

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université A.MIRA-BEJAIA



Faculté des Sciences Exactes
Département Recherche Opérationnelle
Unité de recherche Modélisation et Optimisation de systèmes (LaMOS)

THÈSE
EN VUE DE L'OBTENTION DU DIPLOME DE
DOCTORAT

Domaine : Mathématiques et informatique Filière : Mathématiques appliquées

Spécialité : Recherche Opérationnelle et Aide à la Décision

Présentée par
KHEMICI Mohamed

Thème

**ESTIMATION NON PARAMÉTRIQUE PAR NOYAUX DANS LE CAS DES
DONNÉES DÉPENDANTES**

Soutenue le : 03/07/2024

Devant le Jury composé de :

Nom et Prénom	Grade		
M^{me} LAGHA Karima	Professeur	Univ. de Béjaia	Présidente
M^r ZOUGAB Nabil	Professeur	Univ. de Béjaia	Rapporteur
M^r ADJABI Smail	Professeur	Univ. de Béjaia	Co-rapporteur
M^{me} BARECHE Aicha	Professeur	Univ. de Béjaia	Examinatrice
M^r HAMAZ Abdelghani	Professeur	Univ. de Tizi-Ouzou	Examinateur

Année Universitaire : 2023/2024

** Remerciements **

Je remercie DIEU, de m'avoir accordé santé, courage et patience afin d'accomplir ce travail.

Je tiens à exprimer toute ma reconnaissance à Monsieur ZOUGAB Nabil, Professeur à l'Université de Béjaia, mon directeur de thèse pour sa disponibilité, son écoute, ses conseils et ses encouragements.

Je tiens à exprimer toute ma reconnaissance à Monsieur ADJABI Smail, Professeur à l'Université de Béjaia, mon co-directeur de thèse pour sa disponibilité, son écoute, ses conseils et ses encouragements.

Je tiens à exprimer ma gratitude envers Madame LAGHA Karima, Professeur à l'Université de Béjaia, pour avoir accepté de présider le jury de cette thèse.

Je remercie aussi Monsieur Hamaz Abdelghani, Madame BARECHE Aicha de m'avoir fait l'honneur d'accepter de juger ce travail.

Je leur suis sincèrement reconnaissant de s'être rendus disponibles pour cette soutenance. Vos remarques m'ouvrent de nouveaux horizons et de nouvelles perspectives.

Je remercie ma famille, mes proches et mes amis qui m'ont soutenu durant cette période de préparation de ce travail.

Table des matières

Table des matières	i
Table des figures	iv
Listes des tableaux	vi
Liste des contributions	vii
Introduction générale	1
1 Estimation non paramétrique de la fonction de densité par la méthode du noyau dans le cas des données indépendantes	5
1.1 Introduction	6
1.2 Noyau continu symétrique	6
1.3 Cas continu	8
1.3.1 Estimateur à noyau associé continu	8
1.3.2 Construction du noyau associé	9
1.3.3 Estimateur à noyau asymétrique de la densité de probabilité	11
1.3.4 Propriétés des estimateurs à noyaux asymétriques	15
1.3.5 Convergence en moyenne quadratique (MSE) et en moyenne quadratique intégrée (MISE)	17
1.3.6 Choix du paramètre de lissage	18
1.4 Cas discret	24
1.4.1 Noyaux associés discrets	24
1.4.2 Propriétés de l'estimateur à noyau discret	25
1.4.3 Convergence en moyenne quadratique et en moyenne quadratique intégrée	27
1.5 Conclusion	28
2 Estimation de la densité par la méthode du noyau dans le cas des données dépendantes	29

2.1	Introduction	30
2.2	Contexte de l'étude	30
2.3	Estimateur à noyau asymétrique : cas des données dépendantes	31
2.4	Hypothèses et principaux résultats	32
2.5	Propriétés de l'estimateur à noyau asymétrique	32
2.5.1	MSE de l'estimateur à noyau asymétrique	33
2.5.2	Preuve du Théorème 2.5.1.	33
2.6	Application numérique	34
2.6.1	Simulation	34
2.6.2	Durées conditionnelles auto-régressives	35
2.6.3	Volatilité stochastique	38
2.7	Application sur des données réelles	41
2.8	Conclusion	43
3	Estimation récursive de la densité de probabilité par la méthode du noyau	44
3.1	Introduction	45
3.2	Estimation récursive de la densité de probabilité : cas des données indépendantes	46
3.2.1	Estimation récursive par noyaux symétriques	46
3.2.2	Propriétés des estimateurs récursifs à noyaux symétriques	47
3.2.3	Méthode d'approximation stochastique pour estimer la densité par la méthode du noyau symétrique	49
3.2.4	Estimateur récursif à noyau asymétrique (cas univarié)	53
3.2.5	Propriétés de l'estimateur récursif à noyaux asymétriques	54
3.3	Estimation récursive de la densité par noyau : cas des données dépendantes	55
3.3.1	Estimateur récursif à noyau symétrique cas multivarié	55
3.3.2	Propriétés de l'estimateur récursif à noyau symétrique	55
3.3.3	Estimateur récursif de la densité par noyaux asymétriques gamma généralisé cas univarié	56
3.4	Estimation récursive de la fonction de masse de probabilité par noyau discret : cas des données de comptage	66
3.4.1	Présentation de l'estimateur récursif à noyau discret	66
3.4.2	Principaux résultats	67
3.4.3	Simulations	68
3.4.4	Temps de calcul	73
3.4.5	Illustrations à partir des données réelles	74
3.5	Conclusion	75

Table des matières	iii
Conclusion générale	76
Bibliographie	78
Index	83

Table des figures

2.1	Quelques estimateurs à noyaux asymétriques avec la fonction de densité empirique utilisant le modèle DCA.	37
2.2	Quelques estimateurs des noyaux asymétriques avec la fonction de densité empirique en utilisant le modèle VS.	41
2.3	Comparaison de quelques estimateurs à noyaux asymétriques avec l’histogramme des donnée SNCF	42
3.1	Les estimateurs récursifs et non récursifs pour les noyaux MG et W pour le modèle DCA (D1); dépendance forte (en haut), dépendance faible (au milieu) et indépendance (en bas).	62
3.2	Les estimateurs récursifs et non récursifs pour les noyaux MG et W pour le modèle VS; dépendance forte (en haut), dépendance faible (au milieu) et l’indépendance (en bas).	63
3.3	Temps de calcul (en secondes) des estimateurs non récursif à noyau MG (en trait plein)/récursif (en pointillé) pour le modèle DCA (D1) à partir de $\{X_1, \dots, X_{200+m}\}$	64
3.4	Les estimateurs récursifs et non récursifs par les noyau MG et W pour l’indice des prix à la consommation	65
3.5	Les estimateurs récursifs et non récursifs par les noyau MG et W pour le trafic voyageurs de la SNCF.	66
3.6	La fonction de masse de probabilité (pmf).	70
3.7	Estimation pmf de données poissonnières et géométriques avec $n = 500$ en utilisant l’estimateur récursif et non récursif à noyau discret $DT(a = 2)$	71
3.8	Temps de calcul (en secondes) de l’estimateur à noyau DT non récursif (en trait plein)/récursif (en pointillé) pour le modèle D1 à partir de $\{X_1, \dots, X_{100+k}\}$	74
3.9	Estimateurs récursifs et non récursifs à noyau discret (DT) pour les données 1 du tableau 3.9 avec $n = 210$, et les données 2 du tableau 3.10 avec $n = 82$	75

Liste des tableaux

1.1	Exemples des noyaux continus symétriques.	7
1.2	Quelques noyaux associés continus asymétrique.	15
1.3	Distributions GBS	16
1.4	La forme explicite de $\phi(x, f)$ et $\rho(x, h)$	17
1.5	Constantes de normalisation	17
1.6	Noyaux associé discrets de deuxième ordre.	25
1.7	Noyaux associé discrets de premier ordre (standard).	25
1.8	La forme explicite de $q(x, f)$	26
1.9	Expression du biais de l'estimateur à noyau associé discret standard.	27
2.1	Biais et variance de $\hat{f}(x)$	34
2.2	Les niveaux de dépendance	36
2.3	Densité d'innovation.	36
2.4	Les valeurs de l'ISE moyen (\overline{ISE}) basées sur 100 réplifications pour DCA-D1	38
2.5	Les valeurs de l'ISE moyen (\overline{ISE}) basées sur 100 réplifications pour DCA-D2	39
2.6	Paramètres de niveau de dépendance pour le modèle VS.	39
2.7	Les valeurs de l'ISE moyen (ISE) basées sur 100 réplifications pour VS	40
3.1	Noyaux gamma généralisés (MG, NK et W).	56
3.2	Les valeurs de C et V pour les noyaux MG, NK et W.	57
3.3	Les densités d'innovation.	59
3.4	Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour DCA-D1	59
3.5	Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour DCA-D2	60
3.6	Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour DCA-D3	61
3.7	les valeurs de ISE moyenne (ISE) basées sur 250 réplifications pour VS, le number en gras le ISE le plus petit	61
3.8	Les densités d'innovation.	61
3.9	Statistiques sommaires des données sur le choix du mode de transport	71
3.10	Données de longévité des insectes adultes observés en jours.	72

3.11 ISE×1000 moyenne des noyaux discrets, avec $(\gamma_n) = (1/n)$	72
3.12 ISE×1000 moyenne des noyaux discrets, avec $(\gamma_n) = (1/n)$	73

Liste des contributions

Dans le cadre de cette thèse, nous avons réalisé les contributions scientifiques suivantes :

1. M. KHEMICI, N. ZOUGAB, Y. ZIANE, S. ADJABI (2024). Recursive generalized gamma kernel density estimation for nonnegative dependent data. *Communications in Statistics - Simulation and Computation* : [https : Doi.org/10.1080/03610918.2024.2347930](https://doi.org/10.1080/03610918.2024.2347930). 53(6) : 2976-2987.
2. M. KHEMICI, N. ZOUGAB, C. C. KOKONENDJI, S. ADJABI (2024). Recursive discrete kernel estimator for count Data. (article soumis).
3. M. KHEMICI, N. ZOUGAB, S. ADJABI (2024). Recursive kernel probability mass function estimator for count data. Séminaire Mathématique de Béjaia organisé par l'unité des recherche LaMOS durant l'année 2023-2024.
4. M. KHEMICI, N. ZOUGAB, S. ADJABI (2022). Asymetric kernel density estimation for dependent data. Colloque International MOAD'2022 (Méthodes et Outils d'Aide à la Décision).
5. M. KHEMICI, N. ZOUGAB, S. ADJABI (2022). Gamma family kernel density estimation for α -mixing process. The First International Workshop on Applied Mathematics 1st-IWAM'2022, December 6-8, 2022, Constantine, Algeria.
6. M. KHEMICI, N. ZOUGAB, S. ADJABI (2022). Generalized Birnbaum-Saunders kernel estimators for positive time series (η -weak dependence). 2nd National Conference on Pure and Applied Mathematics NCPAM'2022, December 18 - 19, 2022, Laghouat, Algeria.
7. M. KHEMICI, N. ZOUGAB, S. ADJABI (2022). Estimation par noyau asymétrique récursif. Séminaire Mathématique de Béjaia organisé par l'unité des recherche LaMOS durant l'année 2021-2022.

Introduction générale

L'estimation d'une densité de probabilité sous-jacente à un ensemble fini d'observations est un problème fondamental en statistiques qui a fait l'objet d'une vaste littérature. On retrouve cette problématique dans de nombreux domaines des sciences et techniques tels que le traitement du signal et des images, la mécanique, la robotique, etc. Il existe plusieurs méthodes d'estimation non paramétrique de la fonction densité : Méthode d'estimation par histogramme, les propriétés des estimateurs par histogramme ont été étudiées par Geffroy (1974)[20] et Abou-Jouadé (1976)[65], la méthode d'estimation par les séries orthogonales (développement en séries de Fourier) proposées pour estimer les densités continues. Cette méthode a été développée par Cencov (1962)[14] et étudiée par la suite par plusieurs auteurs Schwartz (1967)[68], Kronmal and Tarter (1968)[46], Wahba (1981)[80] et récemment par Saadi and Adjabi (2009)[66]. On trouve aussi la méthode d'interpolation par les fonctions splines la méthode basée sur les développements en ondelettes proposée par Walter (1992)[81] et Kerkyacharian and Picard (1993)[40]. La méthode qui a rencontré le plus de succès est la méthode du noyau introduite par Rosenblatt (1956)[61] et Parzen (1962)[55]. Ce succès rencontré par l'estimateur à noyau auprès de la communauté des utilisateurs peut essentiellement s'expliquer par : la simplicité de l'expression théorique de l'estimateur, sa convergence dans différents sens et sa flexibilité dans le choix du paramètre de lissage. La théorie de l'estimation constitue l'une des préoccupations centrales des statisticiens.

L'estimation de la densité, par exemple, est un domaine qui a suscité de nombreuses recherches en raison de son application dans divers domaines tels que l'analyse de la régression des séries chronologiques et la théorie de la fiabilité. En effet, dans de nombreux secteurs tels que la médecine, l'économie, la sociologie, l'environnement, les finances et le marketing, des phénomènes complexes se manifestent, difficiles à modéliser. Ces phénomènes sont souvent décrits par une variable aléatoire réelle (v.a.r) inconnue, d'où la nécessité d'estimer sa loi de probabilité. L'estimation récursive de la densité de probabilité d'une séquence des variables aléatoires de même loi (qui ne sont pas nécessairement indépendantes) se réalise à travers un processus à temps discret, noté $(X_t, t \in N)$. Dans ce contexte, les X_t représentent des vecteurs aléatoire dans \mathbb{R}^d suivant la même loi, caracté-

risés par une densité de probabilité inconnue que l'on notera f . Pour élaborer des estimateurs non paramétriques d'une distribution de densité et d'une fonction de régression, il est possible d'utiliser des méthodes d'approximation stochastique. L'application la plus renommée des algorithmes stochastiques dans le domaine des statistiques non paramétriques est mise en évidence dans le travail de Kiefer et Wolfowitz (1952)[41]. Les méthodes d'approximation stochastique permettent de rechercher le zéro z^* d'une fonction inconnue $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, qui peut être difficilement calculé directement. L'algorithme le plus célèbre à cet égard est celui de Robbins et Monro (1951)[60].

Soit (X_1, \dots, X_n) un échantillon de densité de probabilité f . Pour construire un estimateur récursif de f en un point x par la méthode des algorithmes stochastiques, Mokkadem et al.(2009)[52], ont défini un algorithme de recherche du zéro de la fonction $\varphi : y \rightarrow f(x) - y$.

L'estimation non paramétrique de la densité du noyau avec une version non récursive est largement utilisée dans de nombreux domaines d'application, tels que l'économie, l'ingénierie, la biologie, les sciences de l'environnement, etc. Tout comme l'approche non récursive, les estimateurs récursifs de densité du noyau et de régression pour les cas indépendants et non indépendants ont également été étudiés par plusieurs auteurs dans la littérature statistique. Le grand avantage des estimateurs récursifs par rapport à leurs homologues non récursifs est d'ordre informatique, puisqu'ils nécessitent beaucoup moins de calculs pour passer d'un échantillon de taille n à un échantillon de taille $n + 1$. Cette caractéristique est particulièrement cruciale dans l'estimation de la densité, où le nombre de points auxquels la fonction est estimée est généralement élevé. L'estimateur récursif de densité a été introduit à l'origine par Wolverton et Wagner (1969)[85] dans le cas d'un noyau symétrique. Selon la version de Wolverton et Wagner, plusieurs estimateurs récursifs ont été développés par Deheuvels (1974)[19], Wegman et Davies (1979)[84] et Hall et Patil (1994)[31]. Plus tard, une grande classe d'estimateurs récursifs à noyau symétrique a été proposée par Mokkadem et al. (2009)[52] à l'aide de la méthode d'approximation stochastique pour définir une classe d'estimateurs récursifs de la densité du noyau, qui comprend celui introduit par Hall et Patil (1994)[31]. L'utilisation la plus célèbre des algorithmes d'approximation stochastique dans le cadre des statistiques non paramétriques est le travail de Kiefer et Wolfowitz (1952)[41].

L'objectif de cette thèse est d'estimer les fonctions de densité et de masse de probabilité dans le cas des données indépendantes et dépendantes à support non négatif et discret, en utilisant la méthode du noyau, basée sur des estimateurs non récursifs et récursifs. Nous proposons également la méthode de validation croisée pour la sélection du paramètre de lissage.

Dans le cas continu, nous développons un estimateur récursif à noyau gamma généralisé (GG) pour les données dépendantes positives d'un processus stationnaire α mélange. Nous établissons les propriétés biais, variance ainsi que l'erreur quadratique moyenne intégrée (MISE) en tant que propriété globale. Une étude de simulation compare les performances de l'estimateur récursif gamma généralisé(GG) à celles de l'estimateur non récursif, en utilisant des données générées à partir d'un

modèle de durée conditionnelle autorégressive (DCA) et d'un modèle de volatilité stochastique (SV). Deux applications sur des séries temporelles de données positives de l'indice des prix à la consommation et du trafic de passagers sont également présentées par Gouriéroux et al. (1990) [29].

Dans le cas discret, nous proposons un estimateur récursif à noyau discret pour estimer des données de comptage. Nous établissons les propriétés biais, variance ainsi que l'erreur quadratique moyenne intégrée (MISE) en tant que mesure globale. Une étude de simulation compare les performances des estimateurs à noyau discret récursif à ceux des estimateurs non récursifs, utilisant des données générées suivant une densité de probabilité discrète. Deux applications sur des données réelles, concernant le choix du mode de déplacement, sont également présentées par Baker et Greene (2011)[4]. et la longévité des insectes adultes observée en jours, sont également présentées par Senga Kiessé et Mizère (2013)[42].

Cette thèse est composée d'une introduction générale, de deux parties, d'une conclusion générale et une liste de références bibliographiques. La première partie concerne les estimateurs non récursifs (classiques) se divise en deux chapitres.

- Le premier chapitre concerne l'estimation non paramétrique de la fonction de densité (continu, discret) basée sur l'estimateur de type noyau associé. Nous présentons la définition des estimateurs à noyaux associés, puis nous introduisons le principe de construction des noyaux associés. Ensuite, nous présentons les différentes propriétés des estimateurs à noyaux associés.

- Le deuxième chapitre est consacré à l'estimation non paramétrique de la fonction de densité (continu) basée sur des données dépendantes positives. Nous proposons les noyaux asymétriques pour l'estimation de densité dans le contexte des données dépendantes positives, nous déduisons la propriété locale (MSE) pour les données dépendantes d'un processus stationnaire α mélange. Deuxièmement, une étude de Monte-Carlo basée sur des données dépendantes d'un modèle (DCA) et d'un modèle (SV).

- Le troisième chapitre, portant sur l'estimation récursive de la densité de probabilité, aborde tout d'abord les estimateurs récursifs à noyaux symétriques et asymétriques dans le cas des données indépendantes. Nous présentons les estimateurs récursifs, ainsi que leurs propriétés, et nous présentons la méthode d'approximation stochastique pour définir une classe d'estimateurs récursifs de densité du noyaux, et nous exposons les propriétés de cette classe. Dans un deuxième temps, nous nous penchons sur l'estimateur récursif à noyau discret. Nous proposons un estimateur récursif à noyau discret adapté aux données de comptage. Nous déduisons la propriété globale (MISE) pour les données de comptage et réalisons une étude de Monte-Carlo basée sur des données de comptage, comparant ainsi leurs performances avec celles des estimateurs classiques. En troisième lieu,

l'attention se porte sur l'estimateur récursif à noyau asymétrique. Nous proposons un estimateur récursif à noyau asymétrique adapté aux données dépendantes. Nous déduisons la propriété globale (MISE) pour les données dépendantes issues d'un processus stationnaire α mélange. Deuxièmement, une étude de Monte-Carlo basée sur des données dépendantes d'un modèle (DCA) et d'un modèle (SV) est réalisée, comparant ainsi leurs performances avec celles des estimateurs classiques. Nous terminerons ce mémoire par une conclusion générale, quelques perspectives et une liste de références bibliographiques.

Estimation non paramétrique de la fonction de densité par la méthode du noyau dans le cas des données indépendantes

Sommaire

1.1	Introduction	6
1.2	Noyau continu symétrique	6
1.3	Cas continu	8
1.3.1	Estimateur à noyau associé continu	8
1.3.2	Construction du noyau associé	9
1.3.3	Estimateur à noyau asymétrique de la densité de probabilité	11
1.3.4	Propriétés des estimateurs à noyaux asymétriques	15
1.3.5	Convergence en moyenne quadratique (MSE) et en moyenne quadratique intégrée (MISE)	17
1.3.6	Choix du paramètre de lissage	18
1.4	Cas discret	24
1.4.1	Noyaux associés discrets	24
1.4.2	Propriétés de l'estimateur à noyau discret	25
1.4.3	Convergence en moyenne quadratique et en moyenne quadratique intégrée	27
1.5	Conclusion	28

1.1 Introduction

Dans la littérature, la plupart des estimateurs de la densité par la méthode du noyau se résument à l'utilisation des noyaux symétriques (classiques) connus : noyau optimal d'Epanechnov, noyau gaussien, etc.

Cependant, lorsque il s'agit de densité à support borné ou semi-borné, l'utilisation des noyaux symétriques pose un sérieux problème aux bords (biais aux bords) c'est à dire, ils assignent un poids à l'extérieur du support lorsque, le lissage est pris en compte près du bord. Ce problème engendre des estimateurs non consistants. Pour cette raison, plusieurs auteurs Chen (1999)[15], Scaillet (2004)[67], Jin et Kawczak (2003)[38] et récemment Marchant et al. (2013)[50] et Makhouloufi et al. (2021)[47] ont proposé une famille des noyaux asymétriques pour remédier à ce problème. L'utilisation de ces noyaux a permis d'augmenter considérablement la précision de l'estimateur de la densité à support $[0, \infty[$.

Dans ce chapitre, nous présentons les estimateurs à noyaux associés continus (symétriques, asymétriques), et les noyaux discrets. Nous présentons la définition des estimateurs à noyaux associés, puis nous introduisons le principe de construction des noyaux associé. Ensuite, nous présentons les différentes propriétés des estimateurs à noyaux associés. On termine par les méthodes de choix du paramètre de lissage.

1.2 Noyau continu symétrique

Dans cette partie, nous présentons l'estimateur à noyau continu symétrique. Ensuite, nous présentons également les différentes propriétés relatives à cet estimateur : biais, variance et erreur quadratique moyenne intégrée.

Définition 1.2.1. Soit X_1, \dots, X_n un échantillon des variables aléatoires i.i.d. De la fonction de densité inconnue f . L'estimateur non paramétrique de la fonction de densité f introduit par (Rosenblatt en (1956) [54] et Parzen (1962)[55]) est défini par :

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

où h est appelé paramètre de lissage, il joue le rôle du paramètre de dispersion autour de la cible x , et K est une fonction définie sur \mathbb{R} appelée noyau si vérifie les conditions suivantes :

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

TABLEAU 1.1 – Exemples des noyaux continus symétriques.

Noyau	Fonction noyau	Domaine de définition
Cauchy	$K(u) = [\pi(1 + u^2)]^{-1}$	\mathbb{R} ,
Biweight	$K(u) = (15/16) * (1 - u^2)^2$	$[-1, 1]$,
Triangulaire continu	$K(u) = 1 - u $	$[-1, 1]$,
Epanechnikov	$K(u) = (3/4) * (1 - u^2)$	$[-1, 1]$,
Gaussien	$K(u) = (1/\sqrt{2\pi}) * \exp(u^2/2)$	\mathbb{R} .

Propriétés de l'estimateur à noyau symétrique

Dans cette partie, nous présentons les propriétés fondamentales de l'estimateur à noyau symétrique, biais, variance et MISE de l'estimateur $\hat{f}(x)$.

Proposition 1.2.1. (Parzen (1962)[55]). Soit x fixé dans \mathbb{R} . Le biais de l'estimateur $\hat{f}(x)$ est :

$$\text{Biais}(\hat{f}_h(x)) = \frac{1}{2}h^2 f''(x) \int_{\mathbb{R}} t^2 K(t) dt + o(h^2). \quad (1.2)$$

Proposition 1.2.2. (Parzen (1962)[55]). Soit x fixé dans \mathbb{R} . La variance de l'estimateur $\hat{f}(x)$ est :

$$\text{Var}(\hat{f}_h(x)) = \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2(t) dt + o\left(\frac{1}{nh}\right).$$

En utilisant les expressions finales des termes : le biais, la variance et l'erreur quadratique moyenne (*MSE*) en un point x est donnée par :

$$\begin{aligned} \text{MSE}(\hat{f}_h(x)) &= \text{Biais}^2(\hat{f}_h(x)) + \text{Var}(\hat{f}_h(x)), \\ &= \frac{1}{nh} f(x) \int_{\mathbb{R}} K^2(t) dt + \frac{h^4}{4} \{f''(x)\}^2 \left[\int_{\mathbb{R}} t^2 K(t) dt \right]^2 + o\left(\frac{1}{nh} + h^4\right). \end{aligned}$$

L'erreur quadratique moyenne intégrée (*MISE*) est :

$$\begin{aligned} \text{MISE}(\hat{f}_h) &= \int_{\mathbb{R}} \text{MSE}(\hat{f}_h(x)) dx, \\ &= \frac{1}{nh} \int_{\mathbb{R}} K^2(t) dt + \frac{h^4}{4} \int_{\mathbb{R}} \{f''(x)\}^2 dx \left[\int_{\mathbb{R}} t^2 K(t) dt \right]^2 + o\left(\frac{1}{nh} + h^4\right). \end{aligned}$$

Théorème 1.2.1. (Parzen (1962)[55]). Si $\lim_{n \rightarrow \infty} h = 0$, $\lim_{n \rightarrow \infty} nh = \infty$, si le noyau K satisfait les conditions suivantes

$$\int K(y) = 1, \quad \sup |K(y)| < \infty \quad \text{et} \quad \int |K(y)| dy < \infty,$$

alors l'estimateur $\hat{f}(x)$ est consistant en moyenne quadratique, en tout point x pour lequel la densité f est continue, c'est-à-dire $\lim_{n \rightarrow \infty} \text{MSE}\{\hat{f}_h(x)\} = 0$.

