x,h)}$ une variable aléatoire de loi $K_{\theta(x,h)}$, comme suit :

$$\mathbb{E}\left\{\hat{f}_n(x)\right\} = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}_I} K_{\theta(x,h)}(t)f(t)dt = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}_I} f(t)K_{\theta(x,h)}(t)dt = \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\}.$$

Pour la deuxième propriété, partant du fait que $K_{\theta(x,h)}$ est une densité de probabilité, on a $K_{\theta(x,h)}(X_i) \in [0,1]$, pour tout $x \in \mathbb{S}_{\theta(x,h)} \cap \mathbb{T}_I$, $X_i \in \mathbb{T}_I$ et $h > 0$. Puisque les $\left(K_{\theta(x,h)}(X_i)\right)_i$ sont aussi iid, alors $\frac{1}{n}\sum_{i=1}^{n} K_{\theta(x,h)}(X_i) \in [0,1]$. Quant à la dernière propriété, elle se déduit de la Remarque 2.2.3 dans le cas des noyaux associés classiques, puis du Tableau 2.5 dans le cas non-classique.■

Nous donnons dans le Tableau 2.5 les différentes valeurs de Λ dans des cas particuliers des noyaux associés bêta, gamma, gamma inverse, lognormal, gaussien inverse et sa réciproque. Dans chaque cas, nous calculons la masse totale de l'estimateur sur le support $\mathbb{T}_I$ correspondant en utilisant cinq échantillons de variables aléatoires de taille $n = 1000$ suivant la loi normale tronquée sur $\mathbb{T}_I$. Signalons tout de même qu'une étude similaire été menée par Gourieroux & Monfort (2006) dans le cas particulier de bêta et ils ont abouti à la même conclusion.

Table 2.5 – Quelques valeurs de $\Lambda(n,h,K) \simeq 1$ de la Proposition 2.2.2 pour $h = 0.05$ et $n = 1000$

Type de noyauu	*Echan* 1	*Echan* 2	*Echan* 3	*Echan* 4	*Echan* 5
Bêta	1.028727	1.037655	1.042822	1.045671	1.046681
Gamma	0.978753	0.982778	1.009463	1.056642	0.912753
Gamma inverse	0.970302	0.922467	0.976362	0.978275	0.967052
Gaussien inverse	1.247213	0.978136	1.034702	1.003275	1.062435
Gaussien inv. Rec.	0.937992	0.998741	0.993740	0.990672	0.983422
Lognormal	1.056480	0.947233	0.887284	1.139463	0.874912

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

Proposition 2.2.4 *Soit $\widehat{f_n}$ l'estimateur à noyau associé (2.15) de f appartenent à $\mathscr{C}^2(\mathbb{T}_I)$. Pour tout x dans $\mathbb{T}_I$ et $h = h_n > 0$, alors*

$$Biais\left\{\widehat{f_n}(x)\right\} = A_\theta(x,h)f'(x) + \frac{1}{2}\left\{A_\theta^2(x,h) + B_\theta(x,h)\right\}f''(x) + o(h^2). \tag{2.19}$$

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

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{n}f(x)\left\|K_{\theta(x,h)}\right\|_2^2 + o\left(\frac{1}{nh^{r_2}}\right), \tag{2.20}$$

où $r_2 = r_2\left(K_{\theta(x,h)}\right) > 0$ est le plus grand réel tel que $\left\|K_{\theta(x,h)}\right\|_2^2 = \int_{\mathbb{S}_{\theta(x,h)}\cap\mathbb{T}} K^2_{\theta(x,h)}(u)du \leq c_2(x)h_n^{-r_2}$ et $0 < c_2(x) < \infty$.

Signalons que les estimateurs à noyaux associés continus sont aussi (comme dans le cas classique), asymptotiquement sans biais (grâce à la Définition 2.1.3). Cependant, l'expression du biais donné dans cette proposition diffère de celui introduit en (1.4) 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. Pour y remédier, un algorithme de réduction de biais est proposé dans la section 2.2.2.

Démonstration de la Proposition 2.2.4. Pour tout $x \in \mathbb{T}_I$ et $r > 0$, on obtient le biais de $\widehat{f_n}$ en procédant de la manière suivante :

$$\begin{aligned}
Biais\left(\widehat{f_n}(x)\right) &\overset{(i)}{=} \mathbb{E}\left(f(\mathcal{Z}_{\theta(x,h)})\right) - f(x) \\
&\overset{(ii)}{=} f\left\{\mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\} + \frac{1}{2}Var(\mathcal{Z}_{\theta(x,h)})f''\left\{\mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\} - f(x) \\
&\quad + o\left(\mathbb{E}\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right) \\
&\overset{(iii)}{=} f\left\{x + A_\theta(x,h)\right\} + \frac{1}{2}B_\theta(x,h)f''\left\{x + A_\theta(x,h)\right\} - f(x) \\
&\quad + o\left\{B_\theta(x,h)\right\} \\
&\overset{(iv)}{=} A_\theta(x,h)f'(x) + \frac{1}{2}\left\{A_\theta^2(x,h) + B(x,h)\right\}f''(x) + o(h^2).
\end{aligned}$$

L'égalité (i) vient de la Proposition 2.2.2. Le (ii) est obtenu après développement de $f(\mathcal{Z}_{\theta(x,h)})$ en séries de Taylor au voisinage de $\mathbb{E}(\mathcal{Z}_{\theta(x,h)})$ à l'ordre deux. Pour le (iii), on remplace dans (ii) successivement, $f(\mathcal{Z}_{\theta(x,h)})$ par son approximation de Taylor, puis $\mathbb{E}(\mathcal{Z}_{\theta(x,h)})$ par son expression dans la Définition 2.1.3. Enfin, le (iv) s'obtient en approximant $f\left(x + A_\theta(x,h)\right)$ 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 (2.8) et $o\left(\mathbb{E}\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right) = \mathbb{E}\left(o_P\left\{\mathcal{Z}_{\theta(x,h)} - \mathbb{E}(\mathcal{Z}_{\theta(x,h)})\right\}^2\right)$ où $o_P(\cdot)$ est le taux de convergence en probabilité. Concer-

nant la variance on a

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

A partir de (2.16) et (2.19) le second terme se comporte de la manière suivante $I_2 := (1/n)\left[\mathbb{E}\left\{K_{\theta(x,h)}(X_1)\right\}\right]^2 \simeq (1/n)f^2(x) \simeq O(1/n)$ puisque f est bornée pour tout $x \in \mathbb{T}_I$. En utilisant le développement de f en séries de Taylor aux voisinages de x, le premier terme I_1 donne

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

avec

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

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

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

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

Erreurs quadratiques ponctuelles et globales

Comme nous l'avons dit dans le deuxième paragraphe de la Section 1.1 du chapitre précédent, le critère naturel qui permet d'évaluer la similarité de l'estimateur $\widehat{f_n}$ par rapport à la vraie densité f à estimer est le *MSE* dans le cas ponctuel et le *MISE* dans le cas global. Nous les utilisons une fois de plus dans ce paragraphe, pour mesurer l'efficacité de l'estimateur $\widehat{f_n} = \widehat{f_{n,h,K,f}}$ obtenu à l'aide des noyaux associés continus. Nous rappelons ici que le *MSE* et le *MISE* de l'estimateur à noyaux associés continus sont comme dans le cas des noyaux associés classiques (1.5) et (1.8), définis respectivement par

$$MSE(x) = Var\left\{\widehat{f_n}(x)\right\} + Biais^2\left\{\widehat{f_n}(x)\right\}, \tag{2.21}$$

et

$$MISE(n,h,K,f) = \int_{\mathbb{T}_I} Var\left\{\widehat{f_n}(x)\right\}dx + \int_{\mathbb{T}_I} Biais^2\left\{\widehat{f_n}(x)\right\}dx. \tag{2.22}$$

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

$$AMSE = \left[A_\theta(x,h)f'(x) + \frac{1}{2}\left\{A^2_\theta(x,h) + B_\theta(x,h)\right\}f''(x)\right]^2 + \frac{1}{n}f(x)\left\|K_{\theta(x,h)}\right\|_2^2, \tag{2.23}$$

et

$$AMISE(\widehat{f_{n,h,K,f}}) = \int_{\mathbb{T}_I} \left(\left[A_\theta(x,h) f'(x) + \frac{1}{2}\{A_\theta^2(x,h) + B_\theta(x,h)\} f''(x) \right]^2 \right) dx$$
$$+ \int_{\mathbb{T}_I} \frac{1}{n} \left\| K_{\theta(x,h)} \right\|_2^2 f(x) dx. \tag{2.24}$$

La proposition suivante donne la vitesse de convergence de l'estimateur $\widehat{f_n}$ au sens de *AMISE*.

Proposition 2.2.5 *Soit $f \in \mathscr{C}^2(\mathbb{T}_I)$, la densité à estimer. Si les dérivées premières et secondes de f sont bornées alors la fenêtre optimale de lissage qui minimise le AMISE est*

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

avec $r_2 = r_2\left(K_{\theta(x,h)}\right)$ défini dans (2.20).

Démonstration. La Définition 2.1.3 et la Proposition 2.1.5 permettent de dire qu'il existe deux constantes positives $c_1^*(x)$ et $c_1^{**}(x)$ telles que $A_\theta(x,h) \leq c_1^*(x)h$ et $B_\theta(x,h) \leq c_1^{**}(x)h^2$. D'après la Proposition 2.2.4, on a :

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

avec $c_1(x) = \sup_{x\in\mathbb{T}} |c_1^*(x)f'(x) + c_1^{**}(x)f''(x)|$. De (2.21), on déduit que

$$MSE(x) \leq h^2 c_1^2(x) + n^{-1}h^{-r_2}c_2(x). \tag{2.25}$$

En intégrant (2.25), on obtient

$$AMISE(\widehat{f_{n,h,K,f}}) \leq h^2 C_1(x) + n^{-1}h^{-r_2}C_2(x)$$

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

La valeur de h ainsi obtenue, fournit la vitesse de convergence de l'estimateur à noyau associé continu. Pour les noyaux classiques, cette notion a été aussi sujet des travaux de beaucoup d'auteurs dont Sarda & Vieu (1988).

Il y a lieu ici de constater que la présence du terme $A_\theta(x,h)f'(x)$ avec $A_\theta(x,h) \neq 0$ dans l'expression du biais de $\widehat{f_{n,h,K,f}}$ augmente les erreurs de ce dernier. D'où nécessité de la réduction de biais, plus précisément l'annulation de $A_\theta(x,h)$. Ainsi, nous proposons dans le prochain paragraphe un algorithme permettant d'annuler $A_\theta(x,h)$ dans la plus grande partie du support $\mathbb{T}_I$; cette procédure est inspirée de Chen (1999, 2000, 2010).

2.2.2 Réduction de biais et noyaux associés modifiés

Pour réduire le biais de $\widehat{f_{n,h,K,f}}$ défini en (2.19), 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.

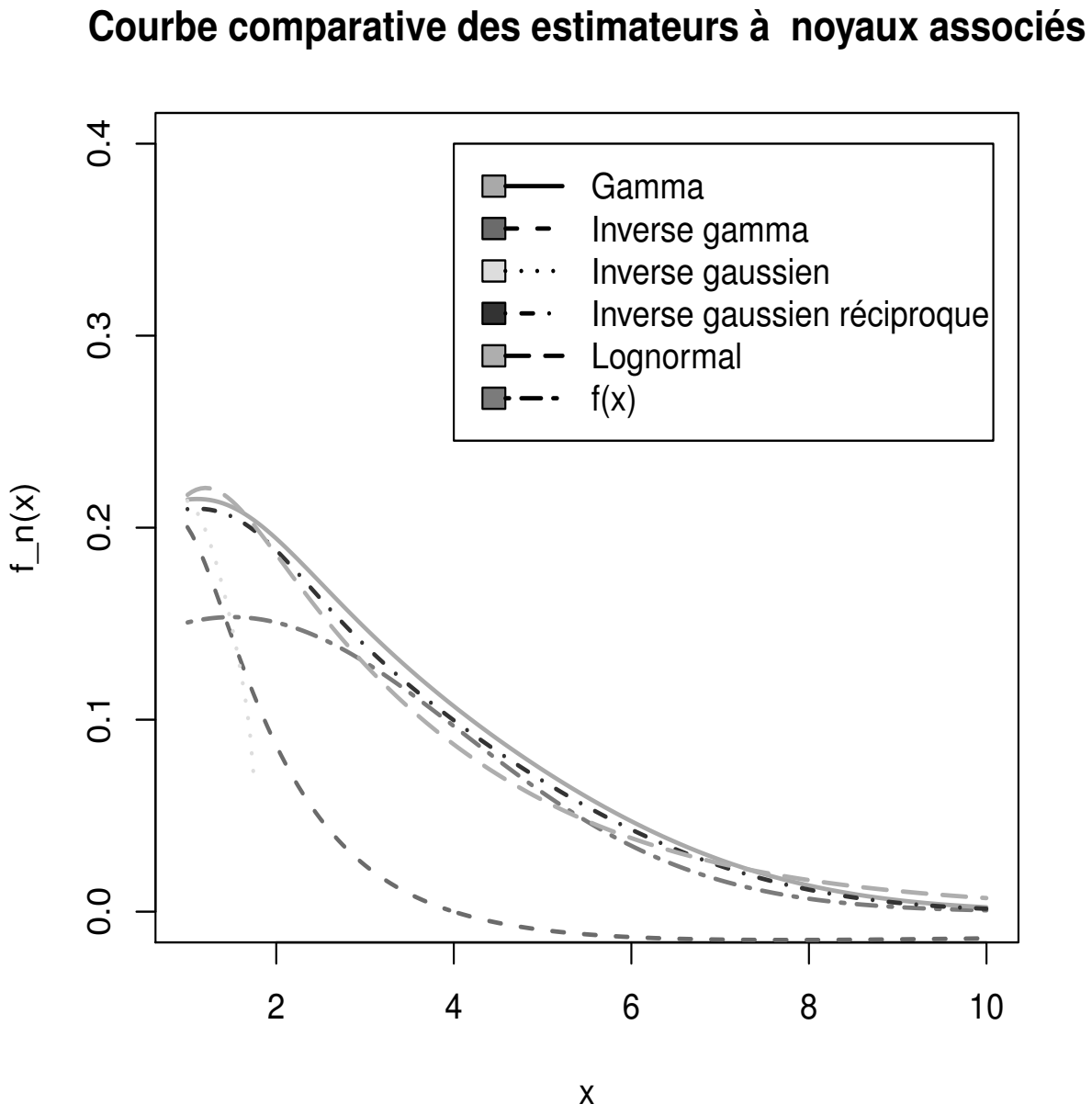

FIGURE 2.6 – Comportements des estimateurs noyaux associés non-classiques de la droite réelle positive.

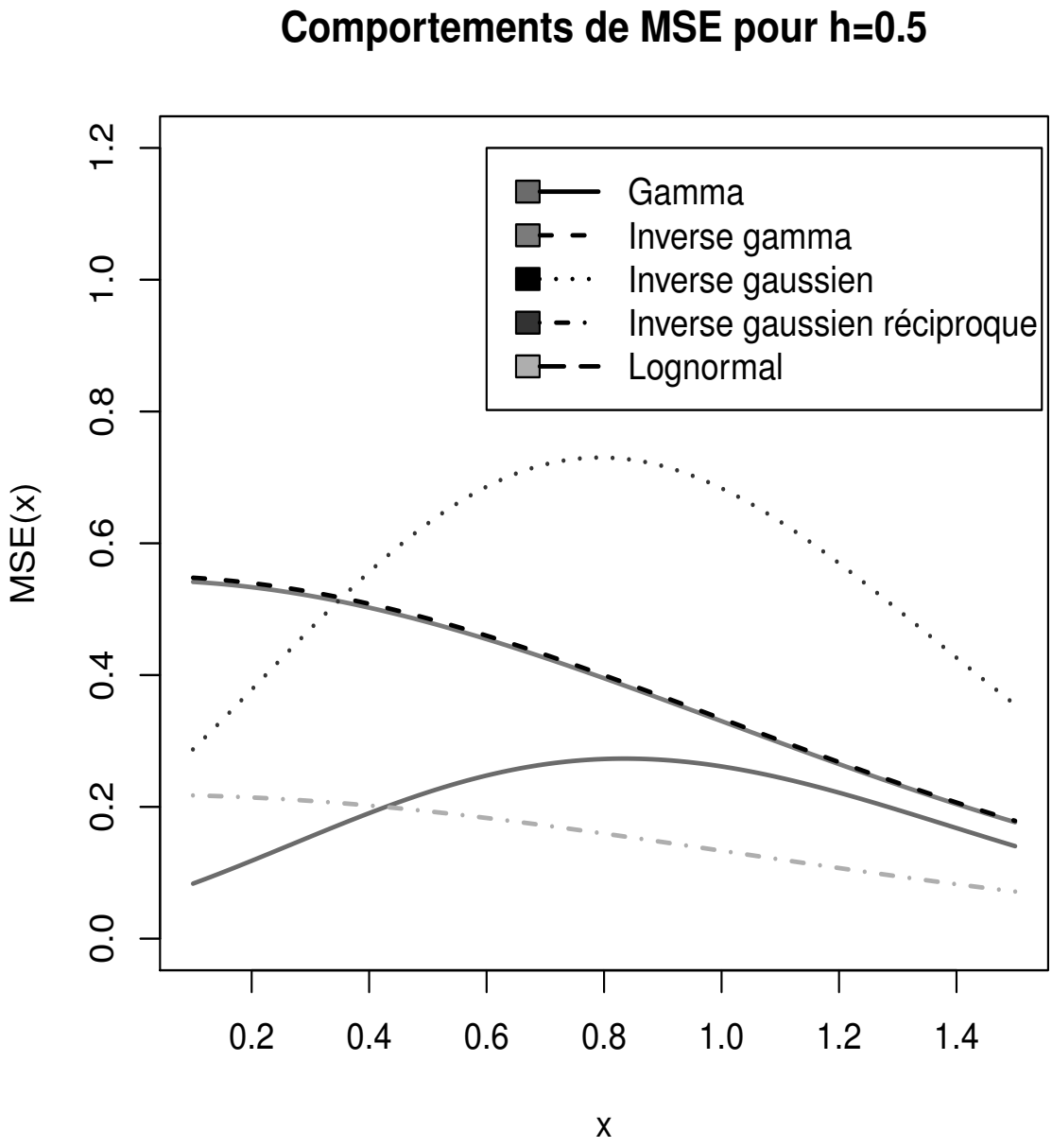

FIGURE 2.7 – Comportements des MSE aux bords dans l'estmation de la densité normale tronquée.

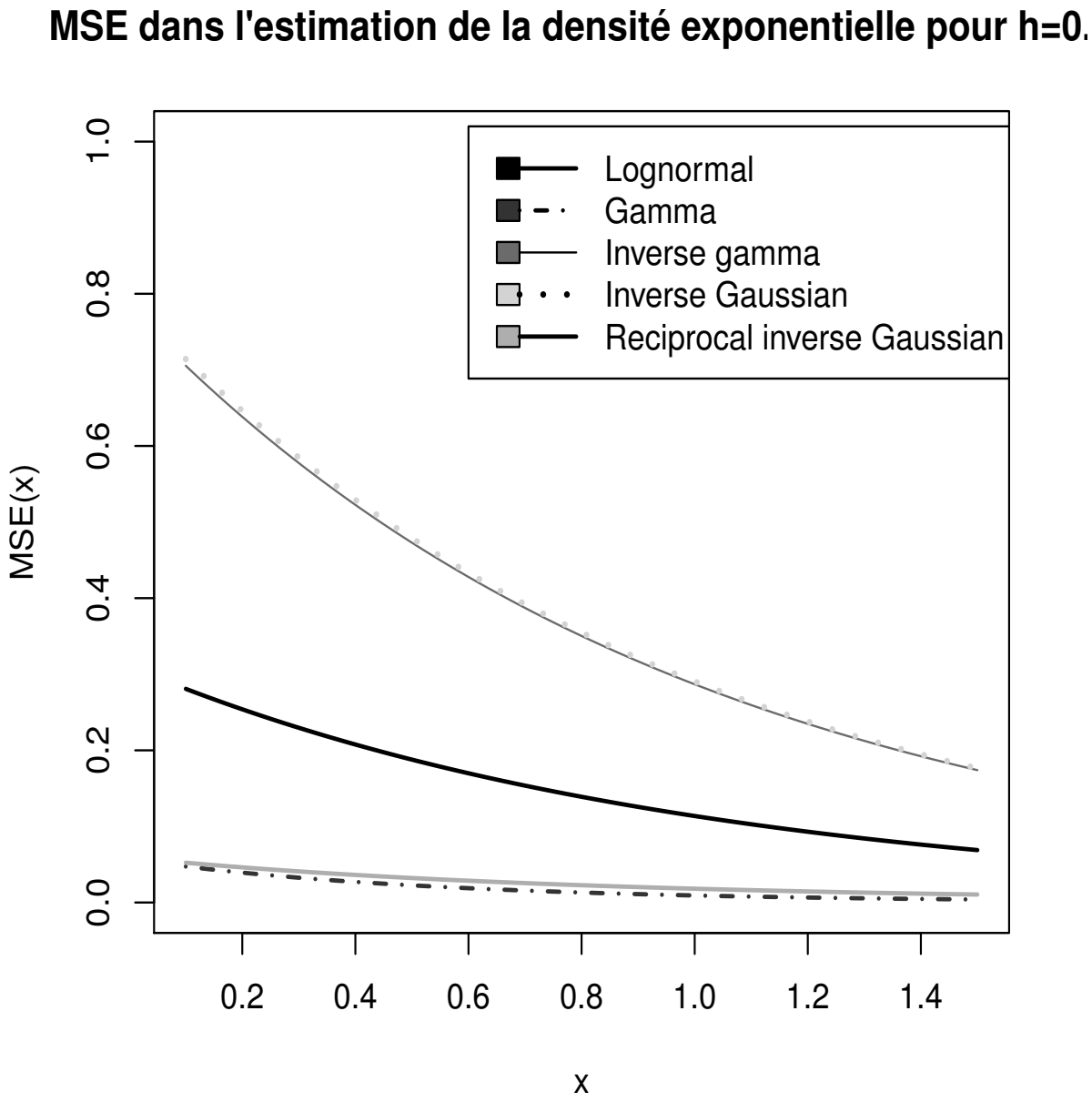

FIGURE 2.8 – Comportements des MSE aux bords dans l'estmation de la densité exponentielle.

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 $\mathbb{T}_I = [t_1, t_2]$ en deux régions d'ordre $\alpha(h) > 0$ avec $\alpha(h) \to 0$ quand $h \to 0$;

(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

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

et

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

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

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

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 $K_{\widetilde{\theta}(x,h)}$ correspondant à $\widetilde{A}(x,h) = \left(\widetilde{A}_{-1}(x,h), \widetilde{A}_0(x,h), \widetilde{A}_{+1}(x,h)\right)$ et $\widetilde{B}(x,h) = \left(\widetilde{B}_{-1}(x,h), \widetilde{B}_0(x,h), \widetilde{B}_{+1}(x,h)\right)$ de sorte que, pour un h fixé,

$$\widetilde{\theta}(x,h) = \left(\widetilde{\theta}_{-1}(x,h), \widetilde{\theta}_0(x,h), \widetilde{\theta}_{+1}(x,h)\right) = \begin{cases} \widetilde{\theta}_{-1}(x,h) & & \text{si } x \in \mathbb{T}_{\alpha(h),-1} \\ \widetilde{\theta}_0(x,h) : & \widetilde{A}_0(x,h) = 0 & \text{si } x \in \mathbb{T}_{\alpha(h),0} \\ \widetilde{\theta}_{+1}(x,h) & & \text{si } x \in \mathbb{T}_{\alpha(h),+1} \end{cases} \tag{2.26}$$

soit continue sur $\mathbb{T}_I$ et constant sur $\mathbb{T}^c_{\alpha(h),0}$ voir Chen (2010). Nous présentons quelques illustrations dans la Section 4.

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

Proposition 2.2.6 *La fonction $K_{\widetilde{\theta}(x,h)}$ issue de (2.26) est un noyau associé de même type $K = K_{\theta(x,h)}$.*

Démonstration. Nous montrons ici que $K_{\widetilde{\theta}(x,h)}$ satisfait les conditions de la Définition 2.1.3. 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 $j \in J = \{-1, 0, +1\}$,

$$x \in \mathbb{T}_{\alpha(h),j} \Rightarrow x \in \mathbb{T}_I \Rightarrow x \in \mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\widetilde{\theta}(x,h)}.$$

Les deux derniers résultats découlent de la Proposition 2.1.6. C'est-à-dire que pour toute variable aléatoire $\mathcal{Z}_{\widetilde{\theta}(x,h)}$ de densité $K_{\widetilde{\theta}(x,h)}$ on a

$$\mathbb{E}(\mathcal{Z}_{\widetilde{\theta}(x,h)}) = x + A_{\widetilde{\theta}(x,h)} \text{ et } Var(\mathcal{Z}_{\widetilde{\theta}(x,h)}) = B_{\widetilde{\theta}(x,h)}$$

avec $A_{\widetilde{\theta}(x,h)}$ et $B_{\widetilde{\theta}(x,h)}$ qui tendent vers 0 quand h tend vers 0. En prenant $\widetilde{A}(x,h) = A_{\widetilde{\theta}(x,h)}$ et $\widetilde{B}(x,h) = B_{\widetilde{\theta}(x,h)}$ il vient que pour tout $j \in J$, $\widetilde{A}_j(x,h) = A_{\widetilde{\theta}_j(x,h)}$ et $\widetilde{B}_j(x,h) = B_{\widetilde{\theta}_j(x,h)}$. En particulier, à partir de (2.26) on obtient $A_{\widetilde{\theta}_0(x,h)} = 0$.■

Ainsi, à partir de la Définition 2.2.1 puis en utilisant le noyau associé modifié $K_{\widetilde{\theta}(x,h)}$, on définit l'*estimateur à noyau associé modifié* $\widetilde{f_n}$ de f par :

$$\widetilde{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{\widetilde{\theta}(x,h)}(X_i). \tag{2.27}$$

La proposition suivante donne les expressions du biais et la variance de $\widetilde{f_n}$.

Proposition 2.2.7 *Soit $\mathbb{T}_I = \cup_{j\in J}\mathbb{T}_{\alpha(h),j}$ ($J = \{-1, 0, +1\}$), le support de la densité f à estimer, $\widehat{f_n}$ et $\widetilde{f_n}$ les estimateurs à noyaux associés continus de f donnés respectivement définis en (2.15) et (2.27). Pour tout $x \in \mathbb{T}_{\alpha(h),0}$ on a*

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

et

$$Var\left\{\widetilde{f_n}(x)\right\} \simeq Var\left\{\widehat{f_n}(x)\right\} \text{ quand } h \to 0.$$

Démonstration. On obtient le premier résultat en remplaçant dans (2.19) respectivement, $\widehat{f_n}$, A_θ et B_θ par $\widetilde{f_n}$, $\widetilde{A}_0$ et $\widetilde{B}_0$. Pour le second résultat, il suffit de montrer que

$$\left\|K_{\theta(x,h)}\right\|_2^2 \simeq \left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \text{ quand } h \to 0.$$

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

$$\left\|K_{\theta(x,h)}\right\|_2^2 \le c_2(x)h^{-r} \text{ et } \left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \le \widetilde{c}_2(x)h^{-r},$$

avec $0 < c_2(x), \widetilde{c}_2(x) < \infty$. En prenant $c(x) = \sup\{c_2(x), \widetilde{c}_2(x)\}$ on a $\left\|K_{\theta(x,h)}\right\|_2^2 \le c(x)h^{-2r}$ et $\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \le c(x)h^{-2r}$. Puisque $c(x)/nh^{2r} = o(n^{-1}h^{-2r})$ alors $\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 \simeq \left\|K_{\theta(x,h)}\right\|_2^2$.■

Contrairement aux cas classiques, le choix des noyaux associés non-classiques est très important. D'abord, il dépend d'une connaissance *a priori* du support $\mathbb{T}_I$ de la densité à estimer, qui doit coïncider avec le support $\mathbb{S}_\theta$ du noyau associé (non-classique). De plus, ce choix doit tenir compte de comportement du noyau associé selon les différentes positions de la cible x sur $\mathbb{T}_I$. Dans le prochain paragraphe, nous proposons un critère pour le choix du noyau associé continu.

Choix du noyau

Nous proposons dans ce paragraphe, un critère pour le choix du noyau essentiellement dans la partie la plus importante du support $\mathbb{T}_I$ en l'occurrence $\mathbb{T}_{\alpha(h),0}$. Ce critère est basé sur la comparaison des erreurs quadratiques moyennes intégrées asymptotiques i.e. le *MISE*. D'après la proposition précédente, nous pouvons écrire l'erreur quadratique moyenne ponctuelle de $\widetilde{f_n}$ sur $\mathbb{T}_{\alpha(h),0}$ notée $AMSE_0$, de la manière suivante :

$$AMSE_0(x) = \frac{1}{4}B^2_{\widetilde{\theta}_0}(x,h)\{f''(x)\}^2 + \frac{1}{n}\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 f(x)$$

par conséquent le $AMISE_0$ est

$$AMISE_0(\widetilde{f_{n,h,K,f}}) = \int_{\mathbb{T}_{\alpha(h),0}} \left[\frac{1}{4}B^2_{\widetilde{\theta}_0}(x,h)\{f''(x)\}^2 + \frac{1}{n}\left\|K_{\widetilde{\theta}(x,h)}\right\|_2^2 f(x)\right] dx.$$

Ainsi, pour des noyaux associés de même support $\mathbb{S}$, le noyau optimal (le noyau qui estime mieux) noté $K_{\theta\, opt}$ dans la région intérieure de $\mathbb{T}_I$ (i.e. sur $\mathbb{T}_{\alpha(h),0}$) est celui qui a le plus petit $AMISE_0$ pour tout $h > 0$. On écrit alors

$$K_{\theta\, opt} = \underset{K_\theta \text{ sur } \mathbb{S}}{\arg\min}\, AMISE_0(\widetilde{f_{n,h,K}}).$$

Remarque 2.2.8 *Lorsque* $\mathbb{T} = \mathbb{R}$, *ceci coïncide avec le cas classique et le noyau optimal est celui d'Epachnikov.*

2.2.3 Illustration sur les noyaux associés non-classiques

Dans cette section, nous présentons dans chacun des cas, le calcul des biais et variance ponctuels de l'estimateur à noyaux associés continus définis dans la Proposition 2.2.4, avec application sur la réduction de biais en passant par la modification du noyau associé construit par le principe mode-dispersion (voir Section 2.2.2). Pour une meilleure compréhension, nous rappelons à chaque fois les ingrédients nécessaires pour ces calculs.