1.3 Cas continu

1.3.1 Estimateur à noyau associé continu

Kokonendji et Senga Kiessé (2007) [45] ont introduit la notion d'un noyau associé $K_{x,h}$ de cible x et de paramètre de lissage h dans le cas discret, et l'ont étendue récemment Kokonendji et libengué (2013)[44] au cas continu.

Définition 1.3.1. Soit \mathbb{S} le support d'une fonction de densité f à estimer. Étant donné $x \in \mathbb{S}$ et $h > 0$, on appelle **noyau associé continu** $K_{x,h}$ toute fonction de densité liée à la variable aléatoire $\kappa_{x,h}$ de support $\mathbb{S}_{x,h}$ contenant au moins x , vérifiant les quatre conditions suivantes :

$$\cup_{x \in \mathbb{S}} \mathbb{S}_{x,h} \supseteq \mathbb{S}, \quad (1.3)$$

$$E(\kappa_{x,h}) \sim x, \quad \text{lorsque } h \rightarrow 0,$$

$$\mathbf{Var}(\kappa_{x,h}) < \infty,$$

$$\mathbf{Var}(\kappa_{x,h}) \rightarrow 0, \quad \text{lorsque } h \rightarrow 0.$$

L'estimateur à noyau associé continu \hat{f} de f sur $\mathbb{S} \subseteq \mathbb{R}$ est alors défini par :

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad x \in \mathbb{S}. \quad (1.4)$$

La cible x et le paramètre de lissage h interviennent dans la définition du noyau associé $K_{x,h}(X_i)$.

Remarque 1.3.1. L'estimateur de la densité f à support non borné avec un noyau K symétrique. Le noyau associé s'écrit $K_{x,h}(\cdot) = (\frac{1}{h})K((x - \cdot)/h)$ dans le cas des noyaux symétriques. Lorsqu'il s'agit des noyaux associés asymétriques par rapport à la cible x cette écriture n'est pas possible, c'est-à-dire le passage de l'estimateur (1.4) à l'estimateur (1.1) n'est pas toujours possible.

Les noyaux associés continus englobent les noyaux symétriques (dit aussi classiques) et les noyaux associés asymétriques (dit aussi non classiques). Dans ce travail, on s'intéresse aux noyaux associés asymétriques.

Notons que si l'on dispose d'un noyau associé, l'estimateur correspondant peut être facilement déduit. Si tel n'est pas le cas, il est possible de construire un noyau associé. Voir la thèse de Libengué (2013)[21] qui a proposé une technique de construction des noyaux associés continus dite

"**mode-dispersion**" dont le principe est de mettre la cible x sur le mode et le paramètre de lissage h sur le paramètre de dispersion.

1.3.2 Construction du noyau associé

Les noyaux associés peuvent se construire à partir de n'importe quel type de noyau représentant une densité de probabilité à deux paramètres ayant un paramètre de **dispersion**. Kokonendji et Libingué (2011)[44], Libingué (2013)[21] ont proposé récemment une technique de construction des noyaux associés continus dits **mode-dispersion**, où le principe est de mettre la cible x sur le mode et le paramètre de lissage h sur le paramètre de dispersion. Pour construire un noyau associé continu $K_{x,h}$, la densité de probabilité à partir de laquelle on construit ce noyau doit vérifier certaines conditions.

Définition 1.3.2. Le type de noyau continu K est une densité de probabilité à carré intégrable $K = k_\theta$ dépendant du paramètre $\theta \in \Theta \subseteq \mathbb{R}^2$ et de support $\mathbb{S}_K = \mathbb{S}_\theta$.

Si on considère le type de noyau k_θ unimodal de mode M_θ et ayant un paramètre de dispersion D_θ . Étant donné Θ bidimensionnel, on retient la notation suivante : $\theta = \theta(\alpha, \beta)$, $M_\theta = M(\alpha, \beta)$ et $D = D(\alpha, \beta)$ avec deux réels positifs.

Principe mode dispersion

Soit $K_{\theta(\alpha,\beta)}$ un type de noyau continu de mode $M(\alpha, \beta)$ et de mesure de dispersion $D(\alpha, \beta)$. On construit un noyau associé $K_{\theta(x,h)}$ à partir de $K_{\theta(\alpha,\beta)}$ en résolvant le système :

$$\begin{cases} M(\alpha, \beta) = x, \\ D(\alpha, \beta) = h. \end{cases} \quad (1.5)$$

Cela conduit à $\theta(x, h) = \theta(\alpha(x, h), \beta(x, h))$, où $\alpha(x, h)$ et $\beta(x, h)$ sont des solutions du système (1.5). Il existe d'autres méthodes de construction des noyaux associés, voir par exemple, Jin et al. (2003)[38] pour les noyaux Birnbaum-Saunders (BS) et Log-Normal(LN).

Exemple de construction des noyaux associés

Donnons maintenant quelques exemples illustrant la construction des noyaux associés asymétriques (non classiques) :

Noyau associé log-normal Jin et al. (2003)[38]

La densité de probabilité de la distribution log-normal est :

$$f_{LN(\mu, \sigma^2)} = \frac{1}{\sqrt{2\pi\sigma y}} \exp\left(-\frac{1}{2\sigma}(\log(y) - \mu)^2\right), \quad (1.6)$$

où $y > 0$ et $\sigma > 0$. Son mode et son paramètre de dispersion sont $M(\mu, \sigma) = e^{\mu - \sigma^2}$ et $D(\mu, \sigma) = \sigma$. En se basant sur le principe mode-dispersion suivant :

$$\begin{cases} M(\mu, \sigma) = e^{\mu - \sigma^2} = x, \\ D(\mu, \sigma) = \sigma = h, \end{cases} \quad (1.7)$$

alors :

$$\begin{cases} \mu = \log(x) + h^2, \\ \sigma = h. \end{cases} \quad (1.8)$$

Le noyau associé log-normal sera alors défini comme suit :

$$K_{(\log(x)+h^2, h)}(y) = \frac{1}{\sqrt{2\pi}hy} \exp\left(\frac{1}{2h}(\log(y) - \log(x) + \sigma^2)^2\right), \quad y \geq 0. \quad (1.9)$$

Noyau associé Birnbaum-Saunders (BS) Jin et al. (2003)[38]

La densité de probabilité de la distribution Birnbaum-Saunders(BS) est :

$$f_{BS(\alpha, \beta)}(y) = \frac{1}{2\alpha} \left(\sqrt{\frac{1}{\beta y}} + \sqrt{\frac{\beta}{y^3}} \right) \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-1}{2\alpha^2} \left(\frac{y}{\beta} + \frac{\beta}{y} - 2\right)\right). \quad (1.10)$$

où : $y > 0$, $\alpha > 0$ et $\beta > 0$. En faisant le changement des variables suivants $\alpha = \sqrt{h}$ et $\beta = x$ on obtient le noyau associé BS donné par :

$$K_{BS(\sqrt{h}, x)}(y) = \frac{1}{2\sqrt{h}} \left(\sqrt{\frac{1}{xy}} + \sqrt{\frac{x}{y^3}} \right) \sqrt{2\pi} \exp\left(\frac{-1}{2\alpha^2} \left(\frac{y}{x} + \frac{x}{y} - 2\right)\right), \quad y > 0, h > 0 \text{ et } x > 0. \quad (1.11)$$

Noyau associé Gamma Chen (2000)[16]

La densité de probabilité de la distribution gamma est :

$$f(y) = \frac{1}{\Gamma(\alpha)} \frac{y^{\alpha-1} e^{-\frac{y}{\beta}}}{\beta^\alpha}, \quad (1.12)$$

avec $y > 0$, $\alpha > 0$ et $\beta > 0$ sont les paramètres de la loi gamma. La moyenne, la variance ainsi que le mode de la loi gamma sont donnés respectivement par :

$$E(X) = \alpha\beta, \quad Var(X) = \alpha\beta^2, \quad Mod(X) = (\alpha - 1)\beta.$$

En faisant le changement de variable suivant :

$$\begin{cases} Mod(X) = (\alpha - 1)\beta = x, \\ \beta = h, \text{ alors } h \neq \sigma(x) = \beta\sqrt{\alpha}. \end{cases} \quad (1.13)$$

alors :

$$\begin{cases} \alpha = \frac{x}{h} + 1, \\ \beta = h. \end{cases} \quad (1.14)$$

Le noyau associé gamma est :

$$K_{Gam(\frac{x}{h}, h)}(y) = \frac{1}{\Gamma(\frac{x}{h} + 1)} \frac{y^{\frac{x}{h}} e^{-\frac{y}{h}}}{h^{\frac{x}{h} + 1}}, \quad y > 0, h > 0, x > 0. \quad (1.15)$$

1.3.3 Estimateur à noyau asymétrique de la densité de probabilité

La définition suivante présente les notions de noyau asymétrique et d'estimateur à noyau asymétrique pour une fonction de densité f inconnue sur le support $[0, \infty[$.

Définition 1.3.3. : Soit X_1, \dots, X_n une suite de variables aléatoires continue i.i.d. Avec une fonction de densité inconnue f . L'estimateur à noyau asymétrique pour des données indépendantes est :

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad (1.16)$$

où : h est le paramètre de lissage et $K_{x,h}$ est un noyau asymétrique continu de cible x et du paramètre de lissage h .

1.3.3.1 Estimateur de la densité à noyau associé bêta

Le noyau bêta a été introduit par Chen (1999)[15] pour estimer des densités à support $[0, 1]$. Il utilise la loi bêta pour construire des noyaux associés continus asymétriques. Deux classes du noyau bêta ont été proposées.

Classe 1 : noyau bêta 1 Chen (1999) [15]

La première version du noyau bêta est sous la forme :

$$K_{B(x,h)}(u) = \frac{u^{\frac{h}{x}}(1-u)^{\frac{(1-x)}{h}}}{B\left\{\frac{h}{x+1}, \frac{(1-x)}{h+1}\right\}} \mathbb{I}_{u \in [0,1]}. \quad (1.17)$$

Si les conditions $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$ sont vérifiés. L'estimateur de la densité à noyau bêta est :

$$\widehat{f}_1(x) = \frac{1}{n} \sum_{i=1}^n K_{B\left(\frac{x}{h+1}, \frac{1-x}{h+1}\right)}(X_i). \quad (1.18)$$

L'erreur quadratique moyenne intégrée (MISE) pour des données indépendantes du noyau bêta 1 est :

$$\begin{aligned} MISE(\widehat{f}_1) &= \int_0^1 \text{Biais}^2(\widehat{f}_1(x)) dx + \int_0^1 \text{Var}(\widehat{f}_1(x)) dx, \\ &= h^2 \int_0^1 \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right)^2 dx, \\ &+ \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^1 (x(1-x))^{-\frac{1}{2}} f(x) dx + o(n^{-1} h^{-\frac{1}{2}} + h^2). \end{aligned} \quad (1.19)$$

Le paramètre de lissage optimal qui minimise le MISE est :

$$h_1^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^1 (x(1-x))^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{4^{\frac{2}{5}} \left[\int_0^1 \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right)^2 dx \right]^{\frac{2}{5}}} n^{-\frac{2}{5}}. \quad (1.20)$$

En remplaçant l'équation (1.20) dans l'équation (1.19), le MISE optimal est :

$$\begin{aligned} MISE^*(\hat{f}_1) &= \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^1 \frac{f(x)}{(x(1-x))^{\frac{1}{2}}} dx \right]^{\frac{4}{5}} \\ &\quad \times \left[\int_0^1 \left((1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x) \right)^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \end{aligned} \quad (1.21)$$

Classe 2 : noyau bêta 2 Chen (1999)[15]

Si les conditions $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$ sont vérifiées. L'estimateur de la densité à noyau bêta est :

$$\hat{f}_2(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}^*(X_i), \quad (1.22)$$

$$K_{x,h}^*(y) = \begin{cases} K_{\rho(x), \frac{1-x}{h}}(y) & \text{pour } y \in [0, 2h[, \\ K_{\frac{x}{h}, \frac{1-x}{h}}(y) & \text{pour } y \in [2h, 1-2h[, \\ K_{\frac{x}{h}, \rho(1-x)}(y) & \text{pour } y \in [1-2h, 1], \end{cases}$$

$$\text{où } \rho(x, h) = 2h^2 + 2.5 - \sqrt{4h^4 + 6h^2 + 2.5 - x^2 - \frac{x}{h}}.$$

L'erreur quadratique moyenne intégrée (MISE) pour des données indépendantes du noyau bêta 2 est :

$$\begin{aligned} MISE(\hat{f}_2) &= \int_0^1 \text{Biais}^2(\hat{f}_2(x)) + \int_0^1 \text{Var}(\hat{f}_2(x)), \\ &= \frac{1}{4} h^2 \int_0^1 (x(1-x)f''(x))^2 dx, \\ &\quad + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^1 (x(1-x))^{-\frac{1}{2}} f(x) dx + o(n^{-1} h^{-\frac{1}{2}} + h^2). \end{aligned} \quad (1.23)$$

Le paramètre de lissage optimal qui minimise le MISE est :

$$h_2^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^1 (x(1-x))^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{\left[\int_0^1 (x(1-x)f''(x))^2 dx \right]^{\frac{2}{5}}} n^{-\frac{2}{5}}. \quad (1.24)$$

En remplaçant l'équation (1.24) dans l'équation (1.23), le MISE optimal est :

$$MISE^* \left(\widehat{f}_2 \right) = \frac{5}{4^{3/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^1 (x(1-x))^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \times \left[\int_0^1 (x(1-x)f''(x))^2 dx \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \quad (1.25)$$

1.3.3.2 Estimateur de la densité à noyau associé gamma

Le noyau gamma a été introduit par Chen (2000)[16] pour estimer des densités à support $[0, \infty[$. Il utilise la loi gamma pour construire des noyaux associés continus asymétriques. Plusieurs classes de noyau gamma ont été proposées.

Classe 1 : noyau gamma 1 Chen (2000)[16]

La première version du noyau gamma est :

$$K_{Gam(\frac{x}{h}, x)}(u) = \frac{u^{\frac{x}{h}} \exp^{-\frac{u}{h}}}{h^{\frac{x}{h}} \Gamma(\frac{x}{h} + 1)}, \quad (1.26)$$

où $\Gamma(\alpha) = \int_0^\infty \exp(-t)t^{\alpha-1}$ est la fonction gamma et h est le paramètre de lissage satisfaisant les conditions $h \rightarrow 0$ et $nh \rightarrow \infty$ quand $n \rightarrow \infty$. L'estimateur à noyau gamma est :

$$\widehat{f}_{Gam}(x) = \frac{1}{n} \sum_{i=1}^n K_{Gam(\frac{x}{h}, x)}(X_i). \quad (1.27)$$

L'erreur quadratique moyenne intégrée (MISE) pour des données indépendantes du noyau gamma 1 est de la forme :

$$\begin{aligned} MISE \left(\widehat{f}_{Gam} \right) &= \int \text{Biais}^2 \left(\widehat{f}_{Gam}(x) \right) + \int \text{Var} \left(\widehat{f}_{Gam}(x) \right), \\ &= h^2 \int_0^\infty (xf'(x) + \frac{1}{2}xf''(x))^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx. \end{aligned} \quad (1.28)$$

Le paramètre de lissage optimal qui minimise le (MISE) est :

$$h_{Gam}^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{4^{\frac{2}{5}} \left[\int_0^\infty (xf'(x) + \frac{1}{2}xf''(x))^2 \right]^{\frac{2}{5}}} n^{-\frac{2}{5}}. \quad (1.29)$$

En remplaçant l'équation (1.29) dans l'équation (1.28), le MISE optimal est :

$$MISE^* \left(\widehat{f}_{Gam} \right) = \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \times \left[\int_0^\infty \left(xf'(x) + \frac{1}{2}f''(x) \right)^2 \right] n^{-\frac{4}{5}}. \quad (1.30)$$

Classe 2 : noyau gamma 2 Chen (2000)[16]

Cette classe de noyau gamma a été proposée pour améliorer les performances (réduire le biais) de l'estimateur à noyau associé gamma 1. La forme du noyau gamma modifié est :

$$K_{(\rho h(x), h)}(u) = \frac{u^{\frac{x}{h}} \exp\left(-\frac{u}{h}\right)}{h^{\rho h(x)} \Gamma(\rho h(x))}. \quad (1.31)$$

Où :

$$\rho h(x) = \begin{cases} \frac{1}{4} \left(\frac{x}{h}\right)^{1/4} + 1 & \text{si } x \in [0, 2h], \\ \frac{x}{h} & \text{si } x \geq 2h. \end{cases}$$

L'estimateur à noyau gamma modifié est :

$$\hat{f}_{Gam-mod}(x) = \frac{1}{n} \sum_{i=1}^n K_{Gam-mod(\rho h(x), h)}(X_i). \quad (1.32)$$

Le MISE est :

$$\begin{aligned} MISE\left(\hat{f}_{Gam-mod}\right) &= \int_0^\infty \text{Biais}^2\left(\hat{f}_{Gam-mod}(x)\right) dx + \int_0^\infty \text{Var}\left(\hat{f}_{Gam-mod}(x)\right) dx, \\ &= \frac{1}{4} h^2 \int_0^\infty (x f'(x))^2 dx + \frac{1}{2\sqrt{\pi}} n^{-1} h^{-\frac{1}{2}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx. \end{aligned} \quad (1.33)$$

Le paramètre de lissage optimal qui minimise le MISE est :

$$h_{Gam-mod}^* = \frac{\left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{2}{5}}}{\left[\int_0^\infty (x f''(x))^2 dx \right]^{\frac{2}{5}}} n^{\frac{-2}{5}}. \quad (1.34)$$

En remplaçant l'équation (1.34) dans l'équation (1.33), le MISE optimal est :

$$MISE^*\left(\hat{f}_{Gam-mod}\right) = \frac{5}{4^{4/5}} \left[\frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^{\frac{4}{5}} \left[\int_0^\infty (x f''(x))^2 dx \right]^{\frac{1}{5}} n^{\frac{-4}{5}}. \quad (1.35)$$

Classe 3 : noyau gamma modifié raffiné

Malec et Schienl (2014)[48] ont introduit deux autres versions du noyau gamma qui sont : noyau gamma modifié 1 et gamma modifié 2.

Les deux versions de gamma modifié raffiné se distinguent du noyau gamma modifié uniquement par leur paramètre $\rho h_1(x)$. La première version de gamma modifié raffiné est paramétrée par :

$$\rho h_1(x) = \begin{cases} \left[\frac{1}{4} \left(\frac{x}{hc}\right)^2 + 1 \right] + [c + 2h(1 - c)] & \text{si } 0 \leq x < 2hc; \\ \frac{x}{hc}(c + 2h - x) & \text{si } 2hc \leq x < 2h; \\ \frac{x}{h} & \text{si } x \geq 2h \quad , 0 < c < 1. \end{cases}$$

La seconde version est paramétrée par :

$$\rho h_2(x) = \begin{cases} \frac{1}{4} \left(\frac{x}{hc}\right)^2 + 1 & \text{si } 0 \leq x < 2hc; \\ \frac{x}{hc} & \text{si } x \geq 2h, \end{cases}$$

où $c \in]0, 1]$ et $h > 0$.

Remarque 1.3.2. Si $c = 1$, on retrouve la paramétrisation originale de chen (2000)[16] dans les deux cas.

Le tableau 1.2 présente quelques noyaux associés continus. On peut se référer aux travaux de Chen (2000) [16], Jin et al. (2003)[38], Scaillet (2004)[67], Hirukawa et Sakudo (2015)[35] et Makhoulfi et al. (2021)[47] pour plus de détails sur ces noyaux ainsi que leurs applications.

TABLEAU 1.2 – Quelques noyaux associés continus asymétrique.

Noyaux	Domaine	Fonction noyau
Bêta (B)	$[0, 1]$	$K(y) = \frac{y^{x/h}(1-y)^{(1-x)/h}}{\mathbf{B}(1+x/h, 1+(1-x)/h)},$
Gamma (G)	$[0, \infty[$	$K(y) = \frac{1}{\Gamma(\frac{x}{h}+1)} \frac{y^{\frac{x}{h}} e^{-y/h}}{h^{\frac{x}{h}+1}},$
Gamma modifié (MG)	$[0, \infty[$	$K(y) = \frac{y^{\frac{x}{h}} \exp^{-\frac{y}{h}}}{h^{\rho h(x)} \Gamma(\rho h(x))}$
Réciproque-Inverse-Gaussien (RIG)	$[0, \infty[$	$K(y) = \frac{1}{\sqrt{2\pi h y}} \exp\left(-\frac{x-h}{2h} \left[\frac{y}{x-h} - 2 + \frac{x-h}{y}\right]\right),$
Lognormal (LN)	$[0, \infty[$	$K(y) = \frac{1}{\sqrt{2\pi h y}} \exp\left(-\frac{1}{2h} (\log(y) - \log(x) + \sigma^2)^2\right),$
Inverse Gaussien (IG)	$[0, \infty[$	$K(y) = \frac{1}{\sqrt{2\pi h y^3}} \exp\left(\frac{-1}{2hx} \left[\frac{y}{x} - 2 + \frac{x}{y}\right]\right),$
Birnbaum-Saunders	$[0, \infty[$	$K(y) = \frac{1}{2\sqrt{h}} \left(\sqrt{\frac{1}{xy}} + \sqrt{\frac{x}{y^3}}\right) \sqrt{2\pi} \exp\left(\frac{-1}{2\alpha^2} \left(\frac{y}{x} + \frac{x}{y} - 2\right)\right),$
Gamma Généralisé (GG)	$[0, \infty[$	$K(y) = \frac{\gamma y^{\alpha-1} \exp\{-\frac{y}{\beta\Gamma(\alpha/\gamma)/\Gamma((\alpha+1)/\gamma)}\}^{\gamma}}{\{\beta\Gamma(\alpha/\gamma)/\Gamma((\alpha+1)/\gamma)\}^{\alpha}\Gamma(\alpha/\gamma)},$
Inverse Gamma (IG)	$[0, \infty[$	$K(y) = \frac{\left(\frac{x(h+x)}{h}\right)^{1+\frac{x}{h}}}{\Gamma(1+\frac{x}{h})} \left(\frac{1}{y}\right)^{2+\frac{x}{h}} \exp\left(\frac{-x(h+x)}{hy}\right),$
Birnbaum Saunders Général (GBS)	$[0, \infty[$	$K(y) = cg \left(\frac{1}{h} \left(\frac{y}{x} + \frac{x}{y} - 2\right)\right) \frac{1}{\sqrt{4h}} \left(\frac{1}{\sqrt{yx}} + \sqrt{\frac{x}{y^3}}\right),$
Log-Symétriques (LS)	$[0, \infty[$	$K_{LS(x, h^{1/2}, r)}(y) = \frac{1}{\sqrt{hy}} r \left(\frac{1}{h} (\log(y) - \log(x))^2\right),$

Où : la constante de normalisation c , et le générateur g . Sont donnés dans le tableau 1.3.

1.3.4 Propriétés des estimateurs à noyaux asymétriques

Dans cette partie, nous présentons une étude générale de propriété biais et variance des estimateur à noyaux associés asymétriques dans le cas des données indépendantes.

Le Biais de l'estimateur $\hat{f}(x)$ est donné par la formule suivante :

TABLEAU 1.3 – Distributions GBS

Distribution	c	$\mu_1(g)$	$g = g(u), u > 0$
Normal	$\frac{1}{\sqrt{2\pi}}$	1	$\exp\left(-\frac{1}{2}u\right),$
Power Exponential (PE)	$\frac{\nu}{2^{2\nu}\Gamma(\frac{1}{2\nu})}$	$\frac{2^{\frac{1}{2\nu}}\Gamma(\frac{3}{2\nu})}{\Gamma(\frac{1}{2\nu})}$	$\exp\left(-\frac{1}{2}u^\nu\right), \nu > 0,$
Student (t)	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}$	$\frac{\nu}{\nu-2}, \nu > 2$	$\left(1 + \frac{u}{\nu}\right)^{-\frac{(\nu+1)}{2}}, \nu > 0.$

1.3.4.1 Biais

Le Biais $\widehat{f}(x)$ est

$$\begin{aligned}
 \text{Biais}\left(\widehat{f}(x)\right) &= E(\widehat{f}(x)) - f(x), \\
 &= \int_{S \cap S_{x,h}} f(t)K_{x,h}(t)dt - f(x), \\
 &= E(f(\kappa_{x,h})) - f(x).
 \end{aligned} \tag{1.36}$$

Le développement de Taylor de $f(\kappa_{x,h})$ au point moyen $E(\kappa_{x,h}) = m_{x,h}$ est :

$$\begin{aligned}
 f(\kappa_{x,h}) &= f(m_{x,h}) + (\kappa_{x,h} - m_{x,h})f'(m_{x,h}) + \frac{1}{2}(\kappa_{x,h} - m_{x,h})^2 f''(m_{x,h}), \\
 \text{Biais}\left(\widehat{f}(x)\right) &= f(m_{x,h}) + \frac{1}{2}E((\kappa_{x,h} - m_{x,h})^2)f''(x) - f(x), \\
 &= f(E(\kappa_{x,h})) + \frac{1}{2}V(\kappa_{x,h})f''(x) - f(x).
 \end{aligned} \tag{1.37}$$

1.3.4.2 Variance

La Variance de $\widehat{f}(x)$ est :

$$\begin{aligned}
 \mathbf{V}\left(\widehat{f}(x)\right) &= \frac{1}{n} \int_{S \cap S_{x,h}} K_{x,h}^2(t)f(t)dt - \frac{1}{n} \left(\int_{S \cap S_{x,h}} K_{x,h}(t)f(t)dt \right)^2, \\
 &= \frac{1}{n} \int_{S \cap S_{x,h}} K_{x,h}^2(t)f(t)dt - \frac{1}{n} \left(\text{Biais}(\widehat{f}(x)) + f(x) \right)^2, \\
 \mathbf{V}\left(\widehat{f}(x)\right) &\rightarrow 0,
 \end{aligned} \tag{1.38}$$

quand

$$\frac{1}{n} \int_{S \cap S_{x,h}} K_{x,h}^2(t)f(t)dt \rightarrow 0.$$

D'une manière plus exacte, l'expression du biais et de la variance de l'estimateur (1.16) avec les différents noyaux définis dans le tableau 1.4 prend respectivement la forme suivante :

$$\text{Biais}(\widehat{f}(x)) = \phi(x, f)h + o(h),$$

$$\text{Var}(\widehat{f}(x)) = p(x, h) \frac{f(x)}{n} + o\left(\frac{1}{n}\right),$$

où la forme explicite de $\phi(x, f)$ et $p(x, h)$ pour chaque noyau est donnée du tableau 1.4.