Estimateurs de densités à noyau bêta étendu

Le noyau associé bêta étendu construit dans la Section 2.1.3 est approprié pour l'estimation des densités f ayant pour support $\mathbb{T}_I = \mathbb{S}_{EB} = [t_0, t_1]$. Considérons $\widehat{f_{n;t_0,t_1}}$ un estimateur de f par le noyau associé bêta étendu. Par utilisation de la Proposition 2.2.4, le calcul de son biais donne :

$$\begin{aligned} Biais\left\{\widehat{f_{n;t_0,t_1}}(x)\right\} &= \frac{h\{(t_0+t_1)-2x\}}{1+2h}f'(x) + \frac{1}{2}\left[\frac{h\{(t_0+t_1)-2x\}}{1+2h}\right]^2 f''(x) \\ &+ \frac{\{x-t_0+h(t_1-t_0)\}\{(t_0-x)+h(t1-t_0)\}}{2(1+2h)^2(1+3h)}f''(x) + o(h^2). \end{aligned}$$

Ce biais est grand par la présence du terme non nul en f' et nous allons l'éliminer en utilisant l'algorithme de réduction de biais proposé dans la Section 2.2.2. Premièrement, nous divisons le support $\mathbb{T}_I$ respectivement en $\mathbb{T}_{\alpha(h),-1} = [t_0, t_0 + \alpha(h)]$, $\mathbb{T}_{\alpha(h),0} =]t_0 + \alpha(h), t_1 - \alpha(h)[$ et $\mathbb{T}_{\alpha(h),+1} = [t_1 - \alpha(h), t_1]$. Dans un second temps, nous considérons la fonction $\psi : \mathbb{T} \to \mathbb{T}$ définie par

$$\psi(z) = \alpha(h)\{z - \alpha(h) + 1\} \text{ pour tout } z \geq 0, \tag{2.28}$$

puis nous construisons $\widetilde{\theta}(x,h) = \widetilde{\theta}(x,h;t_1,t_2)$ comme suit :

$$\widetilde{\theta}(x,h;t_0,t_1) = \begin{cases} \left(\dfrac{\psi(x-t_0)}{(t_1-t_0)h}, \dfrac{x-t_0}{(t_1-t_0)h}\right) & \text{si } x \in [t_0, t_0+\alpha(h)] \\ \left(\dfrac{x-t_0}{(t_1-t_0)h}, \dfrac{t_1-x}{(t_1-t_0)h}\right) & \text{si } x \in]t_0+\alpha(h), t_1-\alpha(h)[\\ \left(\dfrac{t_1-x}{(t_1-t_0)h}, \dfrac{\psi(t_1-x)}{(t_1-t_0)h}\right) & \text{si } x \in [t_1-\alpha(h), t_1]. \end{cases}$$

En exprimant les moyenne et variance de $EB_{\widetilde{\theta}(x,h,t_0,t_1)}$, on obtient des quantités $\widetilde{A}(x,h;t_0,t_1)$ et $\widetilde{B}(x,h;t_0,t_1)$ données par :

$$\widetilde{A}(x,h,t_0;t_1) = \begin{cases} \dfrac{x(x-t_0)+(1-x)\psi(x-t_0)}{x-t_0+\psi(x-t_0)} & \text{si } x \in [t_0, t_0+\alpha(h)] \\ 0 & \text{si } x \in]t_0+\alpha(h), t_1-\alpha(h)[\\ \dfrac{x(t_1-x)+(1-x)\psi(t_1-x)}{t_1-x+\psi(t_1-x)} & \text{si } x \in [t_1-\alpha(h), t_1], \end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{h(x-t_0)\psi(x-t_0)\{x-t_0+\psi(x-t_0)\}^{-2}}{\{x-t_0+\psi(x-t_0)+h\}} & \text{si } x \in [t_0, t_0+\alpha(h)] \\ \dfrac{h(x-t_0)(t_1-x)}{1+h} & \text{si } x \in]t_0+\alpha(h), t_1-\alpha(h)[\\ \dfrac{h(t_1-x)\psi(t_1-x)\}\{(t_1-x)+\psi(t_1-x)\}^{-2}}{\{(t_1-x)+\psi(t_1-x)+h\}} & \text{si } x \in [t_1-\alpha(h), t_1]. \end{cases}$$

Ainsi, le biais réduit de $\widetilde{f}_{n;t_0,t_1}$ sur $]t_0+\alpha(h), t_1-\alpha(h)[$ est donné par :

$$Biais\left\{\widetilde{f}_{n;t_0,t_1}(x)\right\} = \frac{1}{2}h(x-t_0)(t_1-x)f''(x) + o(h^2).$$

Le calcul de la variance de $\widetilde{f}_{n;t_0,t_1}$ est réduit à la détermination de $\left\|EB_{\widetilde{\theta}(x,h;t_0,t_1)}\right\|_2^2$. En partant de sa définition donnée dans la Proposition 2.2.4, on a

$$\begin{aligned} \left\|EB_{\widetilde{\theta}(x,h;t_0,t_1)}\right\|_2^2 &= \int_{[t_0,t_1]} EB^2_{(x,h;t_0,t_1)}(t)dt \\ &= \int_{[t_0,t_1]} \frac{(u-t_0)^{2(x-t_0)/h}(t_1-u)^{2(t_1-x)/h}du}{\{\mathscr{B}(1+(x-t_0)/(t_1-t_0)h, 1+(t_1-x)/(t_1-t_0)h)\}^2} \\ &= \frac{\mathscr{B}(1+2(x-t_0)/(t_2-t_0)h, 1+2(t_1-x)/(t_1-t_0)h)}{\{\mathscr{B}(1+(x-t_0)/(t_1-t_0)h, 1+(t_1-x)/(t_1-t_0)h)\}^2}. \end{aligned} \tag{2.29}$$

Donc, on peut écrire la variance de la forme :

$$Var\left\{\widehat{f}_{n;t_0,t_1}(x)\right\} = \frac{\mathscr{B}\left(\frac{2(x-t_0)}{(t_1-t_0)h}+1, \frac{2(t_1-x)}{(t_1-t_0)h}+1\right)}{\left\{\mathscr{B}\left(\frac{(x-t_0)}{(t_1-t_0)h}+1, \frac{(t_1-x)}{(t_1-t_0)h}+1\right)\right\}^2}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{r_2}}\right), \tag{2.30}$$

Nous allons maintenant approximer cette expression de la variance afin d'exhiber le réel r_2. Considérons la fonction R introduite par Brown & Chen (1998). Elle est définie pour tout $z \geq 0$ comme suit :

$$R(z) = \frac{1}{\Gamma(z+1)}\left(\frac{z}{e}\right)^z \sqrt{2\pi z}. \tag{2.31}$$

C'est une fonction positive, croissante et majorée par 1. En utilisant la relation entre les fonctions bêta et gamma puis en injectant dans (2.29), on trouve

$$\left\|EB_{\widetilde{\theta}(x,h;t_0,t_1)}\right\|_2^2 = \begin{cases} \dfrac{\Gamma(2c+1)}{2^{2c+1}\Gamma^2(c+1)} & \text{si } \dfrac{x-t_0}{(t_1-t_0)h} \text{ et } \dfrac{t_1-x}{(t_1-t_0)h} \to c \\ \dfrac{(x-t_0)^{-1/2}(t_1-x)^{-1/2}}{2\sqrt{\pi}(t_1-t_0)^{-1}h^{1/2}} & \text{si } \dfrac{x-t_0}{(t_1-t_0)h} \text{ et } \dfrac{t_1-x}{(t_1-t_0)h} \to \infty. \end{cases}$$

On obtient alors la variance de $\widetilde{f}_{n;t_1,t_2}(x)$ pour tout $x \in (t_1+\alpha(h), t_2-\alpha(h))$

$$Var\left\{\widetilde{f}_{n;t_1,t_2}(x)\right\} = \frac{(x-t_0)^{-1/2}(t_1-x)^{-1/2}}{2\sqrt{\pi}(t_1-t_0)^{-1}nh^{1/2}}f(x) + o\left(\frac{1}{nh^{1/2}}\right). \tag{2.32}$$

On en déduit que $r_2 = 1/2$.

Le noyau bêta présenté dans Chen(1999) est un cas particulier de ce bêta étendu. On l'obtient en prenant $t_0 = 0$ et $t_1 = 1$. Dans ce cas, son biais est

$$Biais\left\{\widehat{f}_n(x)\right\} = \frac{h-2xh}{1+2h}f'(x) + \frac{1}{2}\left\{\frac{(h-2xh)^2}{(1+2h)^2} + \frac{x(1-x)h+h^2+h^3}{(1+3h)(1+2h)^2}\right\}f''(x) + o(h^2).$$

La première étape de réduction de biais donne : $\mathbb{T}_{\alpha(h),-1} = [0, \alpha(h)]$, $\mathbb{T}_{\alpha(h),0} =]\alpha(h), 1-\alpha(h)[$ et $\mathbb{T}_{\alpha(h),+1} = [1-\alpha(h), 1]$ (e.g. dans Chen(1999), $\alpha(h) = 2h$). Le noyau bêta modifié est obtenu à partir de $\widetilde{\theta}(x,h)$ définie par :

$$\widetilde{\theta}(x,h) = \begin{cases} \left(\dfrac{\psi(x)}{h}, \dfrac{x}{h}\right) & \text{si } x \in [0, \alpha(h)] \\ \left(\dfrac{x}{h}, \dfrac{1-x}{h}\right) & \text{si } x \in]\alpha(h), 1-\alpha(h)[\\ \left(\dfrac{1-x}{h}, \dfrac{\psi(1-x)}{h}\right) & \text{si } x \in [1-\alpha(h), 1], \end{cases}$$

les calculs de $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ donnent respectivement :

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{(1-x)\psi(x)+x^2}{x+\psi(x)} & \text{si } x \in [0, \alpha(h)] \\ 0 & \text{si } x \in]\alpha(h), 1-\alpha(h)[\\ \dfrac{(1-x)\{x-\psi(1-x)\}}{1-x+\psi(1-x)} & \text{si } x \in [1-\alpha(h), 1], \end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{hx\psi(x)}{\{x+\psi(x)\}^2\{x+\psi(x)+h\}} & \text{si } x \in [0,\alpha(h)] \\ \dfrac{hx(1-x)}{1+h} & \text{si } x \in]\alpha(h), 1-\alpha(h)[\\ \dfrac{h(1-x)\psi(1-x)\{1-x+\psi(1-x)\}^{-2}}{\{1-x+\psi(1-x)+h\}} & \text{si } x \in [1-\alpha(h), 1]. \end{cases}$$

Donc le biais réduit à l'intérieur (i.e. pour $x \in]\alpha(h), 1-\alpha(h)[$) est :

$$Biais\left\{\widetilde{f_n}(x)\right\} = \frac{1}{2}\frac{x(1-x)h}{1+h}f''(x) + o(h^2).$$

Sa variance déduite de (2.30) est alors

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{\mathscr{B}\left(1+2x/h, 1+2(1-x)/h\right)}{\{\mathscr{B}\left(1+x/h, 1+(1-x)/h\right)\}^2}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

Sa valeur approchée à l'intérieur s'obtient de (2.32) en posant $t_0 = 0$ et $t_1 = 1$. Ceci donne

$$Var\left\{\widetilde{f_n}(x)\right\} = \frac{1}{2\sqrt{\pi}}\frac{n^{-1}h^{-1/2}}{\{x(1-x)\}^{1/2}}f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

Estimateurs de densités à noyau gamma

D'après la construction du noyau gamma en Section 2.1.3, on a $A(x,h) = h$ et $B(x,h) = x+h$. Par conséquent le biais de cet estimateur est :

$$Biais\left\{\widehat{f_n}(x)\right\} = hf'(x) + \frac{1}{2}\left\{xh + 2h^2\right\}f''(x) + o(h^2). \tag{2.33}$$

Pour le réduire, on divise le support $\mathbb{T}_I = [0,\infty)$ en $\mathbb{T}_{\alpha(h),-1} = [0,\alpha(h)]$ et $\mathbb{T}_{\alpha(h),0} =]\alpha(h),\infty[$ (e.g. dans Chen(2000), $\alpha(h) = 2h$). On modifie le noyau en utilisant la paramétrisation $\widetilde{\theta}(x,h)$ définie par :

$$\widetilde{\theta}(x,h) = \begin{cases} \left(\dfrac{x^2}{2h\alpha(h)} + 1, h\right) & \text{si } x \in [0,\alpha(h)] \\ \left(\dfrac{x}{h}, h\right) & \text{si } x \in]\alpha(h), +\infty[. \end{cases}$$

Ainsi, par les calculs de moyenne et variance on obtient les caractéristiques $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ données respectivement par :

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{x^2 + 2\alpha(h)(h-x)}{2\alpha(h)} & \text{si } x \in [0,\alpha(h)] \\ 0 & \text{si } x \in (\alpha(h),\infty), \end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{h\{x^2 + 2h\alpha(h)\}}{2\alpha(h)} & \text{si } x \in [0,\alpha(h)] \\ xh & \text{si } x \in]\alpha(h), +\infty[. \end{cases}$$

Ce qui conduit au biais réduit de $\widetilde{f_n}$ sur $]\alpha(h), \infty[$ donné par

$$Biais\left\{\widetilde{f_n}(x)\right\} = \frac{1}{2}xhf''(x) + o(h^2).$$

Pour la variance, on calcule d'abord $\left\|GA_{\theta(x,h)}\right\|_2^2$. Ceci donne

$$\begin{aligned}\left\|GA_{\theta(x,h)}\right\|_2^2 &= \int_{\mathbb{R}_+} GA^2_{\theta(x,h)}(u)du \\ &= \frac{1}{\Gamma^2(1+x/h)\,h^2 2^{2x/h}} \int_{\mathbb{R}_+} \left(\frac{2u}{h}\right)^{2x/h} e^{-2u/h}du \\ &= \frac{\Gamma(1+2x/h)}{\Gamma^2(1+x/h)\,2^{1+2x/h}}\end{aligned}$$

Finalement, la variance de l'estimateur est

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{\Gamma(1+2x/h)}{\Gamma^2(1+x/h)\,2^{1+2x/h}}\frac{1}{nh}f(x) + o\left(\frac{1}{nh^{r_2}}\right).$$

L'utilisation de la fonction R définie en (2.31) conduit à

$$Var\left\{\widetilde{f_n}(x)\right\} = \frac{x^{-1/2}}{2\sqrt{\pi}}\frac{1}{n\sqrt{h}}f(x) + o\left(\frac{1}{nh^{1/2}}\right), \quad \forall x \in]\alpha(h), \infty[.$$

Par conséquent $r_2 = 1/2$.

Estimateurs de densités à noyau gamma inverse

Considérons les ingrédients $A(x,h) = 2x^2h/(1-2xh)$ et $B(x,h) = x^3/(1-2xh)^2(1-3xh)$ issus de la construction du noyau gamma inverse en Section 2.1.3. La première partie de la Proposition 2.2.4 permet d'avoir :

$$Biais\left\{\widehat{f_n}(x)\right\} = \frac{2x^2h}{1-2xh}f'(x) + \frac{1}{2}\left\{\frac{4x^4h^2}{(1-2xh)^2} + \frac{x^3h}{(1-2xh)^2(1-3xh)}\right\}f''(x) + o(h^2).$$

Suivant l'algorithme décrit en Section 2.2.2, on réduit ce biais en commençant par diviser le support $\mathbb{T}_I =]0, \infty[$ en $\mathbb{T}_{\alpha(h),-1} = (0, \alpha(h)]$ et $\mathbb{T}_{\alpha(h),0} = (\alpha(h), \infty)$. Puis on modifie le noyau par l'utilisation de la paramétrisation $\widetilde{\theta}(x,h)$ définie par :

$$\widetilde{\theta}(x,h) = \begin{cases} \left(1 + \dfrac{x}{h\alpha^2(h)}, \dfrac{1}{h}\right) & \text{si } x \in (0, \alpha(h)] \\ \left(1 + \dfrac{1}{xh}, \dfrac{1}{h}\right) & \text{si } x \in (\alpha(h), \infty). \end{cases}$$

En calculant respectivement les moyenne et variance du noyau gamma inverse modifié, on déduit les expressions de $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ qui sont :

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{\alpha^2(h) - x^2}{x} & \text{si } x \in]0, \alpha(h)] \\ 0 & \text{si } x \in]\alpha(h), +\infty[\end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{h\alpha^6(h)}{x^2(x-h\alpha^2(h))} & \text{si } x \in]0,\alpha(h)] \\ \dfrac{x^3h}{1-xh} & \text{si } x \in]\alpha(h),+\infty[. \end{cases}$$

Le biais réduit sur $]\alpha(h),+\infty[$ est finalement

$$Biais\left\{\widetilde{f_n}(x)\right\} = \frac{x^3h}{2(1-xh)}f''(x) + o(h^2).$$

Comme dans le cas précédent, la variance s'obtient à partir de $\left\|GI_{\theta(x,h)}\right\|_2^2$. Par une démarche similaire on a

$$\left\|GI_{\theta(x,h)}\right\|_2^2 = 2^{1-2/(xh)}\frac{h\Gamma(-1+2/xh)}{\Gamma^2(-1+1/xh)},$$

ce qui conduit à

$$Var\left\{\widehat{f_n}(x)\right\} = 2^{1-2/(xh)}\frac{h\Gamma(-1+2/xh)}{n\Gamma^2(-1+1/xh)}f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

Sa valeur approchée pour $x \in]\alpha(h),+\infty[$ est alors :

$$Var\left\{\widetilde{f_n}(x)\right\} = \frac{h}{2\sqrt{\pi}}\left(\frac{1}{xh}-1\right)^{3/2}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

On en déduit que $r_2 = 1/2$.

Estimateurs de densités à noyau gaussien inverse

Partant de la construction du noyau gaussien inverse en Section 2.1.3, on a les ingrédients $A(x,h) = x\left\{(1-3xh)^{-1/2}-1\right\}$ et $B(x,h) = x^3\,(1+3xh)^{-3/2}$. En utilisant (2.19) il vient :

$$\begin{aligned} Biais\left\{\widehat{f_n}(x)\right\} &= \left(\frac{x}{\sqrt{1-3xh}} - x\right)f'(x) \\ &\quad + \frac{1}{2}\left\{\left(\frac{x}{\sqrt{1-3xh}} - x\right)^2 + \frac{x^3h}{(1-3xh)^{3/2}}\right\}f''(x) + o(h^2) \end{aligned}$$

Pour réduire ce biais conformément à l'algorithme décrit en Section 2.2.2, on divise le support $\mathbb{T}_I =]0,+\infty[$ en $\mathbb{T}_{\alpha(h),-1} =]0,\alpha(h)]$ et $\mathbb{T}_{\alpha(h),0} =]\alpha(h),+\infty[$. Ensuite, on considère la nouvelle paramétrisation $\widetilde{\theta}(x,h)$ définie par :

$$\widetilde{\theta}(x,h) = \begin{cases} \left(\dfrac{2x^2-\alpha^2(h)}{\alpha(h)}, \dfrac{1}{h}\right) & \text{si } x \in]0,\alpha(h)] \\ \left(x, \dfrac{1}{h}\right) & \text{si } x \in]\alpha(h),+\infty[\end{cases}$$

Par les calculs de biais et variance du noyau gaussien inverse modifié, on déduit les caractéristiques $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ données respectivement par

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{2x^2 - \alpha(h)\{\alpha(h)+x\}}{\alpha(h)} & \text{si } x \in]0,\alpha(h)] \\ 0 & \text{si } x \in]\alpha(h),+\infty[\end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{h\{2x^2 - \alpha(h)\}^3}{\alpha^3(h)} & \text{si } x \in (0,\alpha(h)] \\ x^3 & \text{si } x \in]\alpha(h),+\infty[. \end{cases}$$

Ainsi, le biais réduit sur $]\alpha(h),+\infty[$ est alors

$$Biais\left\{\widetilde{f_n}(x)\right\} = x^3 f''(x) + o(h^2).$$

La variance de l'estimateur gaussien inverse se calcule de la même manière que dans les exemples précédents. On détermine d'abord $\left\|IG_{\theta(x,h)}\right\|_2^2$ comme suit

$$\left\|IG_{\theta(x,h)}\right\|_2^2 = \frac{1}{2\sqrt{\pi h x^3}},$$

On en déduit que la variance de $\widehat{f_n}$ est

$$Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{2\sqrt{\pi h x^3}}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{1/2}}\right),$$

par conséquent on a à l'intérieur

$$Var\left\{\widetilde{f_n}(x)\right\} = \frac{1}{2\sqrt{\pi h x^3}}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{1/2}}\right)$$

et donc $r = 1/2$.

Estimateurs de densités à noyau gaussien inverse réciproque

Les différents ingrédients issus de la construction du noyau gaussien inverse réciproque étant $A(x,h) = (x^2+xh)^{1/2} - x + h$ et $B(x,h) = (x^2+xh)^{1/2} + 2h$ alors d'après (2.19) on a

$$\begin{aligned} Biais\left\{\widehat{f_n}(x)\right\} &= \{(x^2+xh)^{1/2} - x + h\}f'(x) \\ &\quad + \frac{1}{2}\left[\{(x^2+xh)^{1/2} - x + h\}^2 + (x^2+xh)^{1/2} + 2h\right]f''(x) + o(h^2). \end{aligned}$$

L'utilisation de l'algorithme de réduction de biais décrit en Section 2.2.2 conduit à diviser le support $]0,+\infty[$ en $\mathbb{T}_{\alpha(h),-1} =]0,\alpha(h)]$ et $\mathbb{T}_{\alpha(h),0} =]\alpha(h),+\infty[$. Ensuite, on modifie le noyau gaussien inverse réciproque en utilisant la nouvelle paramétrisation

$$\widetilde{\theta}(x,h) = \begin{cases} \left(\dfrac{\alpha(h)}{x^2 - h\alpha(h)}, \dfrac{1}{h}\right) & \text{si } x \in]0,\alpha(h)] \\ \left(\dfrac{1}{x-h}, \dfrac{1}{h}\right) & \text{si } x \in]\alpha(h),+\infty[. \end{cases}$$

Par les calculs de biais et variance du gaussien inverse réciproque et modifié, on déduit les caractéristiques $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ données respectivement par

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{x(x-h\alpha(h)}{\alpha(h)} & \text{si } x \in]0,\alpha(h)] \\ 0 & \text{si } x \in]\alpha(h),+\infty[\end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{h\{x^2+h\alpha(x)\}}{\alpha(h)} & \text{si } x \in]0,\alpha(h)] \\ \dfrac{1}{x-h}+2h & \text{si } x \in]\alpha(h),+\infty[. \end{cases}$$

Ceci permet d'obtenir le biais réduit dans $]\alpha(h),+\infty[$ défini par

$$Biais\left\{\widetilde{f_n}(x)\right\} = \left\{\frac{1}{x-h}+2h\right\} f''(x) + o(h^2).$$

Le calcul de $\left\|RG_{\theta(x,h)}\right\|_2^2$ donne

$$\left\|RG_{\theta(x,h)}\right\|_2^2 = \frac{1}{2\sqrt{\pi h x}}.$$

En utilisant la Proposition 2.2.7, on déduit pour tout $x \in]\alpha(h),+\infty[$

$$Var\left\{\widetilde{f_n}(x)\right\} = Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{2\sqrt{\pi h x}}\frac{1}{n}f(x) + o\left(\frac{1}{nh^{1/2}}\right).$$

On a dans ce cas $r_2 = 1/2$.

Estimateurs de densités à noyau de Pareto

La construction du noyau de Pareto dans la Section 2.1.3 a permis d'avoir $A(x,h) = xh/(1-h)$ et $B(x,h) = (xh)^2/(1-h)^2(1-2h)$. Ceci permet d'écrire le biais de $\widehat{f_n}$ d'après (2.19) de la manière suivante

$$Biais\left\{\widehat{f_n}(x)\right\} = \frac{xh}{1-h}f'(x) + \frac{(xh)^2}{(1-h)(1-2h)}f''(x) + o(h^2).$$

Ce biais est nettement large par rapport aux cas des noyaux associés classiques. L'utilisation de l'algorithme donné dans la Section 2.2.2, permet de constater que les régions de bords sont vides pour ce noyau. Et donc il ne reste qu'à modifier ce noyau en utilisant la nouvelle paramétrisation suivante

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

Ceci conduit à

$$\widetilde{A}(x,h) = 0 \text{ et } \widetilde{B}(x,h) = \frac{(xh)^2}{(1-2h)}$$

et donc le biais réduit est

$$Biais\left\{\widetilde{f_n}(x)\right\} = \frac{(xh)^2}{(1-2h)}f''(x) + o(h^2).$$

La variance de l'estimateur à noyau de Pareto est calculée de la même manière que dans les cas précédents. Le calcul de $\left\|PA_{\theta(x,h)}\right\|_2^2$ donne

$$\left\|PA_{\theta(x,h)}\right\|_2^2 = \frac{1}{2xh}.$$

Ainsi, en utilisant la Proposition 2.2.7, on trouve

$$Var\left\{\widetilde{f_n}(x)\right\} = Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{2xnh}f(x) + o\left(\frac{1}{nh}\right)$$

et donc $r_2 = 1$.

Estimateurs de densités à noyau lognormal

Le noyau lognormal construit en Section 2.1.3 a pour caractéristiques $A_\theta(x,h) = x\left(e^{(3h^2)/2} - 1\right)$ et $B_\theta(x,h) = x^2 e^{(3h^2)}\left(e^{h^2} - 1\right)$. Ainsi, son biais peut être écrit de la forme

$$\begin{aligned} Biais\left\{\widehat{f_n}(x)\right\} &= x\left(e^{(3h^2)/2} - 1\right)f'(x) + \frac{1}{2}\{x\left(e^{(3h^2)/2} - 1\right) + \}^2 f''(x) \\ &\quad + x^2 e^{(3h^2)}\left(e^{h^2} - 1\right)f''(x) + o(h^2). \end{aligned}$$

L'application de l'algorithme décrit en Section 2.2.2 permet de diviser le support $\mathbb{T}_I =]0, +\infty[$ en $\mathbb{T}_{\alpha(h),-1} = [0, \alpha(h)]$ et $\mathbb{T}_{\alpha(h),0} =]\alpha(h), +\infty[$, respectivement. Ensuite, on modifie le noyau en utilisant la nouvelle paramétrisation $\widetilde{\theta}$ définie par :

$$\widetilde{\theta}(x,h) = \begin{cases} \left(\dfrac{\alpha^3(h)}{x^2}\exp\left(-3h^2/2\right), h\right) & \text{si } x \in]0, \alpha(h)] \\ \left(x\exp\left(-3h^2/2\right), h\right) & \text{si } x \in]\alpha(h), +\infty[. \end{cases}$$

Le calcul de $\widetilde{A}(x,h)$ et $\widetilde{B}(x,h)$ donne respectivement

$$\widetilde{A}(x,h) = \begin{cases} \dfrac{\alpha^3(h) - x^3}{x^2} & \text{si } x \in (0, \alpha(h)] \\ 0 & \text{si } x \in (\alpha(h), \infty) \end{cases}$$

et

$$\widetilde{B}(x,h) = \begin{cases} \dfrac{\alpha^6(h)}{x^4}\left(e^{h^2} - 1\right) & \text{si } x \in (0, \alpha(h)] \\ x^2(e^{h^2} - 1) & \text{si } x \in (\alpha(h), \infty). \end{cases}$$

Ainsi, le biais de l'estimateur à noyau lognormal modifié est alors pour tout $x \in]\alpha(h), +\infty[$

$$Biais\left\{\widetilde{f_n}(x)\right\} = x^2(e^{h^2} - 1)f''(x) + o(h^2).$$

Le calcul de $\left\|LN_{\theta(x,h)}\right\|_2^2$ conduit à

$$\left\|LN_{\theta(x,h)}\right\|_2^2 = \frac{1}{2xh\sqrt{\pi}}.$$

En faisant usage de la Proposition 2.2.7, on déduit alors pour tout $x \in]\alpha(h), +\infty[$

$$Var\left\{\widetilde{f_n}(x)\right\} = Var\left\{\widehat{f_n}(x)\right\} = \frac{1}{2xnh\sqrt{\pi}}f(x) + o\left(\frac{1}{nh}\right),$$

Ce qui prouve que $r_2 = 1$. Nous récapitulons dans les Tableau 2.6 et Tableau 2.7 les différents résultats présentés ci-dessus.

Table 2.6 – Quelques noyaux associés modifiés (suite voir Tableau 2.7)

$K_{\widetilde{\theta}(x,h)}$	Bêta : $\psi(z) = \{z - \alpha(h) + 1\}\alpha(h)$ for all $z \geq 0$.			Pareto	Lognormal	
$\mathbb{T}_{\alpha(h),j}$	$[0, \alpha(h)]$	$(\alpha(h), 1 - \alpha(h))$	$[1 - \alpha(h), 1]$	$[x, \infty)$	$[0, \alpha(h)]$	$(\alpha(h), \infty)$
$\widetilde{\theta}(x,h)$	$\left(\frac{\psi(x)}{h}, \frac{x}{h}\right)$	$\left(\frac{x}{h}, \frac{1-x}{h}\right)$	$\left(\frac{1-x}{h}, \frac{\psi(1-x)}{h}\right)$	$\left(x(1-h), \frac{1}{h}\right)$	$\left(\frac{\alpha^3(h)}{x^2} \exp\left(-3h^2/2\right), h\right)$	$\left(x \exp\left(-3h^2/2\right), h\right)$
$A_{\widetilde{\theta}}(x,h)$	$\frac{(1-x)\psi(x) + x^2}{x + \psi(x)}$	0	$\frac{(1-x)\{x - \psi(1-x)\}}{1 - x + \psi(1-x)}$	0	$\frac{\alpha^3(h) - x^3}{x^2}$	0
$B_{\widetilde{\theta}}(x,h)$	$\frac{hx\psi(x)\{x + \psi(x)\}^{-2}}{\{x + \psi(x) + h\}}$	$\frac{hx(1-x)}{1+h}$	$\frac{h(1-x)\{1 - x + \psi(1-x)\}^{-2}}{\{\psi(1-x)\}^{-1}\{1 - x + \psi(1-x) + h\}}$	$\frac{(xh)^2}{(1-2h)}$	$\frac{\alpha^6(h)}{x^4}\left(e^{h^2} - 1\right)$	$x^2(e^{h^2} - 1)$
$\widetilde{r}_2(x)$	1	$1/2$	1	1	$\widetilde{r}_2\{\alpha(h)\}$ (**??**)	1

Table 2.7 – Quelques noyaux associés modifiés (suite du Tableau 2.6)

$K_{\widetilde{\theta}(x,h)}$	Gamma		Gamma inverse		Gaussien inverse		Gaussien inverse réciproque	
$\mathbb{T}_{\alpha(h),j}$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$	$[0,\alpha(h)]$	$(\alpha(h),\infty)$
$\widetilde{\theta}(x,h)$	$\left(\frac{x^2}{2h\alpha(h)}+1,h\right)$	$\left(\frac{x}{h},h\right)$	$\left(1+\frac{x}{h\alpha^2(h)},\frac{1}{h}\right)$	$\left(1+\frac{1}{xh},\frac{1}{h}\right)$	$\left(\frac{\alpha(h)}{x^2-h\alpha(h)},\frac{1}{h}\right)$	$\left(x,\frac{1}{h}\right)$	$\left(\frac{\alpha(h)}{x^2-h\alpha(h)},\frac{1}{h}\right)$	$\left(\frac{1}{x-h},\frac{1}{h}\right)$
$A_{\widetilde{\theta}}(x,h)$	$\frac{x^2+2\alpha(h)(h-x)}{2\alpha(h)}$	0	$\frac{\alpha^2(h)-x^2}{x}$	0	$\frac{2x^2-\alpha(h)\{\alpha(h)+x\}}{\alpha(h)}$	0	$\frac{x(x-h\alpha(h)}{\alpha(h)}$	0
$B_{\widetilde{\theta}}(x,h)$	$\frac{h\{x^2+2h\alpha(h)\}}{2\alpha(h)}$	xh	$\frac{h\alpha^6(h)}{x^2(x-h\alpha^2(h))}$	$\frac{x^3h}{1-xh}$	$\frac{h\{2x^2-\alpha(h)\}^3}{\alpha^3(h)}$	x^3	$\frac{h\{x^2+h\alpha(x)\}}{\alpha(h)}$	$\frac{1}{x-h}+2h$
$\widetilde{r}_2(x)$	1	$1/2$	1	$1/2$	1	$1/2$	1	$1/2$

Nous procédons maintenant au choix de paramètre de lissage.

2.2.4 Problèmes de choix de fenêtre

Le paramètre optimal obtenu dans la Proposition 2.2.5 ne peut être utilisé en pratique car il dépend encore de la densité inconnue. Il existe plusieurs méthodes pour le choix pratique du paramètre h comme nous l'avons décrit dans le chapitre précédent. Dans le cas présent, nous proposons l'utilisation de la méthode de validation croisée par les moindres carrés pour le choix de ce paramètre.

Validation croisée par les moindres carrés

Comme décrit dans le deuxième paragraphe de la Section 1.1.3 du chapitre précédent, considérons $K_{\theta(x,h)}$ le noyau associé construit par la méthode mode-dispersion avec $x \in \mathbb{T}_I$ et $h > 0$. Le paramètre optimal h_{cv} de h est obtenu par

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

où

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

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_i.