TABLEAU 1.4 – La forme explicite de $\phi(x, f)$ et $\rho(x, h)$.

Noyaux	$\phi(x, f)$	$\rho(x, h)$
Bêta (B)	$(1 - 2x)f'(x) + \frac{1}{2}x(1 - x)f''(x)$	$\frac{1}{2\sqrt{\pi x(x-1)h}}$,
Gamma (G)	$f'(x) + \frac{1}{2}xf''(x)$	$\frac{1}{2\sqrt{\pi xh}}$,
Gamma modifié (MG)	$\frac{1}{2}xf''(x)$	$\frac{1}{2\sqrt{2\pi xh}}$,
Gaussien inverse réciproque (RIG)	$\frac{1}{2}xf''(x)$	$\frac{1}{2\sqrt{\pi xh}}$,
Lognormal (LN)	$2xf'(x) + \frac{1}{2}x^2f''(x)$	$\frac{x^{-1}}{4\sqrt{\pi h}}$,
Gaussien inverse (IG)	$\frac{1}{2}x^3f''(x)$	$\frac{x^{-3/2}}{2\sqrt{\pi h}}$,
Birnbaum-Saunders (BS)	$\frac{1}{2}xf'(x) + 2x^2f''(x)$	$\frac{x^{-1}}{\sqrt{2\pi h}}$,
Gamma généralisé (GG)	$\frac{C}{2}xf''(x)$	$\frac{1}{\sqrt{hx}}V(2)$,
Birnbaum Saunders Général (GBS)	$\frac{h}{4}u_1(g)xf'(x) + x^2f''(x)$	$\frac{c^2}{h^{\frac{1}{2}}} \frac{1}{x}$,
Log-Symétriques (LS)	$\frac{E(Z^2)}{2}\{xf'(x) + x^2f''(x)\}$	$\frac{C_{r^2}}{h^{1/2}} \frac{1}{x}$,

Dans le cas log-symétrique on a $0 < |C| < \infty$, $C_{r^2} = \int_{-\infty}^{\infty} r^2(Z^2)$ et $E(Z^2)$ est le moment de second ordre de $Z \sim S(0, 1; r)$ avec un PDF donnée.

Où c_{g^2} (voir tableau 1.3) est une constante de normalisation égale à $\frac{1}{\int_{-\infty}^{\infty} g^2(x^2)dx}$ qui est donnée pour les différents cas particulier du noyau GBS dans le tableau 1.5 :

TABLEAU 1.5 – Constantes de normalisation

Constantes de normalisation	Normal	BS-PE	BS-Student
c_{g^2}	$\frac{1}{\sqrt{\pi}}$	$2^{\frac{1}{2}}\Gamma(\frac{2\nu+1}{2\nu})\cos(\frac{\pi}{2\nu})$	$\frac{\sqrt{\nu}\Gamma(\nu)}{\sqrt{\nu}\Gamma(\frac{1+2\nu}{2})}$

1.3.5 Convergence en moyenne quadratique (MSE) et en moyenne quadratique intégrée (MISE)

Nous donnons ici la convergence au sens de l'erreur quadratique moyenne et au sens de l'erreur quadratique moyenne intégrée. Ces résultats sont donnés par Senga Kiessé (2008)[70]. Nous rappelons d'abord le développement du biais et de la variance de $\widehat{f}(x)$, par la suite nous déduisons les expressions de l'erreur quadratique moyenne et l'erreur quadratique moyenne intégrée.

L'erreur quadratique moyenne est :

$$\begin{aligned}
MSE(\widehat{f}(x)) &= \mathbf{Biais}^2(\widehat{f}(x)) + Var(\widehat{f}(x)), \\
&= \left(f(E(\kappa_{x,h})) + \frac{1}{2}V(\kappa_{x,h})f''(x) - f(x) \right)^2 \\
&\quad + \frac{1}{n} \int_{\mathbb{S} \cap \mathbb{S}_{x,h}} K_{x,h}^2(t) f(t) dt - \frac{1}{n} (\mathbf{Biais}(f(x)) + f(x))^2,
\end{aligned} \tag{1.39}$$

$$\begin{aligned}
MISE(\widehat{f}) &= \int_{\mathbb{S}} MSE(\widehat{f}(x)) dx, \\
&= \int_{\mathbb{S}} \left(f(E(\kappa_{x,h})) + \frac{1}{2}V(\kappa_{x,h})f''(x) - f(x) \right)^2 dx, \\
&\quad + \frac{1}{n} \int_{\mathbb{S}} \left(\int_{\mathbb{S} \cap \mathbb{S}_{x,h}} K_{x,h}^2(t) f(t) dt - (\mathbf{Biais}(\widehat{f}(x)) + f(x))^2 \right) dx.
\end{aligned} \tag{1.40}$$

Proposition 1.3.1. *Senga Kiessé (2008)[70]*

$\frac{1}{n} \int_{\mathbb{S} \cap \mathbb{S}_{x,h}} K_{x,h}^2(t) f(t) dt \rightarrow 0$ Quand $n \rightarrow \infty$ et $h = h(n) \rightarrow 0$ alors,

$$\lim_{n \rightarrow \infty} MISE(\widehat{f}(x)) = 0.$$

1.3.6 Choix du paramètre de lissage

Le paramètre de lissage h joue un rôle important dans l'estimation de la densité par la méthode du noyau. Lorsque le paramètre de lissage est très petit, l'estimateur est très volatil, ce qui conduit à un sous-lissage. Par contre, lorsque le paramètre de lissage est grand, l'estimateur est sur-lissé. Plusieurs méthodes existent pour le choix de h , elles peuvent être classées en deux catégories. La première, est l'approche fréquentiste dont les méthodes plug-in et les méthodes validation croisée. La seconde, est l'approche bayésienne qui regroupe, l'approche globale qui fait appel aux méthodes d'approximation MCMC et l'approche bayésienne variable (locale et adaptative).

1.3.6.1 Méthode plug-in

Les méthodes plug-in sont une approche fréquentiste pour estimer le paramètre de lissage h dans la méthode du noyau. On cherche à minimiser l'erreur quadratique moyenne intégrée donnée par

$$MISE(h) = \int (Biais^2\{\widehat{f}(x)\} + Var\{\widehat{f}(x)\}). \tag{1.41}$$

Cette approche consiste à remplacer les fonctions de densité inconnues par des estimateurs à noyau, tels que la densité de noyau gaussien, pour estimer la densité de probabilité. Pour les noyaux symétriques, la méthode plug-in utilise la règle de Silverman qui estime h en fonction de la variance de l'échantillon et de la dimension de l'espace. Cependant, pour les noyaux symétriques, tels que le noyau Epanechnikov, la règle de Silverman ne s'applique pas directement. Pour estimer h dans ce cas, il est nécessaire de recourir à d'autres méthodes plug-in, telles que la méthode de Sheather et Jones (1991)[71], qui utilisent une approximation de la dérivée du noyau pour estimer h . Cette méthode nécessite de trouver la valeur de h qui minimise la distance de Kullback-Leibler entre la vraie densité et la densité estimée par la méthode du noyau.

Dans le cas d'utilisation des noyaux associés asymétriques, le paramètre de lissage idéal au sens d'AMISE dépend généralement de trois quantités inconnues à travers f, f' et f'' . Ceci rend plus difficile le choix du paramètre de lissage par les méthodes plug-in. La seule méthode qui a été proposée dans ce cas est la méthode de règle de référence par analogie avec le cas des noyaux symétriques. Scaillet (2004)[67] suggère d'estimer les quantités inconnues en remplaçant f par un modèle de référence log-normal de paramètres μ_f et δ_f^2 . Par exemple, les paramètres de lissage obtenus par Scaillet (2004) [67] pour la règle de référence en utilisant le noyau gaussien-inverse et le noyau gaussien-inverse-réciproque sont donnés respectivement par :

$$h_{rdi}^{GI} = \left(\frac{16\delta_f^5 \exp\{\frac{1}{8}(7\delta_f^2 - 20\mu_f)\}}{12 + 68\delta_f^2 + 225\delta_f^4} \right)^{\frac{2}{5}} n^{-\frac{2}{5}},$$

$$h_{rdi}^{GIR} = \left(\frac{16\delta_f^5 \exp\{\frac{1}{8}(7\delta_f^2 - 20\mu_f)\}}{12 + 4\delta_f^2 + \delta_f^4} \right)^{\frac{2}{5}} n^{-\frac{2}{5}},$$

en pratique, on estime les deux paramètres μ_f et δ_f^2 à l'aide des observations X_1, \dots, X_n en utilisant la moyenne et la variance empiriques. Cette méthode a tendance à donner des valeurs très petites pour le paramètre de lissage, ce qui conduit à un sous-lissage. Une étude de simulation menée par Scaillet (2004)[67] montre que le choix du paramètre de lissage en utilisant les noyaux gaussien-inverse et gaussien-inverse-réciproque confirme ce défaut de cette approche.

Lorsque le noyau utilisé est asymétrique, la méthode la plus courante est la méthode de validation croisée, que nous allons présenter.

1.3.6.2 Méthodes de validation croisée

La méthode de validation croisée repose sur l'idée de trouver une fonction de score $CV(h)$ dont le calcul est plus simple que celui de $MISE(h)$. Le paramètre de lissage h sélectionné par cette méthode n'est pas déterministe et dépend des observations. Nous considérons un échantillon i.i.d. X_1, \dots, X_n et un noyau associé $K_{x,h}$. La méthode la plus couramment utilisée consiste à minimiser

un estimateur approprié de $ISE(h)$, donné par :

$$ISE(h) := \int \{\hat{f}(x) - f(x)\}^2 dx = \int \hat{f}(x)^2 dx - 2 \int \hat{f}(x)f(x) dx + \int f(x)^2 dx. \quad (1.42)$$

Puisque $\int f(x)^2 dx$ ne dépend pas du paramètre de lissage h , on peut choisir le paramètre de lissage de façon ce qu'il minimise un estimateur de :

$$CV(h) = \int \hat{f}(x)^2 dx - 2 \int \hat{f}(x)f(x) dx. \quad (1.43)$$

Ne remarquons que $\frac{1}{n} \sum_{i=1}^n f_{h,i}(X_i)$ est un estimateur sans Biais de $\int \hat{f}(x)f(x) dx$

où $\hat{f}_{h,i}(X_i)$ est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation X_i . Le critère à optimiser devient alors

$$CV(h) = \int \left\{ \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) dx - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{i \neq j} K_{X_i,h}(X_j) \right\}. \quad (1.44)$$

Cette approche, également appelée validation croisée non Biaisée, a été proposée par Rudemo (1982)[64] et Bowman (1984)[10] en utilisant des noyaux de la forme $K_{x,h}(\cdot) = \frac{1}{h} K(x - \cdot)/h$. Cette méthode est populaire en raison de sa motivation intuitive et du fait que cet estimateur est asymptotiquement optimal sous certaines conditions. L'optimalité asymptotique de la validation croisée non Biaisée a été obtenue par Charles et Stone (1984)[75]. Cependant, cette méthode présente deux problèmes majeurs. Elle est très variable d'un échantillon à l'autre et présente souvent plusieurs minimums locaux. Deux autres approches ont été proposées pour remédier aux problèmes de.

Validation croisée non biaisée : l'approche de validation croisée biaisée proposée par David et al. (1987)[69] et l'approche de validation croisée lissée proposée par Hall et al. (1992)[30], qui est considérée comme une combinaison de la méthode de validation croisée non biaisée et de la méthode de validation croisée biaisée. Cependant, ces deux méthodes souffrent du problème de variabilité. Nous présentons un bref aperçu de ces deux méthodes.

Validation croisée biaisée : Le critère de validation croisée biaisée a été introduit par Scott et Terrell (1987)[69] pour résoudre les problèmes de validation croisée non biaisée. Le paramètre de lissage basé sur la méthode de validation croisée biaisée est la valeur h_b qui minimise un estimateur de AMISE. Scott et Terrell (1987)[69] proposent d'estimer $\int f''^2$ en utilisant $\int \hat{f}''^2 - \int \frac{K''^2}{(nh^5)}$, où \hat{f}''^2 désigne la dérivée seconde de l'estimateur à noyau \hat{f} . Cet estimateur a été proposé en se basant sur un résultat asymptotique obtenu par les mêmes auteurs.

Lemme 1.3.1. *Supposant que le noyau K satisfait aux conditions suivantes :*

$$\int K''(u) du = 0, \int uK''(u) du = 0, \int u^2 K''(u) du = 2.$$

Alors on a le développement asymptotique

$$E \left\{ \int \widehat{f}''^2 \right\} = \int \widehat{f}''^2 + \frac{K''^2}{nh^5} + o(h^2).$$

Le critère à optimiser devient alors

$$BCV(h) = \frac{h^4 \sigma_k^4}{4} \left[\int \widehat{f}''^2 - \frac{\int K''^2}{nh^5} \right] + \frac{\int K^2(y) dy}{nh}. \quad (1.45)$$

Les résultats des simulations effectuées par Parzen (1990)[53] montrent que la méthode de validation croisée biaisée présente le même point faible que la méthode de validation croisée non biaisée. Cette méthode donne plusieurs minima locaux pour la fonctionnelle cible à minimiser.

Validation croisée lissée : la méthode de validation croisée lissée est une méthode de nature combinatoire, proposée par Hall et al. (1992)[30]. Dans le même esprit que le critère de validation croisée biaisée, les auteurs proposent d'approximer le MISE.

$$MISE(h) = \int \text{Biais}^2\{\widehat{f}(x)\} + \int \text{Var}\{\widehat{f}(x)\}. \quad (1.46)$$

1.3.6.3 Approche bayésienne pour le choix du paramètre de lissage

L'approche bayésienne pour le choix du paramètre de lissage consiste à considérer le paramètre comme une variable aléatoire ayant une distribution de probabilité a priori. Ensuite, en utilisant les observations X_1, \dots, X_n , on met à jour la distribution de probabilité a posteriori du paramètre.

Plus précisément, on définit une distribution a priori pour le paramètre de lissage, généralement une distribution de probabilité de la famille des lois gamma, et on utilise cette distribution pour calculer la distribution a posteriori du paramètre. On peut ensuite utiliser cette distribution pour obtenir une estimation ponctuelle du paramètre (par exemple, la moyenne ou la médiane de la distribution a posteriori) ou un intervalle de confiance pour le paramètre.

L'approche bayésienne présente l'avantage de fournir une mesure de l'incertitude associée à l'estimation du paramètre, contrairement aux méthodes fréquentistes qui ne fournissent généralement qu'une seule estimation ponctuelle du paramètre. Cependant, elle peut être plus complexe à mettre en œuvre et nécessite un choix approprié de la distribution a priori.

En pratique, pour choisir la distribution a priori appropriée, il est souvent nécessaire de faire des hypothèses sur les connaissances préalables que l'on a sur le paramètre de lissage. Si l'on n'a pas d'informations a priori

Principe de l'approche bayésienne

Soit X_1, \dots, X_n un n-échantillon de variables aléatoires i.i.d. Et θ le vecteur des paramètres à estimer dans un modèle statistique. Le principe de l'approche bayésienne est d'associer l'information a priori sur les paramètres et l'information apportée par les données en une information a posteriori

sur les paramètres. Le paramètre θ est considéré comme vecteur aléatoire, et une loi a priori $\pi(\theta)$ de θ est choisie en fonction des connaissances disponibles sur ce dernier avant la prise en compte des observations. En posant $X = (X_1, \dots, X_n)$, la loi a posteriori $\pi(\theta|X)$ du modèle est obtenue en utilisant le théorème de Bayes par :

$$\pi(\theta|X) = \frac{\pi(X|\theta)\pi(\theta)}{\pi(X)},$$

où $\pi(X|\theta)$ est la fonction vraisemblance et $\pi(X) = \int_{\Theta} \pi(X|\theta)\pi(\theta)d\theta$ est la constante de normalisation.

Soit $\hat{\theta}$ un estimateur de θ . Et considérons une fonction de coût quadratique $C(\hat{\theta}-\theta) = (\hat{\theta}(x) - \theta)^2$, avec $(\hat{\theta} - \theta)$ l'erreur d'estimation pour les observations données. L'estimateur qui minimise le coût moyen (appelé le risque bayésien) en utilisant la fonction de coût quadratique est la moyenne a posteriori donnée par

$$\hat{\theta} = E(\theta|X) = \int_{\Theta} \theta \pi(\theta|X) d\theta.$$

Le point le plus délicat de l'approche bayésienne est le choix de la loi a priori, car l'information a priori est souvent insuffisante pour déterminer la loi a priori de façon précise. Les bayésiens suggèrent d'utiliser des priors non informatifs lorsque aucune information n'est disponible a priori. Dans ce cas, on laisse les données conduire l'inférence. Pour certaines distributions des observations, afin de simplifier les calculs, le choix de la distribution a priori peut être fait de telle sorte que la loi a posteriori, ait la même forme que cette loi a priori, on parle alors des familles conjuguées. Il existe trois variantes de l'approche bayésienne ; globale, locale et adaptative.

Approche bayésienne globale

L'approche bayésienne globale a été proposée par Brewer (1998)[11] et Zhang et al. (2006)[87] dans le cas continu avec le noyau gaussien. Cette approche consiste à sélectionner le même paramètre de lissage pour tout l'échantillon. Considérons une suite de variables aléatoires X_1, \dots, X_n de fonction de densité f et de réalisation $x = (x_1, \dots, x_n)$, pour estimer le paramètre de lissage globalement nous suivons les étapes suivantes :

- On détermine l'estimateur de la vraisemblance $\pi(x|h)$ des données X_1, \dots, X_n sachant le paramètre de lissage. L'estimateur de la vraisemblance s'écrit comme suit :

$$\hat{\pi}(x_1, \dots, x_n|h) = \prod_{i=1}^n \hat{f}_h(x_i).$$

- Par la technique de validation croisée, l'estimateur $\hat{f}_h(x_i)$ est obtenue à partir de l'ensemble de points sauf en x_i . Ceci nous donne

$$\hat{\pi}(x_1, \dots, x_n|h) = \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j).$$

- On choisit la loi a priori du paramètre de lissage h notée $\pi(h)$.
- On détermine par le théorème de Bayes la loi a posteriori définie par :

$$\begin{aligned}\hat{\pi}(h|x_1, \dots, x_n) &= \frac{\pi(x_1, \dots, x_n|h) \pi(h)}{\pi(x_1, \dots, x_n)}, \\ &= \frac{\pi(h) \prod_{i=1}^n \hat{f}(x_i)}{\pi(x_1, \dots, x_n)},\end{aligned}$$

où $\pi(x_1, \dots, x_n) = \int \pi(x_1, \dots, x_n|h) \pi(h) dh$. Le calcul de cette intégrale est difficile voir impossible. La loi, a posteriori, sera proportionnelle au produit de la loi a priori et la vraisemblance

$$\hat{\pi}(h|x_1, \dots, x_n) \propto \pi(h) \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_{x_i, h}(x_j).$$

- On estime le paramètre de lissage h , par la moyenne, le mode ou la médiane a posteriori en utilisant la méthode Monte Carlo par Chaîne de Markov (MCMC).

Approche bayésienne locale

L'approche bayésienne globale a été proposée par Gangopadhyay et Cheung (2002)[28] dans le cas continu avec le noyau gaussien. Cette approche est de sélectionner le paramètre de lissage localement, c'est-à-dire estimer h en chaque point x où on désire estimer la densité. Le paramètre de lissage h est considéré comme une variable aléatoire de loi a priori $\pi(h)$, qu'on considère aussi comme paramètre du modèle $\hat{f}_h(x)$. La méthode consiste à utiliser $f(x)$ et construire un estimateur de h pour chaque point x . Par la formule de Bayes, la loi a posteriori de h au point x est donnée par :

$$\pi(h|x) = \frac{f(x)\pi(h)}{\int f(x)\pi(h)dh}.$$

En utilisant l'estimateur $\hat{f}(x)$ de $f(x)$, la loi a posteriori de h devient alors :

$$\hat{\pi}(h|x, X_1, \dots, X_n) dh.$$

L'estimateur de Bayes sous la perte quadratique de h au point x est donné par

$$\hat{h}_n(x) = \int h \hat{\pi}(h|x, X_1, \dots, X_n) dh.$$

Approche bayésienne adaptative

Le principe de cette approche consiste à sélectionner le paramètre de lissage pour chaque observation X_i . Nous présentons les travaux particuliers de Brewer (2000)[12] et Zougab et al. (2013)[89] dans le cas continu. Si on considère les observations X_1, \dots, X_n de fonction de densité inconnue f , l'estimateur à noyau associé adaptatif est donné par

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{x, h_i}(X_i), \quad x \geq 0,$$

tel que K_{x,h_i} est le noyau associé, et h_i est le paramètre de lissage adaptatif associé pour chaque observation X_i .

L'estimateur à noyau associé adaptatif de $f(X_i)$ par validation croisé est donné par :

$$\hat{f}_{h,-i}(X_i) = \hat{f}(X_i | \{X_{-i}\}, h_i) = \frac{1}{n-1} \sum_{j=1, j \neq i}^n K_{X_i, h_i}(X_j).$$

Soit $\pi(h_i)$ la loi a priori de h_i , la loi a posteriori pour chaque h_i par la formule de bayes, prend la forme suivante :

$$\hat{\pi}(h_i | X_i, \{X_{-i}\}) = \frac{\hat{f}(X_i | \{X_{-i}\}, h_i) \pi(h_i)}{\int \hat{f}(X_i | \{X_{-i}\}, h_i) \pi(h_i) dh_i}.$$

L'estimateur de bayes du paramètre de lissage h sous la perte quadratique est donné par :

$$\hat{h}_i = \int h_i \hat{\pi}(h_i | X_i, \{X_{-i}\}) dh_i.$$

1.4 Cas discret

1.4.1 Noyaux associés discrets

Dans cette section, on présente l'essentiel des travaux sur les noyaux associés discrets de Aitchison (1976)[2] et Wang (1981)[82]. On désignera par \mathbb{T}_1 le support de la fonction de masse de probabilité f à estimer. L'ensemble \mathbb{T}_1 sera soit \mathbb{N} , \mathbb{Z} ou un ensemble fini. Étant donné $x \in \mathbb{T}_1$ et $h > 0$, on appelle **noyau associé discret** $K_{x,h}$ toute fonction de masse de probabilité liée à la variable aléatoire $\mathcal{Z}_{x,h}$ de support $\mathbb{S}_{x,h}(\subseteq)\mathbb{N}$ (pouvant ne pas dépendre de x et/ ou h) contenant au moins x , vérifiant les quatres conditions suivantes :

$$x \in \mathbb{S}_{x,h}, \tag{1.47}$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + a(x, h), \tag{1.48}$$

$$\mathbf{V}\{\mathcal{K}_{x,h}\} < \infty \tag{1.49}$$

$$\mathbb{V}ar(\mathcal{Z}_{x,h}) = b(x, h). \tag{1.50}$$

Les quantités $a(x, h)$ et $b(x, h)$ tendent vers 0 quand h tend vers 0.

Définition 1.4.1. Soit X_1, \dots, X_n un n -échantillon i.i.d. d'une variable aléatoire X de la fonction de masse de probabilité inconnue f sur l'ensemble $\mathbb{T}_1 \subseteq \mathbb{N}$. L'estimateur à noyau associé discret est de cette forme :

$$\tilde{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \tag{1.51}$$

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

Nous distinguons deux classes de noyaux discrets. La première classe des noyaux associés discrets, dite, classe de noyaux associés discrets de deuxième ordre, c'est-à-dire, les conditions (1.47)-(1.50) sont vérifiées. Le tableau 1.6 présente quelques noyaux associés discrets de deuxième ordre. La deuxième classe de noyaux associés discrets, dite classe des noyaux associés discrets du premier ordre ou standard. Ce type de noyaux ne vérifie pas la condition (1.50). Le tableau 1.7 présente trois exemples de noyaux associé discrets asymétriques dits standards. On peut se référer aux travaux d'Aitchison et Aitken (1976)[2], Kokonendji et Senga Kiessié (2007)[45], et Wang et Van Ryzine (1981)[82] pour plus de détails sur ces noyaux ainsi que leurs applications.

Les noyaux associés discrets de deuxième ordre et premier ordre

Dans cette partie, nous présentons les noyaux associés de deuxième ordre dans le tableau 1.6 et les noyaux de premier ordre dans le tableau 1.7

TABLEAU 1.6 – Noyaux associé discrets de deuxième ordre.

Noyau	Domaine	Fonction noyau
Triangulaire discret (DT)	\mathbb{N}	$K_{DT(x,h)}(y) = \frac{(a+1)^h - y-x ^h}{P(a,h)}$, où $P(a,h) = (2a+1)(a+1)^h - 2 \sum_{k=0}^a k^h$.
Dirac Discret Uniform (DDU)	$\{0, 1, \dots, m-1\}$	$K_{DDU(x,h)}(y) = (1-h)\mathbb{I}_{y=x} + \frac{h}{m-1}\mathbb{I}_{y \neq x}$.
Wang Vin Ryzin (WVR)	\mathbb{Z}	$K_{WVR(x,h)}(y) = (1-h)\mathbb{I}_{y=x} + \frac{1}{2}(1-h)h^{ y-x }\mathbb{I}_{ y-x \geq 1}$.
Li et Racine (LR)	\mathbb{Z}	$K_{LR(x,h)}(y) = \mathbb{I}_{y=x} + h^{ y-x }\mathbb{I}_{y \neq x}$.
Discrete optimal symmetric kernel (DOSK)	\mathbb{Z}	$K_{(x,h)}(y) = 1 - (2h/3)x\mathbb{I}_{y=x} + \frac{h}{3}\mathbb{I}_{ x-y =1}$,

TABLEAU 1.7 – Noyaux associé discrets de premier ordre (standard).

Noyau	Domaine	Fonction noyau
Binomial	\mathbb{N}	$K_{Bin(x,h)}(y) = \frac{(x+1)!}{y!(x+1-y)!} \left(\frac{x+h}{x+1}\right)^y \left(\frac{1-h}{x+1}\right)^{x+1-y}$.
Binomial négatif	\mathbb{N}	$K_{BinN(x,h)}(y) = \frac{(x+y)!}{x!y!} \left(\frac{x+h}{2x+h+1}\right)^y \left(\frac{x+1}{2x+h+1}\right)^{x+1}$.
Poisson	\mathbb{N}	$K_{Pois(x,h)}(y) = \frac{(x+h)^y e^{-(x+h)}}{y!}$.

1.4.2 Propriétés de l'estimateur à noyau discret

Nous donnons ici la convergence au sens de l'erreur quadratique moyenne intégrée. Ces résultats sont donnés dans Kokonendji et Senga Kiessié (2011)[45]. Nous rappelons d'abord le développement du biais et de la variance de $\tilde{f}(x)$ en utilisant l'estimateur (1.51), par la suite nous déduisons les expressions de l'erreur quadratique moyenne intégrée.

Propriétés : Dans cette partie, nous présentons quelques propriétés qui ont été établies par Kokonendji et Senga Kiessé (2011)[45]. Nous donnons les résultats sur la convergence presque sûre, en moyenne quadratique intégrée et la convergence en loi de l'estimateur à noyau discret.

La proposition suivante est nécessaire pour l'étude des estimateurs à noyaux discrets.

Proposition 1.4.1. (*Kokonendji et Senga Kiessé (2011)[45]*). Soit X_1, \dots, X_n un n -échantillon de v.a. i.i.d avec la fonction de masse inconnue f sur \mathbb{T}_1 . Soit \tilde{f} un estimateur à noyau associé discret de f . Alors, pour $x \in \mathbb{T}_1$ et $h > 0$ on a

$$\mathbf{E}\{\tilde{f}(x)\} = \mathbf{E}\{\mathcal{K}_{x,h}\}, \quad (1.52)$$

où $\mathcal{K}_{x,h}$ est la variable aléatoire de loi $K_{x,h}$ sur \mathfrak{N}_x . De plus, $\tilde{f}(x) \in [0, 1]$ pour $x \in \mathbb{T}_1$ et

$$\sum_x \tilde{f}(x) = C, \quad (1.53)$$

où C est une constante strictement positive et finie.

L'expression du biais de l'estimateur (1.51) avec les différents noyaux discrets du deuxième ordre défini dans le tableau 1.6 prend respectivement la forme suivante :

$$\text{Biais}(\tilde{f}(x)) = q(x, f)h + o(h),$$

où la forme explicite de $q(x, f)$ pour chaque noyau discret du deuxième ordre est donnée dans le tableau 1.8.

TABLEAU 1.8 – La forme explicite de $q(x, f)$.

Noyaux	$q(x, f)$
Triangulaire discret (DT)	$\left(\frac{\log(a+1)}{2} S(a) - 2 \sum_{k=1}^a k^2 \log(k) \right) \frac{f^{(2)}(x)}{2}$
Dirac Discret Uniform (DDU)	$\left(\frac{m}{2} - x - \frac{x}{m-1} \right) f^{(1)}(x) + \left(\frac{m(2m-1)}{6} + x^2 + \frac{x^2}{m-1} - mx \right) \frac{f^{(2)}(x)}{2}$
Wang Van Ryzin (WVR)	$\frac{f^{(2)}(x)}{2}$
Li et Racine (LR)	$2xf^{(1)}(x) - (2x^2 - 2) \frac{f^{(2)}(x)}{2}$
Discrete optimal symmetric kernel (DOSK)	$\frac{f^{(2)}(x)}{3}$

Dans le cas des estimateurs à noyaux associés discrets standard, l'estimateur ne converge pas au sens l'erreur quadratique moyenne intégrée (*MISE*), car pour x fixé, l'estimateur est avec biais c'est-à-dire :

$$\lim_{h \rightarrow 0} \text{Biais}\{\tilde{f}(x)\} \neq 0.$$

le tableau 1.9 présente le biais de l'estimateur (1.51) avec les noyaux associé discrets standards

TABLEAU 1.9 – Expression du biais de l'estimateur à noyau associé discret standard.

Binomial	$hf^{(1)}(x) + \frac{1}{2}(x+h) \left(\frac{1-h}{x+1}\right) f^{(2)} + o(h).$
Binomial négatif	$hf^{(1)}(x) + \frac{1}{2}(x+h) \left(\frac{2x+1+h}{x+1}\right) f^{(2)} + o(h).$
Poisson	$hf^{(1)}(x) + \frac{1}{2}(x+h)f^{(2)}(x) + o(h).$

L'expression de la variance asymptotique de l'estimateur à noyau associé discret est de la forme :

$$\text{Var}(\tilde{f}(x)) = \frac{1}{n}f(x)\{1-f(x)\}K_{x,h}^2(x) + o\left(\frac{1}{n}\right).$$

Convergence du biais

La proposition suivante garantit que l'estimateur à noyau discret du deuxième ordre $\tilde{f}(x)$ est asymptotiquement sans biais en chaque point x

Proposition 1.4.2. *Kokonendji et Senga Kiessé (2011)[45]. Soit $f : \mathbb{T}_1 \rightarrow \mathbb{R}$ une fonction de masse de probabilité et soit $x \in \mathbb{T}_1$ fixé.*

Si $\tilde{f}(x)$ est l'estimateur à noyau discret de f alors

$$\mathbf{E}(\tilde{f}(x)) = \sum_{y \in \mathbb{T}_1 \cap \mathbb{S}_{x,h}} f(y)K_{x,h}(y) \rightarrow f(x) \text{ quand } h \rightarrow 0 \text{ pour } n \rightarrow \infty.$$

Convergence presque sûre

Théorème 1.4.1. *(Senga Kiessé (2008)[70]). Soit $f : \mathbb{T}_1 \rightarrow \mathbb{R}$ une fonction de masse de probabilité et soit $x \in \mathbb{T}_1$ fixé. Sous les conditions (1.47)-(1.50), nous avons*

$$\tilde{f}(x) \xrightarrow{p.s.} f(x) \text{ quand } n \rightarrow \infty, \quad (1.54)$$

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

1.4.3 Convergence en moyenne quadratique et en moyenne quadratique intégrée

Les expressions de l'erreur quadratique moyenne intégrée s'expriment en fonction du biais et de la variance comme suit

$$\text{MSE}(\tilde{f}(x)) = \mathbf{E}\{\tilde{f}(x) - f(x)\}^2 = \text{Biais}^2\{\tilde{f}(x)\} + \text{Var}\{\tilde{f}(x)\}, \quad (1.55)$$

et

$$\text{MISE}(\tilde{f}) = \sum_{x \in \mathbb{T}_1} \mathbf{E}\{\tilde{f}(x) - f(x)\}^2 = \sum_{x \in \mathbb{T}_1} \text{Biais}^2\{\tilde{f}(x)\} + \sum_{x \in \mathbb{T}_1} \text{Var}\{\tilde{f}(x)\}. \quad (1.56)$$

Nous présentons maintenant les résultats de convergence en moyenne quadratique et en moyenne quadratique intégrée de l'estimateur à noyau discret.

Théorème 1.4.2. (Senga Kiessé (2008)[70]). Soit $f : \mathbb{T}_1 \rightarrow \mathbb{R}$ une fonction de masse de probabilité et soit $x \in \mathbb{T}_1$ fixé.

Si $\tilde{f}(x)$ est l'estimateur à noyau discret de f alors

$$\lim_{n \rightarrow \infty} \text{MSE}(\{\tilde{f}(x)\}) = 0.$$

Théorème 1.4.3. (Senga Kiessé (2008)[70]). Soit f une fonction de masse de probabilité sur \mathbb{T}_1 . Alors l'estimateur à noyau discret $\tilde{f}(x)$ de f est tel que, pour $n \rightarrow \infty$ et $h \rightarrow 0$, on a

$$\begin{aligned} \text{MISE}(\tilde{f}) &= \frac{1}{n} \sum_{x \in \mathbb{T}_1} f(x) [\{\text{Pr}(\mathcal{K}_{x,h} = x)\}^2 - f(x)], \\ &+ \sum_{x \in \mathbb{T}_1} \left[f\{\mathbf{E}(\mathcal{K}_{x,h})\} - f(x) + \frac{1}{2} \text{Var}(\mathcal{K}_{x,h}) f^{(2)}(x) \right]^2 + o\left(\frac{1}{n} + h^2\right), \end{aligned}$$

où $f^{(2)}$ est la différence finie donnée par :

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

Convergence en loi

Théorème 1.4.4. (Kokonendji et Senga Kiessé (2011)[43]). Soit $f : \mathbb{T}_1 \rightarrow \mathbb{R}$ une fonction de masse de probabilité et soit $x \in \mathbb{T}_1$ fixé et $\tilde{f}(x)$ est l'estimateur à noyau discret de f . Sous les conditions (1.47)-(1.50), nous avons

$$\frac{\tilde{f}(x) - \mathbf{E}(\tilde{f}(x))}{\sqrt{\text{Var}(\tilde{f}(x))}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow \infty,$$

où $\xrightarrow{\mathcal{L}}$ désigne la convergence en loi et $\mathcal{N}(0, 1)$ est la loi normale standard.

1.5 Conclusion

Dans ce chapitre, nous avons revisité la définition de l'estimateur de la densité de probabilité à noyau classique (symétrique), et du noyau associé (asymétrique) dans le cas continu et discret, lorsque les données sont indépendantes. Nous avons présenté leurs propriétés en termes de biais, de variance, ainsi que MSE et MISE. Nous avons présenté les différentes méthodes pour le choix du paramètre de lissage, à savoir les méthodes Plug-in, validation croisée et l'approche bayésienne avec ses variantes. Dans le chapitre suivant, nous présenterons l'estimateur à noyau associé continu dans les cas de données dépendantes, ainsi qu'une étude d'évaluation de performances basées sur des données simulées à partir des deux modèles (DCA, VS), ainsi que des données réelles.

Estimation de la densité par la méthode du noyau dans le cas des données dépendantes

Sommaire

2.1	Introduction	30
2.2	Contexte de l'étude	30
2.3	Estimateur à noyau asymétrique : cas des données dépendantes	31
2.4	Hypothèses et principaux résultats	32
2.5	Propriétés de l'estimateur à noyau asymétrique	32
2.5.1	MSE de l'estimateur à noyau asymétrique	33
2.5.2	Preuve du Théorème 2.5.1.	33
2.6	Application numérique	34
2.6.1	Simulation	34
2.6.2	Durées conditionnelles auto-régressives	35
2.6.3	Volatilité stochastique	38
2.7	Application sur des données réelles	41
2.8	Conclusion	43

2.1 Introduction

Le but de ce chapitre est de présenter les estimateurs à noyaux asymétriques dans le cas des données dépendantes. Dans un premier temps, nous dérivons la propriété locale (MSE) de l'estimateur présenté pour des données dépendantes d'un processus stationnaire α mélangeant en utilisant les noyaux G, IG, RIG, LN, BS, BSPE et MG. Deuxièmement, nous réalisons une étude de Monte-Carlo basée sur des données dépendantes d'un modèle de durée conditionnelle autorégressive (DCA) et d'un modèle de volatilité stochastique (VS) en utilisant les noyaux asymétriques G, IG, RIG, LN, BS, BSPE et MG. Nous terminons par une étude d'application sur des données réelles, en nous focalisant sur le cas où les données sont dépendantes ; voir Gouriéroux et Monfort (1990)[29]. Cette étude peut être considérée comme un complément à l'estimateur de Bouezmarni et Rombouts (2010)[9], spécifiquement pour les données dépendantes.

2.2 Contexte de l'étude

Soit $(X_t, t \in \mathbb{N})$ un processus à temps discret défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathcal{P})$, tel que les X_t sont des valeurs de \mathbb{R}^+ , ayant la même loi, de densité de probabilité inconnue f .

Nous donnons quelques définitions qui seront utiles dans notre travail. Nous commençons par le coefficient de mélange fort introduit par Rosenblatt (1956)[61] (voir aussi Bouezmarni et Rombouts (2010)[9] et Amiri (2009)[3]).

Définition 2.2.1. Soit $(\Omega, \mathcal{A}, \mathcal{P})$ un espace probabilisé. Le coefficient de α -mélange entre deux tribus F_1^k et F_1^{k+i} de \mathcal{A} est défini par :

$$\alpha(F_1^k, F_1^{k+i}) := \sup_{A \in F_1^k, B \in F_1^{k+i}} |P(A \cap B) - (P(A)P(B))|. \quad (2.1)$$

Nous donnons ensuite une définition de la notion de forte mélangeance pour un processus.

Définition 2.2.2. Un processus $(X_t, t \in \mathbb{N})$ est fortement mélangeant ou α -mélangeant si le coefficient de mélange fort de X défini pour tout $i \geq 0$, par :

$$\alpha(i) := \sup_{t \in \mathbb{N}} \alpha(\sigma(X_s, s \leq t), (\sigma(X_s, s > t + i))), \quad (2.2)$$

tel que $\alpha(i) \rightarrow 0$ lorsque $i \rightarrow +\infty$, et $\sigma(X)$ est σ -algèbre des événements engendrés par X .

Remarque 2.2.1. Les conditions pour qu'un processus soit stationnaire et mélangeant (α -mélangeant) sont importantes en théorie des probabilités et en statistique. Voici une description de ces conditions :

- **Processus Stationnaire** : Un processus est stationnaire si ses propriétés statistiques ne changent pas au cours du temps.
- **Processus α -Mélangeant** : Un processus est α -mélangeant s'il existe une fonction $\alpha(i)$ telle que $\alpha(i) \rightarrow 0$ lorsque $i \rightarrow \infty$, et qui mesure la dépendance entre les événements dans le passé et le futur, séparés par un intervalle de longueur i . Plus précisément, pour un processus $\{X_t\}$, la fonction $\alpha(i)$ quantifie la décroissance de la dépendance entre le passé et le futur à mesure que l'intervalle i augmente.

Nous présentons deux exemples des processus stationnaires et α -mélangeants.

- **Exemple 1** : Un exemple classique de processus α -mélangeant est le processus autorégressif AR(1) (voir Doukhan (1994)[22]) est défini par :

$$X_t = \phi X_{t-1} + \epsilon_t, \quad (2.3)$$

où $|\phi| < 1$ et ϵ_t est un bruit blanc avec variance finie σ^2 .

- **Exemple 2** : Un deuxième exemple de processus α -mélangeant est le processus à durées conditionnelles auto-régressives (DCA), introduit par Carasco et Chen (2002)[13] et défini par la relation suivante :

$$\begin{cases} x_t = \psi_t \epsilon_t; \\ \psi_t = \omega + \alpha_1 x_{t-1} + \alpha_2 \psi_{t-1}. \end{cases} \quad (2.4)$$

Pour que le processus DCA soit stationnaire et α -mélangeant avec décroissance exponentielle, il faut que les conditions suivantes soient remplies :

- $|\alpha_2| < 1$,
- $E(\alpha_2 + \alpha_1 \epsilon_i)^2 < 1$.

Pour plus de détails sur le modèle DCA, voir la Section 2.6.2

2.3 Estimateur à noyau asymétrique : cas des données dépendantes

La définition ci-dessous aborde les concepts de l'estimation de la densité par la méthode du noyau dans le contexte des données dépendantes sur le support $[0, \infty[$.

Définition 2.3.1. Soit X_1, \dots, X_n un processus stationnaire α mélange avec une fonction de densité inconnue f . L'estimateur à noyau asymétrique pour des données dépendantes est :

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad (2.5)$$

où : h est le paramètre de lissage et $K_{x,h}$ est le noyau asymétrique continu de cible x et du paramètre de lissage h .

Dans leur étude, Bouezmarni et Rombouts (2010)[9] ont présenté l'estimateur à noyau asymétrique gamma dans le cas des données dépendantes pour estimer la densité sur l'intervalle $[0, \infty[$. En suivant Bouezmarni et Rombouts (2010)[9], nous présentons les estimateurs à noyaux asymétriques adaptés aux données dépendantes, voir le tableau 1.2 pour quelques noyaux asymétriques.

2.4 Hypothèses et principaux résultats

Le problème du biais de frontière persiste dans l'estimateur de densité à noyau standard pour les fonctions de densité définies sur $[0, \infty[$ lorsque les données sont dépendantes. Dans cette section, on montre que l'estimateur à noyau asymétrique est exempt de biais de frontière dans le cas des données dépendantes, il est possible d'obtenir les mêmes taux de convergence que dans le cas des données indépendantes, de plus on montre que l'estimateur à noyau asymétrique atteint le taux de convergence optimale pour l'erreur quadratique moyenne dans la classe des estimateurs de densité à noyau non négatif. Nous désignons par f_i la densité conjointe de (X_i, X_{i+1}) et $g_i(x, y) = f_i(x, y) - f(x)f(y)$. Nous rappelons quelques conditions utiles pour cette étude. Nous commençons par le coefficient de mélange fort, voir Bouezmarni et Rombouts (2010)[9] et Rosenblatt (1956)[61].

Condition 1. Supposons que g_i satisfasse la condition de Lipschitz,

$$|g_i(\epsilon') - g_i(\epsilon)| \leq l \|\epsilon' - \epsilon\|, \epsilon', \epsilon \in \mathbb{R}^+ \times \mathbb{R}^+, \quad (2.6)$$

pour une certaine constante l .

Nous exigeons que le processus $\{X_t\}_{t \geq 1}$ est α -mélange, voir les définitions 2.2.1 et 2.2.2.

Condition 2. Supposons que $\{X_i\}_{i \geq 1}$ est un processus α -mélange, tel que

$$\alpha(i) \leq \gamma i^{-\beta}, i \geq 1, \quad (2.7)$$

pour certaines constantes positive γ et β .

2.5 Propriétés de l'estimateur à noyau asymétrique

Dans cette section, nous examinons les propriétés de l'estimateur à noyau asymétrique dans le cadre des données dépendantes.

2.5.1 MSE de l'estimateur à noyau asymétrique

Nous supposons que :

Hypothèse 1. f est deux fois continuellement différentiable.

Hypothèse 2. $\lim_{n \rightarrow \infty} h = 0$ et $\lim_{n \rightarrow \infty} nh^{\frac{1}{2}} = \infty$.

Nous présentons ci-dessous le théorème sur l'approximation du MSE.

Théorème 2.5.1. *Par les conditions 1 et 2 et si les hypothèses 1 et 2 sont vérifiées, alors pour la séquence $v_n \cong cte \times h^{-\frac{3}{2\beta}}$ ($\beta > \frac{3}{2}$), le MSE de $\hat{f}(x)$ pour les données dépendantes peut être approximée par :*

$$MSE(\hat{f}(x)) = B^2(x) + V_{x,h}(x) + C(x),$$

où $B(x)$ et $V(x)$ sont les approximations du biais et de la variance de $\hat{f}(x)$ données dans le tableau 2.1 pour un noyau asymétrique et

$$C(x) = \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) cov[K_{x,h}(X_i), K_{x,h}(X_{i+1})],$$

avec $nh^{1/2}C(x) = o(1)$.

2.5.2 Preuve du Théorème 2.5.1.

Le théorème peut être établi en combinant les résultats de Chen (2000)[16], Scaillet (2004)[67], Jin et al. (2003)[38], Marchant et al. (2013) [50], et les résultats de Bouezmarni et Rombouts (2010)[9].

$$\begin{aligned} MSE(\hat{f}(x)) &= \text{Biais}^2(\hat{f}(x)) + \text{Var}(\hat{f}(x)) + C(x), \\ &= B^2(x) + V_{x,h}(x) + C(x), \\ &= B^2(x) + V_{x,h}(x) + \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) cov[K_{x,h}(X_i), K_{x,h}(X_{i+1})], \end{aligned} \tag{2.8}$$

Les expressions de $B(x)$ et $V_{x,h}(x)$ peuvent être dérivées à partir du tableau 2.1.

$$\text{avec, } c^2 = \left(\frac{\nu}{2^{2\nu} \Gamma(\frac{1}{2\nu})}\right)^2, \quad c_g = 2^{\frac{1}{\nu}} \Gamma\left(\frac{2\nu+1}{2\nu}\right).$$

Remarque 2.5.1. Les termes $B(x)$ et $V_{x,h}(x)$ sont les mêmes que dans le cas i.i.d.

Il reste maintenant à calculer le terme de covariance $C(x)$. D'après l'équation (2.6), l'inégalité de Bosq et Blanke (1996) [8] :

$$cov((K_{x,h}(t, X_i)), (K_{x,h}(t, X_{i+1}))) \leq \gamma(l)\alpha(i)^{\frac{1}{3}}, \tag{2.9}$$

TABLEAU 2.1 – Biais et variance de $\widehat{f}(x)$.

Noyau	$B(x) = \text{Biais}(\widehat{f}(x))$	$V_{x,h}(x) = \text{Variance}(\widehat{f}(x))$
G [16]	$h[f'(x) + \frac{x}{2}f''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-1/2}f(x),$
MG [16]	$h[\frac{1}{2}xf''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-1/2}f(x),$
RIG [67]	$h[\frac{1}{2}xf''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-1/2}f(x),$
LN [38]	$h[2xf'(x) + 2x^2f''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-1}f(x),$
IG[67]	$h[\frac{1}{2}x^3f''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-3/2}f(x),$
BS[38]	$h[\frac{1}{2}xf'(x) + \frac{1}{2}x^2f''(x)]$	$\frac{1}{2\sqrt{\pi}}n^{-1}h^{-1/2}x^{-1}f(x),$
BSPE [50]	$\frac{h^2}{2}u_1(g)[xf'(x) + x^2f''(x)]$	$\frac{c^2}{c_q^2nh^{-\frac{1}{2}}x}f(x),$

pour une certaine constante $\gamma(l)$. D'autre part, d'après l'inégalité de Billingsley [8].

$$\text{cov}((K_{x,h}(t, X_i)), (K_{x,h}(t, X_{i+1}))) \leq 4 \|K_{x,h}\|_{\infty}^2 \alpha(i). \quad (2.10)$$

En utilisant la formule de Stirling et le fait que x est le mode du noyau ($K_{x,h}(x)$), alors

$$\|K_{x,h}\|_{\infty}^2 \leq \frac{1}{2\pi}x^{-1}h^{-1}. \quad (2.11)$$

En combinant les équations (2.9), (2.10) et (2.11), nous obtenons pour une certaine séquence v_n ,

$$\begin{aligned} \frac{n}{2}C(x) &\leq \Gamma(l) \sum_{i=1}^{v_n} (1 - \frac{i}{n})\alpha^{1/3}(i) + \frac{2}{\pi}x^{-1}h^{-1} \sum_{i>v_n}^{n-1} (1 - \frac{i}{n})\alpha(i), \\ &\leq \Gamma(l) \sum_{i=1}^{v_n} \alpha^{1/3}(i) + \frac{2}{\pi}x^{-1}h^{-1} \sum_{i>v_n}^{n-1} \alpha(i), (1 - \frac{i}{n}) \leq 1, \\ &\leq \gamma\Gamma(l) \sum_{i=1}^{v_n} (i)^{-\frac{\beta}{3}} + \frac{2}{\pi}x^{-1}h^{-1}\gamma \sum_{i>v_n}^{n-1} (i)^{-\beta}. \end{aligned}$$

On choisit $v_n \cong cte \times h^{-\frac{3}{2\beta}}$, on a les résultats suivants $nh^{\frac{1}{2}}C(x) = o(1)$ pour $\beta > 3/2$. ■

2.6 Application numérique

2.6.1 Simulation

L'objectif principal de cette section est d'examiner les performances des estimateurs à noyau, IG, RIG, LN, BSPE, MG et BS, et de les comparer aux performances de l'estimateur à noyau gamma dans les cas des données dépendantes positives introduit par Bouezmarni et Rombouts

(2010)[9]. Nous considérons les données dépendantes générées à partir d'un modèle de durée conditionnelle autorégressive (DCA) et d'un modèle de volatilité stochastique (VS). Le paramètre de lissage est choisi par $h = h_{ucv}$. Les performances sont comparées à l'aide de l'erreur quadratique intégrée (ISE) donnée par :

$$ISE := \int \left(\widehat{f}(x) - f_{emp}(x) \right)^2 dx. \quad (2.12)$$

Puisque la densité sous-jacente f est inconnue pour les modèles DCA et VS, nous considérons la fonction de densité empirique basée sur 100000 observations pour chaque modèle, nous effectuons 100 répliques pour différents degrés de dépendance et pour des tailles d'échantillon de 50, 100, 500 et 1000. La fonction de densité empirique est définie par :

$$f_{emp}(x) = \frac{\widehat{F}(x + \epsilon) - \widehat{F}(x)}{\epsilon}. \quad (2.13)$$

où $\widehat{F}(x)$ est la distribution empirique, et ϵ est choisi assez petit (très proche de zéro). Dans notre étude nous avons choisi $\epsilon = 0.1$

2.6.2 Durées conditionnelles auto-régressives

L'analyse des intervalles de temps entre les transactions financières revêt une importance capitale pour la compréhension de la microstructure des marchés financiers. Dans leur article, d'Engle et Russell (1998)[25], ont proposé une nouvelle classe de processus ponctuels dont les taux d'arrivées sont dépendants. Ils ont modélisé la durée attendue entre les événements en utilisant une structure autorégressifs et l'ont appelé le modèle de durée conditionnelle autorégressive (DCA). La durée est définie comme la différence entre deux temps d'arrivée consécutifs, soit $x_i = t_i - t_{i-1}$. Le modèle de durée conditionnelle autorégressive (DCA) est défini par Bouezmarni et Rombouts (2010)[9].

$$\begin{cases} x_i = \psi_i \epsilon_i; \\ \psi_i = \omega + \alpha_1 x_{i-1} + \alpha_2 \psi_{i-1}. \\ \psi_1 = 1 \end{cases} \quad (2.14)$$

Les innovations $\epsilon_i, i = 1, \dots, n$ sont des variables aléatoires i.i.d. Sous une hypothèse de densité sur ϵ_i , telle que la distribution gamma généralisé ou la distribution de Burr, le modèle peut être facilement estimé par maximum vraisemblance. Feike et Werker (2004)[23] ont fourni des méthodes d'inférence pour les modèles semi-paramétriques où la distribution de ϵ_i est déterminée uniquement à partir de l'échantillon de données. Bien que la distribution conditionnelle de x_i soit inconnue, des expressions pour les moments conditionnelles sont disponibles, voir Bauwens (2001)[6], et Fernandes (2006)[26]. Pour des extensions et des applications du modèle DCA, voir Engle (2000)[24], Hamilton (2002)[32], Bauwens (2004)[7], ainsi que Bauwens, Giot (2004) [7]. Nous considérons trois niveaux

de dépendance : forte, faible et indépendance. Les valeurs des paramètres correspondant à ces niveaux sont présentées dans le tableau 2.2.