Etudes par simulation des estimateurs à noyaux lognormaux

Cette section se propose de présenter les résultats des études par simulation faites sur trois estimateurs de densité à noyaux lognormaux notés respectivement $\widehat{f_{n,LN}}$, $\widetilde{f_{n,LN}}$ et $f^*_{n,LN}$ qui sont issus des trois noyaux lognormaux LN_θ, $LN_{\widetilde{\theta}}$ et LN_{θ^*}. Le noyau lognormal LN_θ est construit par la méthode mode-dispersion, $LN_{\widetilde{\theta}}$ est sa version modifiée et enfin LN_{θ^*} est la version extraite des travaux de Jin & Kawczak (2003). Le noyau $LN_{\theta(x,h)}$ est défini par

$$LN_{\theta(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(\frac{-1}{2h^2}\left[\log u - \log\left\{x\exp\left(h^2\right)\right\}\right]\right)\mathbf{1}_{u>0}, \quad x>0, h>0. \tag{2.34}$$

Pour $\alpha(h) > 0$, on définit les deux régions de $\mathbb{T}_I =]0, +\infty[$ par $\mathbb{T}_I =]0, \alpha(h)] \cup]\alpha(h), +\infty[$. La version modifiée de $LN_{\theta(x,h)}$ notée $LN_{\widetilde{\theta}(x,h)} := \widetilde{LN}_{\theta(x,h)}$ est donnée par

$$\begin{aligned} LN_{\widetilde{\theta}(x,h)}(u) &= LN_{\widetilde{\theta}_{-1}(x,h)}(u) + LN_{\widetilde{\theta}_0(x,h)}(u) \\ &= \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{\frac{\alpha^3(h)\exp(-h^2/2)}{x^2}\right\}\right]^2\right)\mathbf{1}_{]0,\alpha(h)]}(u) \\ &\quad + \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{x\exp(-h^2/2)\right\}\right]^2\right)\mathbf{1}_{]\alpha(h),+\infty[}(u), \end{aligned} \tag{2.35}$$

où $\widetilde{\theta}_{-1}(x,h)$ et $\widetilde{\theta}_0(x,h)$ sont respectivement les paramétrisations aux bords et à l'intérieur données dans le Tableau 2.6. La troisième version de noyau lognormal LN_{θ^*} proposée par Jin & Kawczak (2003) (sans montrer sa construction) est définie par :

$$LN_{\theta^*(x,h)}(u) = \frac{1}{2u\sqrt{2\pi\log(1+h)}} \exp\left\{-\frac{\log^2(u/x)}{8\log(1+h)}\right\}\mathbf{1}_{(0,\infty)}(u). \tag{2.36}$$

Elle est liée à (2.34) par la relation

$$LN_{\theta^*(x,h)} = LN_{\theta(x\exp(h^2),2\sqrt{\log(1+h)})}, \quad x > 0, h > 0;$$

Ces trois noyaux associés sont positifs et beaucoup plus flexibles que le noyau gamma selon leur construction. Nous illustrons le comportement des trois estimateurs lognormaux sur des échantillons à taille finie à travers les études par simulation. Nous analysons les influences de paramètres de lissage sur les MISEs de ces estimateurs et nous mesurons l'efficacité qui est acquise grâce à la réduction du biais. Ces études sont faites aux bords et à l'intérieur sur deux modèles différents. Le premier modèle est essentiellement constitué de la densité normale tronquée $\mathcal{N}_t(\mu, \sigma; a, b)$ sur l'intervalle $[a, b]$

$$\text{Modèle 1} : X \sim \mathcal{N}_t(0.5, 0.15; a, b);$$

et le second modèle est un mélange des densités normale tronquée et l'exponentielle

$$\text{Modèle 2} : X \sim 0.25\mathcal{N}_t(0.25, 0.15; a, b) + 0.25\mathcal{E}xp(2).$$

Pour chaque Modèle, nous simulons 100 échantillons de taille $n = 50, 100, 500$ ou 1000, en prenant $a = 0.01$ et $b = 0.5$ pour la région de bords puis $a = 0.5$ et $b = 8.5$ pour la région intérieure.

Dans les Tableaux 2.8 et 2.9, nous rapportons les valeurs de fenêtres de lissage optimales (dans le sens de MISE) avec les durées correspondantes. On voit que ces valeurs diminuent en général (pour les deux modèles) lorsque l'on augmente la taille de l'échantillon d'ailleurs, l'estimateur à noyau lognormal modifié $\widetilde{f}_{n,LN}$ présente les plus petites valeurs que les deux autres estimateurs $\widehat{f}_{n,LN}$ et $f^*_{n,LN}$. Cependant, on constate qu'il met plus de temps de calcul que les autres.

Les Tableaux 2.10 et 2.11 fournissent les valeurs des biais et variance de ces estimateurs en certains points sélectionnés, correspondant aux points de bords et de

TABLE 2.8 – Valeurs optimales de h obtenues aux bords par la validation croisée par les moindres carrés

Model	Size	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
	n	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$
1	50	0.121	0.456	0.412	0.360	0.002	2.153
	100	0.111	0.494	0.376	0.416	0.003	3.922
	500	0.110	1.409	0.367	1.090	0.004	7.543
	1000	0.104	2.893	0.351	2.458	0.003	2.829
2	50	0.374	0.554	0.089	0.579	0.088	0.572
	100	0.366	0.553	0.087	0.772	0.087	0.906
	500	0.364	1.976	0.056	2.734	0.055	2.846
	1000	0.327	4.251	0.025	5.858	0.025	5.664

TABLE 2.9 – Valeurs optimales de h obtenues à l'intérieur par la validation croisée par les moindres carrés

Model	Size	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
	n	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$	h_{lscv}	$Time/s$
1	50	0.470	0.394	0.997	0.429	0.996	0.471
	100	0.429	0.410	0.996	0.524	0.996	0.573
	500	0.406	1.151	0.991	1.513	0.993	1.695
	1000	0.400	2.605	0.991	3.456	0.991	3.718
2	50	0.456	0.639	0.997	0.630	0.996	0.622
	100	0.429	0.582	0.996	0.812	996	0.849
	500	0.410	2.021	0.994	2.570	993	2.951
	1000	0.397	4.124	0.992	4.889	991	6.039

l'intérieur, respectivement pour le Modèle 1 et le Modèle 2. On voit que pour les échantillons de petite taille, l'estimateur à noyau lognormal modifié $\widetilde{f}_{n,LN}$ se comporte mieux que $\widehat{f}_{n,LN}$ et $f^*_{n,LN}$ aux bords et il devient moins bon qu'eux lorsqu'on augmente la taille de l'échantillon. Toutefois, en prenant en considération le compromis biais-variance, nous voyons que l'estimateur $\widehat{f}_{n,LN}$ à noyau lognormal construit par la méthode mode-dispersion est meilleur par rapport à l'autre et ce, quelle que soit la taille de l'échantillon. Cela consolide les résultats de la Proposition 2.2.7 de la Section 2.2.2, puisque les trois noyaux lognormaux appartiennent à la même famille. De plus, en observant les biais ponctuels de ces estimateurs, on voit que les biais des deux premiers estimateurs $\widehat{f}_{n,LN}$ et $f^*_{n,LN}$ augmentent de plus en plus à l'intérieur. Cela est dû à la présence de terme non nul $A(x,h)f'(x)$ dans (2.19). Cette nette différence entre leur biais et ceux de $\widetilde{f}_{n,LN}$ montre l'intérêt de l'algorithme proposé pour la réduction du biais en Section 2.2.2. Ainsi, l'estimateur à noyau lognormal modifié $\widetilde{f}_{n,LN}$ reste sans aucun doute le meilleur à l'intérieur.

Table 2.10 – Biais et variances des trois estimateurs à noyaux lognormaux suivant le Modèle 1

Taille	Cible	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
n	x	*Bias*	*Variance*	*Bias*	*Variance*	*Bias*	*Variance*
100	0.25	0.189 e+1	0.450 e-2	−0.207 e+1	0.248 e-2	−0.383 e+1	0.912 e-2
	0.50	−0.111 e+1	0.683 e-2	−0.986 e+0	0.192 e-2	−0.330 e+1	0.820 e-2
	2.50	−0.331 e+1	0.215 e-2	−0.329 e+1	0.533 e-2	−0.390 e+1	0.121 e-2
	3.50	−0.381 e+1	0.265 e-2	−0.410 e+1	0.127 e-2	−0.396 e+1	0.368 e-2
500	0.25	0.126 e+1	0.781 e-2	−0.325 e+1	0.534 e-2	−0.168 e+1	0.091 e-2
	0.50	−0.138 e+1	0.752 e-2	−0.519 e+1	0.186 e-2	−0.122 e+1	0.431 e-2
	2.50	−0.374 e+1	0.504 e-2	−0.370 e+1	0.101 e-2	−0.481 e+1	0.136 e-2
	3.50	−0.347 e+1	0.674 e-2	−0.529 e+1	0.307 e-2	−0.501 e+1	0.477 e-2
1000	0.25	0.123 e+1	0.912 e-2	−0.185 e+1	0.100 e-2	−0.124 e+1	0.081 e-2
	0.50	−0.159 e+1	0.811 e-2	−0.208 e+1	0.487 e-2	−0.113 e+1	0.119 e-3
	2.50	−0.376 e+1	0.785 e-2	−0.372 e+1	0.537 e-2	−0.584 e+1	0.652 e-2
	3.50	−0.372 e+1	0.354 e-2	−0.641 e+1	0.154 e-2	−0.549 e+1	0.845 e-2

Enfin, nous donnons dans le Tableau 2.12 Les MISEs minimales de ces estimateurs. Nous rappelons au lecteur que la fonction MISE théorique est bien approchée par la moyenne des ISEs qui peuvent être définies pour $N = 100$ (nombre d'essais) par

$$\widehat{ISE} = \frac{1}{N}\sum_{n=1}^{N}\int_{\mathbb{I}_I}\left\{\widehat{f_n}(x) - f(x)\right\}^2 dx.$$

Table 2.11 – Biais et variances des trois estimateurs à noyaux lognormaux suivant le Modèle 2

Taille	Cible	$f^*_{n,LN}$		$\widehat{f}_{n,LN}$		$\widetilde{f}_{n,LN}$	
n	x	*Bias*	*Variance*	*Bias*	*Variance*	*Bias*	*Variance*
100	0.25	0.265 e-1	0.137 e-3	0.207 e-1	0.248 e-3	0.191 e-1	0.721 e-3
	0.50	0.273 e-1	0.176 e-3	0.212 e-1	0.292 e-3	0.217 e-1	0.714 e-3
	2.50	0.230 e-1	0.148 e-3	0.195 e-1	0.283 e-3	0.183 e-1	0.698 e-3
	3.50	0.237 e-1	0.132 e-3	0.173 e-1	0.257 e-3	0.162 e-1	0.672 e-3
500	0.25	0.125 e-1	0.342 e-3	0.121 e-1	0.434 e-3	0.258 e-1	0.575 e-3
	0.50	0.141 e-1	0.364 e-3	0.109 e-1	0.486 e-3	0.262 e-1	0.562 e-3
	2.50	0.143 e-1	0.219 e-3	0.108 e-1	0.367 e-3	0.103 e-1	0.740 e-3
	3.50	0.113 e-1	0.205 e-3	0.095 e-1	0.326 e-3	0.091 e-1	0.725 e-3
1000	0.25	0.060 e-1	0.428 e-3	0.058 e-1	0.600 e-3	0.321 e-1	0.379 e-3
	0.50	0.054 e-1	0.454 e-3	0.046 e-1	0.687 e-3	0.347 e-1	0.360 e-3
	2.50	0.038 e-1	0.325 e-3	0.034 e-1	0.437 e-3	0.031 e-1	0.795 e-3
	3.50	0.035 e-1	0.318 e-3	0.021 e-1	0.414 e-3	0.019 e-1	0.781 e-3

Table 2.12 – La moyenne des ISEs pour les trois estimateurs de densité à noyaux lognormaux

Modèle	Taille	$\widehat{ISE}$ aux bords			$\widehat{ISE}$ à intérieur		
	n	$f^*_{n,LN}$	$\widehat{f}_{n,LN}$	$\widetilde{f}_{n,LN}$	$\widehat{f}_{n,LN}$	$f^*_{n,LN}$	$\widetilde{f}_{n,LN}$
1	50	0.146 e-1	0.141 e-2	0.359 e-2	0.184 e-2	0.152 e-2	0.481 e-3
	100	0.120 e-1	0.134 e-2	0.475 e-2	0.179 e-2	0.146 e-2	0.444 e-3
	500	0.116 e-1	0.132 e-2	0.482 e-2	0.168 e-2	0.135 e-2	0.438 e-3
	1000	0.828 e-2	0.121 e-2	0.485 e-2	0.165 e-2	0.132 e-2	0.416 e-3
2	50	0.124 e-2	0.340 e-3	0.277 e-3	0.312 e-3	0.396 e-3	0.808 e-4
	100	0.917 e-3	0.305 e-3	0.288 e-3	0.302 e-3	0.311 e-3	0.786 e-4
	500	0.891 e-3	0.287 e-3	0.308 e-3	0.298 e-3	0.287 e-3	0.751 e-4
	1000	0.837 e-3	0.267 e-3	0.310 e-3	0.272 e-3	0.252 e-3	0.322 e-4

En comparant ces modèles, on peut apprécier que la densité correspondant au premier est sous-estimée et celle correspondant au second est surestimée par ces trois estimateurs. Ceci est dû au fait que nous avons des biais négatifs pour le premier et positifs pour le second modèle, respectivement. Toutefois, selon le modèle 1, les Mises de ces estimateurs sont les plus petits que ceux dans le modèle 2. Les meilleurs estimateurs restent $\widetilde{f_{n,LN}}$ à l'intérieur et $\widehat{f_{n,LN}}$ aux bords.

2.3 Conclusion

Nous avons mis en œuvre dans ce chapitre une nouvelle famille d'estimateurs de densité à support borné ou non. Dans cette famille, l'estimateur est basé sur les noyaux associés qui dépendent intrinsèquement du point d'estimation x et du paramètre de lissage h. Tous les estimateurs à noyaux associés sont sans effets de bord mais ils ont une forme de biais légèrement différente de celle des noyaux dits classiques proposés par Rosenblatt (1956) et Parzen (1962). Nous avons fourni une définition générale et une méthode de construction de ces noyaux associés appelée "mode-dispersion" ainsi qu'un algorithme de la réduction des biais des estimateurs qui y sont issus. Nous avons illustré cette méthode de construction et l'algorithme de réduction du biais sur les noyaux associés non-classiques. Nous avons, en particulier, examiné les performances des trois estimateurs de densités à noyaux lognormaux $\widehat{f_{n,LN}}$, $\widetilde{f_{n,LN}}$ et $f^*_{n,LN}$ (avec $LN_{\theta(x,h)}$ construit par la méthode mode-dispersion, $LN_{\widetilde{\theta}(x,h)}$ sa version modifiée et $LN_{\theta^*(x,h)}$ celui proposé par Jin & Kawczak (2003) sur des échantillons à tailles finies. Dans ces simulations, les fenêtres optimales sont obtenues par la méthode de validation croisée par les moindres carrés et nous avons constaté que l'estimateur de densités à noyau lognormal modifié $\widetilde{f_{n,LN}}$ possède les plus petites valeurs de h, mais avec beaucoup de temps de calcul. En fait, nous avons constaté que l'estimateur de densités à noyau lognormal modifié $\widetilde{f_{n,LN}}$ est sans doute meilleur dans la région intérieure et sa première version $\widehat{f_{n,LN}}$ issue de la méthode mode-dispersion est aussi meilleure dans la région de bords. Cela nous amène à recommander fortement l'utilisation de $\widehat{f_{n,LN}}$ sur le bord et $\widetilde{f_{n,LN}}$ dans la région intérieure.

Chapitre 3

Convergence des estimateurs à noyaux associés continus

L'étude de la similarité entre l'estimateur et la vraie densité à estimer fait l'objet des travaux de nombreux auteurs ces dernières décennies. En ce qui concerne les estimateurs à noyaux classiques, les auteurs comme Parzen (1962), Nadaraya (1965), Devroye (1987) et Sylverman (1986) ont été les premiers à étudier la consistance de ces estimateurs. Il faut noter que pour la plupart, ces auteurs supposent que le noyau classique $\mathcal{K}$ est à variations bornées et que la densité f à estimer est uniformément continue sur $\mathbb{T} = \mathbb{R}$, et ce, indépendamment de x et h.

Depuis les travaux de Chen (1999, 2000) puis Kokonendji & Senga Kiessé (2011), d'autres auteurs comme Bouezmarni & Rolin (2003), Bertin & Klutchnikoff (2011), Zhang (2010), Zhang & Karunamuni (2010) ainsi que Abdous & Kokonendji (2009) ont (dans des cas particuliers) étudié les performances des estimateurs décrits dans le chapitre précédent.

Rappelons que ces estimateurs sont généralement définis par

$$\widehat{f_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T}_I := \text{support}(f), \tag{3.1}$$

où $X_1, X_2, \cdots, X_n$ est une suite de variables aléatoires continues indépendantes et identiquement distribuées (i.i.d.) de densité f de support $\mathbb{T}_I$ (intervalle borné au moins d'un côté), $h = h_n > 0$ est une suite de paramètres de lissage vérifiant $\lim_{n\to\infty} h_n = 0$, tandis que $K_{x,h}(\cdot)$ est le noyau associé i.e une densité de probabilité paramétrée par x et h, de support $\mathbb{S}_{x,h} \subseteq \mathbb{R}$ et vérifiant les conditions suivantes :

$$x \in \mathbb{S}_{x,h}, \tag{3.2}$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + A(x,h), \tag{3.3}$$

$$Var(\mathcal{Z}_{x,h}) = B(x,h), \tag{3.4}$$

où, $\mathcal{Z}_{x,h}$ est une variable aléatoire de densité $K_{x,h}$ sur $\mathbb{S}_{x,h}$ et $A(x,h)$, $B(x,h)$ qui tendent vers 0 quand h tend aussi 0. Signalons que la construction d'un noyau associé continu $K_{x,h}$ se fait à l'aide de la méthode mode-dispersion proposée dans la Section 2.1.2 du chapitre précédent que nous résumons brièvement de la manière suivante. On considère tout

d'abord un type de noyau $K := K_\theta$ (qui est une densité de probabilité paramétrée et de carré intégrable sur $\mathbb{S} = \mathbb{S}_\theta$ avec $\theta \in \Theta \subseteq \mathbb{R}^k$ pour $k \geq 2$). En particulier on choisit $\theta = \theta(a, b)$ avec a et b deux réels positifs. Ensuite on résout les équations $x = M(a, b)$ et $h = D(a, b)$ où $M(a, b)$ et $D(a, b)$ sont, respectivement, l'unique mode et le paramètre de dispersion de $K_{\theta(a,b)}$. Cela conduit au noyau associé $K_{\theta(x,h)} := K_{\theta(a(x,h),b(x,h))}$ de support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$, où $a(x, h)$ et $b(x, h)$ sont les solutions des équations précédentes. Aussi, il a été donné dans la Section 2.2.2 du chapitre précédent, un algorithme de réduction de biais de l'estimateur (3.1). Ceci conduit à une seconde version $\widetilde{f_n}$ de $\widehat{f_n}$, qui a la même variance ponctuelle que $\widehat{f_n}$, mais qui provient de la version modifiée de $K_{\theta(x,h)}$.

Dans ce chapitre, nous présentons une étude générale des propriétés asymptotiques dans la famille des estimateurs à noyaux associés continus $\widehat{f_n}$ ou $\widetilde{f_n}$, qui dépendent intrinsèquement du point d'estimation x et de paramètre de lissage h, pour les fonctions de densité à support $\mathbb{T} = \mathbb{T}_I$ borné au moins d'un côté. Afin de souligner la sensibilité des différents points de l'estimation (aux bords et à l'intérieur), nous proposons d'utiliser des techniques simples, qui obtiennent les mêmes résultats mais avec des hypothèses plus souples que dans le cas classique et qui dépendent fortement de la situation du point d'estimation x et d'une certaine puissance du paramètre de lissage h. L'hypothèse de la continuité uniforme de f sur $\mathbb{T} = \mathbb{R}$ est relaxée à une simple continuité de f sur tout compact de $\mathbb{T}_I$. Ainsi, cela peut aider à décider de choix du noyau associé (modifié ou non) approprié pour une zone donnée de l'estimation.

Nous organisons ce chapitre de la manière suivante. Dans la Section 3.1, nous présentons les résultats concernant les consistances faible et forte ponctuelles et la normalité asymptotique de ces estimateurs, en utilisant $\widehat{f_n}$. Ensuite, nous montrons dans la Section 3.2 les consistances globales avec les normes uniforme et $\mathbb{L}^1$. Nous donnons également quelques remarques relatives à certaines situations fréquentes. Enfin, nous clôturons ce chapitre avec la Section 3.3 qui illustre ces différents résultats sur trois types d'estimateurs à noyaux lognormaux sur $\mathbb{T}_I =]0, +\infty[$.

3.1 Consistance et normalité asymptotique ponctuelles

3.1.1 Consistance faible et forte

Nous présentons ici les deux premiers résultats concernant deux types de convergences ponctuelles de $\widehat{f_n}$ vers f : il s'agit de la convergence presque sûre (notée $p.s$) et la convergence en probabilité (notée $\mathbb{P}$). Nous signalons ici que le résultat de la convergence en probabilité est déduite de celui de la convergence de l'erreur quadratique moyenne (i.e. convergence ponctuelle en norme $\mathbb{L}^2$).

Théorème 3.1.1 *Soit $f \in C^2(\mathbb{T}_I)$ la densité à estimer et $\widehat{f_n}$ son estimateur à noyau associé défini en (3.1). Pour chaque point x fixé dans $\mathbb{T}$ on a alors*

$$\widehat{f_n}(x) \xrightarrow{p.s.} f(x) \ \ \textit{quand } n \to +\infty \ ;$$

de plus, s'il existe un plus grand nombre réel $r_2 = r_2(K_{x,h_n}) > 0$ *tel que*

$$h_n^{r_2} \left\| K_{x,h_n} \right\|_2^2 \leq c_2(x) < +\infty \ \ et \ \lim_{n\to+\infty} nh_n^{r_2} = +\infty,$$

alors

$$\widehat{f_n}(x) \xrightarrow{\mathbb{L}^2} f(x) \ \ quand \ n \to +\infty.$$

Démonstration. Pour le premier résultat, considérons l'inégalité triangulaire

$$\left| \widehat{f_n}(x) - f(x) \right| \leqslant \left| \widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\} \right| + \left| \mathbb{E}\left\{\widehat{f_n}(x)\right\} - f(x) \right|. \tag{3.5}$$

Puisque le terme non-stochastique du second membre de l'inégalité (3.5) représente le biais de $\widehat{f_n}$ alors d'après (2.19), (3.3) et (3.4), on a $\left| \mathbb{E}\left\{\widehat{f_n}(x)\right\} - f(x) \right| = O(h_n)$. Nous allons montrer que le terme stochastique est majoré par une série de Riemann convergente puis nous déduisons la convergence *p.s*. Nous savons que les X_i sont iid et donc $K_{x,h}(X_i)$ sont aussi iid. Comme les noyaux associés $K_{x,h_n}(\cdot)$ sont uniformément continus en $h_n > 0$ et $x \in \mathbb{T}$, par l'inégalité de Hoeffding (1963) on a :

$$\mathbb{P}\left(\left| \sum_{i=1}^{n} \left[K_{x,h_n}(X_i) - \mathbb{E}\left\{ K_{x,h_n}(X_i) \right\} \right] \right| \geqslant \varepsilon \right) \leqslant 2\exp\left(-\frac{2\varepsilon^2}{n} \right),$$

par conséquent

$$\mathbb{P}\left(\left| \widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\} \right| \geqslant \varepsilon \right) \leqslant 2\exp\left\{ -2n\varepsilon^2 \right\}.$$

En appliquant le lemme de Borel-Cantelli, on déduit que $\left| \widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\} \right| \underset{n\to+\infty}{\longrightarrow} 0.$
Le second résultat s'obtient à partir de la décomposition usuelle biais-variance du MSE de $\widehat{f_n}$. D'après ce qui précède, on sait que le biais de $\widehat{f_n}(x)$ est un $O(h_n)$. Par les hypothèses $h_n^{r_2} \left\| K_{x,h_n} \right\|_2^2 \leq c_2(x) < +\infty$ et $\lim_{n\to+\infty} nh_n^{r_2} = +\infty$, sa variance existe et est un $O(n^{-1}h_n^{-r_2})$. Ce qui permet d'obtenir

$$\widehat{f_n}(x) \xrightarrow{\mathbb{L}^2} f(x) \quad \Longrightarrow \quad \widehat{f_n}(x) \xrightarrow{\mathbb{P}} f(x). \blacksquare$$

3.1.2 Normalité asymptotique

Avant de donner ce résultat, nous rappelons dans un lemme, les conditions de Lindeberg pour le théorème de limites centrales dont la preuve est ici omise et que le lecteur peut trouver dans les travaux de Ash & Doléeans-Dade (2000) ainsi que Billingsley (1961).

Lemme 3.1.2 *Soit* $(X_{n,i})_{n\geq1,i=1,\cdots,n}$, *un tableau triangulaire tel que pour tout* $n \geq 1$ *et* $1 \leq i \leq n$,

$$\mathbb{E}\left(X_{n,i}\right) = 0 \tag{3.6}$$

$$\sum_{i=1}^{n} Var\left(X_{n,i}\right) = 1 \tag{3.7}$$

$$(X_{n,i})_{n\geq1,i=1\cdots n} \ \ sont \ iid. \tag{3.8}$$

$$\forall\varepsilon > 0, \max_{i\leq n} \left\{ \mathbb{E}\left(X_{n,i}^2 \mathbb{I}_{\{|X_{n,i}|>\varepsilon\}} \right) \right\} \to 0 \ lorsque \ n \to +\infty. \tag{3.9}$$

Alors

$$\sum_{i=1}^{n} X_{n,i} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ quand } n \to +\infty.$$

Nous faisons usage de ce lemme dans la démonstration du théorème suivant qui fournit les résultats de la normalité asymptotique.

Théorème 3.1.3 *Soit f, la densité à estimer sur $\mathbb{T}_I$ et $\widehat{f_n}$ son estimateur à noyau associé défini en (3.1). Pour chaque point x fixé dans $\mathbb{T}_I$ tel que $f(x) > 0$, et $(h_n)_{n\geq 1}$ une suite de paramètres de lissage telle que $h_n \to 0$ quand $n \to +\infty$, on a*

$$\frac{\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}}{\left[Var\left\{\widehat{f_n}(x)\right\}\right]^{1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \text{ quand } n \to +\infty,$$

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

Démonstration. Pour $x \in \mathbb{T}$ tel que $f(x) > 0$ considérons la suite $U_{n,i}(x,h_n)$ définie par

$$U_{n,i}(x,h_n) = \frac{K_{x,h_n}(X_i) - \mathbb{E}\left\{K_{x,h_n}(X_i)\right\}}{\sqrt{nVar\left\{K_{x,h_n}(X_1)\right\}}}.$$

En sommant sur n on obtient

$$\sum_{i=1}^{n} U_{n,i}(x,h_n) = \frac{\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}}{\left[Var\left\{\widehat{f_n}(x)\right\}\right]^{1/2}}$$

Il est facile de montrer que la suite $(U_{n,i}(x,h_n))_{n\geq 1, i\leq n}$ vérifie les trois premières conditions de Lemme 3.1.2 ; par conséquent nous ne montrons que la dernière condition. Considérons $A = \{\left|U_{n,i}(x,h_n)\right| > \varepsilon\}$. Puisque K_{x,h_n} est un noyau associé (donc uniformément borné en x et h), alors il existe $M > 0$ tel que pour tout $u \in A$ on a $|K_{x,h_n}(u) - \mathbb{E}\left\{K_{x,h_n}(u)\right\}| \leq M$. Il s'ensuit que $U_{n,i}^2(x,h_n) \leq M^2/\left[nVar\left\{K_{x,h_n}(X_1)\right\}\right]$. Enfin,

$$\max_{i\leq n}\left[\mathbb{E}\left\{U_{n,i}^2(x,h_n)\mathbf{1}_A\right\}\right] \leq \frac{M^2}{nVar\left\{K_{x,h_n}(X_1)\right\}} \to 0 \text{ quand } n \to +\infty,$$

ce qui conduit au résultat.■

3.2 Nouvelles convergences globales

Dans cette section, notre idée de convergence globale est liée à la continuité de f au lieu de sa continuité uniforme puisque la continuité sur $\mathbb{T}_I$ notée $C^0(\mathbb{T}_I)$ implique la continuité uniforme sur tout compact de $\mathbb{T}_I$. Nous montrons successivement, dans les paragraphes suivants, les convergences faible et forte globales de $\widehat{f_n}$ par la norme uniforme, puis la convergence en norme $\mathbb{L}^1$. Dans chacun des cas, nous montrons prioritairement la convergence du biais de $\widehat{f_n}$.

3.2.1 Convergence uniforme

La proposition suivante montre la convergence du biais de $\widehat{f_n}$ en norme uniforme.

Proposition 3.2.1 *Soit f une densité continue et bornée sur $\mathbb{T}_I$, et $\widehat{f_n}$ son estimateur à noyau associé donné en (3.1). Pour tout sous ensemble compact I de $\mathbb{T}_I$, on a :*

$$\sup_{x\in I} |\mathbb{E}\{\widehat{f_n}(x)\} - f(x)| \to 0 \quad \textit{quand} \quad n \to +\infty.$$

Démonstration. A partir des équations (3.3) et (3.4) on peut trouver pour tout $\delta > 0$, deux nombres entiers naturels n_0 et $n_1$tels que si on a : $n > n_0$ alors $|\mathbb{E}(\mathcal{Z}_{x,h}) - x| < \delta/2$ et lorsque $n > n_1$ on a $|Var(\mathcal{Z}_{x,h})| < \delta/4$. On peut majorer le biais de $\widehat{f_n}$ comme suit

$$\begin{aligned}
\left|\mathbb{E}\left\{\widehat{f_n}(x)\right\} - f(x)\right| &= \left|\int_{\mathbb{T}} (f(u) - f(x))\, K_{x,h_n}(u)du\right| \\
&\leq \int_{\mathbb{T}} \left|f(u) - f(x)\right| K_{x,h_n}(u)du \\
&\leq \int_{\left|\mathbb{E}(\mathcal{Z}_{x,h_n})-x\right|\leq\delta/2} \left|f(t) - f(x)\right| K_{x,h_n}(u)du \\
&\quad + \int_{\left|\mathbb{E}(\mathcal{Z}_{x,h_n})-x\right|>\delta/2} \left|f(u) - f(x)\right| K_{x,h_n}(u)du. \\
&= J_1 + J_2
\end{aligned}$$

Comme l'ensemble $\{x \in \mathbb{T}; |\mathbb{E}(\mathcal{Z}_{x,h}) - x| \leqslant \delta/2\} \subseteq \mathbb{S}_{x,h}$ et f est uniformément continue sur I, on obtient pour tout $n > n_0$

$$\begin{aligned}
J_1 &= \int_{|u-\mathbb{E}(\mathcal{Z}_{x,h})|\leqslant\delta/2} \left|f(u) - f(x)\right| K_{x,h_n}(u)du \\
&\leqslant \frac{\varepsilon}{2} \int_{|u-\mathbb{E}(\mathcal{Z}_{x,h})|\leqslant\delta/2} K_{x,h_n}(u)du \leqslant \frac{\varepsilon}{2}.
\end{aligned}$$

Par la suite, en utilisant l'inégalité de Bienaymé-Tchebychev, on obtient pour tout $n > n_1$,

$$\begin{aligned}
J_2 &= \int_{|u-\mathbb{E}(\mathcal{Z}_{x,h})|>\delta/2} \left|f(u) - f(x)\right| K_{x,h_n}(u)du \\
&\leqslant \sup_{u\in I} \left|f(u) - f(x)\right| \int_{|u-\mathbb{E}(\mathcal{Z}_{x,h})|>\delta/2} K_{x,h_n}(u)du \\
&\leqslant \frac{2\varepsilon}{\delta^2} |Var(\mathcal{Z}_{x,h})| \leqslant \frac{\varepsilon}{2}.
\end{aligned}$$

Par conséquent, en prenant $n > \max(n_0, n_1)$, on obtient

$$\sup_{x\in I} |\mathbb{E}\{\widehat{f_n}(x)\} - f(x)| \leqslant \varepsilon. \blacksquare$$

Nous montrons dans le théorème suivant, les résultats des convergences faible ($\mathbb{P}$) et forte ($p.s.$) en norme uniforme.