	ω	α_1	α_2
Dépendance forte	0.017	0.11	0.88
Dépendance faible	0.153	0.11	0.80
Indépendance	1.700	0.00	0.00

TABLEAU 2.2 – Les niveaux de dépendance

Dans cette étude, nous considérons deux densités d'innovation : gamma standard et log-normal avec des paramètres illustrés dans le tableau 2.3.

	DCA-D1	DCA-D2
ϵ_i	Gamma($\alpha = 4.91, \beta = 4.91$)	Log-normal($\sigma = 0.01, \mu = 0.3$)
Densités	$x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}$	$\frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(\frac{-1}{2} \left(\frac{\log x - \mu}{\sigma}\right)^2\right)$

TABLEAU 2.3 – Densité d'innovation.

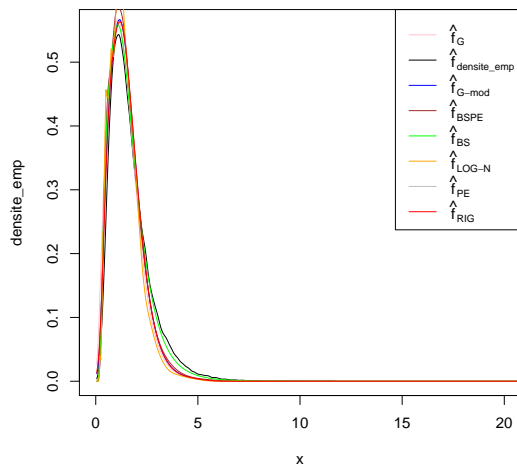
La figure 2.1 présente les courbes des estimations de la densité pour le modèle DCA utilisant les noyaux G, RIG, LN, BS, MG et BSPE, avec trois niveaux de dépendance : forte, faible et indépendance. Les résultats sont donnés pour une taille d'échantillon $n = 500$, et $\xi_i \sim \text{Gamma}(4.92, 4.92)$. Le tableau 2.4 présente les résultats de l'erreur quadratique intégrée (ISE) pour certains estimateurs à noyau asymétrique comparés à l'estimateur à noyau gamma (G) pour le modèle DCA. Pour chaque réplication, l'ISE diminue lorsque la taille de l'échantillon augmente, indépendamment du niveau de dépendance. Pour toutes les tailles d'échantillon, on observe que :

Pour une forte dépendance : L'estimateur à noyau MG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50$. Pour $n = 100$ et 500 , l'estimateur à noyau IG est meilleur que les estimateurs à noyaux asymétriques. De même, l'estimateur à noyau LN est plus efficace que les estimateurs à noyaux $n = 1000$.

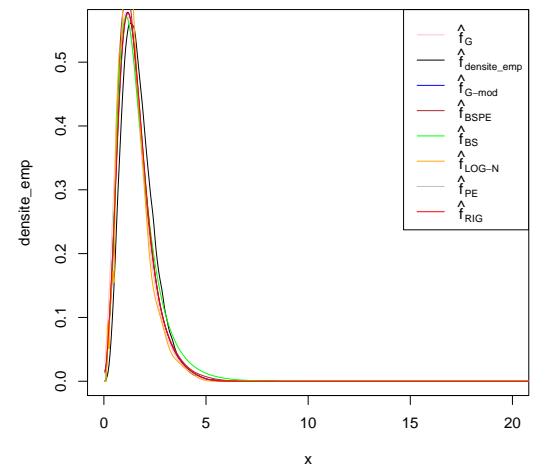
Pour une faible dépendance : L'estimateur à noyau IG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100, 500$ et 1000

Pour l'indépendance : L'estimateur à noyau G est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50$. Pour $n = 100$, l'estimateur à noyau MG est meilleur que les estimateurs à noyaux asymétriques. De même, l'estimateur à noyau IG est plus efficace que les estimateurs à noyaux 500 et $n = 1000$.

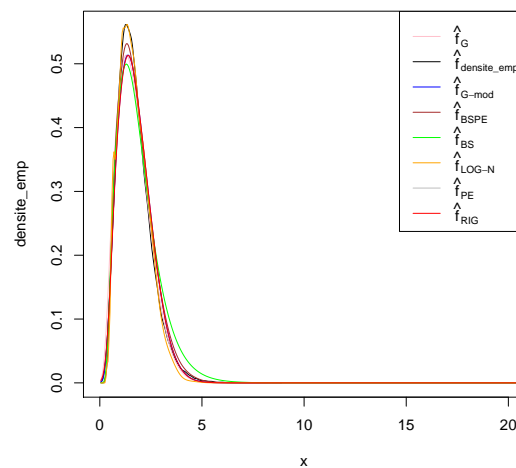
Le tableau 2.5 présente les résultats de l'erreur quadratique intégrée (ISE) pour certains estimateurs à noyau asymétrique comparés à l'estimateur à noyau gamma (G) pour les modèles DCA. Pour chaque réplication, l'ISE diminue lorsque la taille de l'échantillon augmente, indépendamment du niveau de dépendance. On observe que :



(a) Dépendance forte



(b) Dépendance faible



(c) Indépendance

FIGURE 2.1 – Quelques estimateurs à noyaux asymétriques avec la fonction de densité empirique utilisant le modèle DCA.

Pour une forte dépendance : L'estimateur à noyau MG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100$. Pour $n = 500$, l'estimateur à noyau LN est meilleur que les estimateurs à noyaux asymétriques. De même, l'estimateur à noyau G est plus efficace que les estimateurs à noyaux asymétriques $n = 1000$.

Pour une faible dépendance : L'estimateur à noyau MG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100, 500$ et 1000

Pour l'indépendance : L'estimateur à noyau MG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100, 500$ et 1000

Méthode	Noyau	n=50	n=100	n=500	n=1000
Dépendance forte	G	0.03301931	0.01248200	0.00433304	0.00212228
	IG	0.03708251	0.01214169	0.00167025	0.00365833
	RIG	0.03766101	0.01271207	0.00305636	0.00275693
	LN	0.06941590	0.03461377	0.00513096	0.00051436
	BS	0.05184881	0.03757398	0.00970163	0.00265085
	MG	0.03101200	0.02544668	0.00547289	0.00215835
	BSPE	0.03203270	0.02037138	0.00654061	0.00226218
Dépendance faible	G	0.00194289	0.00183213	0.00055676	0.00029023
	IG	0.00101720	0.00056117	0.00022239	0.00003876
	RIG	0.00123893	0.00065165	0.00034258	0.00004566
	LN	0.01041245	0.00203983	0.00038563	0.00023011
	BS	0.00127137	0.00165136	0.00142784	0.00085027
	MG	0.00502474	0.00179221	0.00032198	0.00022325
	BSPE	0.00761552	0.00375600	0.00102867	0.00055744
Indépendance	G	0.00011662	0.00082675	0.00022560	0.00010133
	IG	0.00379541	0.00098342	0.00006166	0.00004972
	RIG	0.00475561	0.00125115	0.00080590	0.00050781
	LN	0.01329811	0.01340214	0.00133199	0.00013509
	BS	0.01340176	0.01332706	0.00133099	0.00131754
	MG	0.00089206	0.00042144	0.00019450	0.00011058
	BSPE	0.00241656	0.00191203	0.00066136	0.00052565

TABLEAU 2.4 – Les valeurs de l'ISE moyen (\overline{ISE}) basées sur 100 réplifications pour **DCA-D1**

2.6.3 Volatilité stochastique

Le modèle de volatilité stochastique (VS), voir par exemple Taylor (2008)[76], est une alternative populaire au modèle de volatilité du type GARCH. L'estimation du modèle VS est moins simple que celle du modèle GARCH, la fonction de vraisemblance impliquant une intégrale de forte dimension, mais il a des liens directs avec la théorie financière, comme le prix des options et des obligations par exemple. Comme discuté par Harvey (1994)[34], le modèle VS peut être interprété comme une approximation en temps discret du modèle GARCH. Processus Orstein-Uhlenbeck a temps continu qui est souvent considéré comme le processus générateur de la volatilité réalisée (voir Barndorff (2002) [5] pour plus de détails) à ce sujet. Le modèle VS de base est donné par :

$$\begin{cases} y_i = \exp\left(\frac{x_i}{2}\right)\mu_i; \\ x_i = \gamma + \delta x_{i-1} + \nu\epsilon_i. \end{cases}$$

Où μ_i et ϵ_i sont des variables aléatoires normales i.i.d. et indépendantes les unes des autres. Il existe diverses extensions et autres formulations du modèle VS de base, voir Shephard (2005)[72] pour plus de détails. Notre intérêt réside dans la densité marginale de la variance. Dans l'étude

Méthode	Noyau	n=50	n=100	n=500	n=1000
Dépendance forte	G	0.18506710	0.13373490	0.05648755	0.00277883
	IG	0.46693370	0.37736720	0.01411313	0.00926124
	RIG	0.76625490	0.46587370	0.02954113	0.00913346
	LN	0.02737507	0.01490732	0.00387449	0.00390815
	BS	0.10792340	0.64784960	0.56728070	0.48779660
	MG	0.00474407	0.00461151	0.00448084	0.00348279
	BSPE	0.18983051	0.09206280	0.05364508	0.03018982
Dépendance faible	G	0.06825731	0.00719895	0.00133342	0.00102902
	IG	0.01496650	0.02553251	0.00564403	0.00663738
	RIG	0.04217519	0.03340141	0.01117332	0.00766492
	LN	0.08929997	0.02878806	0.00586958	0.00162445
	BS	0.09584003	0.06957631	0.02752687	0.01092379
	MG	0.00296957	0.00190395	0.00228694	0.00117548
	BSPE	0.03724414	0.01190574	0.00855247	0.00668118
Indépendance	G	0.03028409	0.00813753	0.00671034	0.00023419
	IG	0.13846190	0.01651651	0.00428371	0.00257743
	RIG	0.22353920	0.08826013	0.02268449	0.00811969
	LN	0.08572216	0.07582349	0.01212289	0.00166365
	BS	0.25791790	0.14008840	0.15704940	0.05221085
	MG	0.00370288	0.00365028	0.00263454	0.00101312
	BSPE	0.02424231	0.00547537	0.00522705	0.00476241

TABLEAU 2.5 – Les valeurs de l’ISE moyen (\overline{ISE}) basées sur 100 réplifications pour **DCA-D2**

de simulation, nous considérons deux densités d’innovation pour μ_i , la distribution normale et la distribution de student avec 5 degrés de liberté. Comme précédemment, nous considérons trois niveaux de dépendance : forte, faible et indépendance, dont les paramètres sont donnés dans le tableau 2.6. Pour illustrer le processus de VS.

	γ	δ	ν
Dépendance forte	0.01	0.98	0.30
Dépendance faible	0.01	0.95	0.30
Indépendance	1.00	0.00	0.00

TABLEAU 2.6 – Paramètres de niveau de dépendance pour le modèle VS.

Le tableau 2.7 présente les résultats de l’erreur quadratique intégrée (ISE) pour certains estimateurs à noyau asymétrique comparés à l’estimateur à noyau gamma (G) pour les modèles VS. Pour chaque réplification, l’ISE diminue lorsque la taille de l’échantillon augmente, indépendamment du niveau de dépendance. On observe que :

Pour une forte dépendance : L’estimateur à noyau G est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100$. Pour $n = 500$, l’estimateur à noyau BS est meilleur que les

estimateurs à noyaux asymétriques. De même, l'estimateur à noyau LN est plus efficace que les estimateurs à noyaux asymétriques $n = 1000$.

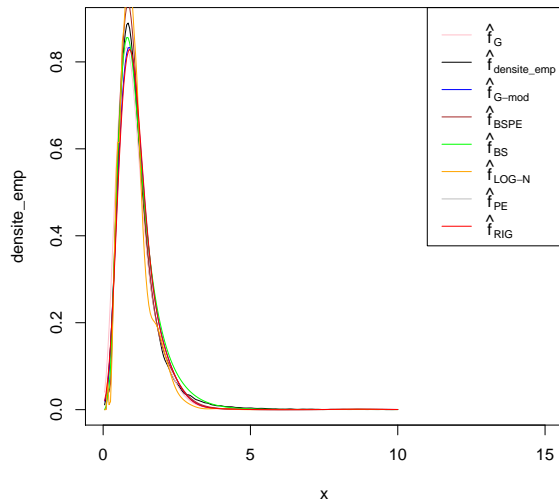
Pour une faible dépendance : L'estimateur à noyau RIG est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50, 100, 500$ et 1000 .

Pour l'indépendance : L'estimateur à noyau LN est meilleur que les estimateurs à noyaux asymétriques lorsque $n = 50$. Pour $n = 100, 500$, l'estimateur à noyau RIG est meilleur que les estimateurs à noyaux asymétriques. De même, l'estimateur à noyau G est plus efficace que les estimateurs à noyaux asymétriques $n = 1000$.

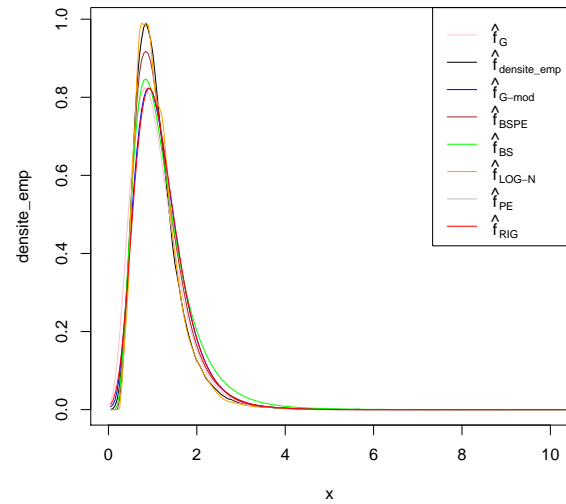
Méthode	Noyau	n=50	n=100	n=500	n=1000 .
Dépendance forte	G	0.07144356	0.05405882	0.04693856	0.04593494
	IG	0.10203280	0.06959852	0.04958614	0.04350670
	RIG	0.09392880	0.06543285	0.04810290	0.04531115
	LN	0.09835047	0.06309846	0.03573257	0.01597062
	BS	0.61685630	0.11904761	0.02329112	0.00809929
	MG	0.63514867	0.64157087	0.65334171	0.65795131
	BSPE	0.58778812	0.56562114	0.54893153	0.51768427
Dépendance faible	G	0.04223314	0.01013281	0.00515694	0.00128628
	IG	0.01961111	0.05546070	0.00173382	0.00156445
	RIG	0.02180124	0.00101757	0.00169805	0.00090602
	LN	0.01015645	0.09339953	0.00881566	0.00131510
	BS	0.90164450	0.36916050	0.10232470	0.01281856
	MG	0.85441140	0.78125508	0.81090276	0.79629044
	BSPE	0.79537440	0.71585054	0.71771000	0.69996671
Indépendance	G	0.02117934	0.00776984	0.00411825	0.00030324
	IG	0.08565430	0.00409344	0.00229338	0.00074186
	RIG	0.05982110	0.00297274	0.00201671	0.00041302
	LN	0.01262010	0.00867296	0.00583447	0.00159354
	BS	0.12151680	0.25452610	0.03194193	0.09474685
	MG	0.30662892	0.32319289	0.30213859	0.29707279
	BSPE	0.32203102	0.20906900	0.18231235	0.17243175

TABLEAU 2.7 – Les valeurs de l'ISE moyen (ISE) basées sur 100 réplifications pour **VS**

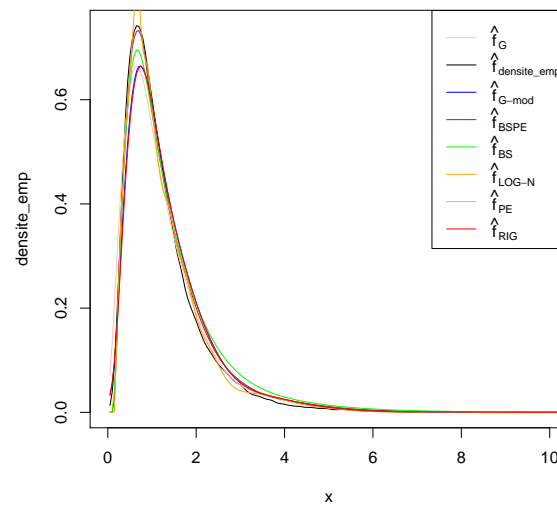
La figure 2.2 présente les tracés des estimations de la densité de probabilité (PDF) du modèle VS en utilisant les noyaux G, RIG, LN, BS, MG et BSPE, avec trois niveaux de dépendance : forte, faible et indépendance. Les résultats sont donnés pour une taille d'échantillon $n = 500$, $\epsilon_i \sim \mathcal{N}(0, 0.3)$, $\mu_i \sim \mathcal{N}(0, 0.3)$.



(a) Dépendance forte



(b) Dépendance faible



(c) Indépendance

FIGURE 2.2 – Quelques estimateurs des noyaux asymétriques avec la fonction de densité empirique en utilisant le modèle VS.

2.7 Application sur des données réelles

Dans cette section, nous utilisons les estimateurs à noyaux asymétriques (G, IG et LN) pour estimer la densité d'une série chronologique positive. Les données portent sur le trafic de voyageurs de la SNCF en seconde classe, exprimé en millions de passagers. Les observations mensuelles couvrent la période de 1963 à 1980, pour une taille d'échantillon $n = 216$ (voir Christian et Gouriéroux (1990)[29] pour plus de détails). La méthode de sélection du paramètre de lissage est

la méthode validation croisée (UCV).

La figure 2.3 présente les courbes d'estimation des données SNCF obtenues à l'aide des noyaux G, IG et LN, comparées à l'histogramme. On observe que l'estimateur à noyau LN surperforme les estimateurs à noyaux asymétriques (G et IG) car il épouse mieux la forme de l'histogramme, en touchant une majorité de ses points.

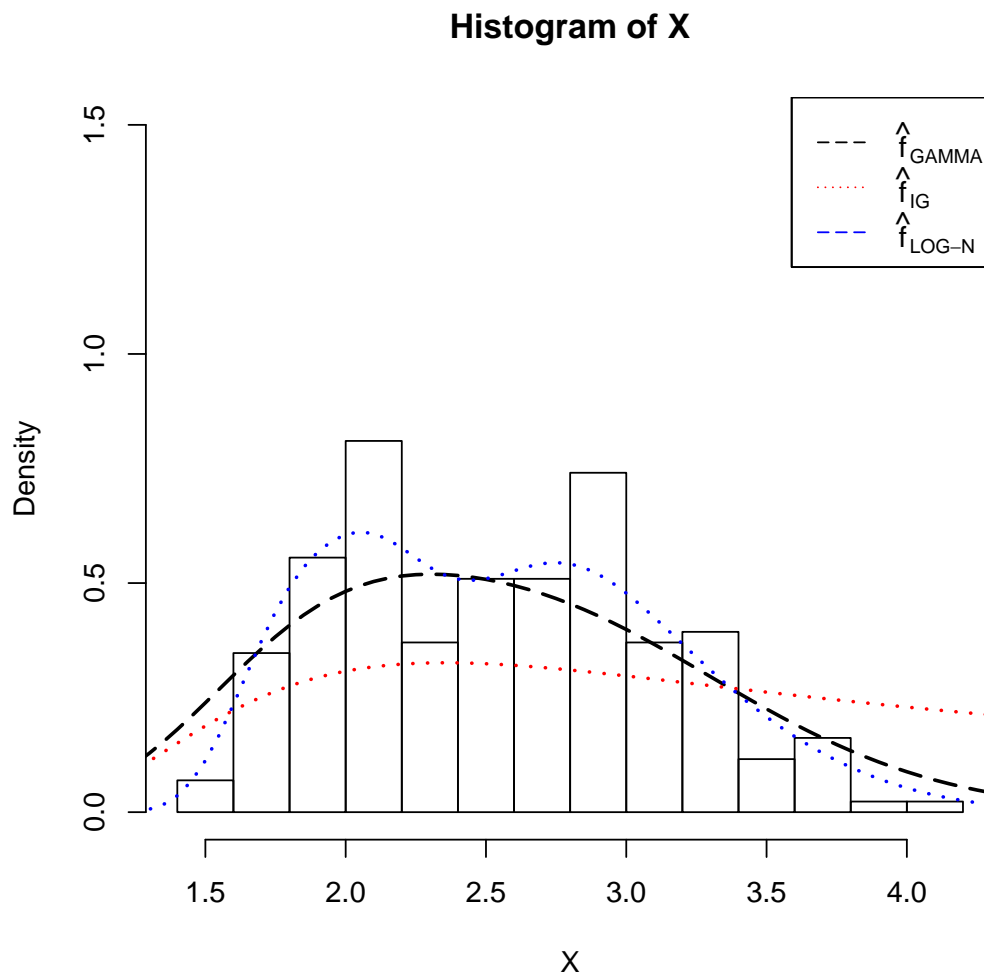


FIGURE 2.3 – Comparaison de quelques estimateurs à noyaux asymétriques avec l'histogramme des donnée SNCF

2.8 Conclusion

Dans ce chapitre, nous avons étudié les performances des estimateurs à noyaux asymétriques dans le cas des données dépendantes. Nous avons calculé l'erreur quadratique moyenne (MSE) d'un processus stationnaire α mélange, et nous avons mené une étude de simulation pour comparer les performances des estimateurs à noyaux IG, RIG, LN, MG, BSPE et BS avec l'estimateur à noyau gamma de Bouezmarni et Rombouts (2010)[9].

Estimation récursive de la densité de probabilité par la méthode du noyau

Sommaire

3.1	Introduction	45
3.2	Estimation récursive de la densité de probabilité : cas des données indépendantes	46
3.2.1	Estimation récursive par noyaux symétriques	46
3.2.2	Propriétés des estimateurs récursifs à noyaux symétriques	47
3.2.3	Méthode d'approximation stochastique pour estimer la densité par la méthode du noyau symétrique	49
3.2.4	Estimateur récursif à noyau asymétrique (cas univarié)	53
3.2.5	Propriétés de l'estimateur récursif à noyaux asymétriques	54
3.3	Estimation récursive de la densité par noyau : cas des données dépendantes	55
3.3.1	Estimateur récursif à noyau symétrique cas multivarié	55
3.3.2	Propriétés de l'estimateur récursif à noyau symétrique	55
3.3.3	Estimateur récursif de la densité par noyaux asymétriques gamma généralisé cas univarié	56
3.4	Estimation récursive de la fonction de masse de probabilité par noyau discret : cas des données de comptage	66
3.4.1	Présentation de l'estimateur récursif à noyau discret	66
3.4.2	Principaux résultats	67
3.4.3	Simulations	68
3.4.4	Temps de calcul	73
3.4.5	Illustrations à partir des données réelles	74
3.5	Conclusion	75

3.1 Introduction

Les estimateurs récursifs à noyau symétrique sont des méthodes d'estimation non paramétrique de densité de probabilité utilisées en traitement du signal et en apprentissage automatique. Ces techniques permettent d'estimer la densité de probabilité à partir d'une série chronologique de données. L'estimation non paramétrique de la densité du noyau avec une version récursive et non récursive est largement utilisée dans de nombreux domaines d'application, tels que l'économie, l'ingénierie, les sciences de l'environnement, etc. Cependant, la version récursive des estimateurs à noyau peut être plus appropriée dans certaines situations, en particulier pour l'estimation de données dépendantes. Le deuxième avantage des estimateurs récursifs par rapport à leurs homologues non récursifs est d'ordre informatique, car elle nécessitait beaucoup moins de calculs pour passer d'un échantillon de taille n à un échantillon de taille $n + 1$. Cette caractéristique est particulièrement cruciale dans l'estimation de la densité, où le nombre de points auxquels la fonction est estimée est généralement élevé. L'estimateur récursif de densité a été introduit à l'origine par Wolverton et Wagner (1969)[85] dans le cas d'un noyau symétrique. Selon la version de Wolverton et Wagner, plusieurs estimateurs récursifs ont été développés par Deheuvels (1974)[19], Wegman et Davies (1979)[84] et Hall Patil (1994)[31]. Plus tard, une grande classe d'estimateurs à noyau symétrique récursif a été proposée par Mokkaïdem et al. (2009)[52] en utilisant la méthode d'approximation stochastique, voir également Slaoui (2019) et Slaoui (2020). Jmaei et al. (2017)[37] ont également proposé l'algorithme d'approximation stochastique et les polynômes de Bernstein pour construire un nouvel estimateur récursif. Par ailleurs, l'étude des propriétés de l'estimateur récursif utilisant un noyau symétrique sous différentes hypothèses de dépendance a été discutée par plusieurs auteurs; voir par exemple Amiri (2009)[3] et Mezhoud et al. (2014)[51]. À notre connaissance, il existe un travail attribué à Kakizawa (2021)[39] sur l'estimateur récursif de correction du biais de frontière utilisant des noyaux asymétriques, et proposé pour des données indépendantes non négatives.