Théorème 3.2.2 *Soit f une densité continue et bornée sur $\mathbb{T}_I$, et $\widehat{f_n}$ son estimateur à noyau associé donné en (3.1). On suppose qu'il existe un plus grand nombre réel $r_0 = r_0(K) > 0$ dépendant du type de noyau K tel que, pour tout $x \in \mathbb{T}_I$,*

$$h_n^{r_0} \int_{\mathbb{S}_{x,h}} |dK_{x,h_n}(s)| \leqslant c_0(x)$$

avec $c_0(x)$ bornée sur tout compact contenant x. Pour tout sous ensemble compact I de $\mathbb{T}_I$, on a :

$$(i) \quad si \lim_{n\to+\infty} nh_n^{2r_0} = +\infty \quad alors \quad \sup_{x\in I} \left|\widehat{f_n}(x) - f(x)\right| \xrightarrow{\mathbb{P}} 0 \quad quand \quad n \to \infty;$$

$$(ii) \quad si \lim_{n\to+\infty} nh_n^{2r_0}/\log n = +\infty \quad alors \quad \sup_{x\in I} \left|\widehat{f_n}(x) - f(x)\right| \xrightarrow{p.s.} 0 \quad quand \quad n \to +\infty.$$

Démonstration. En utilisant l'inégalité triangulaire définie en (3.5) et la Proposition 3.2.1 on a

$$\sup_{x\in I} \left|\widehat{f_n}(x) - f(x)\right| \leqslant \sup_{x\in I} \left|\widehat{f_n}(x) - \mathbb{E}\{\widehat{f_n}(x)\}\right|. \tag{3.10}$$

Nous allons dans ce qui suit montrer respectivement que

$$\sup_{x\in I} \left|\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}\right| \xrightarrow{\mathbb{P}} 0 \quad \text{quand} \quad n \to +\infty,$$
$$\sup_{x\in I} \left|\widehat{f_n}(x) - \mathbb{E}\left\{\widehat{f_n}(x)\right\}\right| \xrightarrow{p.s.} 0 \quad \text{quand} \quad n \to +\infty.$$

Considérons F et $F_n = (1/n)\sum_{i=1}^n \delta_{X_i}$, respectivement les fonctions de répartitions théoriques et empiriques de X_i. A partir de F et F_n, on peut écrire

$$\begin{aligned}\widehat{f_n}(x) &= \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} K_{x,h_n}(u)dF_n(u) \\ &= K_{x,h_n}(u)F_n(u)\Big|_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} - \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} F_n(u)dK_{x,h_n}(u),\end{aligned} \tag{3.11}$$

et

$$\mathbb{E}\{\widehat{f_n}(x)\} = K_{x,h_n}(u)F(u)\Big|_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} - \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} F(u)dK_{x,h_n}(u). \tag{3.12}$$

Puisque d'après Glivenko-Cantelli F_n converge presque sûrement vers F quand n tend vers $+\infty$, alors la différence de (3.11) et (3.12) conduit à

$$\left|\widehat{f_n}(x) - \mathbb{E}\{\widehat{f_n}(x)\}\right| \leqslant \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} |F_n(u) - F(u)|\, dK_{x,h_n}(u).$$

En utilisant (3.10) et l'hypothèse $h_n^{r_0}\int_{\mathbb{S}_{x,h}} |dK_{x,h_n}(s)| \leqslant c_0(x)$, on obtient

$$\begin{aligned}\sup_{x\in I\subseteq\mathbb{T}_I} \left|\widehat{f_n}(x) - f(x)\right| &\leqslant \left(\sup_{x\in I\subseteq\mathbb{T}_I} |F_n(x) - F(x)|\right)\left(\sup_{x\in I\subseteq\mathbb{T}_I} \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} \left|dK_{x,h_n}(u)\right|\right) \\ &\leqslant C_0 h_n^{-r_0} \sup_{x\in I\subseteq\mathbb{T}_I} |F_n(x) - F(x)|\end{aligned}$$

avec $C_0 = \sup_{x\in I\subseteq\mathbb{T}_I} c_0(x)$. Pour un $h_n > 0$ donné, l'utilisation de l'inégalité de Massart (1990) (qui est une extension de celle de Dvoretzky-Kiefer-Wolfowitz) donne, pour tout $\varepsilon > 0$,

$$\mathbb{P}(\sup_{x\in I\subseteq\mathbb{T}_I} |F_n(x) - F(x)| \geqslant \varepsilon) \leqslant 2\exp\left(-2\varepsilon^2 nh_n^{2r_0}/C_0^2\right) \to 0;$$

et le résultat (i) en découle.

Pour la convergence *p.s.*, on a d'après (3.10) et l'inégalité précédente :

$$\sum_{n\geqslant 1} \mathbb{P}\left(\sup_{x\in I\subseteq\mathbb{T}_I} |\widehat{f_n}(x) - f(x)| \geqslant \varepsilon\right) \leqslant 2\sum_{n\geqslant 1} \exp\left(-2\varepsilon^2 nh_n^{2r_0}/C_0^2\right).$$

En prenant $\varepsilon \geqslant C_0 h_n^{r_0}\sqrt{(\log n)/n}$ et n assez grand, on obtient

$$\sum_{n\geqslant 1} \exp\left(-2\varepsilon^2 nh_n^{2r_0}/C_0^2\right) \simeq o\left(n^{-(1+\delta)}\right), \forall\delta > 0.$$

Ainsi, on a la convergence presque complète qui implique le résultat (ii).■

3.2.2 Convergence en norme L^1

Nous donnons maintenant les résultats de convergence en norme $\mathbb{L}^1$ par rapport à la mesure de Lebesgue sur $\mathbb{T}_I$. Tout comme dans la section précédente, nous montrons dans la proposition suivante que le biais de $\widehat{f_n}$ converge en norme $\mathbb{L}^1$. Ensuite, nous montrons dans un théorème les convergences faible et forte en norme $\mathbb{L}^1$.

Proposition 3.2.3 *Soit f une densité continue et bornée sur $\mathbb{T}_I$, et $\widehat{f_n}$ son estimateur à noyau associé donné en (3.1). Alors :*

$$\int_{\mathbb{T}_I} \left|\mathbb{E}\{\widehat{f_n}(x)\} - f(x)\right| dx \to 0 \text{ quand } n \to +\infty.$$

Démonstration. D'une manière similaire à Devroye (1987), nous utilisons la décomposition suivante

$$\left\|\mathbb{E}\left(\widehat{f_n}\right) - f\right\|_1 \leqslant \left\|\mathbb{E}\left(\widehat{f_n}\right) - \mathbb{E}\left(\widehat{f_n^\star}\right)\right\|_1 + \left\|\mathbb{E}\left(\widehat{f_n^\star}\right) - f^\star\right\|_1 + \left\|f^\star - f\right\|_1,$$

où $f^\star$ est continue et est telle que $\left\|(f - f^\star)\mathbf{1}_I\right\|_1 < \varepsilon, \forall\varepsilon > 0$ et $f^\star = 0$ sur le complémentaire de I défini par $I^c := \mathbb{T}_I \setminus I$. Pour un événement A, nous désignons par $\mathbf{1}_A$ la fonction indicatrice de A qui prend la valeur 1 si l'événement A est réalisé et 0 sinon. En considérant $\widehat{f_n^\star} = \mathbf{1}_I\widehat{f_n}$ on a d'une part

$$\begin{aligned}\left\|\mathbb{E}\left(\widehat{f_n}\right) - \mathbb{E}\left(\widehat{f_n^\star}\right)\right\|_1 &= \int_{\mathbb{T}_I}\left|\int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} K_{x,h_n}(u)\left\{f(u) - \mathbf{1}_I(x)f(u)\right\}du\right|dx \\ &\leqslant \Lambda\int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I}\left|f(u) - f^\star(u)\right|du \leqslant \Lambda\varepsilon\end{aligned}$$

où $\Lambda := \int_{\mathbb{T}_I} K_{x,h_n}(u)dx \simeq 1, \forall u$, est la masse totale du noyau associé K_{x,h_n}. D'autre part, on a

$$\begin{aligned}\left\|\mathbb{E}\left(\widehat{f_n^\star}\right) - f^\star\right\|_1 &\leqslant \int_I \left|\int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} K_{x,h_n}(u)\left\{f^\star(u) - f^\star(x)\right\}du\right| dx \\ &\leqslant \varepsilon \int_I \int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} K_{x,h_n}(u)dxdu \leqslant \Lambda\ell(I)\varepsilon,\end{aligned}$$

où $\ell(I)$ est la longueur de l'intervalle I. Puisque

$$\left\|f^\star - f\right\|_1 = \left\|(f - f^\star)\mathbf{1}_I\right\|_1 < \varepsilon,$$

on en déduit que

$$\left\|\mathbb{E}\left(\widehat{f_n}\right) - f\right\|_1 \leqslant [\Lambda\{1 + \ell(I)\} + 1]\,\varepsilon.\blacksquare$$

Enfin, nous allons donner dans le théorème suivant les deux derniers résultats de consistance, faible et forte de $\widehat{f_n}$ en norme $\mathbb{L}^1$.

Théorème 3.2.4 *Soit f une densité continue et bornée sur $\mathbb{T}_I$, et $\widehat{f_n}$ son estimateur à noyau associé donné en (3.1). On suppose qu'il existe un plus grand nombre réel $r_1 = r_1(K) > 0$ dépendant du type de noyau K tel que, pour tout $x \in \mathbb{T}_I$,*

$$h_n^{r_1} \int_{\mathbb{S}_{x,h}} |dK_{x,h_n}(s)| \leqslant c_1(x)$$

avec $c_1(x)$ intégrable sur tout compact contenant x. Alors on a :

$$(i) \quad si \lim_{n\to+\infty} nh_n^{2r_1} = +\infty \quad alors \int_{\mathbb{T}_I} \left|\widehat{f_n}(x) - f(x)\right| dx \xrightarrow{\mathbb{P}} 0 \quad quand \quad n \to +\infty\,;$$

$$(ii) \quad si \lim_{n\to+\infty} nh_n^{2r_1}/\log n = +\infty \quad alors \quad \int_{\mathbb{T}_I} \left|\widehat{f_n}(x) - f(x)\right| dx \xrightarrow{p.s.} 0 \quad quand \quad n \to +\infty.$$

Démonstration. Considérons l'inégalité triangulaire

$$\left\|\widehat{f_n} - f\right\|_1 \leqslant \left\|\widehat{f_n} - \mathbb{E}\left(\widehat{f_n}\right)\right\|_1 + \left\|\mathbb{E}\left(\widehat{f_n}\right) - f\right\|_1. \tag{3.13}$$

D'après la proposition précédente, le second terme du deuxième membre de (3.13) converge vers 0, autrement dit

$$\left\|\mathbb{E}\left(\widehat{f_n}\right) - f\right\|_1 \longrightarrow 0 \quad \text{quand} \quad n \to +\infty.$$

Donc, il nous suffit de montrer que le terme stochastique peut être majoré par une série de Riemann convergente. Puisque par hypothèse $\int_{\mathbb{S}_{x,h}\cap\mathbb{T}} \left|dK_{x,h_n}(u)\right| \leqslant c_1(x)h_n^{-r_1}$ avec

$c_1(x)$ intégrable, on a :

$$\begin{aligned}\left\|\widehat{f_n} - \mathbb{E}\left(\widehat{f_n}\right)\right\|_1 &= \int_{\mathbb{T}_I} \left|\widehat{f_n}(u) - \mathbb{E}\left\{\widehat{f_n}(u)\right\}\right| dx \\ &\leqslant \int_{\mathbb{T}_I}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} |F_n(u) - F(u)| \left|dK_{x,h_n}(u)\right| dx \\ &\leqslant \sup_{u\in\mathbb{T}_I} |F_n(u) - F(u)| \int_{\mathbb{T}_I}\int_{\mathbb{S}_{x,h}\cap\mathbb{T}_I} \left|dK_{x,h_n}(u)\right| dx \\ &\leqslant C_1 h_n^{-r_1} \sup_{u\in\mathbb{T}_I} |F_n(u) - F(u)|\end{aligned}$$

avec $C_1 = \displaystyle\int_{\mathbb{T}_I} c_1(x)dx \neq C_0$. L'inégalité de Massart (1990) permet d'écrire

$$\begin{aligned}\mathbb{P}\left(\left\|\widehat{f_n} - \mathbb{E}\left(\widehat{f_n}\right)\right\|_1 \geqslant \varepsilon\right) &\leqslant \mathbb{P}\left(\sup_{u\in\mathbb{T}_I} |F_n(u) - F(u)| \geqslant \frac{h_n^{r_1}\varepsilon}{C_1}\right) \\ &\leqslant 2\exp\left(-2\frac{\varepsilon^2}{C_1^2} n h_n^{2r_1}\right);\end{aligned}$$

ce qui conduit à la consistance faible. Par la suite, en utilisant le résultat précédent puis en sommant à l'infini on a

$$\sum_{n\geqslant 1} \mathbb{P}\left(\left\|\widehat{f_n} - \mathbb{E}\left(\widehat{f_n}\right)\right\|_1 \geqslant \varepsilon\right) \leqslant 2\sum_{n\geqslant 1} \exp\left(-\frac{2\varepsilon^2}{C_1^2} n h_n^{2r_1}\right).$$

En prenant $\varepsilon \geqslant \dfrac{C_1}{h_n^r}\sqrt{\dfrac{\log n}{n}}$ et n assez grand on trouve

$$\sum_{n\geq 1} \exp\left(-\frac{2\varepsilon^2}{C_1^2} n h_n^{2r_1}\right) \simeq o(n^{-(1+\delta)}), \quad \forall \delta > 0.$$

Ce qui permet d'obtenir la convergence presque complète qui implique le résultat (*ii*).■

La remarque suivante vient compléter notre idée de faire une étude avec des conditions souples pour souligner la sensibilité des noyaux associés continus, qui dépendent fortement du point d'estimation.

Remarque 3.2.5 *Soit $x \in \mathbb{T}_I$.*

(*i*) *Si $c_0(x)$ du Théorème 3.2.2 garantissant un type de "variation ponctuelle bornée" du noyau associé continu est intégrable sur tout compact contenant x alors $r_1 = r_0$, où r_1 est donné dans le Théorème 3.2.4.*

(*ii*) *La fonction $c_2(x)$ dans le Théorème 3.1.1 traduit le fait que le noyau associé continu soit ponctuellement de carré intégrable.*

(*iii*) *Pour les noyaux (associés) classiques $\mathcal{K}$, on vérifie facilement que $c_j(x) = c_j$ dans les Théorèmes 3.1.1, 3.2.2 et 3.2.4, sont des constantes pour $j = 0, 1, 2$; Elles donnent un même sens respectivement aux hypothèses de "carré intégrable de $\mathcal{K}$" pour $j = 2$, "variations bornées de $\mathcal{K}$" pour $j = 0$, et "variation intégrable de $\mathcal{K}$" pour $j = 1$. En outre on a $r_0 = r_1 = r_2 = 1$.*

3.3 Étude des estimateurs à noyaux lognormaux

Nous examinons ici trois familles d'estimateurs à noyaux associés lognormaux.

Tout d'abord, nous rappelons les propriétés de base des types de noyaux lognormaux dont nous avons besoin. La fonction de densité du noyau lognormal à deux paramètres $a \in \mathbb{R}$ et $b > 0$ est définie comme suit :

$$LN_{\theta(a,b)}(u) = \frac{1}{ub\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\log u - a}{b}\right)^2\right\} \mathbf{1}_{(0,\infty)}(u).$$

D'après sa construction dans la Section 2.1.3, la méthode mode-dispersion conduit d'une part à la paramétrisation

$$\theta(x,h) = \left(\log x + h^2, h\right) \text{ pour tout } x \in]0,+\infty[\text{ et } h > 0.$$

Ce qui a permis de calculer les caractéristiques suivants

$$A_\theta(x,h) = x\left(\exp(3h^2/2) - 1\right) \quad \text{et} \quad B_\theta(x,h) = x^2 \exp(3h^2)\left(\exp h^2 - 1\right)$$

pour tout $x \in \mathbb{S}_{LN_{\theta(x,h)}} =]0,+\infty[= \mathbb{S}_{LN_{\theta(a,b)}}$.

Pour estimer une densité de probabilité inconnue f de support $]0,+\infty[$, on peut utiliser le noyau associé lognormal $LN_{\theta(x,h)}$ issu de la méthode mode-dispersion et qui est défini par

$$LN_{\theta(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(\frac{-1}{2h^2}\left[\log u - \log\left\{x \exp\left(h^2\right)\right\}\right]\right) \mathbf{1}_{u>0}, \quad x > 0, h > 0.$$

Suivant l'algorithme de réduction de biais, on peut d'après la première étape considérer $\alpha_1(h) > 0$ et on définit les régions de bord et de l'intérieur par $\mathbb{T} =]0,+\infty[=]0,\alpha_1(h)] \cup]\alpha_1(h),+\infty[$, puis on modifie le noyau $LN_{\theta(x,h)}$ en $LN_{\widetilde{\theta}(x,h)} := \widetilde{LN}_{\theta(x,h)}$ et défini par

$$LN_{\widetilde{\theta}(x,h)} = LN_{\widetilde{\theta}_{-1}(x,h)} + LN_{\widetilde{\theta}_0(x,h)}$$

où la paramétrisation du bord gauche notée $\widetilde{\theta}_{-1}(x,h)$ est donnée par

$$\widetilde{\theta}_{-1}(x,h) = \theta\left(x^{-2}\alpha_1^3(h)\exp\left(-3h^2/2\right), h\right) \text{ pour } x \in]0,\alpha_1(h)] =: \mathbb{T}_{-1}$$

et celle de l'intérieur notée $\widetilde{\theta}_0(x,h)$ est donnée par

$$\widetilde{\theta}_0(x,h) = \theta\left(x\exp\left(-3h^2/2\right), h\right) \text{ pour } x \in]\alpha_1(h),+\infty[=: \mathbb{T}_0.$$

Cependant, une autre version du noyau lognormal donnée par

$$LN_{\theta^*(x,h)} = LN_{\theta\left(x\exp h^2, 2\sqrt{\log(1+h)}\right)}$$

avait été proposée par Jin & Kawczak (2003). En utilisant (3.1) nous obtenons les trois estimateurs $\widehat{f_n}$, $\widetilde{f_n}$ et f_n^* de f correspondant aux trois noyaux lognormaux LN_θ, $LN_{\widetilde{\theta}}$ et LN_{θ^*}, respectivement.

Dans ce qui suit, nous allons procéder à la détermination de : $\widehat{r_j} = r_j(\widehat{f_n})$, $\widetilde{r_j} = r_j(\widetilde{f_n})$ et $r_j^* = r_j(f_n^*)$, respectivement des estimateurs à noyaux lognormaux LN_θ, $LN_{\widetilde{\theta}}$ et LN_{θ^*} comme dans les Théorèmes 3.1.1, 3.2.2 et 3.2.4 avec $j = 0,1,2$.

Calcul de $\widehat{r_0}$ et $\widehat{r_1}$

Tout d'abord, signalons que les calculs de $\widehat{r_0}$ et $\widehat{r_1}$ sont presque similaires et se font à l'aide de la dérivée première du noyau associé utilisé. Pour le cas de l'estimateur à noyau lognormal construit par la méthode mode-dispersion $LN_{\theta(x,h)}(\cdot)$ on procède de la manière suivante. Partant du fait que $LN_{\theta(x,h)}(\cdot)$ puisse s'écrire sous la forme

$$LN_{\theta(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left[\frac{\log(u) - \log\{x\exp(h^2)\}}{h}\right]^2\right)\mathbf{1}_{[0,+\infty[}(u),$$

on détermine sa fonction dérivée $dLN_{\theta(x,h)}$ donnée par

$$\left|dLN_{\theta(x,h)}(u)\right| = \left|\frac{-1}{h^2u}\log(u/x)LN_{\theta(x,h)}(u)\right|.$$

En majorant son intégrale on a par conséquent

$$\begin{aligned}\int \left|dLN_{\theta(x,h)}(u)\right| &= \int \left|\frac{-1}{h^2u}\log(u/x)LN_{\theta(x,h)}(u)\right| du \\ &\leqslant \frac{\kappa}{h^2}\int \left|\frac{2}{u}\log(x)\right| du \\ &\leqslant h^{-2}c_0(x)\end{aligned}$$

avec $c_0(x) = \kappa\log^2(x)$ qui en une fonction bornée et intégrable sur tout compact $I \ni x$. D'après le point (i) de la Remarque 3.2.5, on a $c_0(x) = c_1(x)$ et donc $\widehat{r_0} = \widehat{r_1} = 2$.

Calcul de $\widehat{r_2} = \widehat{r_2}(x)$

Le calcul de $\widehat{r_2}$ se fait à partir de (2.20). Plus précisément en exprimant explicitement la norme carrée du noyau associé en question. Dans le cas présent, nous procédons en faisant le changement de variable $v = u^2/x^2$ et on a

$$\begin{aligned}\left\|LN_{\theta(x,h)}\right\|_2^2 &= \int LN^2_{\theta(x,h)}(u)du \\ &\overset{(v=u^2/x^2)}{=} \frac{1}{2xh\sqrt{2\pi}}\mathbb{E}\left(\mathcal{Y}^{-1/2}\right) \\ &= \frac{1}{2xh\sqrt{2\pi}}\exp(-h^2/8) \\ &\leqslant \frac{1}{2xh\sqrt{2\pi}}\end{aligned}$$

avec $\mathcal{Y}$ une variable aléatoire de loi $LN_{\theta(1,h/\sqrt{2})}$. On obtient alors $r_2(x) = 1$ car $e^w \leqslant 1$ pour $w \leqslant 0$.

Calcul de $\widetilde{r}_0$ et $\widetilde{r}_1$

Rappelons que le noyau lognormal modifié $LN_{\widetilde{\theta}(x,h)}$ est défini par

$$\begin{aligned} LN_{\widetilde{\theta}(x,h)}(u) = & \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{\frac{\alpha^3(h)\exp(-h^2/2)}{x^2}\right\}\right]^2\right)\mathbf{1}_{]0,\alpha(h)]}(u) \\ & + \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{x\exp(-h^2/2)\right\}\right]^2\right)\mathbf{1}_{]\alpha(h),+\infty[}(u). \end{aligned} \tag{3.14}$$

Le calcul de sa fonction dérivée donne

$$dLN_{\widetilde{\theta}(x,h)}(u) = \frac{1}{h^2}\left|\frac{1}{u}\log\left\{\frac{ux^2}{\alpha(h)}\right\}LN_{\widetilde{\theta}_{-1}(x,h)}(u) + \frac{1}{u}\log\left(\frac{u}{x}\right)LN_{\widetilde{\theta}_0(x,h)}(u)\right|.$$

En calculant l'intégrale de $dLN_{\widetilde{\theta}(x,h)}$ puis en la majorant on a successivement

$$\begin{aligned} \int \left|dLN_{\widetilde{\theta}(x,h)}(u)\right| &= \frac{1}{h^2}\int\left|\frac{1}{u}\log\left\{\frac{ux^2}{\alpha(h)}\right\}LN_{\widetilde{\theta}_{-1}(x,h)}(u) + \frac{1}{u}\log\left(\frac{u}{x}\right)LN_{\widetilde{\theta}_0(x,h)}(u)\right| du \\ &\leqslant \frac{\kappa_{-1}}{h^2}\int\left|\frac{2}{u}\log(x) + \frac{1}{u}\log(u)\right|\mathbf{1}_{]0,\alpha(h)]}(u)du \\ &\quad + \frac{\kappa_0}{h^2}\int\left|\frac{2}{u}\log(x)\right|\mathbf{1}_{]\alpha(h),+\infty[}(u)du \\ &\leqslant h^{-2}\widetilde{c}_0(x) \end{aligned}$$

avec $\widetilde{c}_0(x) = \dfrac{5\kappa_{-1}}{2}\log^2(x)\mathbf{1}_{]0,\alpha(h)]}(x) + \kappa_0\log^2(x)\mathbf{1}_{]\alpha(h),+\infty[}(x)$ qui est une fonction bornée et intégrable sur tout compact $I \ni x$ et donc $c_0(x) = c_1(x)$. Une fois de plus d'après le point (i) de la Remarque 3.2.5, on a $\widetilde{r}_0 = \widetilde{r}_1 = 2$.

Calcul de $\widetilde{r}_2 = \widetilde{r}_2(x)$

Ici, nous allons déterminer le réel $\widetilde{r}_2$ seulement aux bords $]0,\alpha(h)]$ noté $\widetilde{r}_{2,-1}$ puisque c'est le cas délicat et le plus intéressant. Autrement dit, nous utilisons uniquement le premier membre de l'équation (3.14) pour déterminer ce réel. Rappelons que la version modifiée du noyau lognormal au bord est définie par

$$LN_{\widetilde{\theta}_{-1}(x,h)}(u) = \frac{1}{uh\sqrt{2\pi}} \exp\left(-\frac{1}{2h^2}\left[\log(u) - \log\left\{\frac{\alpha^3(h)\exp(-h^2/2)}{x^2}\right\}\right]^2\right)\mathbf{1}_{]0,\alpha(h)]}(u)$$

Pour la détermination de $\widetilde{r}_{2,-1}$, nous allons maintenant calculer explicitement la norme carrée de $\left\|LN_{\widetilde{\theta}_{-1}(x,h)}\right\|_2^2$ en faisant le changement de variables $v = u^{\sqrt{2}}$ puis nous la

majorons. On obtient successivement

$$\begin{aligned}
\left\|LN_{\widetilde{\theta}_{-1}(x,h)}\right\|_2^2 &\overset{(v=u^{\sqrt{2}})}{=} \int_0^{\alpha(h)} \frac{v^{-\sqrt{2}/2}}{vh^2\pi\sqrt{8}} \exp\left(-\frac{1}{2h^2}\left[\log(v) - \log\left\{\frac{\alpha^{3\sqrt{2}}(h)}{x^{2\sqrt{2}}e^{\sqrt{2}h^2/2}}\right\}\right]^2\right) dv \\
&\leqslant \frac{1}{2h\sqrt{2\pi}} \mathbb{E}\left(\mathcal{Y}^{-\sqrt{2}/2}\right) \\
&= \frac{1}{2h\sqrt{2\pi}} \left\{\frac{x^2}{\alpha^3(h)}\right\}^{\sqrt{2}/2} \exp\left(\frac{\sqrt{2}+1}{4}h^2\right) \\
&\leqslant \frac{x^{2\sqrt{2}}\left[1 + \{(\sqrt{2}+1)/4\}h^2\right]}{2h\{\alpha(h)\}^{3\sqrt{2}/2}\sqrt{\pi}},
\end{aligned}$$

où $\mathcal{Y}$ est une variable aléatoire de loi $LN_{\theta(x^*,h)}$ avec $x^* = x^{-2\sqrt{2}}\{\alpha(h)\}^{3\sqrt{2}} \exp(-\sqrt{2}h^2/2 - h^2)$. Ce calcul nous a permis de savoir qu'en fait $\widetilde{r}_{2,-1}(x) = \widetilde{r}_{2,-1}\{x, \alpha(h)\} > 0$ dépend à la fois du point d'estimation et de la limitation de la région bord. Par exemple pour $\alpha(h) = a_1 h^{\alpha_1}$ on a $\widetilde{r}_{2,-1}(x) = (3\alpha_1\sqrt{2} + 2)/2 > 0$ pour $\alpha_1 > -\sqrt{2}/3$. Notons aussi que si $\alpha(h)$ est une constante, alors $\widetilde{r}_{2,-1} = 1$.

Calcul de r_0^* et r_1^*

Nous procédons ici au calcul des réels r_0^* et r_1^* en utilisant le noyau lognormal proposé par Jin & Kawczak (2003). Ce noyau, comme nous l'avons dit un peu plus haut, est défini par

$$\begin{aligned}
LN_{\theta^*(x,h)}(u) &= \frac{1}{2u\sqrt{2\pi\log(1+h)}} \exp\left\{-\frac{\log^2(u/x)}{8\log(1+h)}\right\} \mathbf{1}_{[0,+\infty[}(u) \\
&= LN_{\theta(x\exp(h^2), 2\sqrt{\log(1+h)})}(u).
\end{aligned}$$

Le calcul de sa fonction dérivée donne

$$dLN_{\theta^*(x,h)}(u) = \left\{-\frac{1}{u} + \frac{\log(x) - \log(u)}{32u\log^2(1+h)}\right\} LN_{\theta^*(x,h)}(u)$$

Une majoration de l'intégrale de $dLN_{\theta^*(x,h)}$ permet d'avoir successivement

$$\begin{aligned}
\int \left|dLN_{\theta^*(x,h)}(u)\right| &= \int \left|\left\{-\frac{1}{u} + \frac{\log(x) - \log(u)}{32u\log^2(1+h)}\right\} LN_{\theta^*(x,h)}(u)\right| du \\
&\leqslant \frac{\kappa^*}{32\log^2(1+h)} \int \left|\frac{2}{u}\log(x)\right| du \\
&\leqslant \frac{1}{\log^2(1+h)} c_0^*(x) \\
&\simeq \frac{1}{h^2} c_0^*(x)
\end{aligned}$$

car aux voisinages de 0, on a $\log(1+h) \simeq h$. Ici, la quantité $c_0^*(x)$ vaut $c_0^*(x) = \dfrac{\kappa^*}{32}\log^2(x)$ et c'est une fonction bornée et intégrable sur tout compact $I \ni x$, par conséquent $c_0^*(x) = c_1^*(x)$. Par le point (i) de la Remarque 3.2.5, on a $r_0^* = r_1^* = 2$.

Calcul de $r_2^* = r_2^*(x)$

Nous montrons dans ce paragraphe la détermination du réel $r_2^* = r_2^*(x)$ dans l'estimation de densité par le noyau lognormal $LN_{\theta^*(x,h)}$ proposé par Jin & Kawczak (2003). Comme dans les cas précédents, nous allons exprimer explicitement la norme carrée de $LN_{\theta^*(x,h)}$ en faisant le changement de variables $v = u^{\sqrt{2}}/x^{\sqrt{2}}$. On obtient alors

$$\begin{aligned}
\left\|LN_{\theta^*(x,h)}\right\|_2^2 &\overset{(v=u^{\sqrt{2}}/x^{\sqrt{2}})}{=} \frac{1}{4x\sqrt{2\pi\log(1+h)}}\mathbb{E}\left(\mathcal{Y}_*^{-1/\sqrt{2}}\right) \\
&= \frac{1}{4x\sqrt{2\pi\log(1+h)}}\exp\left\{\frac{1}{2}\log(1+h)\right\} \\
&= \frac{1}{4x\sqrt{2\pi\log(1+h)}}\exp\{w_*(h)\} \\
&\leqslant \frac{1}{4x\sqrt{2\pi h}}
\end{aligned}$$

où $\mathcal{Y}_*$ est une variable aléatoire réelle de loi $LN_{\theta^*(1,h)} = LN_{\theta(\exp(h^2),2\sqrt{\log(1+h)})}$ et $w_*(h)$ définie par $w_*(h) = -h^2\sqrt{2}/2 - (2\sqrt{2}-1)\log(1+h)$, par conséquent on a $r_2^*(x) = 1/2$.

Nous résumons dans le Tableau 3.1 les différentes valeurs de r_0, r_1, r_2, $\widetilde{r}_0$, $\widetilde{r}$ et $\widetilde{r}_2$ obtenues sur les estimateurs à noyaux associés non-classiques comme bêta et sa version étendue, gamma et son inverse, gaussien inverse et sa réciproque, Pareto ainsi que le lognormal. Nous signalons également que pour les estimateurs à noyaux associés classiques nous avons obtenu ces différentes valeurs en utilisant la forme générale $K_{x,h}(\cdot) = (1/h)\mathcal{K}\{(x-\cdot)/h\}$. Aussi, nous avons donné les résultats avec le noyau lognormal de Jin & Kawczak (2003) ainsi que ceux de Bouezmarni & Scaillet (2005) sur gaussien inverse et sa réciproque.

Table 3.1 – Tableau récapitulatif des différents r dans les Théorème 3.1.1, Théorème 3.2.2 et Théorème 3.2.4.

$K = K_{\theta(a,b)}$	$\mathbb{S}_K$	$\mathbb{T}$	r_0	$\widetilde{r}_0$	r_1	$\widetilde{r}_1$	$r_2(x)$	$\widetilde{r}_{2,-1}(x)$
Bêta	$[0,1]$	$[0,1]$	1	1	1	1	1/2	1
Bêta étendu	$[t_1,t_2]$	$[t_1,t_2]$	1	1	1	1	1/2	1
Gamma	$[0,+\infty[$	$[0,+\infty[$	1	1	1	1	1/2	1
Gamma inverse	$]0,+\infty[$	$]0,1/h[$	1		1		1/2	
Gaussien inverse	$]0,+\infty[$	$]0,1/(3h)[$	1	1	1	1		
Gaussien inverse réc.	$]0,+\infty[$	$]0,+\infty[$	1		1			
Pareto	$]a,+\infty[$	$]x,+\infty[$	1	1	1	1	1	1
Log-normale	$]0,+\infty[$	$]0,+\infty[$	2	2	2	2	1	$\widetilde{r}_2(\alpha)$
Log-normale JK'03	$]0,+\infty[$	$]0,+\infty[$	2	–	2	–	1/2	–
Classiques $\mathcal{K}$	$x-h\mathbb{S}$	$\mathbb{R}$	1	–	1	–	1	–
Gauss. inv. BS'05	$]0,+\infty[$	$]0,+\infty[$	5/2	–	5/2	–	1/2	–
Gauss. inv. réc. BS'05	$]0,+\infty[$	$]0,+\infty[$	5/2	–	5/2	–	1/2	–

Chapitre 4

Propriétés minimax des estimateurs à noyaux associés

Ce chapitre se propose d'étudier les propriétés minimax des estimateurs de densités par les noyaux associés présentés dans le Chapitre 2. Nous rappelons ici que les noyaux associés sont rigoureusement définis par une famille de fonctions de densités $\mathscr{K} = \{K_{x,h}\}_{x\in\mathbb{T},h>0}$, paramétrée par le point d'estimation x et le paramètre de dispersion h sur un support $\mathbb{S}_{x,h} \subseteq \mathbb{R}$ et qui vérifie les relations suivantes :

$$x \in \mathbb{S}_{x,h} \tag{4.1}$$

$$\mathbb{E}\left(\mathcal{Z}_{x,h}\right) = x + A(x,h) \tag{4.2}$$

$$Var\left(\mathcal{Z}_{x,h}\right) = B(x,h), \tag{4.3}$$

où $\mathcal{Z}_{x,h}$ est une variable aléatoire de loi $K_{x,h}$, et les deux caractéristiques $A(x,h)$ et $B(x,h)$ de $K_{x,h}$ tendent vers 0 quand $h = h_n$ tend vers 0 (quand $n \to +\infty$ par la suite). Notons que :

- Le support $\mathbb{T}$ de la densité inconnue f, à estimer est considéré connu par le statisticien et doit coïncider avec le support $\mathbb{S}_{x,h}$ du noyau associé afin d'éviter les effets de bords.
- x est un paramètre de position (le mode), et h est un paramètre de dispersion.
- $K_{x,h}$ est de plus en plus concentré sur x quand h tend vers 0.

Dans le cas particulier où $\mathbb{T} = \mathbb{R}$, on a

$$K_{x,h}(\cdot) = (1/h)\,\mathcal{K}\left\{(\cdot - x)/h\right\}, \tag{4.4}$$

où la fonction noyau $\mathcal{K}$ de support $\mathbb{S}$ est le noyau classique présenté dans la Section 1.1 du premier chapitre de cette thèse. D'après (4.1) et l'expression (4.4) on a

$$\mathbb{S}_{x,h} = x - h\mathbb{S},\ A(x,h) = 0 \text{ et } B(x,h) = h^2\sigma^2_{\mathcal{K}}, \tag{4.5}$$

où $\sigma^2_{\mathcal{K}} > 0$ est la variance de $\mathcal{K}$. Par exemple pour les noyaux populaires comme le noyau gaussien et celui d'Epanechnikov on a $\mathbb{S}_{x,h} = \mathbb{R}$ et $\mathbb{S}_{x,h} = [x-h, x+h]$, respectivement.

Comme nous l'avons présenté dans le Chapitre 2, pour construire un noyau associé $K_{x,h}$ on considère un type de noyau $K := K_\theta$ (qui est une densité de probabilité paramétrée et de carré intégrable) sur $\mathbb{S} = \mathbb{S}_\theta$ avec $\theta \in \Theta \subseteq \mathbb{R}^k$ pour $k \geq 2$. En particulier,

on choisit $\theta = \theta(a,b)$ avec a et b des réels positifs et $K := K_\theta$ un noyau uni-modal. Ainsi, on utilise la méthode "mode-dispersion" incluant les cas particuliers de beta et gamma. Cette méthode consiste à résoudre les deux équations : $x = M(a,b)$ et $h = D(a,b)$ où $M(a,b)$ et $D(a,b)$ sont respectivement l'unique mode et le paramètre de dispersion de $K_{\theta(a,b)}$. Ce qui permet d'obtenir le noyau associé $K_{\theta(x,h)} := K_{\theta(a(x,h),b(x,h))}$ de support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$, où $a(x,h)$ et $b(x,h)$ sont les solutions des équations précédentes.

Étant donnée la famille $\mathscr{K}$ des noyaux associés, un estimateur issu de ces noyaux est défini par :

$$\widehat{f_{n,h}}(x) = \frac{1}{n}\sum_{k=1}^{n} K_{x,h}(X_k), \quad \forall x \in \mathbb{T}, \forall h > 0. \tag{4.6}$$

Rappelons que le but principal lorsqu'on fait de l'estimation d'une densité f, est de choisir, parmi tous les estimateurs possibles $\widehat{f_{n,h}}$, celui qui est le plus proche de f (le meilleur). Pour ceci, plusieurs approches s'offrent pour juger la qualité d'un estimateur parmi lesquelles figure l'approche minimax. Dans cette approche, on fixe prioritairement un espace fonctionnel sur lequel on définit un risque. Puis, on cherche le meilleur estimateur autrement dit celui dont le risque est minimal. Cette approche suppose une connaissance *a priori* assez forte sur la densité à estimer plus précisément sur son support et sa régularité.

Ce chapitre s'organise de la manière suivante : dans la Section 4.1, nous présentons le modèle sur lequel nous travaillons. Ensuite, nous donnons les hypothèses techniques (tant sur le noyau que sur la densité) de notre travail. Par la suite, nous présentons nos résultats résumés en quatre propositions et un théorème.

4.1 Description du modèle

Considérons une suite de variables aléatoires $X_1, \ldots, X_n$ iid et de densité de probabilité $f : \mathbb{T} \to \mathbb{R}$, où $\mathbb{T} \subseteq \mathbb{R}$ est le support de la densité f. Notre objectif est de construire une procédure d'estimation de f en supposant que son support $\mathbb{T}$ est connu par le statisticien.

Plus précisément, nous voulons construire une procédure, utilisant les noyaux associés continus précédemment décrits, qui sélectionne automatiquement le paramètre de lissage afin d'estimer f. Pour ce faire, considérons les notations suivantes.

4.1.1 Risque minimax

Soit $\mathbb{F}$ la classe des fonctions définies par

$$\mathbb{F} \subseteq \left\{ f : \mathbb{T} \to \mathbb{R} : f \geq 0, \int_{\mathbb{T}} f = 1, \|f\|_\infty \leq f_\infty \right\}, \tag{4.7}$$

où f_∞ est un nombre réel positif fixé.

Puisque nous nous sommes intéressés à l'estimation de densité, nous allons nous concentrer sur la perte $\mathbb{L}^1$ qui est étroitement liée à la distance où variation totale entre les lois de probabilité sous-jacentes.

Si $\widehat{f_{n,h}}$ est un estimateur arbitraire de f, nous définissons son risque par rapport à f de la manière suivante

$$R_n(\widehat{f_{n,h}}, f) = \mathbb{E}_f^n \|\widehat{f_{n,h}} - f\|_1 \tag{4.8}$$

où $\|\cdot\|_1$ est la norme standard sur $\mathbb{L}^1$ et $\mathbb{E}_f^n$ est la moyenne par rapport à la mesure de probabilité sur $\mathbb{T}$.

Ensuite, nous définissons le risque maximal sur la classe $\mathbb{F}$ de la manière suivante :

$$R_n(\widehat{f_{n,h}}, \mathbb{F}) = \sup_{f \in \mathbb{F}} R_n(\widehat{f_{n,h}}, f). \tag{4.9}$$

L'approche minimax consiste à trouver un estimateur (ou de manière équivalente au choix de fenêtre de lissage) qui atteint asymptotiquement et à une constante multiplicative près, le taux de convergence théorique suivante :

$$r_n(\mathbb{F}) = \inf_{\widehat{f_{n,h}}} R_n(\widehat{f_{n,h}}, \mathbb{F}), \tag{4.10}$$

où l'infimum est pris sur tous les estimateurs.

Les inconvénients de l'approche minimax sont bien connus. En général, l'estimateur qui atteint la vitesse de convergence minimax sur une boule d'un espace fonctionnel classique (par exemple les espaces d'Hölder, de Sobolev ou Besov) dépend explicitement des paramètres de régularité de cet espace. De plus, en situation pratique, ces paramètres ne sont pas connus par le statisticien et, par conséquent, l'estimateur ne peut pas être construit.

Il existe une autre approche dite "adaptive" qui propose une tentative de solution pour remédier à cet inconvénient. Elle suppose que l'espace $\mathbb{F}$ est de la forme $\mathbb{F} = \cup_\varkappa \mathbb{F}_\varkappa$ où les espaces $\mathbb{F}_\varkappa$ sont tels que le taux $r_n(\mathbb{F}_\varkappa)$ tend vers 0 plus vite que $r_n(\mathbb{F})$. L'idée principale est de construire une seule procédure d'estimation qui permet d'obtenir le taux de $r_n(\mathbb{F}_\varkappa)$ simultanément sur chaque espace $\mathbb{F}_\varkappa$.

Plus précisément, une procédure d'estimation $f_{n,h}^*$ est appelée adaptative sur l'espace $\{\mathbb{F}_\varkappa\}_\varkappa$ si les conditions suivantes sont vérifiées :

$$\limsup_{n\to\infty} \sup_{\varkappa} \sup_{f \in \mathbb{F}_\varkappa} r_n^{-1}(\mathbb{F}_\varkappa) R_n(f_{n,h}^*, \mathbb{F}_\varkappa) < +\infty.$$

Cette approche est très intéressante mais nous ne l'abordons pas dans ce travail. Le lecteur peut se référer à Bertin & Klutchnikoff (2013) pour s'inspirer du cas particulier des estimateurs à noyau bêta où les auteurs ont procédé au choix du paramètre de lissage par la procédure de Lepski modifié.

Nous donnons dans le paragraphe suivant les différentes hypothèses techniques avec lesquelles nous allons travailler.

4.1.2 Hypothèses techniques

Hypothèses sur la densité

Dans la suite on supposera que la densité inconnue f appartient à une boule de l'espace de Hölder. Le paramètre de régularité de cet espace est désigné par $s > 0$ et le rayon de la boule est désigné par $L > 0$.

Ici et par la suite, on suppose que $\mathbb{T}$ est un ensemble connexe de la droite réelle. Soit $m_s = \sup\{\ell \in \mathbb{N} : \ell < s\}$. On dit que g appartient à $\mathbb{F}(s, L)$ si g est une densité de probabilité qui est m_s fois différentiable et telle que :

$$\forall (u, v) \in \mathbb{T}^2, \quad \left|g^{(m_s)}(u) - g^{(m_s)}(v)\right| \leq L|u - v|^{s-m_s}$$

où les dérivées $g^{(k)}$ sont définies pour $1 \leq k \leq m_s$ par :

$$g^{(k)}(u) = \lim_{\substack{v \to u \\ v \in \mathbb{T}\setminus\{u\}}} \frac{g^{(k-1)}(v) - g^{(k-1)}(u)}{v - u}.$$

Il est bien connu (voir [Tsybakov, 2004] par exemple) qu'en vertu de ces conditions, la quantité suivante est bien définie et finie :

$$\mathrm{f}_\infty(s, L) = \sup_{1 \leq k \leq m_s} \sup_{g \in \mathbb{F}(s,L)} \|g^{(k)}\|_\infty.$$

Hypothèses sur le noyau associé

Maintenant, nous présentons quelques conditions que la famille $\mathscr{K} = \{K_{x,h}\}_{x \in \mathbb{T}, h \in (0,1]}$ doit satisfaire afin d'obtenir les résultats minimax.

Condition H1(A) Nous supposons ici que pour tout noyau associé $K_{x,h}$ appartenant à la famille $\mathscr{K}$ il existe une fonction $A : \mathbb{T} \to \mathbb{R}_+^*$ vérifiant :

$$\forall (x, h) \in \mathbb{T} \times (0, 1], \quad \left|\mathbb{E}\left(\mathcal{Z}_{x,h}\right) - x\right| \leqslant h^2 A(x)$$

En d'autres termes, on dira que le "biais" de $K_{x,h}$ est de l'ordre de h^2. Cette condition va permettre de contrôler le biais de l'estimateur issu du noyau associé $K_{x,h}$.

Condition H2(B) Pour tout noyau associé $K_{x,h}$ appartenant à la famille $\mathscr{K}$, il existe une fonction $B : \mathbb{T} \to \mathbb{R}_+^*$ vérifiant

$$\forall (x, h) \in \mathbb{T} \times (0, 1], \quad \mathrm{Var}\left(\mathcal{Z}_{x,h}\right) \leqslant h^2 B(x),$$

Autrement dit, la "variance" de $K_{x,h}$ est de l'ordre de h^2. Cette condition interviendra dans le contrôle de la variance de l'estimateur à noyau $K_{x,h}$.

Condition H3(C) Enfin, pour tout noyau associé $K_{x,h}$ appartenant à la famille $\mathscr{K}$, il existe une fonction C définie par $C : \mathbb{T} \to \mathbb{R}_+^*$ telle que :

$$\forall (x, h) \in \mathbb{T} \times (0, 1], \quad \|K_{x,h}\|_2^2 \leqslant h^{-1} C(x)$$

Cette condition garantit que le noyau associé $K_{x,h}$ appartienne à $\mathbb{L}^2(\mathbb{T})$. De plus, sa norme $\mathbb{L}^2$ n'explose pas plus vite que h^{-1}.

4.2 Résultats

Dans cette section, on suppose que $s \in]0,1]$ et $L > 0$ sont deux nombres réels connus par le statisticien. De plus, on suppose que la densité inconnue f appartient à $\mathbb{F}(s,L)$.

4.2.1 Contrôle du risque

Nous procédons ici à une étude plus large du risque de $\widehat{f_n}$ pour des régularités comprises entre 0 et 1 en utilisant la norme $\mathbb{L}^1$. Soit $f \in \mathbb{F}(s,L)$ et $h \in (0,1]$ le risque de l'estimateur $\widehat{f_{n,h}}$ défini en (4.8) peut s'écrire sous la forme

$$\begin{aligned} R_n(\widehat{f_{n,h}}, f) &= \mathbb{E}_f^n \|\widehat{f_{n,h}} - f\|_1 \\ &= \mathbb{E}_f^n \left\{ \int_0^1 |\widehat{f_{n,h}}(u) - f(u)| du \right\} \qquad (4.11) \\ &\leqslant I_1(n,h) + I_2(n,h) \qquad (4.12) \end{aligned}$$

où

$$I_1(n,h) = \int_{\mathbb{T}} |\mathcal{B}_h(x)|\, dx \quad \text{avec} \quad \mathcal{B}_h(x) = \mathbb{E}_f^n \widehat{f_{n,h}}(x) - f(x)$$

et

$$I_2(n,h) = \int_{\mathbb{T}} \mathbb{E}_f^n |\mathcal{Z}_h(x)|\, dx \quad \text{avec} \quad \mathcal{Z}_h(x) = \widehat{f_{n,h}}(x) - \mathbb{E}_f^n \widehat{f_{n,h}}(x).$$

Le terme $|\mathcal{B}_h(x)|$ est appelé le biais ponctuel tandis que $I_1(n,h)$ est le biais intégré. De même, $\mathbb{E}_f^n|\mathcal{Z}_h(x)|$ est le terme stochastique ponctuel tandis que $I_2(n,h)$ est le terme stochastique intégré.

Considérons $L : \mathbb{T} \to [0,+\infty[$, $x \mapsto L(x)$. Notre objectif est de contrôler les termes précédents indépendamment de $f \in \mathbb{F}(s,L(x))$ en vue de contrôler le risque maximal sur cette classe de fonctions. Pour ce faire, nous allons décomposer ce travail en quatre propositions et un théorème. Dans toute la suite, $\mathscr{K} = \{K_{x,h}\}_{x\in\mathbb{T}, h\in(0,1]}$ est une famille donnée de noyaux associés continus et $F(s,L(x))$ est la classe $F(s,L)$ d'Hölder dont la constante de Lipschitz L dépend du point d'estimation x.

Avant de donner nos résultats nous allons énoncer le lemme suivant.

Lemme 4.2.1 *Pour tout réel* $0 < s < 1$, $a \geqslant 0$ *et* $x > 0$ *on a*

$$(a+x)^s \leqslant a^s + x^s. \qquad (4.13)$$

Démonstration. Pour $a = 0$ ce résultat est évident. Supposons $a > 0$, pour $0 < s < 1$, considérons l'inégalité

$$\left(1 + \frac{x}{a}\right)^s \leqslant 1 + \left(\frac{x}{a}\right)^s.$$

Posons W la fonction définie par $W(u) = (1+u)^s - u^s - 1$. On vérifie facilement que W est décroissante et admet pour limite 0 en 0. Ce qui montre que $W < 0$ d'où le résultat.■

La proposition suivante montre que le biais ponctuel de $\widehat{f_{n,h}}$ est majoré.

Proposition 4.2.2 (Majoration du biais ponctuel) *Considérons $f \in \mathbb{F}\{s, L(x)\}$, la densité à estimer et $K_{x,h}$ un noyau associé appartenant à $\mathscr{K}$ tel que les hypothèses H1(A) et H2(B) soient vérifiées. Pour $0 < s \leq 1$, on a*

$$|\mathcal{B}_h(x)| \leq h^s L(x) \left\{ B^{\frac{s}{2}}(x) + h^s A^s(x) \right\}. \tag{4.14}$$

Démonstration. Tout d'abord, remarquons que

$$\begin{aligned} |\mathcal{B}_h(x)| &= \left| \mathbb{E}\left\{ f(\mathcal{Z}_{x,h}) \right\} - f(x) \right| \\ &\leqslant \mathbb{E} \left| \left\{ f(\mathcal{Z}_{x,h}) \right\} - f(x) \right| \\ &\leqslant L(x) \mathbb{E} \left(\left| \mathcal{Z}_{x,h} - x \right|^s \right). \end{aligned}$$

En utilisant l'inégalité de Jensen puis en intercalant $\mathbb{E}\left(\mathcal{Z}_{x,h}\right)$ dans le second membre de l'inégalité précédente on a :

$$\begin{aligned} |\mathcal{B}_h(x)| &\leqslant L(x) \left[\mathbb{E} \left(\left| \mathcal{Z}_{x,h} - \mathbb{E}\left(\mathcal{Z}_{x,h}\right) \right| + \left| \mathbb{E}\left(\mathcal{Z}_{x,h}\right) - x \right| \right) \right]^s \\ &\leqslant L(x) \left[\sqrt{\mathbb{E} \left| \mathcal{Z}_{x,h} - \mathbb{E}\left(\mathcal{Z}_{x,h}\right) \right|^2} + \left| \mathbb{E}\left(\mathcal{Z}_{x,h}\right) - x \right| \right]^s. \end{aligned}$$

L'utilisation de l'inégalité de Markov et le résultat (4.13) du lemme précédent ainsi que les hypothèses H1(A) et H2(B) implique que pour tout $0 < s \leq 1$:

$$\begin{aligned} |\mathcal{B}_h(x)| &\leqslant \left[h \left\{ B^{\frac{1}{2}}(x) + hA(x) \right\} \right]^s \\ &\leqslant h^s \cdot \left\{ B^{\frac{s}{2}}(x) + h^s A^s(x) \right\}. \blacksquare \end{aligned}$$

La proposition suivante montre que le biais intégré est aussi majoré.

Proposition 4.2.3 (Majoration du biais intégré) *On considère $f \in \mathbb{F}\{s, L(x)\}$, la densité à estimer et $K_{x,h}$ un noyau associé appartenant à $\mathscr{K}$ tel que les hypothèses H1(A) et H2(B) soient vérifiées. Supposons de plus que :*

$$B^*(s) = \int_{\mathbb{T}} L(x) B^{\frac{s}{2}}(x) dx < +\infty \quad \textit{et} \quad \int_{\mathbb{T}} L(x) A^s(x) dx < +\infty. \tag{4.15}$$

Pour $0 < s \leq 1$, on a

$$I_1(n,h) \leqslant h^s B^*(s) \cdot \{1 + o(1)\}.$$

Démonstration. D'après le résultat de la Proposition 4.2.2, on a

$$|\mathcal{B}_h(x)| \leqslant h^s L(x) \left\{ B^{\frac{s}{2}}(x) + h^s A^s(x) \right\}.$$

En intégrant membre à membre l'inégalité précédente on obtient

$$\begin{aligned} I_1(n,h) &\leqslant \int_{\mathbb{T}} \left[h^s L(x) \left\{ B^{\frac{s}{2}}(x) + h^s A^s(x) \right\} \right] dx \\ &\leqslant h^s \int_{\mathbb{T}} L(x) B^{\frac{s}{2}}(x) \left\{ 1 + \frac{h^{2s} A^{2s}(x)}{B^s(x)} \right\}^{1/2} dx. \end{aligned}$$

Par application du théorème de convergence dominée de Lebesgue et en utilisant les hypothèses (4.15), on a le résultat.■

Le résultat suivant concerne le terme stochastique ponctuel. On y montre que ce dernier est majaoré.

Proposition 4.2.4 (Majoration du terme stochastique ponctuel) *Soit $f \in \mathbb{F}\{s, L(x)\}$, la densité à estimer et $K_{x,h}$ un noyau associé appartenant à $\mathscr{K}$ tel que l'hypothèse H3(C) soit vérifiée. De plus notons*

$$\widetilde{K}_{x,h}(\cdot) = \frac{K^2_{x,h}(\cdot)}{\|K_{x,h}\|_2^2}.$$

Considérons la famille $\mathcal{L} = \{\widetilde{K}_{x,h}\}_{x\in\mathbb{T},h\in(0,1]}$. Cette famille contient des densités appartenant de $\mathscr{K}$ et satisfait les hypothèses H1($\widetilde{A}$) et H2($\widetilde{B}$). On a alors :

$$\mathbb{E}^n_f|\mathcal{Z}_h(x)| \leq \left[\frac{C(x)}{nh}\left\{f(x) + h^s\mu_{s,h}(x)\right\}\right]^{1/2}, \tag{4.16}$$

où

$$\mu_{h,s}(x) = L(x)\left\{\widetilde{B}^{\frac{s}{2}}(x) + h^s\widetilde{A}^s(x)\right\}. \tag{4.17}$$

Démonstration. Pour commencer, introduisons la notation suivante pour tout $k \in \{1, \ldots, n\}$:

$$\eta_k(x) = K_{x,h}(X_k) - \mathbb{E}^n_f K_{x,h}(X_k).$$

En utilisant cette notation, on peut écrire le terme $\mathbb{E}^n_f|\mathcal{Z}_h(x)|$ sous la forme

$$\begin{aligned}\mathbb{E}^n_f|\mathcal{Z}_h(x)| &= \mathbb{E}^n_f\left|\frac{1}{n}\sum_{k=1}^n \eta_k(x)\right| \\ &\leq \sqrt{\frac{\mathbb{E}^n_f|\eta_1(x)|^2}{n}}.\end{aligned}$$

La dernière inégalité est obtenue par l'inégalité de Cauchy et grâce au fait que les variables aléatoires $\eta_k(x)$ sont iid. Étudions maintenant le terme $\mathbb{E}^n_f|\eta_1(x)|^2$. On a

$$\begin{aligned}\mathbb{E}^n_f|\eta_1(x)|^2 &\leqslant \mathbb{E}^n_f K^2_{x,h}(X_1) \\ &\leqslant \int_{\mathbb{T}\cap\mathbb{S}_{x,h}} K^2_{x,h}(u)f(u)du \\ &= \|K_{x,h}\|_2^2 \int_{\mathbb{T}\cap\mathbb{S}_{x,h}} L_{x,h}(u)f(u)du \\ &= \|K_{x,h}\|_2^2\mathbb{E}f(\zeta_{x,h}),\end{aligned}$$

où $\zeta_{x,h}$ est une variable aléatoire de loi $L_{x,h}$. En utilisant l'hypothèse H3(C) on obtient :

$$\mathbb{E}^n_f|\eta_1(x)|^2 \leqslant h^-1C(x)\left\{f(x) + \left|\mathbb{E}\left\{f(\zeta_{x,h})\right\} - f(x)\right|\right\}.$$

Comme $\mathcal{L} = \{L_{x,h}\}_{x\in\mathbb{T},h\in(0,1]}$ satisfait H1(A) et H2(B), on applique la Proposition 4.2.2 au terme $\left|\mathbb{E}\{f(\zeta_{x,h})\} - f(x)\right|$ qui est inférieur à $Lh^s\mu_{s,h}(x)$, et le résultat s'en déduit.■

La proposition suivante va montrer la majoration du deuxième terme du second membre de l'inégalité (4.8).

Proposition 4.2.5 (Majoration du terme stochastique intégré) *Considérons les hypothèses de la Proposition 4.2.4 ensuite supposons que similairement à la Proposition 4.2.3, on a*

$$\int_{\mathbb{T}} \widetilde{A}(x)dx < +\infty, \quad \int_{\mathbb{T}} \widetilde{A}^s(x)dx < +\infty \tag{4.18}$$

et

$$(2^{s-1} \vee 1) \int_{\mathbb{T}} \widetilde{B}^{\frac{s}{2}}(x)dx < +\infty. \tag{4.19}$$

De plus, en supposant que

$$C^*\{s, L(x)\} = \left\{ \sup_{f\in\mathbb{F}(s,L)} \int_{\mathbb{T}} C(x)f(x)dx \right\}^{1/2} < +\infty$$

et

$$\int_{\mathbb{T}} C(x)\mu_{s,h}(x)dx < +\infty. \tag{4.20}$$

on a :

$$I_2(n,h) \leqslant \frac{C^*\{s, L(x)\}}{\sqrt{nh}} \cdot (1 + o(1)).$$

Démonstration. Partant du résultat de la proposition précédente on a :

$$\mathbb{E}_f^n|\mathcal{Z}_h(x)| \leqslant \left[\frac{C(x)}{nh} \{f(x) + h^s\mu_{s,h}(x)\} \right]^{1/2}$$

En intégrant membre à membre on a

$$\begin{aligned}\int_{\mathbb{T}} \mathbb{E}_f^n|\mathcal{Z}_h(x)|dx &\leqslant \frac{1}{\sqrt{nh}} \int_T \left[C(x)\{f(x) + L(x)h^s\mu_{s,h}(x)\}\right]^{1/2} dx \\ &\leqslant \frac{1}{\sqrt{nh}} \left[\int_{\mathbb{T}} C(x)\{f(x) + L(x)h^s\mu_{s,h}(x)\} dx \right]^{1/2}\end{aligned}$$

En prenant le $\sup_{f\in\mathbb{F}(s,L)}$ puis par application du théorème de convergence dominée de Lebesgue en tenant compte des hypothèses (4.18), (4.19) et (4.20), on a le résultat.■

En utilisant toutes les définitions et notations introduites précédemment, on est en mesure d'énoncer maintenant le résultat minimax. Nous rappelons au lecteur que ceci montre implicitement qu'il existe un choix optimal de paramètre h permettant à l'estimateur à noyau associé correspondant d'atteindre la vitesse minimax de convergence de $\mathbb{F}\{s, L(x)\}$. Mais seulement ce choix de paramètre h dépend de $L(x)$ et s.

Théorème 4.2.6 *Soit $\mathscr{K} = \{K_{x,h}\}_{x\in\mathbb{T},h\in(0,1]}$ une famille de noyaux associés continus. Supposons que les hypothèses des Propositions 4.2.2– 4.2.5 sont vérifiées. Alors il existe un paramètre optimal de lissage $h_n\{s, L(x)\}$ tel que l'estimateur à noyau associé (issu de $\mathscr{K}$) construit avec $h_n\{s, L(x)\}$ satisfait :*

$$R_n\left(\widehat{f}_{n,h_n(s,L(x))}, \mathbb{F}\{s, L(x)\}\right) \leqslant D\{s, L(x)\} L^{1/(2s+1)}(x) n^{-s/(2s+1)} \cdot \{1 + o(1)\},$$

où $D\{s, L(x)\}$ est une constante définie dans la démonstration. De plus l'expression de $h_n\{s, L(x)\}$ est de la forme

$$h_n\{s, L(x)\} = \left[\frac{\varkappa\{s, L(x)\}}{nL^2(x)}\right]^{1/(2s+1)},$$

où $\varkappa\{s, L(x)\}$ est explicitement donnée dans la démonstration.

Démonstration. En prenant ensemble les résultats des Propositions 4.2.3 et 4.2.5 on obtient :

$$R_n\left(\widehat{f}_{n,h_n(s,L(x))}, \mathbb{F}\{s, L(x)\}\right) \leq \left[h^s B^*(s) + \frac{C^*\{s, L(x)\}}{\sqrt{nh}}\right]\{1 + o(1)\}.$$

La partie droite de cette inégalité peut être optimisée en choisissant h :

$$h = h_n\{s, L(x)\} = \left[\frac{\varkappa\{s, L(x)\}}{nL^2(x)}\right]^{1/(2s+1)}$$

avec

$$\varkappa\{s, L(x)\} = \left[\frac{C^*\{s, L(x)\}}{2sB^*(s)}\right]^2.$$

Enfin, on peut facilement prouver le résultat escompté avec le h choisi et en prenant

$$D\{s, L(x)\} = B^*(s)\left[\varkappa\{s, L(x)\}\right]^{s/(2s+1)} + C^*\{s, L(x)\}\left[\varkappa\{s, L(x)\}\right]^{-1/2(2s+1)}.$$

Ceci achève la démonstration.■

4.2.2 Illustrations

Dans ce paragraphe, nous examinons les propriétés minimax des estimateurs à noyaux classiques d'une manière générale et les estimateurs à noyaux non-classiques précisément dans le cas de bêta, pour des densités appartenant à la classe $\mathbb{F}\{s, L(x)\}$ avec $0 < s \leqslant 1$. Dans chacun des cas, nous identifions à partir des caractéristiques du noyau associé correspondant, les fonctions $A(x)$, $B(x)$ et $C(x)$ données dans les hypothèses techniques de la Section 4.1.2. Puis, nous procédons aux calculs des différents ingrédients pouvant permettre le contrôle du risque minimax.