L'objectif de ce chapitre est de proposer des estimateurs récursifs à noyau (discret, asymétrique), pour les données indépendantes et dépendantes. Ensuite, nous examinerons les propriétés de chaque estimateur proposé. Nous menons une analyse de Monte Carlo, basée sur des applications et des simulations réalisées sous R sur des modèles dépendants, mettra en lumière les avantages de l'approche récursive par rapport à l'estimation non récursive en termes de temps de calcul et de ISE.

3.2 Estimation récursive de la densité de probabilité : cas des données indépendantes

3.2.1 Estimation récursive par noyaux symétriques

Soit $(X_i, i \in \mathbb{N})$ une séquence des vecteurs aléatoires indépendants identiquement distribués i.i.d. Les X_i sont des vecteurs dans \mathbb{R}^d , $d \geq 1$ ayant la même loi, de densité de probabilité inconnue f . Pour estimer la densité inconnue de manière récursive pour chaque valeur de x , Wolverton et Wagner (1969)[85] proposent la première forme de l'estimateur récursif à noyau symétrique donné par :

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n h_i^{-d} K \left(\frac{x - X_i}{h_i} \right), \quad K : \mathbb{R}^d \rightarrow \mathbb{R} \quad (3.1)$$

où : h_i est une suite de nombres réels satisfaisants les conditions suivantes :

- a) $h_n \leq \dots \leq h_1 \leq 1$,
- b) $\sum_{i=1}^{\infty} \frac{h_n}{n} < \infty$,
- c) $\sum_{i=1}^{\infty} \frac{1}{h_n^d n^2} < \infty$.

Dans le cas i.i.d, Davies (1973)[17], Deheuvels (1974)[19] et Wegman et Davies (1979) [84] étudient la famille $f_n^H(x)$:

$$f_n^H(x) = \left[\sum_{i=1}^n h_i H(h_i) \right]^{-1} \sum_{i=1}^n H(h_i) K \left(\frac{x - X_i}{h_i} \right). \quad (3.2)$$

L'estimateur obtenu dans l'expression de $f_n^H(x)$, pour $H(u) = 1$, est connu dans la littérature sous le nom d'estimateur de Deheuvels et s'écrit en dimension d sous la forme :

$$f_n^{DHV}(x) = \frac{1}{\sum_{i=1}^n h_i^d} \sum_{i=1}^n K \left(\frac{x - X_i}{h_i} \right), \forall x \in \mathbb{R}^d. \quad (3.3)$$

Aussi, Wegman et Davies (1979) [84] étudient l'estimateur récursif suivant :

$$f_n^{DW}(x) = \frac{1}{n \sqrt{h_n^d}} \sum_{i=1}^n \frac{1}{h_i^{d/2}} K \left(\frac{x - X_i}{h_i} \right), \forall x \in \mathbb{R}^d. \quad (3.4)$$

La dernière versions de l'estimateur récursive à noyau symétrique a été introduite par Amiri (2009)[3] et est donnée par :

$$f_n^l(x) = \frac{1}{\sum_{i=1}^n h_i^{d(1-l)}} \sum_{i=1}^n \frac{1}{h_i^{dl}} K\left(\frac{x - X_i}{h_i}\right), \forall x \in \mathbb{R}^d (l \in [0, 1]). \quad (3.5)$$

Deheuvels (1974)[19] établit la convergence en moyenne quadratique de la famille $f_n^H(x)$ défini par l'équation (3.2) et donne les conditions nécessaires et suffisantes pour ça convergence presque sûre. Wolverton et Wagner (1969)[85] établissent les vitesses de convergence presque sûre exacte dans les cas $l = 0$, $l = 1/2$ et $l = 1$. Aussi, Isogai (1984)[36] établit sous certaines conditions, la normalité asymptotique pour $l = 1$ dans le cas i.i.d.

3.2.2 Propriétés des estimateurs récursifs à noyaux symétriques

Les propositions suivantes, établies par Deheuvels (1973)[18], s'appliquent dans le cas univarié pour les estimateurs $f_n^H(x)$. Amiri (2009)[3] généralise en dimension d quelconque.

Nous considérons également des noyaux K vérifiant les hypothèses :

Hypothèse F.1.

- (i) : $K : \mathbb{R}^d \rightarrow \mathbb{R}$ une densité de probabilité, strictement positive, symétrique et bornée ;
- (ii)

$$\lim_{\|x\| \rightarrow \infty} \|x\|^d K(x) = 0; \forall x \in \mathbb{R}^d,$$

$$\int_{\mathbb{R}^d} |v_i v_j| K(v) < \infty, i, j = 1, \dots, d.$$

Hypothèse F.2.

- (i) $h_n \downarrow 0$ et $nh_n^{d+2} \rightarrow \infty$ lorsque $n \rightarrow \infty$
- (ii) : pour tout $r \in]-\infty, d + 2]$,

$$r \in]-\infty, d + 2] \quad B_{n,r} = \frac{1}{n} \sum_{i=1}^{\infty} \left(\frac{h_i}{h_n}\right)^r \rightarrow B_r < \infty \quad \text{lorsque } n \rightarrow \infty.$$

Proposition 3.2.1. Deheuvels (1973)[18]

On suppose que K est une densité positive bornée sur \mathbb{R} et il existe deux réels $\alpha \geq 2$, $\beta \geq 0$ tel que :

$$\lim_{n \rightarrow \infty} t^\beta \int_{|y|>t} |y|^\alpha K(y) dy = 0, \quad t \in \mathbb{N}$$

et,

f est 2 fois uniformément différentiable sur \mathbb{R} .

Si : $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n h_i^{1-2l}}{(\sum_{i=1}^n h_i^{1-l})^2} = 0$ ou plus généralement $\sum_{i=1}^{\infty} h_i^{1-l} = \infty$, alors :

$$E[f_n^l(x) - f(x)] = \frac{B_{n,2-l}}{B_{n,1-l}} h_n f'(x) \int_{\mathbb{R}} y K(y) dy + \frac{B_{n,3-l}}{B_{n,1-l}} h_n^2 f''(x) \int_{\mathbb{R}} y^2 K(y) dy + o(h_n^2).$$

Proposition 3.2.2. *Deheuvels (1973)[18]*

On suppose que K est une densité positive bornée sur \mathbb{R} et il existe deux réels $\alpha \geq 2$, $\beta \geq 0$ tel que :

$$\lim_{n \rightarrow \infty} t^\beta \int_{|y|>t} |y|^\alpha K(y) dy = 0,$$

Si :

$$\int_{|y|>t} K^p(y) dF(x + uy) - \int_{|y|>t} K^p(y) f(x) dx = o(1),$$

et,

$$\int_{|y|\leq t} K^p(y) [f(x + uy) - f(x)] dy = o(1),$$

$\forall p > 1$ uniformément en $u > 0$ et x , alors :

$$\text{Var}(f_n^l(x)) = \frac{B_{n,1-2l}}{B_{n,1-l}^2 n h_n} f(x) \int_{\mathbb{R}} K^2(y) dy - \frac{B_{n,2(1-l)}}{n B_{n,1-l}^2} f(x)^2 + o(n),$$

car

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=2}^n h_i^{1-2l}}{(\sum_{i=2}^n h_i^{1-l})^2} = 0.$$

Nous pouvons maintenant donner le biais, la variance et la MSE asymptotiques de $f_n^l(x)$ en fonction du paramètre l .

Théorème 3.2.1. *Amiri (2009)[3]*

Sous les hypothèses F.1 et F.2, on a :

(i) pour tout $l \in [0, 1]$,

$$h_n^{-4} E[f_n^l(x) - f(x)]^2 \longrightarrow \left[\frac{B_{d(1-l)+2}}{B_{d(1-l)}} \right]^2 b_f^2(x),$$

lorsque $n \rightarrow \infty$, avec :

$$b_f(x) = \frac{1}{2} \sum_{1 \leq i, j \leq d} \frac{\partial^2 f}{\partial x_i \partial x_j} \int_{\mathbb{R}^d} v_i v_j K(v) dv. \quad (3.6)$$

(ii) pour tout $l \in [0, 1]$,

$$n h_n^d \text{Var}(f_n^l(x)) \rightarrow \delta_l^2(x),$$

lorsque $n \rightarrow \infty$, avec :

$$\delta_l^2(x) = \frac{B_{d(1-2l)}}{B_{1-l}^2} f(x) \int_{\mathbb{R}^d} K^2(x) dx. \quad (3.7)$$

3.2.3 Méthode d'approximation stochastique pour estimer la densité par la méthode du noyau symétrique

Mokkadem et al. (2009)[52] définissent une large classe d'estimateurs récursifs, incluant tous les estimateurs récursifs précédents, voir aussi Jmaei et al. (2017)[37]. Ils appliquent la méthode d'approximation stochastique pour définir une classe d'estimateurs récursifs de la densité. L'utilisation la plus célèbre des algorithmes d'approximation stochastique dans le cadre des statistiques non paramétriques est celle de Kiefer et Wolfowitz (1952)[41], qui ont développé un algorithme permettant d'approcher la fonction de régression maximisant une fonction de régression. Les algorithmes d'approximation stochastique ont été introduits par Révész (1973)[58] pour estimer une fonction de régression et par Tsybakov (1990)[78] pour estimer le mode de la densité de probabilité.

Shéma de Robbins-Monro dans l'estimation récursive de la densité (cas univarié).

Les méthodes d'approximation stochastiques permettent la recherche du zéro θ^* d'une fonction inconnue $\varphi : \mathbb{R} \rightarrow \mathbb{R}$, qui peut être difficilement calculé directement. L'algorithme le plus connu est celui de Robbins et Monro (1951)[60]. On procède de la façon suivante :

- (i) Choisir $\theta_0 \in \mathbb{R}$;
- (ii) Pour $n \geq 1$, on construit la suite (θ_n) par la relation récursive

$$\theta_n = \theta_{n-1} + \gamma_n W_n,$$

où W_n est une observation de la fonction φ au point θ_{n-1} et (γ_n) est une suite de réels positifs qui tend vers zéro appelée pas de l'algorithme.

Soit (X_1, \dots, X_n) un échantillon de densité de probabilité f . Pour construire un estimateur récursif de f en un point x par la méthode des algorithmes stochastiques, Mokkadem et al. (2009)[52] ont défini un algorithme de recherche du zéro de la fonction $\varphi : y \rightarrow f(x) - y$. La procédure est la suivante :

- (i) Soit $f_0(x) \in \mathbb{R}$ fixé ;
- (ii) Pour $n \geq 1$, on a

$$\hat{f}_n(x) = f_{n-1}(x) + \gamma_n W_n(x),$$

où $W_n(x)$ est une observation de la fonction φ au point $f_{n-1}(x)$. On peut estimer $f(x)$ par $Z_n(x) = h_n^{-1}K\left(\frac{x-X_n}{h_n}\right)$, ce qui permet de poser $W_n(x) = Z_n(x) - f_{n-1}(x)$. Ainsi, l'estimateur de la densité f en un point x s'écrit récursivement sous la forme :

$$\hat{f}_n(x) = (1 - \gamma_n)f_{n-1}(x) + \gamma_n h_n^{-1}K\left(\frac{x - X_n}{h_n}\right). \quad (3.8)$$

Notons que si on pose $\gamma_n = (n^{-1})$, alors l'estimateur \hat{f}_n donné par l'algorithme de Robbins et Monro(1951)[60] devient

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n h_i^{-1}K\left(\frac{x - X_i}{h_i}\right). \quad (3.9)$$

dans ce cas, \hat{f}_n est l'estimateur récursif introduit par Wolverton et Wagner (1969)[85]. D'autre part, dans le cas où $\gamma_n = h_n [\sum_{i=1}^n h_i]^{-1}$

$$\hat{f}_n(x) = \frac{1}{\sum_{i=1}^n h_i} \sum_{i=1}^n h_i^{-1}K\left(\frac{x - X_i}{h_i}\right). \quad (3.10)$$

Pour donner les caractéristiques de l'estimateur \hat{f}_n , on introduit une classe de suites à variations régulières que nous utiliserons dans nos hypothèses tout au long des chapitres suivants.

Définition 3.2.1. Soit $\gamma \in \mathbb{R}$ et $(v_n)_{n \geq 1}$ une suite des nombres réels positive. On dit que $(v_n) \in GS(\gamma)$ si

$$\lim_{n \rightarrow \infty} n \left[1 - \frac{v_{n-1}}{v_n} \right] = \gamma. \quad (3.11)$$

La condition (3.11) a été introduite par Galambos et Seneta (1973)[27] pour définir les suites à variations régulières.

Hypothèses

(H1) : $K : \mathbb{R} \rightarrow \mathbb{R}$ est la fonction noyau continue, bornée et vérifiant les propriétés de la définition (3.2.1)

$$\text{(H2)} : \begin{cases} i) & (\gamma_n) \in GS(-\gamma) & \text{avec } \gamma \in]\frac{1}{2}, 1], \\ ii) & (h_n) \in GS(-c) & \text{avec } c \in]0, \frac{c}{2}], \\ iii) & \lim_{n \rightarrow \infty} (n\gamma_n) \in]\min\{2\gamma, \frac{1-\gamma}{2}\}, \infty]. \end{cases}$$

(H3) : f est bornée, deux fois différentiables et f'' est bornée. L'hypothèse **iii)(H2)** implique que la limite de $(n\gamma_n)^{-1}$ est finie. On note alors $\xi = \lim_{n \rightarrow \infty} (n\gamma_n)^{-1}$

(H.4) : pour tout $y \geq 0$, le j^{eme} moment brut $\mu_j(y) = \int_0^\infty \mu^j p(u; y) du$ existe pour $j = 1, 2, 4$, avec

$\sup_{y \geq 0} (\hat{\mu}'_j)/(1+y)^k < \infty$ pour $k = 1, 2$.

(H.5) : Pour $\sup_{x \geq 0} u_h(x) \leq L_k h^{-1}$ pour une constante $L_k > 0$ (indépendante de h), et
 2. Chaque fois que $x > 0$, $u_h(x) \leq L'_k (hx)^{-1/2}$ pour une constante $L'_k > 0$ (indépendante de h et x).

(H.6) : étant donné $\tau \in (0, 1)$, et pour tout suffisamment petit $h > 0$,

$$\int_0^\infty \int_{h^{-\tau}}^\infty K(s, h, x) f(s) = o(h^{\tau(k+1)}) ds dx,$$

à condition que $\int_0^\infty s^{k+1} f(s)$ existe pour une certaine constante $k > 0$.

(H.7) : f est bornée et deux fois continûment différentiable, où f, f' et f'' sont bornés.

(H.8) : $(\gamma_n) \in GS(-\gamma), (h_n) \in GS(-c)$ et $\lim_{n \rightarrow \infty} (n\gamma_n)^{-1} = \xi$ pour une certaine constante $0 < a \leq 1, c > 0$ et $\xi \geq 0$ avec

$$\begin{cases} (i) & 0 < c < 2\gamma \text{ et } 0 \leq \xi < \max\{1/c, \frac{2}{(\gamma-c/2)}\}, \\ (ii) & 0 < c < 2\gamma \text{ et } 0 \leq \xi < \frac{2}{(\gamma-c/2)}, \\ (iii) & 0 < c < 2\gamma \text{ et } 0 \leq \xi < \frac{2}{\gamma}. \end{cases} \quad (3.12)$$

H.9 : f'' satisfait la condition de Holder d'ordre $\eta \in (0, 1)$, c'est-à-dire qu'il existe une constante $L > 0$ telle que $|f''(u) - f''(v)| \leq L|u - v|^\eta$ pour tout $u, v \geq 0$.

Le lemme suivant sera utilisé dans ce chapitre.

Lemme 3.2.1. *Mokkadem et al. (2009) [52]. Soit $h_n \in GS(-c)$ et $\gamma_n \in GS(-\gamma)$ pour certaines constantes $c \in \mathbb{R}$ et $0 < \gamma \leq 1$. Supposons que $\lim_{n \rightarrow \infty} (n\gamma_n)^{-1} = \xi \geq 0$ existe. Etant donné c constante $m > 0$, si $m - c\xi > 0$, que.*

$$\lim_{n \rightarrow \infty} h_n \Pi_n^m \sum_{i=1}^n \Pi_i^{-m} \frac{\gamma_i}{h_i} = \frac{1}{m - c\xi}, \quad \Pi_i^{-m} = \prod_{i=1}^n (1 - \gamma_i)^m$$

et, pour toute constante $c^* \in \mathbb{R}$ et toute séquence (d_n) tel que $d_n \rightarrow 0$.

$$\lim_{n \rightarrow \infty} h_n \Pi_n^m \left(\sum_{i=1}^n \Pi_i^{-m} \frac{\gamma_i}{h_i} d_i + c^* \right) = 0.$$

Nous donnons d'abord le biais et la variance de l'estimateur \hat{f}_n , puis l'erreur quadratique intégrée (MISE) ainsi que le choix théorique du pas et paramètre de lissage qui la minimise. Enfin, on rappelle la vitesse de convergence ponctuelle et uniforme de l'estimateur $\hat{f}_n(x)$.

3.2.3.1 Biais et variance de l'estimateur $\hat{f}_n(x)$

Proposition 3.2.3. *Mokkadem et al. (2009) [52]*

Supposons que les hypothèses (H1)-(H3) soient vérifiées et que f'' soit continue au point x .

1. Si $0 < c \leq \frac{\gamma}{5}$, alors

$$E[\hat{f}_n(x)] - f(x) = \frac{1}{2(1-2c\xi)} h_n^2 \delta_k^2 f''(x) + o(h_n^2),$$

si $c > \frac{\gamma}{5}$, alors

$$E[\hat{f}_n(x)] - f(x) = o\left(\sqrt{\gamma_n h_n^{-1}}\right).$$

2. Si $c > \frac{\gamma}{5}$ alors

$$\text{Var}[\hat{f}_n(x)] = \frac{1}{2(1-2c\xi)} \frac{\gamma_n}{h_n} f(x) \int_R K^2(u) du + o\left(\frac{\gamma_n}{h_n}\right),$$

si $0 < c \leq \frac{\gamma}{5}$, alors

$$\text{Var}[\hat{f}_n(x)] = o(h_n^4). \quad \delta_k^2 = \int_R u^2 K(u) du$$

3.2.3.2 Erreur Quadratique Moyenne Intégrée de l'estimateur $\hat{f}_n(x)$

Dans la proposition suivante, on donne le MISE (Erreur quadratique moyenne intégrée) de l'estimateur \hat{f}_n .

Proposition 3.2.4. *Mokkadem et al. (2009)[52]*

Supposons que les hypothèses (H1)-(H3) soient vérifiées et que f'' soit continue au point x .

1. Si $0 < c \leq \frac{\gamma}{5}$, alors

$$\text{MISE}[\hat{f}_n] = \frac{1}{2(1-2c\xi)} h_n^4 \delta_k^4 \int_R f''(x) dx + o(h_n^4).$$

2. Si $c = \frac{\gamma}{5}$, alors

$$\text{MISE}[\hat{f}_n] = \frac{1}{2(1-2c\xi)} h_n^4 \delta_k^4 \int_R f''(x) dx + \frac{1}{2-(1-c)\xi} \frac{\gamma_n}{h_n} \int_R f(x) dx \int_R K^2(u) du + o\left(h_n^4 + \frac{\gamma_n}{h_n}\right).$$

Si $c > \frac{\gamma}{5}$, alors

$$\text{MISE}[\hat{f}_n] = \frac{1}{2(1-2c\xi)} h_n^4 \delta_k^4 \int_R f''(x) dx + \frac{1}{2-(1-c)\xi} \frac{\gamma_n}{h_n} \int_R f(x) dx \int_R K^2(u) du + o\left(\frac{\gamma_n}{h_n}\right).$$

Jemai et al. (2018)[37] infère le choix optimal du pas qui permet de minimiser le MISE de $\hat{f}_n(x)$ à partir de ce corollaire.

Corollaire 3.4.1. Supposons que les hypothèses (H1)-(H3) soient vérifiées et que f'' soit continue sur \mathbb{R} . Pour minimiser le MISE de \hat{f}_n , le pas γ_n doit être choisi dans $GS(-1)$ et tel que $\lim_{n \rightarrow \infty} n\gamma_n = 1$ et la fenêtre h_n doit être égale à.

$$h_n^* = \left(\left[\frac{3 \int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du}{10 \delta_k^4 \int_{\mathbb{R}} (f''(x))^2 dx} \right]^{\frac{1}{5}} \gamma_n^{\frac{1}{5}} \right).$$

Dans ce cas le MISE de \hat{f}_n est de la forme ;

$$MISE(\hat{f}_n) = \frac{5^{\frac{11}{5}}}{4^{\frac{7}{5}} 3^{\frac{6}{3}}} \left[\delta_K^2 \int_{\mathbb{R}} (f''(x))^2 dx \right]^{\frac{2}{5}} \left[\int_{\mathbb{R}} f(x) dx \int_{\mathbb{R}} K^2(u) du \right]^{\frac{4}{5}} n^{-\frac{4}{5}} + o(n^{-\frac{4}{5}}).$$

3.2.3.3 Vitesse de convergence de l'estimateur \hat{f}_n

On donne dans le Théorème suivant la vitesse de convergence faible de l'estimateur \hat{f}_n .

Théorème 3.2.2. *Mokkadem et al. (2009)[52]* Supposons que les hypothèses (H1)-(H3) soient vérifiées. Pour $x \in \mathbb{R}$ tel que $f(x) > 0$ et f'' soit continue au point x , on a.

1. S'il existe $t \geq 0$ tel que $\frac{h_n^5}{\gamma_n} \xrightarrow[n \rightarrow \infty]{} t$, alors

$$\sqrt{\frac{h_n^5}{\gamma_n}} [\hat{f}_n - f(x)] \xrightarrow[n \rightarrow \infty]{l} N \left(\frac{t^{1/2}}{2(1-2c\xi)} \delta_k^2 f''(x), \frac{1}{2-(1-c)\xi} \int_{\mathbb{R}} K^2(u) du f(x) \right).$$

2. Si $\frac{h_n^5}{\gamma_n} \xrightarrow[n \rightarrow \infty]{} +\infty$, alors

$$\frac{1}{h_n^2} [\hat{f}_n - f(x)] \xrightarrow[n \rightarrow \infty]{P} \frac{t^{1/2}}{2(1-2c\xi)} \delta_k^2 f''(x).$$

3.2.4 Estimateur récursif à noyau asymétrique (cas univarié)

Soit X_1, X_2, \dots, X_n une suite de variables aléatoires indépendantes et identiquement distribuées dont les valeurs sont dans $[0, \infty[$, ayant la même loi de densité de probabilité inconnue f . Pour estimer la densité inconnue de manière récursive pour chaque valeur de x , Kakizawa (2021)[39] a proposé l'estimateur récursif à noyaux asymétriques donné par ;

$$\hat{f}_0(x) = 0, \text{ et } \hat{f}_n(x) = (1 - \gamma_n)\hat{f}_{n-1} + \gamma_n K(X_n; x, h_n) \quad \text{pour } n = 1, 2, \dots, \quad (3.13)$$

où $K(y; x, h)$ est un noyau asymétrique, le pas γ_n et le paramètre de lissage h_n sont des séquences de nombres réels positifs tendant vers zéro lorsque $n \rightarrow \infty$.

3.2.5 Propriétés de l'estimateur récursif à noyaux asymétriques

Les théorèmes suivants sont établis par Kakizawa (2021)[39].

Théorème 3.2.3. *Kakizawa (2021)[39]. Soit $\hat{f}_n(x)$ l'estimateur récursif à noyau asymétrique défini par l'équation (3.13). Pour une cible donnée x , si les hypothèses (H.4), (H.5), (H.7), (H.8.i) et (H.9) sont vérifiées, le biais de l'estimateur récursif à noyau asymétrique est :*

$$\text{Biais}(\hat{f}_n(x)) = \begin{cases} \frac{1}{(1-c\xi)}(h_n B(x)) + o(h_n), & \text{si } 0 < c \leq (2/5)\gamma; \\ o((\gamma_n h_n^{-1/2})^{1/2}), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases}$$

Avec,

$$B(x) = \xi_{1,1} f'(x) + \frac{\xi_{2,1}}{2} x f''(x), \quad \xi_{1,1}, \quad \xi_{2,1} > 0.$$

Théorème 3.2.4. *Kakizawa (2021)[39]. Soit $\hat{f}_n(x)$ l'estimateur récursif à noyau asymétrique défini par l'équation (3.13). Pour une cible donnée x , si les hypothèses (H.4), (H.5), (H.7) et (H.8.i) sont vérifiées, la variance de l'estimateur récursif à noyau asymétrique est :*

$$\text{Var}(\hat{f}_n(x)) = \begin{cases} o(h_n^2), & \text{si } 0 < c \leq (2/5)\gamma; \\ \frac{\gamma_n h_n^{-1/2} V(x)}{2 - (\gamma - c/2)\xi} + o(\gamma_n h_n^{-1/2}), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases}$$

Avec,

$$V(x) = \xi \frac{f(x)}{x^{1/2}}, \quad \xi > 0.$$

Théorème 3.2.5. *Kakizawa (2021)[39]. Soit $\hat{f}_n(x)$ l'estimateur récursif de densité utilisant le noyau asymétrique défini par l'équation (3.13). Pour une cible donnée x , et si les hypothèses (H.4), (H.5), (H.7), (H.8.ii) et (H.9) sont vérifiées, la normalité asymptotique de l'estimateur récursif à noyau asymétrique est :*

Si $(2/5)\gamma < c < 2\gamma$, $\lim_{n \rightarrow \infty} (\gamma_n^{-1} h_n^{\frac{5}{2}}) = 0$ alors :

$$\frac{\hat{f}_n(x) - E(\hat{f}_n(x))}{\sqrt{\text{Var}(\hat{f}_n(x))}} \rightarrow^d N(0, (2 - (\gamma - c/2)\xi)^{-1} V(x)).$$

Si $c = \frac{2}{5}\gamma$, $\lim_{n \rightarrow \infty} (\gamma_n^{-1} h_n^{\frac{5}{2}}) = \omega \geq 0$ alors :

$$\frac{\hat{f}_n(x) - E(\hat{f}_n(x))}{\sqrt{\text{Var}(\hat{f}_n(x))}} \rightarrow^d N(\omega^{\frac{1}{2}}(2 - (1 - \frac{2}{5}\gamma)\xi)^{-1} B(x), (2 - (1 - \frac{2}{5}\gamma)\xi)^{-1} V(x)).$$

Théorème 3.2.6. *Kakizawa (2021)[39]. Si les hypothèses (H.4), (H.6), (H.7) et (H.8.i) sont vérifiées, et si $\int_0^\infty \{f'(x)\}^2 dx$, $\int_0^\infty \{x f''(x)\}^2 dx$, et $\int_0^\infty x^{k+1} f(x) dx$ existent pour une certaine constante $k > (6 + \eta)/\eta$, où $\eta \in (0, 1)$. Alors, la MISE de \hat{f}_n est :*

$$MISE(\hat{f}_n) = \begin{cases} \frac{(h_n I^B)^2}{(1-c\xi)} + o(h_n^2), & \text{si } 0 < c \leq (2/5)\gamma, \\ \frac{(h_n I^B)^2}{\{1-(2/5)\gamma\xi\}^2} + \frac{\gamma_n h_n^{-1/2}}{2-(4/5)\gamma\xi} I^V + o(h_n^2 + \gamma_n h_n^{-1/2}), & \text{si } c = (2/5)\gamma, \\ \frac{\gamma_n h_n^{-1/2}}{2-(\gamma-c)\xi} I^V + o(\gamma_n h_n^{-1/2}) + o(\gamma_n h_n^{-1/2}), & \text{si } (2/5)\gamma < c < (2\gamma). \end{cases}$$

3.3 Estimation récursive de la densité par noyau : cas des données dépendantes

3.3.1 Estimateur récursif à noyau symétrique cas multivarié

Soit X_1, \dots, X_n un processus stationnaire α -mélange avec une fonction de densité inconnu f . Pour estimer la densité inconnue de manière récursive pour chaque valeur de x , Amiri (2009)[3] proposé un estimateur récursif à noyau symétrique donné par l'équation (3.5). Remarquons que la famille (3.5) contient les estimateurs les plus utilisés $f^{WW}(x)$, $f^{DHV}(x)$ et la version asymptotiquement non biaisée de $f^{DW}(x)$, qui correspondent respectivement aux cas où $l = 0$, $l = 1$ et $l = 1/2$. Elle possède également l'avantage de s'exprimer sous une forme simple permettant la comparaison des différents estimateurs.

3.3.2 Propriétés de l'estimateur récursif à noyau symétrique

Le théorème suivant établi par Amiri (2009)[3] s'applique dans le cas multivarié pour les estimateurs f_n^l . Il a établi le MSE de $f_n^l(x)$ en fonction du paramètre l .

Théorème 3.3.1. *Amiri (2010)[3]*

Sous les hypothèses F.1, F.2 et le condition (2.7) :

(i) *pour tout $l \in [(\frac{d-2}{2d})^+, 1]$:*

$$nh_n^d \text{Var}(f_n^l(x)) \rightarrow \delta_l^2(x),$$

lorsque $n \rightarrow \infty$, $\beta > 2$, avec δ_l^2 défini par l'équation (3.7).

(ii) *Si $d \geq 3$ et $l \in [0, \frac{d-2}{2d}]$, la conclusion de (i) reste encore vraie si $\beta > \frac{d+2}{2}$*

(iii) *pour tout $l \in [(\frac{d-2}{2d})^+, 1]$, (resp. Pour tout $l \in [0, \frac{d-2}{2d}]$, si $d \geq 5$), si $\beta > 2$ (resp. $\beta > \frac{d+2}{2}$), le choix :*

$$h_n = C_n n^{-\frac{1}{d+4}}, C_n \downarrow c > 0,$$

entraîne que :

$$n^{\frac{4}{d+4}} E[f_n^l(x) - f(x)]^2 \rightarrow c^4 \left(\frac{4 + dl}{2 + dl} \right)^2 b_f^2(x) + \frac{(4 + dl)^2 f(x) \|K\|_2^2}{2c^d (4 + d)(2 + dl)},$$

lorsque $n \rightarrow \infty$, pour les valeurs respectives de l en tout point où $f(x) > 0$, où $b_f(x)$ est défini par l'équation (3.6).

L'estimation récursive par noyau de la densité sous dépendance faible établie par Mazhoud et al (2010)[51] est applicable dans le cas univarié pour les estimateurs f_n^l d'Amiri (2009)[3]. Leur étude détaille les erreurs quadratiques moyennes asymptotiques de $f_n^l(x)$ en fonction du paramètre l . Leurs applications et simulations, réalisées sur des modèles à dépendance faible (décalage de Bernoulli), mettent en évidence les avantages de l'approche récursive par rapport à l'estimation classique, notamment en termes de temps de calcul et de variance.

3.3.3 Estimateur récursif de la densité par noyaux asymétriques gamma généralisé cas univarié

Présentation de l'estimateur

Soit X_1, \dots, X_n un processus stationnaire α mélange non négatif avec une densité inconnu f au point x . Notre objectif est d'estimer de manière non paramétrique et récursive la densité $f(x)$ de support $[0, \infty[$. L'estimateur récursive de la densité basé sur les noyaux GG de Hirukawa et Sakudo (2015)[35] est :

$$\hat{f}_0(x) = 0, \text{ et } \hat{f}_n(x) = (1 - \gamma_n)\hat{f}_{n-1}(x) + \gamma_n K_{GG}(X_n; x, h_n) \text{ pour } n = 1, 2, \dots, \tag{3.14}$$

où le pas γ_n et le paramètre de lissage h_n sont des séquences de nombres réels positifs tendant vers zéro lorsque $n \rightarrow \infty$.

$$K_{GG}(X_i; x, h) = \frac{\delta X_i^{\kappa-1} \exp[-\{\frac{X_i}{\beta\Gamma(\kappa/\delta)/\Gamma((\kappa+1)/\delta)}\}^\delta]}{\{\beta\Gamma(\kappa/\delta)/\Gamma((\kappa+1)/\delta)\}^\kappa \Gamma(\kappa/\delta)}, \quad x \geq 0, \tag{3.15}$$

est le noyau GG avec $(\kappa, \beta, \sigma) = (\kappa_h(x), \beta_h(x), \sigma_h(x))$ sont des fonctions continues du point de conception x qui satisfait certaines conditions (voir Hirukawa et Sakudo (2015)[35] pour plus de détails). Dans cette étude, nous considérons exclusivement trois cas de noyaux GG (noyaux MG, NK et W) présentés dans le tableau 3.1.

TABLEAU 3.1 – Noyaux gamma généralisés (MG, NK et W).

Noyau	Forme explicite	Paramètres
MG	$K_{MG}(y; x, h) = \frac{y^{\kappa-1} \exp\{-y/(\beta/\kappa)\}}{(\beta/\kappa)^\kappa \Gamma(\kappa)}, y \geq 0.$	$(\kappa, \beta, \delta) = \begin{cases} (\frac{x}{h}, x, 1) & \text{pour } x \geq 2h; \\ (\frac{x^2}{4h^2} + 1, \frac{x^2}{4h} + h, 1) & \text{pour } x \in [0, 2h). \end{cases}$
NK	$K_{NK}(y; x, h) = \frac{2(\kappa/2)^{\kappa/2}}{[(\kappa/2)\beta\Gamma(\kappa/2)/\Gamma((\kappa+1)/2)]^2 \Gamma(\kappa/2)} y^{2(\kappa/2)-1} \times \exp[-\frac{\kappa/2}{(\kappa/2)[\beta\Gamma(\kappa/2)/\Gamma((\kappa+1)/2)]^2} y^2], y \geq 0.$	$(\kappa, \beta, \delta) = \begin{cases} (\frac{x}{h}, x, 2) & \text{pour } x \geq 2h; \\ (\frac{x^2}{4h^2} + 1, \frac{x^2}{4h} + h, 2) & \text{pour } x \in [0, 2h). \end{cases}$
W	$K_W(y; x, h) = \frac{\kappa}{\beta\Gamma(1+1/\kappa)} \left[\frac{y}{\beta\Gamma(1+1/\kappa)} \right]^{\kappa-1} \times \exp[-[\beta/\Gamma(1+1/\kappa)]^\kappa], \quad y \geq 0$	$(\kappa, \beta, \delta) = \begin{cases} (\sqrt{\frac{x}{h}}, x, \sqrt{\frac{x}{h}}) & \text{pour } x \geq 2h; \\ (\frac{x}{2h} + 1, \frac{x^2}{4h} + h, \frac{x}{2h} + 1) & \text{pour } x \in [0, 2h). \end{cases}$

3.3.3.1 Propriétés de convergence

Nous rappelons maintenant quelques conditions, pour se faire on utilise d’une part de la définition (3.2.1) et d’une autre part le coefficient de mélange fort défini par l’équation (2.4) et l’équation(2.7), (pour plus de détails voir Bouezmarni et Rombouts (2010)[9]).

Proposition 3.3.1. *Soit $\widehat{f}_n(x)$ l’estimateur récursif à noyau GG défini par l’équation (3.14). Supposons que les hypothèses (H.3), (H.8) et (H.9) sont vérifiés. Nous avons, pour un $x > 0$ donné,*

(i) *le biais de l’estimateur récursif à noyau GG est :*

$$Biais(\widehat{f}_n(x)) = \begin{cases} \frac{C}{2(1-c\xi)}xf''(x)h_n + o(h_n), & \text{si } 0 < c \leq (2/5)\gamma, \\ o((\gamma_n h_n^{-1/2})^{1/2}), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases} \tag{3.16}$$

(ii) *La variance de l’estimateur récursif à noyau GG est :*

$$Var(\widehat{f}_n(x)) = \begin{cases} o(h_n^2), & \text{si } 0 < c < (2/5)\gamma; \\ \frac{V}{2-(\gamma-c/2)\xi}\gamma_n h_n^{-1/2}x^{-1/2}f(x) + o(\gamma_n h_n^{-1/2}), & \text{si } (2/5)\gamma \leq c < 2\gamma. \end{cases} \tag{3.17}$$

Où les constantes C et V sont données dans le tableau 3.2 pour les noyaux MG, NK et W.

TABLEAU 3.2 – Les valeurs de C et V pour les noyaux MG, NK et W.

noyau	C	V
MG	1	$1/(2\sqrt{\pi})$
W	$\pi^2/12$	$1/(2\sqrt{2})$
NK	1/2	$1/(\sqrt{2\pi})$

Preuve de la proposition 3.3.1. *Premièrement, le biais est démontré, comme dans le cas indépendant et identiquement distribué (i.i.d.), pour l’estimateur récursif à noyau asymétrique plus général (voir Kakizawa (2021)[39]). Deuxièmement, la variance peut être établie même pour des données dépendantes non négatives, en combinant les résultats de Bouezmarni et Rombouts (2010)[9], Hirukawa et Sakudo (2015)[35] et Kakizawa (2021)[39].* ■

3.3.3.2 Critère MISE

Nous utilisons l’erreur quadratique moyenne intégrée (MISE) comme critère de la propriété globale. Si $IB^2 = \int_0^\infty \{xf''(x)\}^2 dx$ et $IV = \int_0^\infty \{x^{-\frac{1}{2}}f(x)\}dx$ existent, alors le MISE asymptotique (AMISE) de l’estimateur récursif à noyau GG est :

$$AMISE(\widehat{f}_n) = \begin{cases} \frac{h_n^2 C}{(1-c\xi)^2} IB^2 + o(h_n^2), & \text{si } 0 < c < (2/5)\gamma; \\ \frac{h_n^2 C}{(1-\frac{2}{5}\gamma\xi)^2} IB^2 + \frac{\gamma_n h_n^{-1/2} v}{2-(4/5)\gamma\xi} IV + o(h_n^2 + \gamma_n h_n^{-1/2}), & \text{si } c = (2/5)\gamma. \\ \frac{\gamma_n h_n^{-1/2} V}{2-(\gamma-c)\xi} IV + o(\gamma_n h_n^{-1/2}) + o(\gamma_n h_n^{-1/2}), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases}$$

Notons que pour la minimisation d'AMISE(\widehat{f}_n), la constante c doit être imposée comme $c = \frac{2}{5}a$. La valeur du paramètre de lissage qui minimise l'AMISE(\widehat{f}_n) est :

$$h_n^* = \left[\frac{V(1 - \frac{2}{5}\gamma\xi)}{2C^2} \gamma_0 \right]^{2/5} \left[\frac{\int_0^\infty x^{-\frac{1}{2}} f(x) dx}{\int_0^\infty \{x f''(x)\}^2 dx} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \tag{3.18}$$

Où $(\gamma_n) = (\gamma_0 n^{-1})$, c'est-à-dire $\lim_{n \rightarrow \infty} n \gamma_n = \gamma_0$. Par conséquent, l'AMISE optimale est :

$$AMISE^{opt}(\widehat{f}_n) = \frac{5}{4} \left[\left[\frac{\gamma_0 V}{2} \right]^2 C (1 - \frac{2}{5}\gamma\xi)^{-3} \right]^{\frac{2}{5}} \left[\int_0^\infty \{x f''(x)\}^2 dx \left[\int_0^\infty x^{-\frac{1}{2}} f(x) dx \right]^4 \right]^{\frac{1}{5}} n^{-\frac{4}{5}}. \tag{3.19}$$

3.3.3.3 Applications numériques

Le but principal de cette section, est d'évaluer les performances de l'estimateur récursif proposé en utilisant les noyaux : MG, W et NK. Pour ce faire, une étude de simulation est réalisée, basée sur les échantillons des variables aléatoires dépendantes générées à partir d'un modèle de durée conditionnelle autorégressive (DCA) et d'un modèle de volatilité stochastique (VS) (pour plus de détails, voir Bouezmarni et Rombouts (2010)[9]). Considérons la densité empirique des deux modèles (DCA et VS) sur la base de 100000 observations. Nous avons comparé l'estimateur récursif proposé pour les noyaux MG, W et NK avec l'estimateur non récursif à noyaux MG, W et NK, en utilisant différentes tailles d'échantillon $n = 100, 250, 500$ et 1000 et pour un nombre de réplifications $N = 250$. Cette évaluation des performances est basée sur le critère de comparaison Integrated Squared Error (ISE) défini par :

$$ISE := \int (f_n(x) - f_{emp}(x))^2 dx, \tag{3.20}$$

où : f_n est l'estimateur récursif ou non récursif du noyau, et $f_{emp}(x)$ est la fonction de densité empirique défini par l'équation (2.13).

la taille du pas est fixée par $\gamma_n = \frac{1}{n}$ et le paramètre de lissage est sélectionné par $h_n = hn^{-\frac{2}{5}}$: où h est choisi en utilisant l'approche classique de validation croisée non-biaisée (UCV).

Durées conditionnelles autorégressives

Le modèle DCA est défini par :

$$\begin{cases} x_i = \psi_i \epsilon_i; \\ \psi_i = \omega + \alpha_1 x_{i-1} + \alpha_2 \psi_{i-1}, \end{cases}$$

où : $\epsilon_i, \psi_i, i = 1, \dots, n$ sont des variables aléatoires i.i.d qui peuvent être générées à partir de densités connues. Dans notre cas, nous avons considéré trois densités : exponentielle, log-normale et gamma, présentées dans le tableau 3.3. Nous considérons trois niveaux de dépendance : forte, faible et indépendance. Les valeurs des paramètres pour ces niveaux sont données dans le tableau 2.2. Les erreurs quadratiques intégrées moyennes (sur 250 répétitions) pour le modèle DCA, où ϵ_i est généré à partir de $G(4,91,4,91)$, $LN(0,01, 0,3)$ et $Exp(1,02)$ est indiqué dans les tableaux 3.4, 3.5 et 3.6.

	DCA-D1	DCA-D2	DCA-D3
ϵ_i	Gamma($\alpha = 4.91, \beta = 4.91$)	Log-normal($\sigma = 0.01, \mu = 0.3$)	Exponentiel ($\lambda = 1.02$)
Densities	$x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)}$	$\frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} \exp\left(\frac{-1}{2} \left(\frac{\log x - \mu}{\sigma}\right)^2\right)$	$\lambda e^{-\lambda x}$

TABLEAU 3.3 – Les densités d’innovation.

- Les valeurs du ISE moyen diminuent lorsque la taille de l’échantillon augmente pour toutes les méthodes.
- Dans le cas où $\epsilon \sim \text{gamma}(\mathbf{D1})$, on remarque que : l’estimateur non récursif est meilleur que l’estimateur récursif dans le cas des données dépendantes.

		DCA-D1				
Niveau	Estimateur	$n=100$	250	500	1000	
Dépendance forte	f_{MG-REC}	0.03966	0.02760	0.00838	0.00212	
	f_{MG}	0.03601	0.02644	0.00847	0.00215	
	f_{W-REC}	0.03918	0.02767	0.00821	0.00210	
	f_W	0.03587	0.02577	0.00837	0.00253	
	f_{NK-REC}	0.04037	0.02825	0.008572	0.00243	
	f_{NK}	0.03803	0.02838	0.00841	0.00251	
Dépendance faible	f_{MG-REC}	0.00659	0.00273	0.00045	0.00021	
	f_{MG}	0.00650	0.00279	0.00043	0.00022	
	f_{W-REC}	0.00662	0.00271	0.00046	0.00019	
	f_W	0.00707	0.00268	0.00056	0.00023	
	f_{NK-REC}	0.00716	0.002912	0.00059	0.00026	
	f_{NK}	0.00706	0.00285	0.00057	0.00028	
Indépendance	f_{MG-REC}	0.00157	0.00048	0.00025	0.00015	
	f_{MG}	0.00092	0.00041	0.00019	0.00011	
	f_{W-REC}	0.00165	0.00050	0.00023	0.00017	
	f_W	0.00125	0.00043	0.00015	0.00009	
	f_{NK-REC}	0.00184	0.00052	0.00033	0.00023	
	f_{NK}	0.00179	0.00047	0.00024	0.00019	

TABLEAU 3.4 – Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour DCA-D1.

Niveau	Estimateur	DCA-D2			
		$n = 100$	250	500	1000
Dépendance forte	f_{MG-REC}	0.03732	0.02511	0.00778	0.00195
	f_{MG}	0.03405	0.02509	0.00817	0.00207
	f_{W-REC}	0.03715	0.02618	0.00791	0.00188
	f_W	0.03651	0.02513	0.00803	0.00203
	f_{NK-REC}	0.04173	0.02925	0.009572	0.00343
	f_{NK}	0.03711	0.02718	0.00841	0.00321
Dépendance faible	f_{MG-REC}	0.00673	0.00285	0.00048	0.00023
	f_{MG}	0.00732	0.00360	0.00055	0.00029
	f_{W-REC}	0.00669	0.00266	0.00044	0.00021
	f_W	0.00731	0.00358	0.00053	0.00027
	f_{NK-REC}	0.00731	0.00322	0.00063	0.00030
	f_{NK}	0.00801	0.00384	0.00072	0.00038
Indépendance	f_{MG-REC}	0.00169	0.00051	0.00028	0.00017
	f_{MG}	0.00218	0.00058	0.00033	0.00025
	f_{W-REC}	0.00173	0.00054	0.00029	0.00019
	f_W	0.00227	0.00061	0.00035	0.00027
	f_{NK-REC}	0.00202	0.00058	0.00035	0.00025
	f_{NK}	0.00259	0.00063	0.00039	0.00031

TABLEAU 3.5 – Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour **DCA-D2**.

- Le cas où $\epsilon \sim \text{log-normal}(\mathbf{D2})$, on remarque que : l'estimateur récursif est meilleur que l'estimateur non récursif avec les noyaux MG, NK et W, sauf pour un niveau de dépendance forte pour une taille d'échantillon $n = 100, 250$.
- Dans le cas où $\epsilon \sim \text{exponentiel}(\mathbf{D3})$, on remarque que : l'estimateur récursif est meilleur que l'estimateur non récursif pour $n = 1000$.

La figure 3.1 présente les graphiques des estimations de la densité basées sur une seule estimation d'un modèle DCA simulé dans le cas où $\epsilon \sim \text{gamma}(4.92, 4.92)$, en utilisant les estimateurs récursifs et non récursifs pour les noyaux MG et W. Les résultats sont donnés pour la taille de l'échantillon $n = 500$ et 1000000 observations pour la densité empirique, Ils montrent que l'estimateur récursif est meilleur que l'estimateur non récursif pour une dépendance forte et faible.

Volatilité stochastique

Le modèle VS est défini par :

$$\begin{cases} y_i = \exp\left(\frac{x_i}{2}\right)\mu_i; \\ x_i = \gamma + \delta x_{i-1} + \nu\epsilon_i, \end{cases}$$

où : μ_i et ϵ_i $i = 1, \dots, n$ sont des variables aléatoires, i.i.d et peuvent être générées à partir de densité connue présentées dans le tableau 3.8.

Nous considérons trois niveaux de dépendance : forte, faible et indépendance. Les valeurs des paramètres pour ces niveaux sont données dans le tableau 2.6.

Nous pouvons voir que pour le modèle VS, l'estimateur récursif est meilleur que l'estimateur non récursif pour tous les niveaux dépendances forte et faible pratiquement pour tous les noyaux considérés. Le tableau 3.7 illustre les résultats.

		DCA-D3			
Niveau	Estimateur	$n=100$	250	500	1000
Dépendance forte	f_{MG-REC}	0.04212	0.03312	0.01430	0.00961
	f_{MG}	0.04012	0.03192	0.01450	0.01020
	f_{W-REC}	0.04175	0.03268	0.01531	0.00986
	f_W	0.03651	0.02513	0.01470	0.01060
	f_{NK-REC}	0.04192	0.03315	0.01605	0.01056
	f_{NK}	0.03861	0.03138	0.01631	0.01132
Dépendance faible	f_{MG-REC}	0.00691	0.00421	0.00313	0.00238
	f_{MG}	0.00682	0.00405	0.00320	0.00241
	f_{W-REC}	0.00689	0.00419	0.00308	0.00237
	f_W	0.00706	0.00422	0.00314	0.00025
	f_{NK-REC}	0.00727	0.00426	0.00319	0.00246
	f_{NK}	0.00710	0.00405	0.00312	0.00235
Indépendance	f_{MG-REC}	0.00215	0.00095	0.00056	0.00023
	f_{MG}	0.00208	0.00091	0.00059	0.00026
	f_{W-REC}	0.00223	0.00107	0.00062	0.00027
	f_W	0.00219	0.00101	0.00066	0.00029
	f_{NK-REC}	0.00242	0.00112	0.00069	0.00033
	f_{NK}	0.00245	0.00107	0.00065	0.00037

TABLEAU 3.6 – Les valeurs de ISE moyenne (\overline{ISE}) basées sur 250 réplifications pour DCA-D3.

		VS			
Niveau	Estimateur	$n=100$	250	500	1000
Dépendance forte	f_{MG-REC}	0.05620	0.04507	0.03984	0.02624
	f_{MG}	0.06351	0.05415	0.04533	0.03513
	f_{W-REC}	0.05161	0.04733	0.03736	0.02569
	f_W	0.05873	0.05536	0.04601	0.03285
	f_{NK-REC}	0.06202	0.04871	0.04002	0.03219
	f_{NK}	0.06500	0.05638	0.04955	0.04012
Dépendance faible	f_{MG-REC}	0.04716	0.03656	0.02504	0.01204
	f_{MG}	0.05854	0.04125	0.03109	0.01962
	f_{W-REC}	0.04563	0.03623	0.02373	0.01301
	f_W	0.05412	0.04562	0.02947	0.02013
	f_{NK-REC}	0.05532	0.04237	0.02746	0.01421
	f_{NK}	0.06218	0.05592	0.03579	0.02129
Indépendance	f_{MG-REC}	0.00865	0.00621	0.00319	0.00158
	f_{MG}	0.00968	0.00669	0.00321	0.00172
	f_{W-REC}	0.00893	0.00612	0.00306	0.00156
	f_W	0.01023	0.00656	0.00315	0.00168
	f_{NK-REC}	0.01095	0.00730	0.00375	0.00187
	f_{NK}	0.01032	0.00727	0.00380	0.00191

TABLEAU 3.7 – les valeurs de ISE moyenne (ISE) basées sur 250 réplifications pour VS, le number en gras le ISE le plus petit

VS	
μ_i	Student(5)
ϵ_i	normal($\mu = 0, \sigma = 1$)

TABLEAU 3.8 – Les densités d’innovation.