Estimateurs à noyau classique

Nous présentons dans ce paragraphe les résultats minimax des estimateurs à noyaux associés classiques (en utilisant la forme générale du noyau associé classique). Rappelons que les noyaux associés classiques sont définis par

$$K_{x,h}(\cdot) = \frac{1}{h}\mathcal{K}\left(\frac{x-\cdot}{h}\right). \tag{4.21}$$

où $\mathcal{K}(\cdot)$ est un noyau classique vérifiant la Définition 1.1.1. Il a été démontré dans la Proposition 2.1.5 de la Section 2.1 que

$$\left|\mathbb{E}\left(\mathcal{Z}_{x,h}\right) - x\right| = 0 \text{ et } Var(\mathcal{Z}_{x,h}) = h^2\sigma^2_{\mathcal{K}}. \tag{4.22}$$

Ce qui permet de déduire les quantités $A(x)$ et $B(x)$ de l'hypothèse technique comme suit :

$$A(x) = 1 \text{ et } B(x) = \sigma^2_{\mathcal{K}}; \text{ ce qui implique que } A^s(x) = 1 \text{ et } B^{\frac{s}{2}}(x) = \sigma^s_{\mathcal{K}}.$$

On remarque tout de suite qu'elles sont des constantes indépendantes de x. En remarquant que $L(x)$, $A^s(x)$ et $B^{\frac{s}{2}}(x)$ sont des quantités finies, on a

$$B^*(s) = \int_{\mathbb{R}} L(x)B^{\frac{s}{2}}(x)dx = \sigma^s_{\mathcal{K}}\int_{\mathbb{R}} L(x)dx < +\infty$$

et

$$\int_{\mathbb{R}} L(x)A^s(x)dx = \int_{\mathbb{R}} L(x)dx < +\infty.$$

Par application de la Proposition 4.2.3, on obtient une majoration du biais intégré issu de la décomposition du risque minimax. On peut explicitement l'écrire sous la forme :

$$I_1(n,h) = \int_{[0,1]} \left|\mathbb{E}^n_f \widehat{f}_{n,h}(x) - f(x)\right| dx \leqslant \left\{h^s\sigma^s_{\mathcal{K}}\int_{\mathbb{R}} L(x)dx\right\}\{1 + o(1)\}. \tag{4.23}$$

Pour ce qui est de la partie stochastique de la décomposition du risque minimax, considérons le noyau associé

$$\widetilde{K}_{x,h}(\cdot) = \frac{K^2_{x,h}(\cdot)}{\|K_{x,h}\|^2_2} = \frac{1}{\|\mathcal{K}\|^2_2}\mathcal{K}^2\left(\frac{x-\cdot}{h}\right).$$

Ce qui permet de constater que $\widetilde{K}_{x,h}$ est une densité de probabilité dans la même famille que $\mathcal{K}$ (par conséquent sa moyenne est nulle et sa variance est finie, voir Définition 1.1.1). Considérons $\widetilde{\mathcal{Z}_{x,h}}$ la variable aléatoire de loi $\widetilde{K}_{x,h}$, par changement de variables $t = (x-u)/h$ on a :

$$\begin{aligned}
\mathbb{E}\left(\widetilde{\mathcal{Z}_{x,h}}\right) &= \frac{1}{\|\mathcal{K}\|^2_2}\int_{\mathbb{R}} u\mathcal{K}^2\left(\frac{x-u}{h}\right)du \\
&= \frac{1}{\|\mathcal{K}\|^2_2}\int_{\mathbb{R}} (x-ht)\mathcal{K}^2(t)\,dt \\
&= x - h\mu_{\widetilde{K_{x,h}}}.
\end{aligned}$$

Signalons ici que nous avons grossièrement simplifié le h issu de la différentielle du avec celui q'on obtient dans le calcul de $\|\mathcal{K}\|_2^2$. Puisque $\widetilde{K_{x,h}}$ est de même famille que $\mathcal{K}$, on a

$$\left|\mathbb{E}\left(\widetilde{\mathcal{Z}_{x,h}}\right) - x\right| \leqslant h^2\widetilde{A}(x) \quad \text{avec} \quad \widetilde{A}(x) = 1. \tag{4.24}$$

Le calcul de la variance donne successivement,

$$\begin{aligned} Var\left(\mathcal{Z}_{x,h}\right) &= \mathbb{E}\left(\widetilde{\mathcal{Z}_{x,h}}^2\right) - \left\{\mathbb{E}\left(\widetilde{\mathcal{Z}_{x,h}}\right)\right\}^2 \\ &= \frac{1}{\|\mathcal{K}\|_2^2}\int_{\mathbb{R}}(x-ht)^2\mathcal{K}^2(t)\,dt - x^2 \\ &= h^2\sigma_{\widetilde{K_{x,h}}} - 2xh\mu_{\widetilde{K_{x,h}}} \\ &\leqslant h^2\sigma^2_{\widetilde{K_{x,h}}} \end{aligned} \tag{4.25}$$

d'où $\widetilde{B}(x) = \sigma^2_{\widetilde{K_{x,h}}}$. Ainsi, on obtient :

$$\mu_{h,s}(x) = L(x)\left\{\widetilde{B}^{\frac{s}{2}}(x) + h^s\widetilde{A}^s(x)\right\} = L(x)\left\{\sigma^s_{\widetilde{K_{x,h}}} + h^s\right\} < +\infty.$$

De plus, on sait d'après la condition H3(C) de la Section 4.1.2 que $\|K_{x,h}\|_2^2 \leqslant h^{-1}C(x)$. Dans le cas présent $C(x) = C$ est une constante indépendante de x. En considérant le fait que A^s et $B^{\frac{s}{2}}$ et $L(x)$ sont des quantités finies puis en remarquant que

$$C^*\{s, L(x)\} = \left\{\sup_{f\in\mathbb{F}(s,L)}\int_{\mathbb{R}} Cf(x)dx\right\}^{1/2} < +\infty$$

et

$$\int_{\mathbb{R}} C(x)\mu_{h,s}(x)dx = \int_{\mathbb{R}}\left\{L(x)\sigma^s_{\widetilde{K_{x,h}}} + h^s\right\}dx < +\infty,$$

une application de la Proposition 4.2.5, conduit à la majoration du terme stochastique intégré. C'est à dire

$$I_2(n,h) = \int_{[0,1]}\mathbb{E}^n_f\left|\widehat{f_{n,h}}(x) - \mathbb{E}^n_f\widehat{f_{n,h}}(x)\right|dx \leq \frac{C^*\{s, L(x)\}}{\sqrt{nh}}(1+o(1)). \tag{4.26}$$

En utilisant (4.23), (4.26) et le Théorème 4.2.6, on peut trouver un paramètre optimal de lissage

$$h_n\{s, L(x)\} = \left[\frac{\varkappa\{s, L(x)\}}{nL^2(x)}\right]^{1/(2s+1)},$$

avec

$$\varkappa\{s, L(x)\} = \left[\frac{C^*\{s, L(x)\}}{2sB^*(s)}\right]^2.$$

pour lequel les estimateurs à noyaux associés classiques des densités f, appartenant à la classe des fonctions $\mathbb{F}\{s, L(x)\}$ avec $0 < s \leqslant 1$, atteigne la vitesse minimax.

Estimateurs à noyau bêta

Tout d'abord, nous rappelons au lecteur que les propriétés minimax des estimateurs à noyau bêta ont été l'objet des travaux de Bertin & Klutchnikoff (2011). Ces auteurs ont étudié lesdites propriétés en utilisant la perte $\mathbb{L}^p$ et pour des régularités s appartenant à $]0,2]$. Nous allons dans ce travail utiliser la perte $\mathbb{L}^1$ et en particulier pour des régularités $0 < s \leqslant 1$.

Les caractéristiques du noyau associé bêta (construit par la méthode mode-dispersion) définies dans la Section 2.1.3, donnent

$$\left|\mathbb{E}\left(\mathcal{Z}_{x,h}\right) - x\right| = \frac{h(1-2x)}{1+2h} \leqslant \frac{h^2}{2}(1-x) \tag{4.27}$$

et

$$Var\left(\mathcal{Z}_{x,h}\right) = \frac{h\{x(1-x)+h+h^2}{(1+3h)(1+2h)^2} \leqslant \frac{h^2}{3}x(1-x). \tag{4.28}$$

Par conséquent on a :

$$A(x) = \frac{1}{2}(1-x) \quad et \quad B(x) = \frac{1}{3}x(1-x).$$

Puisque $0 < L(x) < +\infty$ et de plus les fonctions $B^{\frac{s}{2}}(x)$ et $A^s(x)$ sont bornées sur $[0,1]$ alors on a

$$B^*(s) = \int_{[0,1]} L(x)\left(\frac{1}{3}\right)^{\frac{s}{2}} \{x(1-x)\}^{\frac{s}{2}}\, dx < +\infty \quad \text{et} \quad \int_{[0,1]} L(x)\left\{\frac{1}{2}(1-x)\right\}^s dx < +\infty.$$

Par application de la Proposition 4.2.3, on obtient une majoration du biais intégré obtenu à partir de la décomposition du risque minimax. C'est à dire

$$I_1(n,h) = \int_{[0,1]} \left|\mathbb{E}_f^n \widehat{f_{n,h}}(x) - f(x)\right| dx \leqslant h^s B^*(s)\left\{1+o(1)\right\}. \tag{4.29}$$

Pour la majoration de la partie stochastique intégrée issue de la décomposition du risque minimax, considérons le noyau associé $\widetilde{BE}_{x,h}$ défini par $\widetilde{BE}_{x,h}(\cdot) = BE^2_{x,h}(\cdot)/\|BE_{x,h}\|_2^2$. En calculant $BE^2_{x,h}(\cdot)$ on obtient

$$BE^2_{x,h}(u) = \frac{\|BE_{x,h}\|_2^2}{\mathcal{B}\left(\frac{2x}{h}+1, \frac{2(1-x)}{h}\right)} u^{\frac{2x}{h}}(1-u)^{\frac{2(1-x)}{h}}$$

où $\mathcal{B}(\cdot,\cdot)$ est la fonction bêta d'Euler. Il s'ensuit que

$$\widetilde{BE}_{x,h}(u) = \frac{1}{\mathcal{B}\left(\frac{2x}{h}+1, \frac{2(1-x)}{h}\right)} u^{\frac{2x}{h}}(1-u)^{\frac{2(1-x)}{h}} = BE_{x,\frac{h}{2}}(u).$$

On en déduit que

$$\widetilde{A}(x,h) = A\left(x, \frac{h}{2}\right) \leqslant \frac{1}{4}h^2 A(x) \quad \text{et} \quad \widetilde{B}(x,h) = B\left(x, \frac{h}{2}\right) \leqslant \frac{1}{4}h^2 B(x),$$

par conséquent

$$\widetilde{A}(x) = \frac{1}{4}A(x) \quad \text{et} \quad \widetilde{B}(x) = \frac{1}{4}B(x)$$

et donc

$$\begin{aligned}\mu_{h,s}(x) &= \left(\frac{1}{4}\right)^{\frac{s}{2}} \left\{B^{\frac{s}{2}}(x) + \left(\frac{1}{4}\right)^{\frac{s}{2}} h^s A^s(x)\right\} L(x) \\ &= \left\{\frac{1}{4}(1-x)\right\}^{\frac{s}{2}} \left\{\left(\frac{x}{3}\right)^{\frac{s}{2}} + \left(\frac{1}{4}\right)^{s} h^s (1-x)^{\frac{s}{2}}\right\} L(x). \end{aligned} \tag{4.30}$$

De plus d'après la condition H3(C) de la Section 4.1.2, on a $\|BE_{x,h}\|_2^2 \leqslant h^{-1}C(x)$ avec $C(x) = \left\{2\sqrt{\pi x(1-x)}\right\}^{-1}$. Puisque les fonctions A^s et $B^{\frac{s}{2}}$ sont bornées sur $[0,1]$ et $0 < L(x) < +\infty$ en remarquant que les quantités

$$C^*\{s, L(x)\} = \left\{\sup_{f\in\mathbb{F}(s,L)} \int_{[0,1]} \frac{1}{2\sqrt{\pi x(1+x)}} f(x)dx\right\}^{1/2}$$

et

$$\left(\frac{1}{4}\right)^{\frac{s+1}{2}} \frac{1}{\sqrt{\pi}} \int_{[0,1]} \frac{1}{2\sqrt{x}} (1-x)^{\frac{s-1}{2}} \left\{\left(\frac{x}{3}\right)^{\frac{s}{2}} + \left(\frac{1}{4}\right)^{s} (1-x)^{\frac{s}{2}}\right\} L(x)f(x)dx$$

sont finies alors on obtient la majoration du terme stochastique intégré par application de la Proposition 4.2.5. C'est-à-dire

$$I_2(n,h) = \int_{[0,1]} \mathbb{E}_f^n \left|\widehat{f_{n,h}}(x) - \mathbb{E}_f^n \widehat{f_{n,h}}(x)\right| dx \leq \frac{C^*\{s, L(x)\}}{\sqrt{nh}}(1+o(1)). \tag{4.31}$$

De (4.29), (4.31) et d'après le Théorème 4.2.6, il existe un paramètre optimal de lissage

$$h_n\{s, L(x)\} = \left[\frac{\varkappa\{s, L(x)\}}{nL^2(x)}\right]^{1/(2s+1)},$$

avec

$$\varkappa\{s, L(x)\} = \left[\frac{C^*\{s, L(x)\}}{2sB^*(s)}\right]^2.$$

pour lequel les estimateurs à noyaux bêta des densités f, appartenant à la classe des fonctions $\mathbb{F}\{s, L(x)\}$ avec $0 < s \leqslant 1$, atteigne la vitesse minimax.

Chapitre 5

Estimateurs à noyaux associés mixtes et applications

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 :

$$f(x) = \sum_{j=1}^{p} \beta_j f_j(x) \mathbf{1}_{\mathbb{T}_j}(x), \tag{5.1}$$

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 (supposées connues dans ce travail pour simplifier) sur chaque composante $\mathbb{T}_j$ partitions de $\mathbb{T}$ (La Figure 5.2 présente une représentation graphique de f pour $p = 2$).

FIGURE 5.1 – Un exemple de $\mathbb{T}$.

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 :

$$\mathbb{T} = \left(\cup_j \mathbb{T}_{I_j}\right) \bigcup \left(\cup_{j'} \mathbb{T}_{N_{j'}}\right), \tag{5.2}$$

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 $\mathbb{T}_N$

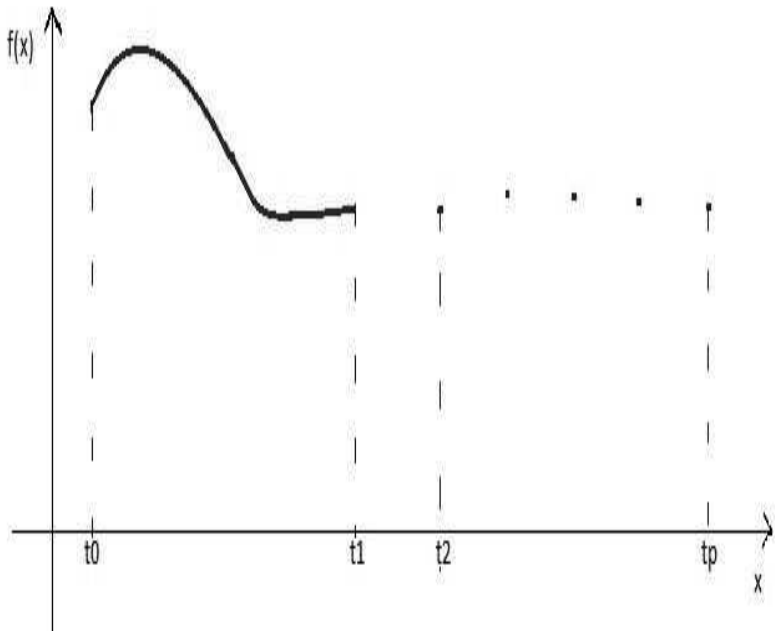

FIGURE 5.2 – Exemple d'une densité mixte.

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 écrivons $\mathbb{T}$ de la manière suivante :

$$\mathbb{T} = T_I \cup \mathbb{T}_N = [t_0, t_1] \cup \{t_2, t_3, ...\}. \tag{5.3}$$

Sur les échelles de temps, plusieurs outils d'analyse ont été proposés pour travailler à la fois sur les intervalles et les ensembles discrets.

Ce chapitre se propose d'en faire usage dans le but d'estimer par les méthodes non-paramétriques les fonctions de type (5.1) sur le support $\mathbb{T}$ défini en (5.3) tout en respectant la structure topologique de ce dernier. La fonction définie en (5.1) 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) et Jørgensen et Kokonendji (2011, 2013). Enfin, l'une dernière des raisons est leur flexibilité dans l'utilisation de l'analyse unifiant le discret et le continu que nous détaillons dans la Section 5.1.

Rappelons au lecteur que lorsque $\mathbb{T}$ est homogène (i.e. restreint à T_I ou à $\mathbb{T}_N$ que nous notons simplement T_j), le noyau associé (continu ou discret) est une f.d.m.p. $K_{x,h}^{[j]}$ paramétrée par le point d'estimation x et le paramètre de lissage h, sur le support $\mathbb{S}_{x,h}^{[j]}$

et qui vérifie

$$x \in \mathbb{S}_{x,h}^{[j]}, \tag{5.4}$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + A_j(x,h), \tag{5.5}$$

$$Var(\mathcal{Z}_{x,h}) = B_j(x,h), \tag{5.6}$$

où $A_j(x,h)$ et $B_j(x,h)$ tendent vers 0 lorsque h tend 0 et $\mathcal{Z}_{x,h}$ est une variable aléatoire de loi $K_{x,h}^{[j]}$.

Pour une suite $X_1, X_2, \cdots, X_n$ de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de densité inconnue f_j sur le support $\mathbb{T}_j$, l'estimateur $\widehat{f_{n_j}}$ de $f_j \in \mathscr{C}^2(\mathbb{T}_j)$ à noyau associé $K_{x,h}^{[j]}$ est de la forme :

$$\widehat{f_{n_j}}(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbf{1}_{\mathbb{T}_j}(x) \mathbf{1}_{\mathbb{T}_j}(X_j), \tag{5.7}$$

où n_j est le nombre d'observations tambant dans $\mathbb{T}_j$. Le lecteur peut consulter les articles de Chen (1999, 2000), Kokonendji & Zocchi (2010), Kokonendji & Senga Kiessé (2011) puis Kokonendji & Libengué (2013) pour une présentation détaillée avec des multiples exemples. Le biais de cet estimateur est défini par

$$Biais\left\{\widehat{f_{n_j}}(x)\right\} = \left[A_j(x,h) f_j^{(1)}(x) + \frac{1}{2}\left\{A_j^2(x,h) + B_j(x,h)\right\} f_j^{(2)}(x)\right] \mathbf{1}_{\mathbb{T}_j}(x) + o(h^2), \tag{5.8}$$

où f_j^k avec $k \in \{1,2\}$ sont les dérivées du premier/second ordre en x (lorsque $\mathbb{T}_j$ est un intervalle) ou bien les différences finies du premier/second ordre en x (lorsque $\mathbb{T}_j$ est un ensemble discret). La variance de $\widehat{f_{n_j}}$ est

$$Var\left\{\widehat{f_{n_j}}(x)\right\} = \frac{1}{n_j} f_j(x) \left\|K_{x,h}^{[j]}\right\|_2^2 \mathbf{1}_{\mathbb{T}_j}(x) + o\left(\frac{1}{nh^{r_2}}\right) \tag{5.9}$$

avec

$$\left\|K_{x,h}^{[j]}\right\|_2^2 = \begin{cases} \displaystyle\int_{\mathbb{S}_{x_j,h} \cap \mathbb{T}_j} K_{x,h}^{[j]2}(u)du & \text{si} \quad \mathbb{T}_j = \mathbb{T}_I \\ \sum_{y \in \mathbb{S}_{x_j,h} \cap \mathbb{T}_j} K_{x,h}^{[j]2}(y) & \text{si} \quad \mathbb{T}_j = \mathbb{T}_N \end{cases}$$

et $r_2 = r_2(K_{x,h}^{[j]})$ est le plus grand réel dépendant du noyau associé $K_{x,h}^{[j]}$ tel que $nh^{r_2} \to +\infty$ lorsque $n \to +\infty$.

Nous organisons ce chapitre de la manière suivante. Nous rappelons dans la première section les outils d'analyses sur les échelles de temps, essentiellement ceux qui sont nécessaires pour la suite de ce travail. Puis, nous présentons dans la deuxième section les noyaux associés mixtes ainsi que leurs estimateurs. Nous donnons les propriétés de ces estimateurs en portant une attention particulière à l'étude des biais et variance ainsi que les erreurs quadratiques moyennes ponctuelles et intégrées. Enfin, nous terminons ce chapitre par des applications.

5.1 Calculs sur des échelles de temps

A présent, nous allons présenter les outils d'analyse sur les échelles de temps. Le lecteur peut se rapporter à Hilger (1988, 1990), Bohner & Peterson (2001, 2003) puis Sanyal (2008) pour plus de détails.

5.1.1 Définitions et notations

Nous donnons ici la définition des opérateurs de saut proposé par Hilger (1988) 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 5.1.1 *[Hilger, 1988] Soit $\mathbb{T}$ une échelle de temps.*

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

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

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

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

Ce deux opérateurs nous permettent de procéder à la classification des points de $\mathbb{T}$ comme dans le Tableau 5.1. Nous signalons au passage que cette nomenclature est spécifique à ce travail. Le lecteur peut toutefois consulter l'article de Agarwal et Bohner (1999) ainsi que les livres de Bohner & Peterson (2001, 2003) pour avoir une idée claire de cette classification.

Table 5.1 – Classification des points de $\mathbb{T}$

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)	$\sigma(t) > t$

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 une 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 5.1.2 *On considère $x \in \mathbb{T}_N$ et $k \in \mathbb{N}$.*

1– *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 $\mathbb{T}_N$.*

2– *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 attirons ici l'attention du lecteur sur cette notion qui sera reprise dans le paragraphe concernant la dérivabilité. Signalons également au passage que 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 $f : \mathbb{T}_j \to \mathbb{R}$, est dite continue sur $\mathbb{T}_j$ lorsqu'on a :

$$\forall \varepsilon > 0,\ \exists \mathcal{V}_{x_j} \ :\ s \in \mathcal{V}_{x_j} \cap \mathbb{T}_j \Rightarrow |f(s) - f(x_j)| < \varepsilon$$

où $\mathcal{V}_{x_j}$ est un voisinage de x_j selon la topologie induite sur $\mathbb{T}_j$. Lorsque $\mathbb{T}_j = \mathbb{T}_I$ (un intervalle), cette notion se confond avec la continuité usuelle sur $\mathbb{R}$ et lorsque $\mathbb{T}_j = \mathbb{T}_N$ (un ensemble discret) ceci coïncide avec la continuité des fonctions discrètes. Compte tenu de la structure particulière des échelles de temps, la notion de dérivabilité est la plus difficile à définir et fait appel aux différents opérateurs de la Définition 5.1.1. Nous résumons dans le prochain paragraphe les différentes techniques de dérivabilité.

5.1.2 Dérivabilité unifiée

La définition suivante présente l'outil d'unification des notions de dérivées à gauche ou à droite (respectivement notions de différences finies décentrées à gauche et à droite) en un point $x \in \mathbb{T}$.

Définition 5.1.3 *Soit $f : \mathbb{T} \to \mathbb{R}$, $\mathbb{T}^\kappa$ et $\mathbb{T}_\kappa$ des sous ensembles de $\mathbb{T}$, respectivement semi ouvert à droite et à gauche.*

*1-) On dit que f est **Δ-dérivable** au point x si : $\forall \varepsilon > 0,\ \exists \eta > 0\ :\ \forall s \in \mathcal{V}_\eta(x) =]x - \eta, x + \eta[\cap \mathbb{T} \Rightarrow$*

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

*2-) De même, on dit que f est **∇-dérivable** en x si : $\forall \varepsilon > 0,\ \exists \eta' > 0\ :\ \forall s \in \mathcal{V}_{\eta'}(x) =]x - \eta', x + \eta'[\cap \mathbb{T} \Rightarrow$*

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

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

Cette définition s'écrit aussi de la manière suivante :

$$f^\Delta(x) = \frac{f\{\sigma(x)\} - f(s)}{\sigma(x) - s} = \lim_{\substack{x \to s \\ x > s}} \frac{f(x) - f(s)}{x - s}, \tag{5.10}$$

et

$$f^{\nabla}(x) = \frac{f\{\rho(x)\} - f(s)}{\rho(x) - s} = \lim_{\substack{x \to s \\ x < s}} \frac{f(x) - f(s)}{x - s}. \tag{5.11}$$

Lorsque $\mathbb{T} = \mathbb{T}_I$ alors (5.10) et (5.11) représentent respectivement les notions usuelles de dérivabilité à droite et à gauche de x. Cependant, lorsque $\mathbb{T} = \mathbb{T}_N$ alors (5.10) et (5.11) 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 5.1.4 *Soient $f : \mathbb{T} \to \mathbb{R}$ et $x \in \mathbb{T}_\kappa^\kappa$ (où $\mathbb{T}_\kappa^\kappa$ 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 :*

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

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 :

$$f(x) = \sum_{j}^{k} \frac{(x - s)^j}{j!} f^{(j)}(s) + o\{(x - s)^k\}, \tag{5.12}$$

avec

$$f^{(j)}(x) = \frac{1}{\sigma(x) - \rho(x)} \left[\{\sigma(x) - x\} f^{\Delta^j}(x) + \{x - \rho(x)\} f^{\nabla^j}(x)\right], \tag{5.13}$$

où $f^{\Delta^j}(x)$ et $f^{\nabla^j}(x)$ sont respectivement **Δ-dérivée** et **∇-dérivée** de f d'ordre j. Dans (5.13), la quantité $f^{(j)}$ unifie les notions classiques de dérivée j^{eme} sur $\mathbb{T}_I$ et de différences finies d'ordre j sur $\mathbb{T}_N$. Précisons que pour les points discrets d'ordres k_1 à droite et k_2 à gauche, (5.13) s'écrit de la forme :

$$f^{(k_1,k_2)}(x) = \frac{1}{\sigma(x) - \rho(x)} \left[\{\sigma(x) - x\} f^{\Delta^{k_1}}(x) + \{x - \rho(x)\} f^{\nabla^{k_2}}(x)\right], \quad \forall k_1,\ k_2. \tag{5.14}$$

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

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x+1) + 2f(x) - 2f(x-1) + f(x-2)\}/2 & \text{si } x \in \mathbb{N} \setminus \{0,1\} \\ \{f(3) - 2f(2) + 2f(1) - f(0)\}/2 & \text{si } x = 1 \\ \{f(2) - 2f(1) + f(0)\}/2 & \text{si } x = 0. \end{cases} \tag{5.15}$$

Ce qui est nettement différent du résultat de Kokonendji & Senga Kiessé (2011). Pour ces auteurs, la quantité (5.15) est exprimée sous la forme :

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\}/4 & \text{si } x \in \mathbb{N} \setminus \{0,1\} \\ \{f(3) - 3f(1) + 2f(0)\}/2 & \text{si } x = 1 \\ \{f(2) - 2f(1) + f(0)\}/2 & \text{si } x = 0. \end{cases} \tag{5.16}$$

On remarque par là que les différences finies centrées à deux pas sont appliquées aux points x, discrets d'ordre supérieur ou égal à deux. Mais dans le cas des points 0 et 1, les différences finies décentrées à un pas vers la droite sont appliquées puis exceptionnellement les décentrées à un pas vers la gauche sont appliquées au point 1.

Il faut noter que le calcul de la dérivée d'ordre k d'une fonction f en un point x de $\mathbb{T}_N$ nécessite une connaissance a priori de l'ordre de discrétisation de ce point. Par exemple dans le cas précédent, on remarque que le point 2 est discret d'ordre 2 à gauche et d'ordre infini à droite, cependant le point 1 est discret d'ordre 1 à gauche d'ordre infini à droite tandis que le point 0 est seulement discret à droite.

Définition 5.1.5 *Une fonction f est dite de classe $\mathscr{C}^k(\mathbb{T})$ si et seulement si $f^{(k)}$ existe et est continue.*

Enfin, 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 5.1.6 *Une fonction $F : \mathbb{T} \to \mathbb{R}$ est appelée Δ-primitive de $f : \mathbb{T} \to \mathbb{R}$ 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 :

$$F^{\Delta}(x) = F^{\nabla}\{\sigma(x)\} \tag{5.17}$$
$$F^{\nabla}(x) = F^{\Delta}\{\rho(x)\}. \tag{5.18}$$

Ainsi, on a le théorème suivant :

Théorème 5.1.7 (Hilger, 1988) *Soient a et b des points de $\mathbb{T}$ tels que $a < b$ et f une fonction $\triangle$-mesurable sur $\mathbb{T}$. L'intégrale de f sur $\mathbb{T}$ satisfait les propriétés suivantes :*

$$\int_a^b f(t) \triangle t = \begin{cases} \displaystyle\int_a^b f(t)dt & si \quad \mathbb{T} = \mathbb{T}_I \\ \displaystyle\sum_a^b f(t) & si \quad \mathbb{T} = \mathbb{T}_N. \end{cases} \tag{5.19}$$

Pour la démonstration, le lecteur peut se référer à Hilger (1988) ou Bohner & Peterson (2001, pages 26-34). On obtient les même propriétés en ce qui concerne les fonctions **∇-mesurables** en utilisant l'une des relations (5.17) et (5.18).

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

5.2 Définitions et propriétés

Cette section fournit les définitions d'une densité mixte, d'un type de noyau mixte ainsi que le noyau associé mixte proprement dit. Ensuite, elle présente les estimateurs à noyaux associés mixtes suivis de leurs premières propriétés fondamentales.

Définition 5.2.1 *Soit $\mathbb{T} = \cup_j \mathbb{T}_j$ un ensemble constitué d'intervalles et des parties discrètes. Une fonction f de support $\mathbb{T}$ est dite* ***densité mixte*** *si pour tout $x \in \mathbb{T}$ on a :*

$$f(x) = \sum_j \beta_j f_j(x) \mathbf{1}_{\mathbb{T}_j}(x)$$

où les β_j sont des réels positifs tels que $\sum_j \beta_j = 1$ et les f_j sont des fonctions de densité et/ou masse de probabilité.

Remarque 5.2.2 *Une densité mixte peut être considérée comme un "modèle de mélange mixte". A travers le mot mixte, nous voulons insister sur le fait que le mélange contient à la fois des données discrètes et continues. Lorsque les réels β_j sont inconnus, l'estimation de f est alors "semi-paramétrique". Dans notre travail, nous les supposons connus. Ce qui nous permet d'estimer f par la méthode non-paramétrique des noyaux associés.*

5.2.1 Noyaux associés mixtes

La définition suivante va préciser ce qu'on appelle type de noyau mixte.

Définition 5.2.3 *Un* ***type de noyau mixte*** *K_θ de support $\mathbb{S}_\theta$ est une combinaison convexe des types de noyaux continus et discrets $K_\theta^{[j]}$ de supports $\mathbb{S}_\theta^{[j]}$. On écrit alors*

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

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

Nous précisons ici que le support $\mathbb{S}_\theta$ 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 :

$$\mathbb{S}_\theta = \bigcup_j \mathbb{S}_\theta^{[j]}. \tag{5.20}$$

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

Nous allons maintenant définir le noyau associé mixte.

Définition 5.2.4 *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 dispersion h, et constitué à la fois de noyaux associés continus et discrets. Il s'écrit :*

$$K_{\theta(x,h)}(\cdot) = \sum_{j=1}^{p} \beta_j K_{x,h}^{[j]}(\cdot)\mathbf{1}_{\mathbb{S}_{x,h}^{[j]}}(x)\mathbf{1}_{\mathbb{T}_j}(\cdot), \tag{5.21}$$

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

$$\mathbb{S}_{\theta(x,h)} = \bigcup_{j=1}^{p} \mathbb{S}_{x,h}^{[j]} \tag{5.22}$$

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

On vérifie facilement que $K_{\theta(x,h)}(\cdot)$ en (5.21) satisfait les conditions des noyaux associés (5.4)–(5.6) que l'on peut récrire de la forme

$$x \in \mathbb{S}_{\theta(x,h)}, \tag{5.23}$$
$$\mathbb{E}(\mathcal{Z}_{x,h_\beta}) = x + A_\theta(x,h), \tag{5.24}$$
$$Var(\mathcal{Z}_{x,h_\beta}) = B_\theta(x,h). \tag{5.25}$$

où $A_\theta(x,h) = \sum_{j=1}^{p} A_j(x,h)\mathbf{1}_{\mathbb{S}_{x,h}^{[j]}}(x)$ et $B_\theta(x,h) = \sum_{j=1}^{p} B_j(x,h)\mathbf{1}_{\mathbb{S}_{x,h}^{[j]}}(x)$ tendent vers 0 lorsque h tend 0 et $\mathcal{Z}_{\theta(x,h)}$ est une variable aléatoire de loi $K_{\theta(x,h)}$.

Quelques observations méritent d'être faites avant de continuer.

Remarque 5.2.5 (*i*) *La construction du noyau associé mixte $K_{\theta(x,h)}$ dépend de celle des noyaux associés continus et discrets $K_{x,h}^{[j]}$ qui le constituent.*

(*ii*) *Le paramètre de lissage h peut être choisi local (i.e. différemment sur chaque $\mathbb{T}_j$ selon le noyau associé utilisé tout comme il peut être choisi globalement en tenant seulement compte de $K_{\theta(x,h)}$.*

(*iii*) *Le noyau associé mixte $K_{\theta(x,h)}(\cdot)$ 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.*

(*iv*) *Enfin, le support $\mathbb{S}_{\theta(x,h)}$ doit être égale à $\mathbb{T}$ pour éviter les effets de bords. Ceci veut implicitement dire que toutes les composantes $\mathbb{S}_{x,h}^{[j]}$ et $\mathbb{T}_j$ respectivement de $\mathbb{S}_{\theta(x,h)}$ et $\mathbb{T}$ sont deux à deux égales.*

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

5.2.2 Estimateur de densités mixtes

Soient $X_1, X_2, \cdots, X_n$ une suite de variables aléatoires *i.i.d.* à densité mixte inconnue f en (5.1) et de support $\mathbb{T} = \cup_j \mathbb{T}_j$. L'estimateur $\widehat{f}$ à noyau associé mixte $K_{\theta(x,h)}$ de f est défini par

$$\widehat{f}(x) = \sum_{j=1}^{p} \frac{\beta_j}{n_j} \sum_{i=1}^{n_j} K_{x,h}^{[j]}(X_i) \mathbf{1}_{\mathbb{S}_{x,h}^{[j]}}(x) \mathbf{1}_{\mathbb{T}_j}(X_i), \tag{5.26}$$

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

Cet estimateur peut être écrit encore sous la forme :

$$\widehat{f}(x) = \sum_{j=1}^{p} \beta_j \widehat{f_j}(x) \mathbf{1}_{\mathbb{T}_j}(x), \tag{5.27}$$

où les $\widehat{f_j}$ sont les estimateurs à noyaux associés $K_{x,h}^{[j]}(\cdot)\mathbf{1}_{\mathbb{S}_{x,h}^{[j]}}(x)\mathbf{1}_{\mathbb{T}_j}(\cdot)$ 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. Nous allons dans la proposition suivante donner ces propriétés.

Proposition 5.2.6 *Soit $\widehat{f}$, l'estimateur à noyaux associés mixtes $K_{\theta(x,h)}$ de f en (5.1). Pour toute variable aléatoire $\mathcal{Z}_{\theta(x,h)}$ de loi $K_{\theta(x,hh)}$, $\widehat{f}$ vérifie les propriétés suivantes*

$$\mathbb{E}\left\{\widehat{f}(x)\right\} = \mathbb{E}\left\{f(\mathcal{Z}_{\theta(x,h)})\right\}, \tag{5.28}$$

$$\int_{\mathbb{T}} \widehat{f_n}(x)\Delta x =: \Lambda(n, h, K_\theta) \text{ n'est pas toujours égale à } 1. \tag{5.29}$$

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

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

Pour le second résultat, on a :

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

D'après le point (*iii*) de la Remarque 5.2.5, 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.■

Le lecteur peut se réferer à la Proposition 2.2.2 de ce travail pour plus de précision. Signalons aussi que l'on peut retrouver le même résultat en utilisant la ∇**-mesure**.

5.3 Etudes de biais et variances

Nous présentons dans cette section, les biais et variances de l'estimateur à noyau associé mixte. Puis, nous procédons au calcul des erreurs quadratiques moyennes asymptotiques respectivement ponctuelles et globales.

Proposition 5.3.1 *Soit $f \in \mathscr{C}^2(\mathbb{T})$ une densité mixte du type défini en (5.1) et $\widehat{f}$ son estimateur à noyau associé mixte en (5.26). Pour tout x dans $\mathbb{T}$ et $h = h_n > 0$, on a*

$$Biais\left\{\widehat{f}(x)\right\} = A_\theta(x,h)f^{(1)}(x) + \frac{1}{2}\left\{A^*_\theta(x,h) + B_\theta(x,h)\right\}f^{(2)}(x) + o(h^2). \tag{5.30}$$

avec $A_\theta(x,h) = \sum_j \beta_j A_j(x,h)$, $\quad B_\theta(x,h) = \sum_j \beta_j B_j(x,h)$ $\quad$ et $\quad A^*_\theta(x,h) = \sum_j \beta_j A_j^2(x,h)$.
Si en plus, f est bornée sur $\mathbb{T}$ alors

$$Var\left\{\widehat{f}(x)\right\} = \sum_{j=1}^{p} \frac{\beta_j^2}{n_j} f_j(x) \left\|K^{[j]}_{x,h}\right\|_2^2 \mathbf{1}_{\mathbb{T}_j}(x) + o\left(\frac{1}{nh^{r_2}}\right), \tag{5.31}$$

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

Démonstration. Pour le premier résultat, considérons $x \in \mathbb{T}$. En utilisant (5.26), on a :

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

Puisque les supports $\mathbb{S}_{x,h}^{[j]}$ et $\mathbb{T}_j$ sont égaux (voir point *iv* de la Remarque 5.2.5), alors on a

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

où les $f^{(j)}$ pour $j\in\{1,2\}$ sont définies dans (5.13) puis $A_\theta(x,h)=\sum_j\beta_jA_j(x,h)$, $B_\theta(x,h)=\sum_j\beta_jB_j(x,h)$ et $A_\theta^*(x,h)=\sum_j\beta_jA_j^2(x,h)$. Montrons maintenant le second résultat. Partant de la définition de $\widehat{f}$, ona :

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

En remplaçant $Var\left\{\widehat{f_j}(x)\right\}$ par sa valeur donnée dans (5.9), on obtient le résultat. ■

Nous voulons attirer l'attention du lecteur ici sur ce résultat. Comme démontré dans la Proposition 2.2.4, 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 (5.12). Aussi, la norme $\left\|K_{x,h}^{[j]}\right\|_2^2$ se calcule en utilisant l'intégrale définie en (5.19).

Risque quadratique moyen asymptotique

Les deux résultats précédents nous permettent de procéder au calcul des erreurs quadratiques moyennes asymptotiques ponctuelles et intégrées. En combinant (5.30) et (5.31) on a :

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

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

$$AMISE(x) = \int_{\mathbb{T}} \left[A_\theta(x,h) f^{(1)}(x) + \frac{1}{2} \left\{ A^*_\theta(x,h) + B_\theta(x,h) \right\} f^{(2)}(x) \right] \Delta x$$
$$+ \sum_{j=1}^{p} \frac{\beta_j^2}{n_j} \int_{\mathbb{T}_j} f_j(x) \left\| K^{[j]}_{x,h} \right\|_2^2 \Delta x. \tag{5.32}$$

Problème du choix de fenêtre

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.

Telle que présentée dans la Section 1.1.3 du Chapitre 1, l'approche de validation croisée par les moindres carrés nous amène à considérer le noyau associé mixte $K_{\theta(x,h)}$ pour tout $x \in \mathbb{T} = \cup_j \mathbb{T}_j$ et $h > 0$. Le paramètre optimal h_{cv} de h est obtenu par

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

où

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

et $\widehat{f}_{j,-i}(u) = (n_j - 1)^{-1} \sum_{i \neq k=1}^{n_j} K^{[j]}_{u,h}(X_k) \mathbf{1}_{\mathbb{S}^{[j]}_{x,h}}(u) \mathbf{1}_{\mathbb{T}_j}(X_k)$ est calculé à partir de $\widehat{f_j}(u)$ sans l'observation X_i.

5.4 Application 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

$$f(x) = 0.65 f_1(x)\mathbf{1}_{[2,5]}(x) + 0.35 f_2(x)\mathbf{1}_{\{6,7,\cdots,20\}}(x) \tag{5.33}$$

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 $\mathbb{T}_2 = \{6,7,\cdots,20\}$.

Compte tenu du manque d'informations générales sur sa forme et son appartenance à une quelconque famille de lois paramétriques, l'estimation de f définie en (5.33) 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

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

avec $\Delta = \mathscr{B}([\{x-t_0\}/\{(t_1-t_0)h\}]+1, [\{t_1-x\}/\{(t_1-t_0)h\}]+1)$. 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

$$EB_{x,h;2,5}(u) = \frac{(u-2)^{(x-2)/3h}(5-u)^{(5-x)/3h3}}{3^{1+1/h}\mathscr{B}([\{x-2\}/3h]+1, [\{5-x\}/3h]+1)}\mathbf{1}_{[2,5]}(u),$$

Il a pour caractéristiques

$$A_1(x,h;2,5) = \frac{2h(4-x)}{1+2h} \tag{5.34}$$

$$B_1(x,h;2,5) = \frac{-x^2+7x+9h+9h^2-10}{(1+2h)^2(1+3h)}. \tag{5.35}$$

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

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

et sa variance est donnée par

$$Var\left\{\widehat{f_1}(x)\right\} = \begin{cases} \dfrac{\Gamma(2c+1)}{2^{2c+1}\Gamma^2(c+1)}\dfrac{1}{n}f(x) & \text{si} \quad \dfrac{x-5}{3h} \quad \text{et} \quad \dfrac{5-x}{3h} \to c \\ \dfrac{3(x-2)^{-1/2}(5-x)^{-1/2}}{2\sqrt{\pi}nh^{1/2}}f(x) & \text{si} \quad \dfrac{x-5}{3h} \quad \text{et} \quad \dfrac{5-x}{3h} \to +\infty. \end{cases}$$

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

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

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

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

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

$$A_2(x,h) := A(6,20;h) = \frac{1}{D(6,20,h)}\left\{189 + 7^{-h}\left(\sum_{k=1}^{6} k^{h+1} - 3^{-h}\sum_{k=1}^{20} k^{h+1}\right)\right\}, \tag{5.36}$$

$$B_2(x,h) := B(6,20;h) = \frac{1}{D(6,20,h)}\left\{2961 - 7^{-h}\left(\sum_{k=1}^{6} k^{h+2} + 3^{-h}\sum_{k=1}^{20} k^{h+2}\right)\right\}. \tag{5.37}$$

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

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

et sa variance est donnée par

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

avec $R_n\left(x;6,20,h\right) \to 0$ quand $n \to +\infty$ (se référer à Kokonendji & Zocchi, 2010).

Enfin, le noyau associé mixte $K_{\theta(x,h)}$ constitué des noyaux associés bêta étendu et triangulaire général est donné par

$$K_{\theta(x,h)}(\cdot) = 0.65EB_{x,h;2,5}(\cdot)\mathbf{1}_{[2,5]}(x)\mathbf{1}_{[2,5]}(\cdot) + 0.35\mathcal{DT}_{x,6,20,h}(\cdot)\mathbf{1}_{\{6,7,\cdots,20\}}(x)\mathbf{1}_{\{6,7,\cdots,20\}}(\cdot). \tag{5.38}$$

Ses caractéristiques sont données par

$$A_\theta(x,h) = 0.65A_1(x,h)\mathbf{1}_{[2,5]}(x) + 0.35A_2(x,h)\mathbf{1}_{\{6,7,\cdots,20\}}(x) \tag{5.39}$$

$$B_\theta(x,h) = 0.65B_1(x,h)\mathbf{1}_{[2,5]}(x) + 0.35B_2(x,h)\mathbf{1}_{\{6,7,\cdots,20\}}(x) \tag{5.40}$$

où les quantités $A_1(x,h)$, $B_1(x,h)$, $A_2(x,h)$ et $B_2(x,h)$ sont respectivement donnés en (5.34), (5.35), (5.36) et (5.37).

L'estimateur $\widehat{f}$ de la densité mixte f définie en (5.33) à noyau associé mixte $K_{\theta(x,h)}$ en (5.38) est donné par

$$\widehat{f}(x) = \frac{0.65}{n_1}\sum_{i=1}^{n_1} EB_{x,h;2,5}(X_i)\mathbf{1}_{[2,5]}(x)\mathbf{1}_{[2,5]}(X_i) + \frac{0.35}{n_2}\sum_{i=1}^{n_2} \mathcal{DT}_{x,6,20,h}(X_i)\mathbf{1}_{\{6,7,\cdots,20\}}(x)\mathbf{1}_{\{6,7,\cdots,20\}}(X_i). \tag{5.41}$$

Son biais est défini par

$$Biais\left\{\widehat{f}(x)\right\} = A_\theta(x,h)f^{(1)}(x) + \frac{1}{2}\left\{A^*_\theta(x,h) + B_\theta(x,h)\right\}f^{(2)} + o(h^2) \tag{5.42}$$

où les quantités A_θ et B_θ sont définies respectivement en (5.39) et (5.40), et la quantité $A^*_\theta(x,h)$ est donnée par $A^*_\theta(x,h) = 0.65A_1^2(x,h)\mathbf{1}_{[2,5]}(x) + 0.35A_1^2(x,h)\mathbf{1}_{\{6,7,\cdots,20\}}(x)$. De plus $f^{(1)}$ et $f^{(2)}$ sont calculées à partir de (5.13).

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

$$\begin{aligned} Var\left\{\widehat{f}(x)\right\} &= \frac{0.65^2}{n_1}f_1(x)\left\|EB_{x,h;2,5}\right\|_2^2\mathbf{1}_{[2,5]}(x) + \frac{0.35^2}{n_2}f_2(x)\left\|\mathcal{DT}_{x,6,20,h}\right\|_2^2\mathbf{1}_{\{6,7,\cdots,20\}}(x) + o\left(\frac{1}{nh^{r_2}}\right) \\ &= \frac{1.2675(x-2)^{-1/2}(5-x)^{-1/2}}{2\sqrt{\pi}n_1h^{1/2}}f_1(x)\mathbf{1}_{[2,5]}(x) \\ &\quad + \frac{0.1225}{n_2\left\{D(6,20,h)\right\}^2}f_2(x)\left[1-f_2(x)\right]\mathbf{1}_{\{6,7,\cdots,20\}}(x) + o\left(\frac{1}{nh^{1/2}}\right) \end{aligned} \tag{5.43}$$

Ici $r_2 = 1/2$.

Nous présentons dans le Tableau 5.2, les différentes valeurs de biais, variance et AMSE de $\widehat{f}$. Ces résultats montrent que la variance de l'estimateur à noyau associé mixte pour le modèle de mélange gaussien-Poisson diminue en général lorsque la taille de l'échatillon 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.

TABLE 5.2 – Biais et variances de l'estimateur à noyaux associés mixtes de la densité définie en (5.33)

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

Conclusion et perspectives

Nous avons proposé dans cette thèse, une méthode non-paramétrique d'estimation de densités partiellement discrètes et continues (dites densités mixtes).

Récapitulatif

Tout d'abord, nous avons brièvement fait l'état des méthodes existantes, plus précisément celles des estimateurs à noyaux continus (classiques) de Rosenblatt (1956) et Parzen (1962), et des estimateurs à noyaux associés discrets de Kokonendji & Senga Kiessé (2011). Ce faisant, nous avons essentiellement développé les notions que nous améliorons dans notre travail en insistant plus précisément sur les notions, d'erreurs quadratiques moyennes et intégrés, les consistances ponctuelles et globales ainsi les différentes techniques de choix du paramètre de lissage h. il faut noter aussi que dans ce travail de rappel, nous avons proposé 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.

Ensuite, nous avons présenté la méthode des noyaux associés continus. Cette nouvelle méthode d'estimation englobe celle proposée par Rosenblatt (1956) puis améliorée par Parzen (1962). Les noyaux associés continus sont en général des densités de probabilité, paramétrées par le point d'estimation x et le paramètre de dispersion h. La méthode " mode-dispersion" a été proposée pour leur construction qui consiste à résoudre les équations $x = M(a, b)$ et $h = D(a, b)$ où $M(a, b)$ et $D(a, b)$ sont, respectivement, l'unique mode et le paramètre de dispersion de $K_{\theta(a,b)}$ (qui est une densité de probabilité paramétrée et de carré intégrable). Cela conduit au noyau associé $K_{\theta(x,h)} := K_{\theta(a(x,h),b(x,h))}$ de support $\mathbb{S}_{\theta(x,h)} = \mathbb{S}_{\theta(a(x,h),b(x,h))}$, où $a(x, h)$ et $b(x, h)$ sont solutions des équations précédentes. Nous avons illustré ceci sur plusieurs noyaux associés non-classiques de la littérature. Dans cette illustration, nous avons constaté que certains noyaux associés à l'exemple des noyaux de Weibul et de Birnbaum-Sanders ne peuvent pas être construits par la méthode mode-dispersion. Aussi, d'autres noyaux associés comme les noyaux gamma inverse et gaussien inverse présentent des insuffisances lorsqu'on sort du domaine de contraintes liées à leur construction.

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 à

partitionner le à $\mathbb{T}$ 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 des exemples sur plusieurs estimateurs à noyaux associés non-classiques.

De là, nous avons procédé à des études par simulations à travers trois estimateurs à noyaux lognormaux dont le premier est construit par la méthode mode-dispersion, le deuxième est sa version modifiée et le dernier est celui proposé par Jin & Kawczak (2003). Dans ces études, nous avons dans un premier temps mis en exergue le choix du paramètre de lissage par la méthodes de validation croisée par les moindres carrés et dans second temps, nous avons comparé les biais, variance et les MISEs de ces trois estimateurs tant à l'intérieur qu'aux bords. Toutefois, nous avons constaté que l'estimateur à noyau lognormal modifié est le meilleur que les deux autres dans la région intérieure et aussi il a un bon comportement aux bords lorsque la taille de l'échatillon est faible. Ce qui a finalement confirmé nos résultats théoriques.

Par ailleurs, nous avons étudié les consistances ponctuelles puis globales des estimateurs à noyaux associés continus en utilisant des différentes normes. Ces études nous ont permis de savoir que le comportement de ces estimateurs dépend d'une part du point d'estimation x, et d'autre part d'une certaine puissance du paramètre de lissage h. Aussi, nous avons montré les propriétés minimax de ces estimateurs pour des densités appartenant à la classe d'Hölder.

Enfin, nous avons combiné les noyaux associés discrets de Kokonendji & Senga Kiessé (2011) 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.

Extension aux cas mixtes multivariés

L'idée ici est de considérer une densité mixte multivariée f à estimer de support $\mathbb{T}_d \left(\subseteq \mathbb{R}^d\right)$ avec $d \in \{2, 3, \ldots\}$. Nous signalons au passage que le support $\mathbb{T}_d$ est une échelle de temps multidimensionnelle. Il peut être alternativement discret et continu dans chacune des directions, ou encore uniquement discret sur une direction et continu sur une autre ainsi de suite. Ensuite on définit pour tout $x = (x_1, \ldots, x_d)^T \in \mathbb{T}_d$ et $\mathbf{H_d}$ est une matrice de fenêtres symétrique et définie positive, le noyau associé mixte multivarié comme étant une densité de probabilité $K_{\theta(x,\mathbf{H_d})}$, dépendant de paramètres $\theta \in \Theta \subseteq \mathbb{R}^{d(d+3)/2}$ et de support $\mathbb{S}_{\theta(x,\mathbf{H_d})}$ satisfaisant les conditions suivantes :

$$\begin{aligned}
&x \in \mathbb{S}_{\theta(x,\mathbf{H_d})},\\
&\mathbb{E}\left(\mathcal{Z}_{\theta(x,\mathbf{H}_d)}\right) = x + A_\theta(x, \mathbf{H_d})\\
&\mathrm{Cov}\left(\mathcal{Z}_{\theta(x,\mathbf{H_d})}\right) = \mathbf{B}_\theta(x, \mathbf{H_d}),
\end{aligned}$$

où $\mathcal{Z}_{\theta(x,\mathbf{H_d})}$ est un vecteur aléatoire de densité $K_{\theta(x,\mathbf{H_d})}$, $A_\theta(x, \mathbf{H_d}) \to 0$ et $\mathbf{B}_\theta(x, \mathbf{H_d}) \to \mathbf{0_d}$ lorsque $\mathbf{H_d} \to \mathbf{0_d}$. Notons que $\mathbb{S}_{\theta(x,\mathbf{H_d})}$ doit nécessairement coïncider avec $\mathbb{T}_d$ afin d'éviter

les effets de bord.

Ainsi, pour une suite $X_1, \ldots, X_n$ de vecteurs aléatoires i.i.d, de densité f inconnue sur $\mathbb{T}_d$. L'estimateur à noyau mixte multivarié $\widehat{f_n}$ de la densité mixte f est défini par :

$$\widehat{f_n}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{\theta(x,\mathbf{H_d})}(X_i), \quad \forall x \in \mathbb{T}_d. \tag{5.44}$$

De là on peut exprimer son biais et sa variance sous la forme de biais et variance dans les cas univarié donnés respectivement en (5.30) et (5.31).

Remarques finales

Les travaux ainsi réalisés offrent de nombreuses perspectives intéressantes.

Tout d'abord concernant les noyaux associé non-classiques, Il serait intéressant de comparer la performance de tous les estimateurs à noyaux associés de supports $]0, +\infty[$ tant aux bords qu' à l'intérieur avec les densités vérifiant les conditions de Schuster comme traité dans Zhang (2010). Puis déterminer un équivalent du noyau d'Epanechnikov dans le cas associé non-classique.

Puisque la fenêtres h joue un rôle très important dans la performance des estimateurs de densités à noyaux associés, un sujet intéressant serait d'étudier ses sélecteurs automatiques en utilisant les critères bayésien ou les méthodes adaptatives de Lepski modifiées.

Un autre sujet intéressant est d'étendre la méthode mode-dispersion de construction des noyaux associés aux cas des données multidimensionnelles à support bornés. Ici l'aspect fondamental est la notion de dispersion mutivariée.

En ce qui concerne les noyaux associés mixtes, il est intéressant d'étudier prioritairement le cas où les coefficients β_j sont inconnus puis examiner les différentes techniques existantes pour le choix du paramètre de lissage. Ensuite étudier finement les différents effets de bords tant au passage du continu au discret qu'aux extrémités. Enfin, traiter le problème de choix de fenêtre avc d'autre méthodes de la littérature.

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 Mathematics* **35**, 3-22.

[3] Aitchison, J. & Aitken, C.G.G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika* **63**, 413-420.

[4] Ash, R.B. & Doléans-Dade, C.A. (2000). *Probability and Measure Theory*. Academic Press, California.

[5] Balakrishnan, N., Gupta, R.C., Kundu, D., Leiva, V. & Sanhueza, A. (2011). On some mixture models base on Birnbaum-Saunder distribution and associated inference. *Journal of Statistical Planning and Inference* **141**, 2175-2190.

[6] Berlinet, A. & Devroye, L. (1989). A comparison of kernel density estimates. *Publications de l'Institut de Statistique de l'Université de Paris* **38**, 03-59.

[7] Bertin, K. & Klutchnikoff, N. (2011). Minimax properties of beta kernel estimators. *Journal of Statistical Planning and Inference* **141**, 2287-2297.

[8] Billingsley, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society* **5**, 788-792.

[9] Birnbaum, Z.W. & Saunders, S.C. (1969). A new family of life distributions. *Journal of Applied Probability* **6**, 319-327.

[10] Bohner, M. & Peterson, A. (2001). *Dynamic Equations on Time Scales*. Birkhäuser Boston Inc., Boston.

[11] Bohner, M., & Peterson, A. (2003). *Advances in Dynamic Equations on Time Scales*. Birkhäuser Boston Inc., Boston.

[12] Bouezmarni, T. & Rolin, J.M. (2003). Consistency of the beta kernel density function estimator. *The Canadian Journal of Statistics* **31**, 89-98.

[13] Bouezemarni, T. & Roumbots, J.V.K. (2010). Nonparametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference* **140**, 139-152.

[14] Bouezemarni, T. & Scaillet, O. (2005). Consistency of Asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* **21**, 390-412.

[15] Bouezemarni, T., Karunamuni, R.J. & Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology* **2**, 191-212.

[16] Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimate. *Biometrika* **71**, 352-360.

[17] Brewer, M.J. (1998). A modelling approach for bandwidth selection in kernel density estimation. *In : Proceedings of COMPAST Physica Verlag, Heidelberg*, 203-208.

[18] Brewer, M.J. (2000). A Baysian model for local smoothing in kernel density estimation. *Statistics and Computing* **10**, 299-309.

[19] Brown, B.M. & Chen, S.X. (1999). Beta Bernstein smoothing for regression curves with compact support. *Scandinavian Journal of Statistics* **26**, 47-59.

[20] Chen, S.X. (1999). Beta kernel estimators for density functions. *Computational Statistics and Data Analysis* **31**, 131-145.

[21] Chen, S.X. (2000). Gamma kernel estimators for density functions. *Annals of Institute for Statistics and Mathematics* **52**, 471-480.

[22] Chen, S.X. (2010). *Personal communications.*

[23] Cline, D.B.H. & Hart, J.D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics* **1** 69-84.

[24] Cowling, A. & Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society. Series B.* **3**, 551–563.

[25] Cuevas, A., Cao, R. & Gonzalez-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics and Data Analysis* **17**, 235-239.

[26] Devroye, L. (1987). *A Course in Density Estimation*. Birkhäuser, Boston.

[27] Epanechnikov, V. (1969). Nonparametric estimates of a multivariate probability density. *Theory of Probability and its Applications* **14**, 153–158.

[28] Gasser, T. & Müller, H.G. (1979). Kernel estimation of regression functions. Smoothing techniques for curve estimation. *Proceedings of Workshop, Heidelberg, April 2–4, 1979*.

[29] Gasser, T., Müller, H.G. & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society* **2**, 238-252.

[30] Gourieroux, C. & Montfort, A. (2006). (Non) consistency of the beta kernel estimator for recovery rate distribution. *Preprint of Centre de Recherche en Economie et Statistique*, **N°2006-31**.

[31] Grubbs, F.E. (1962). Attempts to validate certain PERT statistics or picking on PERT. *Operations Research* **10**, 912–915.

[32] Gustafson, J., Hagmann, M., Nielsen, J.P. & Scaillet, O. (2007). Local transformation kernel density estimation of loss distributions. *Journal of Business and Economic Statistics*.

[33] Habbema, J.D.F., Hermans, J. & Vanden Broee, K. (1974). A stepwise discriminant analysis program using density estimation. In *Compstat* 1974, Ed. G. Bruckmnann, 101-110. Physica Verlag, Vienna.

[34] Hilger, S. (1998), *Ein Maßkettenkalkül mit Anwendung auf Zentrumsmannigfal- tigkeiten*. Ph.D. Thesis of the University of Würzburg : Würzburg.

[35] Hilger, S. (1990). Analysis on measure chains — a unified approach to continuous and discrete calculus. *Results Math* **18**, 18–56.

[36] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association* **58**, 13-30.

[37] Jin, X. & Kawczak, J. (2003). Birnbaum-Saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*. **4**, 103-1024.

[38] Jones, M.C. & Foster, P.J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistics Sinica* **4**, 1005-1013.

[39] Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall, London.

[40] Jørgensen, B. & Kokonendji, C.C. (2011). Dispersion models for geometric sums. *Brazilian Journal of Probability and Statistics* **25**, 263-293.

[41] Jørgensen, B. & Lauritzen, S.L. (2000). Multivariate dispersion models. *Journal of Multivariate Analysis* **74**, 267–281.

[42] Jørgensen, B., Goegebeur, Y. & Martnez, J. R. (2010). Dispersion models for extremes. *Extremes* **13**, 399-437.

[43] Karunamuni, R.J. & Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *The Canadian Journal of Statistics* **33**, 497-509.

[44] Karunamuni, R.J & Alberts, T. (2006). A locally adaptive transformation method of boundary correction in kernel density estimation. *Journal of Statistical Planning and Inference* **136**, 2936-2960.

[45] Karunamuni, R.J. & Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator. *Statistics and Probability Letters* **78**, 499-507.

[46] Karunamuni, R.J., Sriram, T.N. & Wu, J. (2006). Asymptotic normality of an adaptive kernel density estimator for finite mixture models. *Statistics and Probability Letters* **76**, 211-220.

[47] Karunamuni, R. J. & Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator. *Statistics and Probability Letters* **5**, 499-507.

[48] Kokonendji, C.C. & Libengué. F.G. (2011). Méthode des noyaux associés continus et estimation de densité. *Procedings du 43 ème Journée de Statistique de la SFdS*, 86-91, Tunis.

[49] Kokonendji, C.C. & Libengué. F.G. (2012). Covergences des estimateurs à noyaux associés de densité. *Journées de Statistique de la SFdS*, 5 pages, Bruxelles. http ://-jds2012.ulb.ac.be/showabstract.php ?id=98

[50] Kokonendji, C.C. & Senga Kiessé, T. (2011). Discrete associated kernels method and extensions. *Statistical Methodology*, **8**, 497-516.

[51] Kokonendji, C.C., Senga Kiessé, T. & Balakrishnan, N. (2009). Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference* **139**, 3625-3638.

[52] Kokonendji, C.C., Senga Kiessé, T. & Zocchi, S.S. (2007). Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics* **19**, 241-254.

[53] Kokonendji, C.C. & Zocchi, S.S. (2010). Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters* **80**, 1655-1662.

[54] Kuruwita, C.N., Kulasekera, K.B. & Padgett, W.J. (2010). Density estimation using asymmetric kernels and Bayes bandwidths with censored data. *Journal of Statistical Planning and Inference* **140**, 1765-1774.

[55] Markovich, N. (2007). *Nonparametric Analysis of Univariate Heavy-Tailed Data : Research and Practice*. Wiley and sons, Moscow.

[56] Marron, J.S. (1987). A comparison of cross-validation techniques in density estimation. *The Annals of Statistics* **15**, 152–162.

[57] Marron, J.S. & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society* **4**, 653-671.

[58] Marsh, L.C. & Mukhopadhyay, K. (1999). Discrete Poisson kernel density estimation with an application to wildcat coal strikes. *Applied Economics Letters* **6**, 393-396.

[59] Massart, P. (1990). The tight constant in Dvoretzky-Kiefer-Wolfowitz inequality. *Annals of Probability* **18**, 1269-1283.

[60] Monfort, A. (1997). *Cours de Statistique Mathématique*. Economica, Paris.

[61] Nadaraya, E.A. (1965). On nonparametric estimation of density function and regression. *Theoretical Probability for Applications* **10**, 186-190.

[62] Park, B.U., & Marron, J. S. (1992). On the use of pilot estimators in bandwidth selection. *Journal of Nonparametric Statistics* **3**, 231-240.

[63] Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics* **33**, 1065–1076.

[64] R Development Core Team, (2012). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http ://www.R-project.org.

[65] Regneir, E. (2005). Activity completion time in PERT and scheduling network simulation. *Defense Resources Management Institute News Letters* **2005.04.08**, 9 pages.

[66] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics* **27**, 832–837.

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

[68] Sanyal, S. (2008). Stochastic dynamic equations. PhD thesis, Missouri University of Sciences and Technology.

[69] Sarda, P. & Vieu, P. (1986). Régression non paramétrique et prédiction d'un processus markovien. *Cahiers du Centre d'Études de Recherche Opérationnelle*, **1-3**, 203–209.

[70] Sarda, P. & Vieu, P. (1988). Vitesses de convergence d'estimateurs non-paramétriques d'une régression et de ses dérivées. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique*, **2**, 83–86.

[71] Sarda, P. & Vieu, P. (1991). Smoothing parameter selection in hazard estimation. *Statistics and Probability Letters* **5**, 429–434.

[72] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. *Journal of Nonparametric Statistics* **16**, 217–226.

[73] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Communication in Statistics Theory and Methods* **5**, 1123-1136.

[74] Scott, D.W., Tapia, R.A. & Thompson, J.R. (1977). Kernel density estimation revisited. *Nonlinear Analysis* **1**, 339-372.

[75] Scott, D.W. (1992). *Multivariate Density Estimation*. John Willey & Sons, Texas.

[76] Senga Kiessé, T. (2009). *Approche non-parametrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat en Statistique, Université de Pau. http ://tel.archives-ouverts.fr/tel-00372180/fr/.

[77] Senga Kiessé, T., Libengué, F.G., Zocchi, S.S. & Kokonendji, C.C. (2010). The R package for general discrete triangular distributions, http ://www.crn.r-project.org/web/packages/TRIANGG/index.html.

[78] Senga Kiessé, T. & Rivoire, M. (2010). Discrete semiparametric regression models with associated kernel and applications. *Journal of Nonparametric Statistics* **23**, 927-941.

[79] Seshadri, V. (1993). *The Inverse Gaussian Distribution : A Case Study in Exponential Families.* Oxford University Press, New-York.

[80] Sheather, J. & Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society Series B*, **53**, 683-690.

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

[82] Simonof, J.S. (1996). *Smoothing Methods in Statistics*. Springer, New York.

[83] Simonof, J.S. & Tutz, G. (2000). Smoothing methods for discrete data. *In : smoothing and regression : Approaches, computation, and application (ed. M.G. Shimek)*. pp. 193-228. Wiley, New York.

[84] Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistitics* **12**, 1285–1297.

[85] Tiago de Oliviera, J. (1963). Estatistica de densidades : resultados assintoticos *Revista de la Facultad de Ciencias, Universidad de Lisboa*, **A9**, 111-206.

[86] Tsybakov, A.B. (2004). *Introduction à l'Estimation Non-paramétrique*. Springer, Paris.

[87] Youndjé, E., Sarda, P. & Vieu, P. (1996a). Optimal smooth hazard estimates. *Test*, **2**, 379–394.

[88] Youndjé, E., Sarda, P. & Vieu,P. (1996b). Choix de paramètres de lissage en estimation de densité conditionnelle. *Publications de l'Institut de Statistique de Paris* **38**, 57-80.

[89] Zhang, S. (2010). A note on the performance of the gamma kernel estimators at the boundary. *Statistics and Probability Letters* **7-8**, 548-557

[90] Zhang, S. & Karunamuni, R.J. (2010). Boundary performance of the beta-kernal estimator. *Journal of Nonparametric Statistics* **22**, 81–104

[91] Zhang, S., Karunamuni, R.J. & Jones, M.C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association* **94**, 1231-1241.

[92] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics* **24**, 783-795.

[93] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2013a). A Bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice* **7**, 8-23.

[94] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2013b). Adaptive smoothing in associated kernel discrete functions estimation using Bayesian approach. *Journal of Statistical Computation and Simulation* **83**, DOI : 10.1080/00949655.2012.686615.

[95] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2013c). Bayesian approach in nonparametric count regression with binomial kernel. Communications in Statistics - Simulation and Computation **42**, DOI : 10.1080/03610918.2012.725145.

Bibliographie

Annexe A

Programme sous R

A.1 Noyaux associés continus classiques

Nous présentons ici le code R pour la représentation graphique des noyaux classiques sur un même graphique.

```
################################### Epanechnikov
epa=function(y){
  T=rep(0,length(y));

  {for (j in 1:length(y))

     {if (y[j]>=-1 & y[j]<=+1)

        {T[j]= (1-(abs(y[j]))^2)*(3/4)

        }
     }
 }

 return(T)
 }

################################### biweigh
biweigh=function(y){
  T=rep(0,length(y));
   {for (j in 1:length(y))
      {if (y[j]>=-1 & y[j]<=+1)
          {T[j]= (15/16)*(1-((y[j]))^2)^2
          }
      }
   }

 return(T)
```

```
   }

#####################################gaussien
gaussian=function(y){
  T=rep(0,length(y));
   {for (j in 1:length(y))
      {if (y[j]>=-1 & y[j]<=+1)
          {T[j]= (1/(2*pi))*exp(((-1)/2)*(y[j])^2)

          }
     }
  }

 return(T)
}

##################################uniforme
unif=function(h,y){
  T=rep(0,length(y));
     {for (j in 1:length(y))
         {if (y[j]>=-1 & y[j]<=1) #
            {T[j]= 1/2

            }
         }
     }

 return(T)
   }

#######################################"Graphiques
 u=epa(h,y)
 w=gaussian(h,y)
 l=unif(h,y)
 s=biweigh(h,y)
 plot(y,s,type="l",lwd=2,col="red",xlab="x", ylab="density")
 lines(y,u,type="l",lwd=2,col="green")
 lines(y,l,type="l",lwd=2,col="magenta")
 lines(y,w,type="l",lwd=2,col="blue")
  legend("topright",c("Epanechnikov ","Gaussien","Uniforme",
  "Biweigh"),lwd=2,col=c("red","green","magenta","blue"),inset = .02)
```

A.2 Noyaux associés continus non-classiques : bêta étendu

Nous présentons ici nos programmes concernant les différents noyaux associés continus non-classiques construit par la méthode mode-dispersion plus particulièrement le cas du noyau bêta étendu qui n'est pas prédéfini dans R. Ici, nous faisons varier a et b et nous construisons le type de noyau bêta étendu correspondant de la manière suivante :

```
x=seq(2,5,0.01)
a=0.5
b=0.5
BE_1=(1/((beta(a,b))*(3^{(a+b-1)})))*((x-2)^{(a-1)})*((5-x)^{(b-1)})
a=5
b=1
BE_2=(1/((beta(a,b))*(3^{(a+b-1)})))*((x-2)^(a-1))*((5-x)^(b-1))
plot(x,BE_1,xlab="y",ylab="Beta(a,b;2,5)(y)",main="Type de noyau bêta
étendu",cex.lab=1.5,cex.axis=1.5,pch=".",lty=1,lwd=2,col="red")
lines(x, BE_2, lwd=2,lty=2,col="green")
op <- par(bg="white")
legend(2.7,1.8,c( "a=b=0.5","a=5,b=1"),
pch=c(".",."), lty=c(1,2),col=c("red","green"),lwd=c(2,2),cex = 1.5)
```

Après la construction par la méthode mode-dispersion, la représentation graphique pour un réel x fixé est donnée à partir du programme suivant

```
x=2.8    #on fixe x
Vb=seq(2,5,0.01)  # On fait varier l'aléatoire
h=0.1 #pour un bien h choisi
BE_1=(1/((beta(((x-2)/(3*h))+1,((5-x)/(3*h))+1))*(3^(a+b-1))))
*((Vb-2)^((x-2)/(3*h)))*((5-Vb)^((5-x)/(3*h)))

h=0.09 #pour un autre h bien choisi
BE_2=(1/((beta(((x-2)/(3*h))+1,((5-x)/(3*h))+1))*(3^(a+b-1))))
*((Vb-2)^((x-2)/(3*h)))*((5-Vb)^((5-x)/(3*h)))

plot(Vb,BE_1,xlab="y",ylab="Beta(x,h;2,5)",main="Noyau associé bêta
étendu pour x=2.7",cex.lab=1.5,cex.axis=1.5,pch=".",lty=1,lwd=2,col="red")
lines(Vb, BE_2, lwd=2,lty=2,col="green")
op <- par(bg="white")
 legend(3.8,6000,c( "h=0.09","h=0.1"),
 pch=c(".",."), lty=c(1,2),col=c("red","green"),lwd=c(2,2),cex = 1.5)
```

Le lecteur peut éventuellement fixer h puis faire varier x.

A.3 Estimateurs à noyaux associés continus

A.3.1 Programme général d'estimation

Nous donnons ici le cas particulier d'estimateur à noyau gamma. Le lecteur peut remplacer le noyau gamma par n'importe quels autres noyaux associés non-classiques.

Estimeteurs à noyau gamma

```
Est_Gam=function(V,x,h){
out=(1/length(V))*sum(dgamma(V,(x/h)+1,1/h))
out
}
V=rgamma(100,1.5,2.6)
x=1.5
h=0.025
test=Est_Gam(V,x,h)
test
x=seq(0,18,0.05)
F_est=0
for (i in 1:length(x)){
F_est[i]=Est_Gam(V,x[i],h)
}
F_est
f=dgamma(x,1.5,2.6)
plot(x,f,type="l")
lines(x,F_est,type="l",col="red")
```

Calcul de la constante de normalisation

```
############ Calcul de la constante ###########
y=runif(1000,0,1)
F_Est1=0
for (i in 1:length(y)){
F_Est1[i]=Est_Gam(V,18*y[i],h)}
#################################
cal_cte=function(alp, bet){
V=rgamma(100,alp, bet)
h=0.5
y=runif(1000,0.1,1)
F_Est1=0
for (i in 1:length(y)){
F_Est1[i]=Est_Gam(V,18*y[i],h)}
F_Est1=sum(F_Est1)
C1=(18/length(y))*F_Est1
```

```
C1}
Cte=0
alpV=c(1.5,1.9,2.3,2.5,2.8,3.1,3.4,3.7,4.1,4.5)
betV=c(2.6,2.8,2.9,3.5,3.9,4.5,4.7,5.2,5.7,6)
for (i in 1:length(alpV)){
Cte[i]=cal_cte(alpV[i],betV[i])}
Cte
```

A.3.2 Calcul général de Biais, Variance et MSE

Ici, nous désignons la densité f à estimer par la densité normale tronquée.

```
########## densité  normale tronquée  entre a et b############
sim_gauss_trq=function(n,mu,sig, a,b){
vec_sim=qnorm(pnorm(a,mu,sig)+(pnorm(b,mu,sig)-pnorm(a,mu,sig))
*runif(n,0,1),mu,sig)}
V1=sim_gauss_trq(50000,0,1,0,1)
V1
plot(sort(V1), dnorm(sort(V1), 0, 1), type="l",col="red") # Rep. graph.

##################### Calcul de biais ############################
Bias=function(A,B,x,h){
out0=((-(2*x))/(sqrt(2*pi)))
out1=exp(-((x^2)/2))
out2=((-2*(x+1))/(sqrt(2*pi)))
f1=out0*out1
f2=out2*out1
out=(A*f1)+((A^2)+B)*((f2)/2)
out}

#################### Calcul de la variance ########################
Var=function(K1,x,h){
y=runif(1000,0,1)
m=length(y)
f=(2/(sqrt(2*pi)))*exp(-(x^2)/2)
out0=(alph)/m
out1=(f/m)
out2=sum((K1)^2)
out=out0*out1*out2
out}

######################## Calcul du MSE ############################
MSEE=function(A,B,K,x,h){
out0=((Bias(A,B,x,h))^2)+Var(K,x,h)
mse=out0
mse}
```

```
MSE=((Bias(A,B,x,h))^2)+Var(V,K1,x,h)
MSE
```

A.3.3 Cas particuliers

```
############  Lognormal mode-dispersion ########################
LN1=function(V,x,h){
out0=(1/(V*h*sqrt(2*pi)))
out1=(1/2)*(((1/h)*(log(V/x))-h)^2)
out2=exp(-out1)
out=out0*out2
out}

A1=x*(exp(((3/2)*(h^2))-1))
B1=(x^2)*(exp((3*(h^2))))*(exp(h^2)-1)
K1=LN1(V,x,h)

########################  Lognormal   modifié #####################
LN2=function(V,x,h){
out0=(1/(V*h*sqrt(2*pi)))
out1=(1/2)*((((1/(2*h))*(2*log(V)-h^2))-((1/(h*alph))*(x*log(alph))))^2)
out2=exp(-out1)
out=out0*out2
out}

A2=((alph)^(x/alph))-x
B2=(exp(h^2)-1)*(exp(((2*x*log(alph))/alph)+(h^2)/2))
K2=LN2(V,x,h)
```

On peut éventuellement déduire la cas de lognormale de Jin & Kawczk (2003) à partir de première version de lognormal.

```
########################  Gamma mode-dispersion  ####################
A3=h
B3=x*h+h^2
K3=dgamma(V,(x/h)+1,h)

########################### Gamma modifié  ###########################
A4=(x^2+2*(alph)*(h-x))/(2*(alph))
B4=(h*((x^2)+2*h*(alph)))/(2*(alph))
K4=dgamma(V,(((x^2)/(2*h*alph))+1),h,log=FALSE)
```

```
########################  Inverse Gamma   mode-dispersion ###########
GI1=function(V,x,h){
out0=(1/(gamma((1/(x*h))-1)))
out1=(h^(1-(1/(x*h))))
out2=(V^(-(1/(x*h))))
out3=exp(-(1/(h*V)))
out=out0*out1*out2*out3
out}
A5=(2*(x^2)*h)/(1-2*x*h)
B5=(h*x^3)/(((1-2*x*h)^2)*(1-3*x*h))
K5=GI1(V,x,h)

########################## Inverse Gamma modifié  ####################
GI2=function(V,x,h){
out0=(1/(gamma((x/(h*alph))+1)))
out1=(h^(-(x/(h*alph))-1))
out2=(V^(-(2+(x/(h*(alph)^2)))))
out3=exp(-(1/(h*V)))
out=out0*out1*out2*out3
out}
A6=((((alph)^2)-x^2)/x)
B6=(h*((alph)^6))/((x^2)*(x-h*((alph)^2)))
K6=GI2(V,x,h)


###################### Inverse Gaussien mode-dispersion #############
IG1=function(V,x,h){
out0=sqrt(1-3*x*h)
out1=(1/((sqrt(2*pi*h*(V^3)))))
out2=((out0)/(2*x*h))
out3=(((V*out0)/x)-2+(x/(V*out0)))
out4=exp(-out2*out3)
out=out1*out4
out}

A7=x*((1/(sqrt(1-3*x*h)))-1)
B7=((x^3)*h)/((1-3*x*h)^(3/2))
K7=IG1(V,x,h)

################# Inverse Gaussien modifié  ########################
IG2=function(V,x,h){
out0=(1/((sqrt(2*pi*h*(V^3)))))
out1=(alph)/(2*(x^2)-(alph)^2)
out2=(out0/(2*h))
out3=(V/out1-2+(out1/V))
out4=exp(-out2*out3)
out=out0*out4
```

```
out}
A8=(2*(x^2)-alph*(alph+x))/(alph)
B8=h*(((2*x^2-(alph)^2)^3)/((alph)^3))
K8=IG2(V,x,h)

################ Inverse Gaussien  Rec. mode-dispersion ##########
GIR1=function(V,x,h){
out0=sqrt(x^2+x*h)
out1=(V/out0)-2+(out0/V)
out=(1/(sqrt(2*pi*h*V)))*exp(-(out0/(2*h))* out1)
out}
A9=(((x^2+x*h)^(1/2))-x+h)
B9=h*((((x^2)+x*h)^(1/2))+2*h)
K9=GIR1(V,x,h)

################ Rec. Inverse Gaussien modifié  ##################
GIR2=function(V,x,h){
out0=(1/(sqrt(2*pi*h*V)))
out1=(alph)/(x^2-(alph)*h)
out1=((V*out1)-2+(1/(V*out0)))
out2=(out1)/(2*h)
out3=exp(-out2*out1)
out=out0*out3
out}

A10=(x*(x-alph*h))/(alph)
B10=(h*(x^2+h*(alph)))/(alph)
K10=GIR2(V,x,h)


############  Repr. graph. de MSE version  mode-dispersion #########
x=seq(0.1,alph,0.001)
mse1=0
mse3=0
mse5=0
mse7=0
mse9=0
for(i in 1:length(x)){
print(i)
mse1[i]=MSEE(A1,B1,K1,x[i],h)
mse3[i]=MSEE(A3,B3,K3,x[i],h)
mse5[i]=MSEE(A5,B5,K5,x[i],h)
mse7[i]=MSEE(A7,B7,K7,x[i],h)
mse9[i]=MSEE(A9,B9,K9,x[i],h)}
mse1
mse3
mse5
```

```
mse7
mse9

plot(x,mse1,xlab="x",ylab="MSE(x)",ylim=c(0,1),type="l",lwd=2)
lines(x,mse3,type="l",lwd=2,col="red")
lines(x,mse5,type="l",lwd=2,col="blue")
lines(x,mse7,type="l",lwd=2,col="green")
lines(x,mse9,type="l",lwd=2,col="yellow")

######## Représentation graphique de MSE de la version modi fiée #######
x=seq(0.1,alph,0.001)
mse2=0
mse4=0
mse6=0
mse8=0
mse10=0
for(i in 1:length(x)){
print(i)
mse2[i]=MSEE(A1,B1,K1,x[i],h)
mse4[i]=MSEE(A3,B3,K3,x[i],h)
mse6[i]=MSEE(A5,B5,K5,x[i],h)
mse8[i]=MSEE(A7,B7,K7,x[i],h)
mse10[i]=MSEE(A9,B9,K9,x[i],h)}
mse2
mse4
mse6
mse8
mse10

pdf('msee_1.pdf')
plot(x,mse3,xlab="x",ylab="MSE(x)",ylim=c(0,1.1),main="MSE behaviors for
h=0.5",type="l",lwd=2)
lines(x,mse5,type="l",lwd=2,col="grey",lty=2)
lines(x,mse7,type="l",lwd=2,col="black",lty=3)
lines(x,mse9,type="l",lwd=2,col="grey",lty=4)
lines(x,mse1,type="l",lwd=2,col="black",lty=5)
legend(0.6,1.1,c("Gamma","Inverse gamma","Inverse gaussien","Inverse gaussien
réciproque","Lognormal"),lwd=c(2,2,1,1,2),lty=c(1,2,3,4,5),
 c("black","grey","black","grey","black"))
dev.off()
```

A.3.4 Choix de h par la validation croisée

Le programme donné ici est dans le cas particulier du noyau bêta. Le lecteur peut toutefois remplacer le noyau bêta un noyau non-classique de son choix tout en faisant

attention au support de la densité à estimer.

```
 ############## Programme de validation croisée ##################
alp=2
beta=1.5
V=rbeta(200,alp,beta,ncp=0)###########
h=0.1####################  pour test
####################calcul de CV0 ################################
CV_0=function(x,h){
out=(1/length(V))*sum(dbeta(V,(x/h)+1, ((1-x)/h)+1, ncp=0,log = FALSE))}
test=CV_0(0.9,0.5)
test

##################### CV1 ###########################
y=runif(1000,0,1)
m=length(y)
m
CV_1=function(h){
b=0
for(i in 1:m){
print(i)
b[i]=CV_0(y[i],h)}
out1=(1/m)*sum(b[i]^2)
}
test=CV_1(h)
test

################### CV2 #######################
CV_2=function(h){
n=length(V)
a=0
b=matrix(0,nrow=n, ncol=n)

for(i in 1:length(V)){
b[i,]=dbeta(V, (V[i]/h)+1, ((1-V[i])/h)+1, ncp=0,log = FALSE)
a[i]=sum(b[i,])-dbeta(V[i],(V[i]/h)+1, ((1-V[i])/h)+1, ncp=0,log = FALSE)}
out=(2/(n*(n-1)))*sum(a)}
a=CV_2(h)
a

######################## CV final ####################
CV_3=function(h){
out=CV_1(h)-CV_2(h)}

##############optimisation avec durée ##############
hcv=function(){
```

```
h_cv=optimize(CV_3, c(0.001, 1), tol = 0.005, maximum = FALSE)
}
hcv()
system.time(hcv())

h_cv=hcv()

h_cv=h_cv\$minimum
h_cv

################# Repr. graph. de CV ####################
h=seq(0.001,1,by=0.01)
CV=0
for(j in 1:length(h)){
print(j)
CV[j]=CV_3(h[j])}
CV
plot(h,CV, type="l")
```

A.4 Programme de simulation avec réplication : Cas de ISE

Ce programme est un cas particulier de simulation avec réplication utilisant le noyau bêta. Selon le besoin on peut remplacer le noau bêta par le noyau approprié.

```
####################### simulation avce réplication ##################
replic=function(n,Nsim,alph,bet){
n=100#######################Taille de l'échantillon
Nsim=50 ############Nombre de simulation
Vsim=matrix(0,nrow=Nsim,ncol=n)
for (j in 1:Nsim){
Vsim[j,]=rbeta(n,alph,bet)}
Vsim}
Vsim=replic(100,50,1.5,2.5)
Vsim

################### Simulation de ISE ######################
Est_Bet=function(V,x,h){
out=(1/length(V))*sum(dbeta(V,(x/h)+1,((1-x)/h)+1),ncp=0,log=FALSE)
out}

######################### ISE ################################
ISE=function(V,h){
u=runif(1000,0,1)
```

```
out0=0
for (i in 1:length(u)){
out0[i]=Est_Bet(V,u[i],h)}
out0
out1=dbeta(u,1.5,2.5,ncp=0,log=FALSE)
out=mean((out1-out0)^2)}
test=ISE(Vsim[1,],0.05)
test
ISEsim=0
for(j in 1:Nsim){
ISEsim[j]=ISE(Vsim[j,],0.05)}
ISEsim
```

A.5 Estimateurs à noyaux associés mixtes

```
 ####################Calcul de l'indicatrice################
Ind_1=function(x,a,b){ifelse(x>=a & x<=b,1,0)}
a=2
b=5
x1=1
x2=3
Ind_1(x1,a,b)
Ind_1(x2,a,b)

E=6:20
Ind_2=function(X,E){ifelse(X%in%E,1,0)}
x3=4
x4=7
Ind_2(x3,E)
Ind_2(x4,E)

############# Calcul du noyau associé beta etendu ###############
h=0.25
x=2
BE_1=function(V,x,h){
out1=(1/((beta(((x-2)/(3*h))+1,((5-x)/(3*h))+1))))
out2=(((3^(a+b-1))))*((V-2)^((x-2)/(3*h)))
out3=((5-V)^((5-x)/(3*h)))
out=out1*out2*out3}

######################## Calculs des caractéritiques ############
A1=(2*h*(4-x))/(1+2*h)
B1=((-x^2)+7*x+9*h+9*(h^2)-10)/(((1+2*h)^2)*((1+3*h)))

####### Calcul du noyau de poisson triangulaire general ##########
```

```
############ Calcul de la constante de normalisation #############
D=function(h)
   {out0=0
       for (i in 1:6)
       {out0=sum(i^h)
       }
    out1=0
    for (i in 1:20)
       {out1=sum(i^h)
       }
   out=27-(7^(-h))*out0+(21^(-h))*out1
   }
test=D(0.5)
test

######################Noyau associé triangulaire general#########
y=12
E1=6:y
Ind_3=function(X,E1){ifelse(X%in%E1,1,0)}
x3=4
x4=7
Ind_3(x3,E1)
Ind_3(x4,E1)

E2=y:20
Ind_4=function(X,E2){ifelse(X%in%E2,1,0)}
x5=4
x6=7
Ind_4(x3,E2)
Ind_4(x4,E2)

DT=function(V,x,h)
   {out0=(1/D(h))
    out1=(1-((x-V)/7)^h)*Ind_3(x,E1)
    out2=(1-((V-x)/7)^h)*Ind_4(x,E2)
    out=out0*(out1+out2)
   }
V=6:20
x=7
y=x
h=0.5
test=DT(V,x,h)
test

#########################Calcul des caractéristiques ##########
A2=function(h)
```

```
    {out0=0
     for (i in 1:6)
       {out0=sum(i^(h+1))
       }
    out1=0
    for (i in 1:20)
       {out1=sum(i^(h+1))
       }
   out=(1/D(h))*(189+7^(-h)*out0-(21^(-h))*out1)
    }
test=A2(h)
test

B2=function(h)
    {out0=0
     for (i in 1:6)
       {out0=sum(i^(h+2))
       }
    out1=0
    for (i in 1:20)
       {out1=sum(i^(h+2))
       }
   out=(1/D(h))*(2961+7^(-h)*out0-(21^(-h))*out1)
    }
test=B2(h)
test

######################### Calcul de A_theta #######################
Att=0.65*A1*Ind_1(x,2,5)+0.35*A2(h)*Ind_2(x,E)
Att

Btt=0.65*B1*Ind_1(x,2,5)+0.35*B2(h)*Ind_2(x,E)
Btt

Astt=0.65*(A1^2)*Ind_1(x,2,5)+0.35*((A2(h))^2)*Ind_2(x,E)
Astt
#########################Calcul des dérivee #######################
f1=(1/(sqrt(2*pi)))*exp((-1/2)*(x^2))*Ind_1(x,2,5)
f1

f11=(-x/(sqrt(2*pi)))*exp((-1/2)*(x^2))*Ind_1(x,2,5)
f11

f12=(((x^2)-1)/(sqrt(2*pi)))*exp((-1/2)*(x^2))*Ind_1(x,2,5)
f12

f2=function(x)
```

```
{(((lda)^x)/factorial(x))*exp(-lda)}
lda=3
f21=(f2(x+1)-f2(x-1))*Ind_2(x,E)
f21

f22=(1/2)*(f2(x+2)-2*f2(x+1)+2*f2(x)-2*f2(x-1)+f2(x-2))
f22

f_1=0.65*f11+.035*f21
f_2=0.65*f12+.035*f22

##########################Calcul de biais ##########################

Biaisfn=Att*f_1+(1/2)*(Astt+Btt)*f_2
#################################################################
x=2.5
V=6:20
y=x
h=0.5

Att=0.65*A1*Ind_1(x,2,5)+0.35*A2(h)*Ind_2(x,E)
Att

Btt=0.65*B1*Ind_1(x,2,5)+0.35*B2(h)*Ind_2(x,E)
Btt

Astt=0.65*(A1^2)*Ind_1(x,2,5)+0.35*((A2(h))^2)*Ind_2(x,E)
Astt

Biaisfn=Att*f_1+(1/2)*(Astt+Btt)*f_2
Biaisfn

#############################################################
x=4.5
V=6:20
y=x
h=0.5

Att=0.65*A1*Ind_1(x,2,5)+0.35*A2(h)*Ind_2(x,E)
Att

Btt=0.65*B1*Ind_1(x,2,5)+0.35*B2(h)*Ind_2(x,E)
Btt

Astt=0.65*(A1^2)*Ind_1(x,2,5)+0.35*((A2(h))^2)*Ind_2(x,E)
Astt
```

```
Biaisfn=Att*f_1+(1/2)*(Astt+Btt)*f_2
Biaisfn

#################################################################
x=7
V=6:20
y=x
h=0.5

Att=0.65*A1*Ind_1(x,2,5)+0.35*A2(h)*Ind_2(x,E)
Att

Btt=0.65*B1*Ind_1(x,2,5)+0.35*B2(h)*Ind_2(x,E)
Btt

Astt=0.65*(A1^2)*Ind_1(x,2,5)+0.35*((A2(h))^2)*Ind_2(x,E)
Astt

Biaisfn=Att*f_1+(1/2)*(Astt+Btt)*f_2
Biaisfn

##################################################################
x=19
V=6:20
y=x
h=0.5

Att=0.65*A1*Ind_1(x,2,5)+0.35*A2(h)*Ind_2(x,E)
Att

Btt=0.65*B1*Ind_1(x,2,5)+0.35*B2(h)*Ind_2(x,E)
Btt

Astt=0.65*(A1^2)*Ind_1(x,2,5)+0.35*((A2(h))^2)*Ind_2(x,E
Astt

Biaisfn=Att*f_1+(1/2)*(Astt+Btt)*f_2
Biaisfn

########################## Calcul de variance #####################
lda=3
Varfn1=function(x,h,n1,n2)
     {
       out0=2*(sqrt((x-2)*(5-x)*(pi)*h*n1))
       out1=(1.2675*f1)*Ind_1(x,2,5)
       out=(out0/out1)
     }
```

```
Varfn2=function(x,h,n1,n2)
     {out2=(0.1225)*(1-f2(x))*f2(x)*Ind_2(x,E)
      out3=((D(h))^2)*n2
      out=(out2/out3)
     }
Varfn=Varfn1(x,h,n1,n2)+varfn2(x,h,n1,n2)

########################## Calcul du AMSE #######################
AMSE=(Biaisfn)^2+Varfn
```

Résumé

Nous présentons dans cette thèse, l'approche non-paramétrique par noyaux associés mixtes, pour les densités à supports partiellement continus et discrets. Nous commençons par rappeler d'abord les notions essentielles d'estimation par noyaux continus (classiques) et noyaux associés discrets. Nous donnons la définition et les caractéristiques des estimateurs à noyaux continus (classiques) puis discrets. Nous rappelons aussi les différentes techniques de choix de paramètres de lissage et nous revisitons les problèmes de supports ainsi qu'une résolution des effets de bord dans le cas discret. Ensuite, nous détaillons la nouvelle méthode d'estimation de densités par les noyaux associés continus, lesquels englobent les noyaux continus (classiques). Nous définissons les noyaux associés continus et nous proposons la méthode mode-dispersion pour leur construction puis nous illustrons ceci sur les noyaux associés non-classiques de la littérature à savoir bêta et sa version étendue, gamma et son inverse, gaussien inverse et sa réciproque le noyau de Pareto ainsi que le noyau lognormal. Nous examinons par la suite les propriétés des estimateurs qui en sont issus plus précisément le biais, la variance et les erreurs quadratiques moyennes ponctuelles et intégrées. Puis, nous proposons un algorithme de réduction de biais que nous illustrons sur ces mêmes noyaux associés non-classiques. Des études par simulations sont faites sur trois types d'estimateurs à noyaux lognormaux. Par ailleurs, nous étudions les comportements asymptotiques des estimateurs de densité à noyaux associés continus. Nous montrons d'abord les consistances faibles et fortes ainsi que la normalité asymptotique ponctuelle. Ensuite nous présentons les résultats des consistances faibles et fortes globales en utilisant les normes uniformes et L^1. Nous illustrons ceci sur trois types d'estimateurs à noyaux lognormaux. Par la suite, nous étudions les propriétés minimax des estimateurs à noyaux associés continus. Nous décrivons d'abord le modèle puis nous donnons les hypothèses techniques avec lesquelles nous travaillons. Nous présentons ensuite nos résultats minimax tout en les appliquant sur les noyaux associés non-classiques bêta, gamma et lognormal. Enfin, nous combinons les noyaux associés continus et discrets pour définir les noyaux associés mixtes. De là, les outils d'unification d'analyses discrètes et continues sont utilisés, pour montrer les différentes propriétés des estimateurs à noyaux associés mixtes. Une application sur un modèle de mélange des lois normales et de Poisson tronquées est aussi donnée. Tout au long de ce travail, nous choisissons le paramètre de lissage uniquement avec la méthode de validation croisée par les moindres carrés.

Mots clés **: Convergence, Densité mixte, Échelles de temps, Effet de bords, Estimation non-paramétrique par noyau, Modèle de mélange, noyau uni-modal, Paramètre de dispersion, Validation croisée**

Abstract

We present in this thesis, the non-parametric approach using mixed associated kernels for densities with supports being partially continuous and discrete. We first start by recalling the essential concepts of classical continuous and discrete kernel density estimators. We give the definition and characteristics of these estimators. We also recall the various technical for the choice of smoothing parameters and we revisit the problems of supports as well as a resolution of the edge effects in the discrete case. Then, we describe a new method of continuous associated kernels for estimating density with bounded support, which includes the classical continuous kernel method. We define the continuous associated kernels and we propose the mode-dispersion for their construction. Moreover, we illustrate this on the non-classical associated kernels of literature namely, beta and its extended version, gamma and its inverse, inverse Gaussian and its reciprocal, the Pareto kernel and the kernel lognormal. We subsequently examine the properties of the estimators which are derived, specifically, the bias, variance and the pointwise and integrated mean squared errors. Then, we propose an algorithm for reducing bias that we illustrate on these non-classical associated kernels. Some simulations studies are performed on three types of estimators lognormal kernels. Also, we study the asymptotic behavior of the continuous associated kernel estimators for density. We first show the pointwise weak and strong consistencies as well as the asymptotic normality. Then, we present the results of the global weak and strong consistencies using uniform and L^1 norms. We illustrate this on three types of lognormal kernels estimators. Subsequently, we study the minimax properties of the continuous associated kernel estimators. We first describe the model and we give the technical assumptions with which we work. Then we present our results that we apply on some non-classical associated kernels more precisely beta, gamma and lognormal kernel estimators. Finally, we combine continuous and discrete associated kernels for defining the mixed associated kernels. Using the tools of the unification of discrete and continuous analysis, we show the different properties of the mixed associated kernel estimators. All through this work, we choose the smoothing parameter using the least squares cross-validation method.

Keywords**: Asymmetric kernel, Boundary effect, Convergence, Cross-validation, Dispersion parameter, Mixed density, Mixture model, Nonparametric kernel estimation, Time-scales, Unimodal kernel**

Mathematics Subject Classification**: 62G07, 62G20, 62G99, 62E15**