La figure 3.2 présente les graphiques des estimations de la densité basées sur une seule estimation d’un modèle VS simulé en utilisant les estimateurs récursifs et non récursifs pour les noyaux MG et NK avec trois niveaux de dépendance. Les résultats sont donnés pour une taille d’échantillon $n = 500$ et $100\ 000$ observations pour la densité empirique. Ils montrent que l’estimateur récursif est

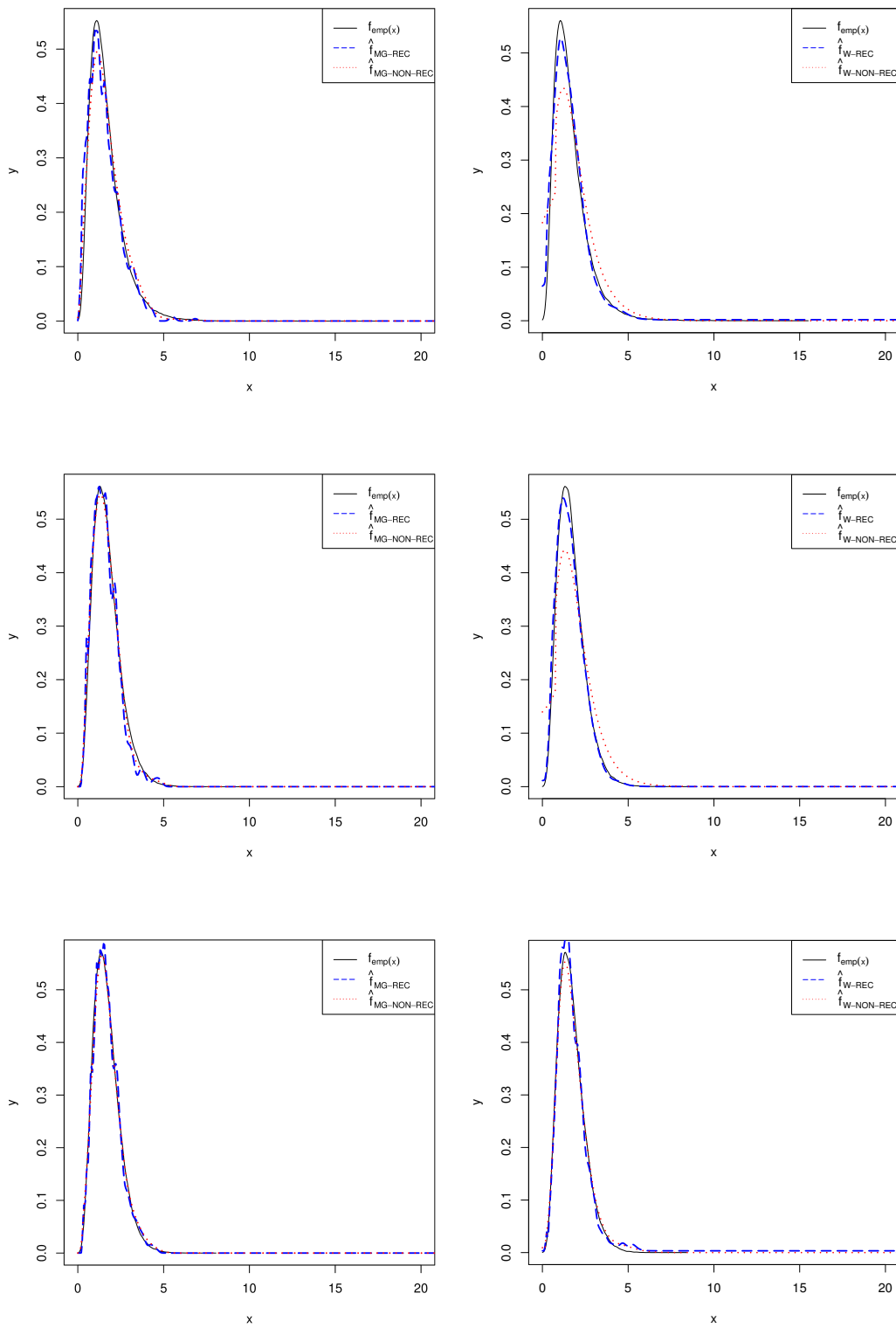


FIGURE 3.1 – Les estimateurs récursifs et non récursifs pour les noyaux MG et W pour le modèle DCA (D1); dépendance forte (en haut), dépendance faible (au milieu) et indépendance (en bas).

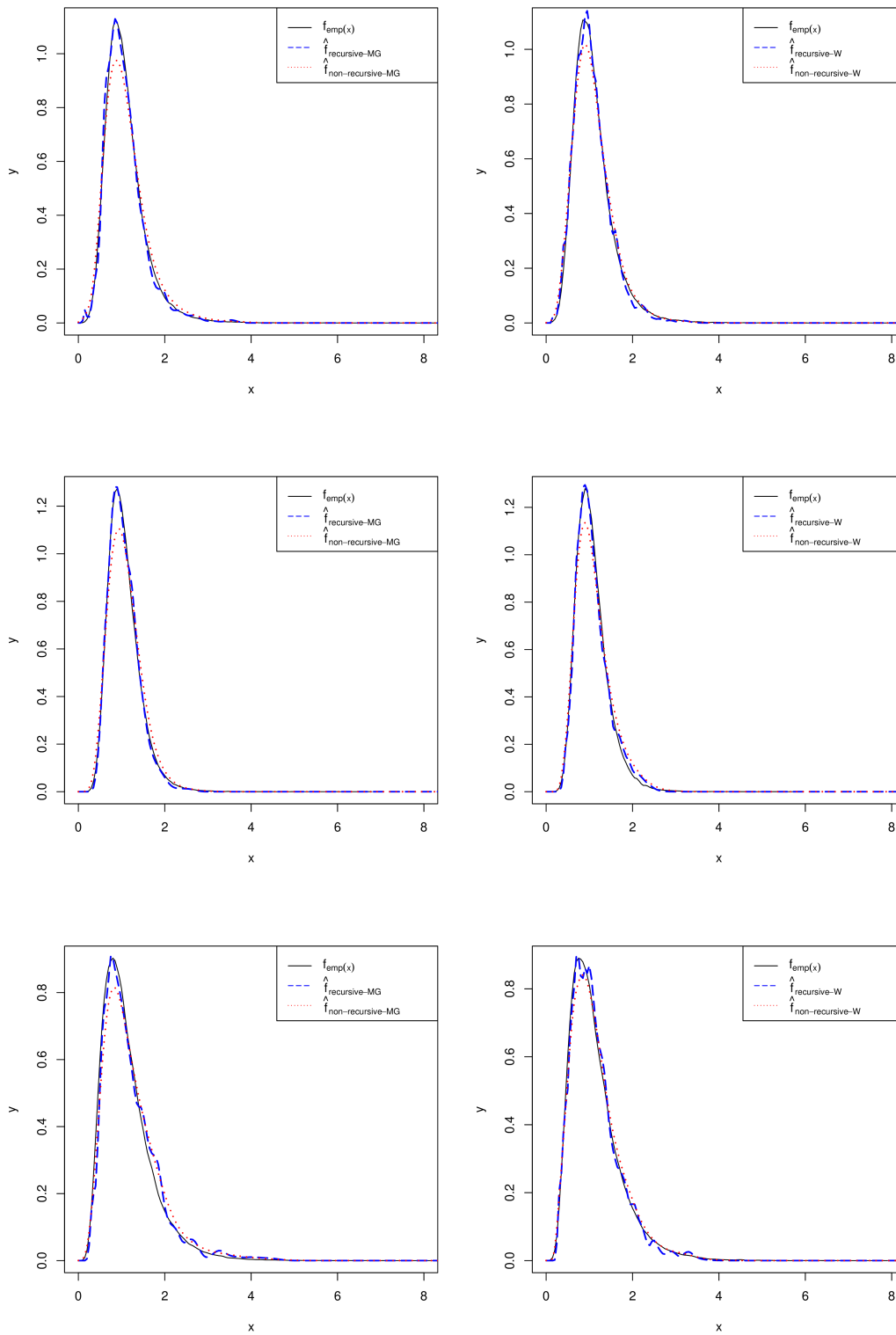


FIGURE 3.2 – Les estimateurs récursifs et non récursifs pour les noyaux MG et W pour le modèle VS; dépendance forte (en haut), dépendance faible (au milieu) et l'indépendance (en bas).

meilleure que l'estimateur non récursif pour une dépendance forte et faible.

3.3.3.4 Temps de calcul

Nous avons effectué une analyse comparative des temps de calcul nécessaires pour évaluer à la fois les estimateurs récursifs et non récursifs utilisant l'approche UCV pour la sélection des paramètres de lissage. Dans un premier temps, nous avons fixé un échantillon $\{X_1, \dots, X_{200}\}$ du modèle DCA (D1) pour dériver les paramètres basés sur la méthode UCV; en particulier, $h_n = h_{CV,n}$ pour le cas non récursif et $h_i = h_{CV,n}i^{-2/5}$ pour le cas récursif. Ensuite, en générant des observations supplémentaires X_{n+1}, \dots, X_{n+N} , nous avons calculé les estimateurs à noyau MG sur la base de l'échantillon complet $\{X_1, \dots, X_{n+m}\}$ pour chaque $m = 1, \dots, N$. Avec $(n, N) = (200, 200)$, la figure 3.3 illustre le temps de calcul (en secondes) nécessaires pour calculer les estimateurs non récursifs (ligne continue) et les estimateurs récursifs (ligne pointillée) en utilisant $h_{n+m} = h_{CV,n+m}$ et $h_i = h_{CV,n}i^{-2/5}$, respectivement. Ces calculs ont été exécutés sur un ordinateur personnel standard fonctionnant à 3.0 GHz avec 8GB RAM en utilisant le logiciel R.

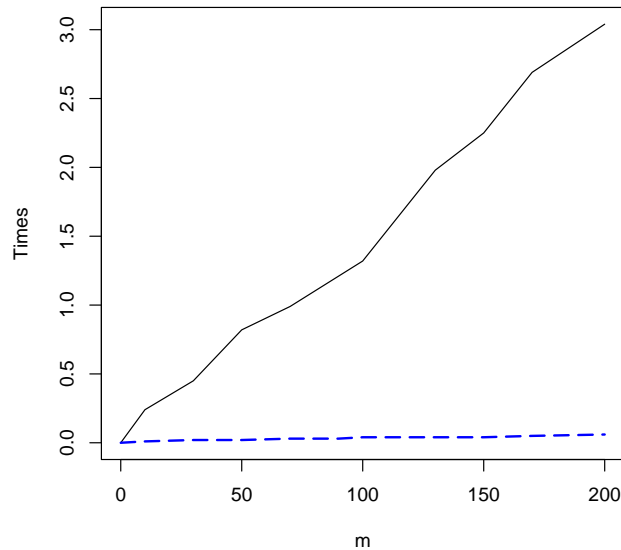


FIGURE 3.3 – Temps de calcul (en secondes) des estimateurs non récursif à noyau MG (en trait plein)/récursif (en pointillé) pour le modèle DCA (D1) à partir de $\{X_1, \dots, X_{200+m}\}$.

Remarque 3.3.1. Il est évident que la vitesse de l'estimateur récursif à noyau asymétrique est très élevée.

3.3.3.5 Application sur les données réelles

Nous analysons deux ensembles de données réelles. Tout d’abord, nous utilisons les estimateurs récursifs et non récursifs par les noyaux asymétriques MG et W pour estimer la densité de l’indice des prix à la consommation. Ces données consistent en $n = 107$ observations et représentent l’augmentation relative de l’indice des prix à la consommation entre les mois de février 1970 et décembre 1978, (France). La deuxième application réelle concerne le trafic de passagers de deuxième classe de la SNCF (Société nationale des chemins de fer français). Les observations mensuelles ($n = 216$) portent sur la période (1963 – 1980) ; voir Gouriéroux et Monfort (1990)[29]. Pour les estimateurs récursifs par les noyaux asymétriques MG et W, la taille du pas est fixée à $\gamma_n = \frac{1}{n}$, et le paramètre de lissage est sélectionné par $h_n = hn^{-\frac{2}{5}}$, où h est choisi par la méthode UCV. Pour les non récursifs à noyau asymétrique MG et W, le paramètre de lissage est également obtenu par la méthode UCV.

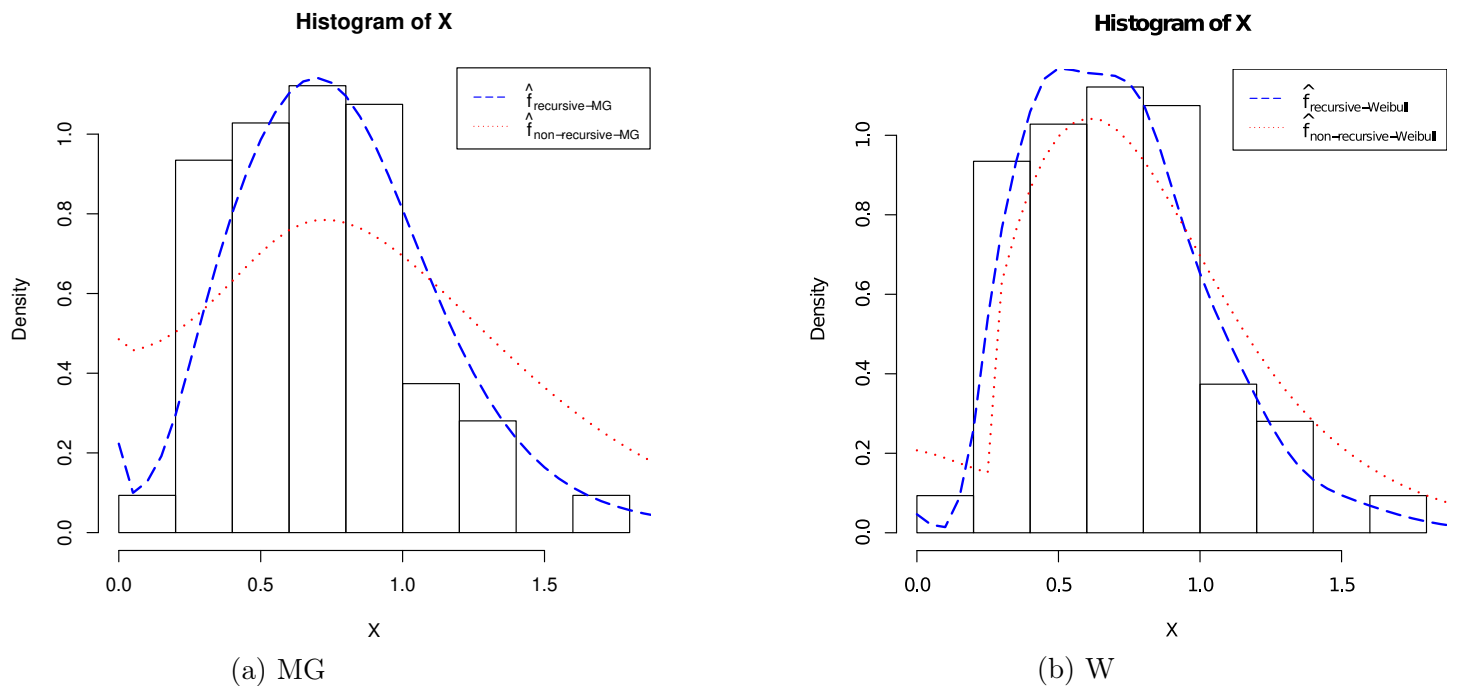


FIGURE 3.4 – Les estimateurs récursifs et non récursifs par les noyau MG et W pour l’indice des prix à la consommation

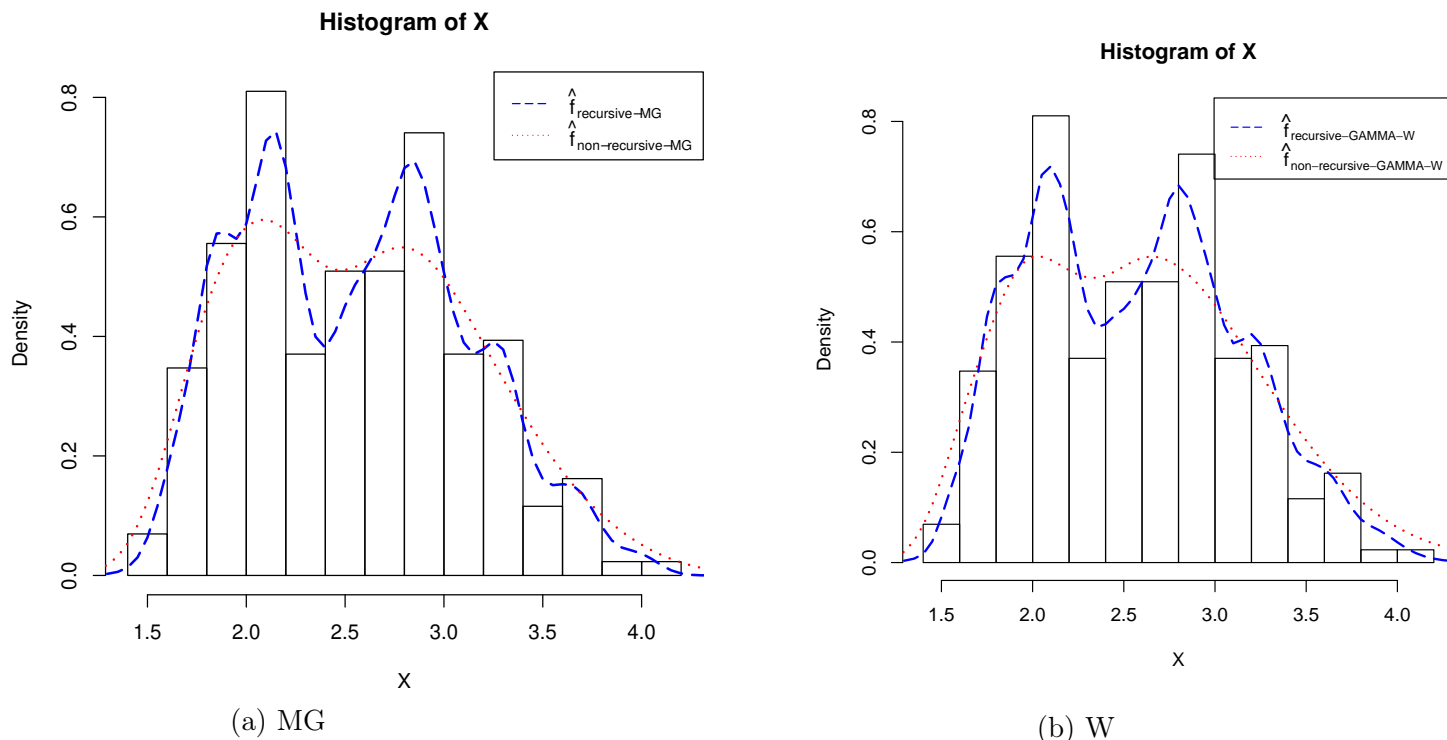


FIGURE 3.5 – Les estimateurs récursifs et non récursifs par les noyau MG et W pour le trafic voyageurs de la SNCF.

Les figures 3.5 et 3.4 montrent les estimateurs récursifs et non récursifs par les noyaux MG et W pour les données de l'indice des prix à la consommation et le trafic de passagers de la SNCF en deuxième classe, respectivement. Nous pouvons constater que les estimateurs récursifs et non récursifs sont capables de reproduire l'uni-modalité et la bi-modalité des données. Ces applications à des ensembles de données réelles montrent les bonnes performances de l'estimateur récursif GG.

3.4 Estimation récursive de la fonction de masse de probabilité par noyau discret : cas des données de comptage

3.4.1 Présentation de l'estimateur récursif à noyau discret

Considérons X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes et identiquement distribuées de fonction de masse de probabilité inconnue f à support \mathbb{T} (\mathbb{N}, \mathbb{Z} où $\mathbb{T} = 0, 1, \dots, N$), pour estimer $f(x)$, où $x \in \mathbb{T}$, nous fixons $f_0(x) = 0$, et

$$\tilde{f}_i(x) = (1 - \gamma_i)\tilde{f}_{i-1}(x) + \gamma_i K(X_i; x, h_i) \quad \text{pour } i = 1, 2, \dots, \quad (3.21)$$

où : le pas γ_n et la paramètre de lissage h_n sont des séquences de nombres réels positifs qui tendent vers zéro lorsque $n \rightarrow \infty$.

3.4.2 Principaux résultats

Proposition 3.4.1. *Soit $\tilde{f}_n(x)$ l'estimateur récursif de la densité utilisant le noyau discret défini par l'équation (3.21). Si les hypothèses (H.2) et (H.8) sont vérifiées. Nous avons, pour un $x \in \mathbb{T}$ donné,*

$$\text{Biais}(\tilde{f}_n(x)) = \begin{cases} \frac{1}{(1-c\xi)}(h_n q(x, f)) + o(h_n), & \text{si } 0 < c \leq (2/5)\gamma; \\ o(\gamma_n h_n), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases}$$

où $q(x, f)$ est donné dans le tableau 1.8

$$\text{Var}(\tilde{f}_n(x)) = \begin{cases} o(\gamma_n), & \text{si } 0 < c < (2/5)\gamma; \\ \frac{\gamma_n}{2-(\gamma-c/2)\xi}(f(x)\{1-f(x)\})K_{L(x, h_n)}^2(X) + o(\gamma_n), & \text{si } (2/5)\gamma \leq c < 2\gamma. \end{cases}$$

Preuve de la proposition 3.4.1. La preuve découle directement de la combinaison des résultats d'Aboubacar et Kokonendji (2024)[1] et de Kokonendji et Senga Kiessé (2011)[44]. ■

3.4.2.1 Critère de MISE

L'erreur quadratique moyenne intégrée (MISE) de l'estimateur récursif de la fonction de masse de probabilité à noyau discret est :

$$\text{MISE}(\tilde{f}_n) = \begin{cases} \frac{h_n^2 \sum_{x \in \mathbb{T}} \phi_{(x, f)}^2}{(1-c\xi)^2} + o(h_n^2), & \text{si } 0 < c \leq (2/5)\gamma; \\ \frac{h_n^2 \sum_{x \in \mathbb{T}} \phi_{(x, f)}^2}{\{1-(2/5)\gamma\xi\}^2} + \frac{\gamma_n}{2-(4/5)\gamma\xi} \left\{ \sum_{x \in \mathbb{T}} f(x)(1-f(x))K_{L(x, h_n)}^2(X) \right\} + o(h_n^2 + \gamma_n), & \text{si } c = (2/5)\gamma; \\ \frac{\gamma_n}{2-(\gamma-c/2)\xi} \left\{ \sum_{x \in \mathbb{T}} f(x)(1-f(x))K_{L(x, h_n)}^2(X) \right\} + o(\gamma_n), & \text{si } (2/5)\gamma < c < 2\gamma. \end{cases}$$

En minimisant $\text{MISE}(\tilde{f}_n)$ pour l'estimateur récursif à noyau discret, la constante c doit être choisie comme $c = (2/5)\gamma$. En outre, en laissant $\gamma_n = (\gamma_0 n^{-1})$ pour une constante $\gamma_0 > 0$, le paramètre de lissage (h_n) qui minimise $\text{MISE}(\tilde{f}_n)$ est :

$$\hat{h}_n = \arg \min_{h > 0} \{\text{MISE}(\tilde{f}_n)\}. \quad (3.22)$$

Nous notons ici que l'erreur quadratique intégrée moyenne $\text{MISE}(\tilde{f}_n)$ dépend des fonctions inconnues f , $f^{(1)}$, $f^{(2)}$ et du noyau K_{x, h_n} . Par conséquent, le développement d'une formule globale pour le paramètre de lissage optimal h_n est impossible. Pour pallier cet inconvénient, nous avons décidé de sélectionner le paramètre de lissage par $h_n = h n^{-1}$, où h est choisi selon la technique de validation croisée sans biais (UCV).

Proposition 3.4.2. *Soit $\tilde{f}_n(x)$ l'estimateur récursif de la fonction de masse de probabilité à noyau défini par l'équation (3.21). Pour une cible donnée x , si les hypothèses (H.2) et (H.8) sont vérifiées,*

la normalité asymptotique de l'estimateur récursif de la fonction de masse de probabilité à noyau est :

Si $(2/5)\gamma < c < 2\gamma$ $\lim_{n \rightarrow \infty} (\gamma_n^{-1} h_n^{\frac{5}{2}}) = 0$ alors :

$$\frac{\tilde{f}_n(x) - E(\tilde{f}_n(x))}{\sqrt{\text{Var}(\tilde{f}_n(x))}} \rightarrow^d N(0, (2 - (\gamma - c/2)\xi)(f(x)\{1 - f(x)\}K_{x,h}^2(x))),$$

Si $c = \frac{2}{5}\gamma$, $\lim_{n \rightarrow \infty} (\gamma_n^{-1} h_n^{\frac{5}{2}}) = \omega \geq 0$ alors :

$$\frac{\tilde{f}_n(x) - E(\tilde{f}_n(x))}{\sqrt{\text{Var}(\tilde{f}_n(x))}} \rightarrow^d N(\omega^{\frac{1}{2}}(2 - (1 - \frac{2}{5}\gamma)\xi)^{-1}q(x, f), (2 - (1 - \frac{2}{5}\gamma)\xi)^{-1}(f(x)\{1 - f(x)\}K_{x,h}^2(x))).$$

Preuve de la Proposition 3.4.2. La preuve découle directement de la combinaison des résultats d'Aboubacar et Kokonendji (2024)[1], Kokonendji et Senga Kiessé (2011)[44]. ■

3.4.3 Simulations

Une étude de simulation à été menée pour évaluer l'efficacité de l'estimateurs récursifs à noyaux discrets (WVR, DDU, DT, LR et DOSK). Cette étude est basée sur des échantillons de variables aléatoires indépendantes et identiquement distribuées (i.i.d) générés suivant plusieurs fonctions de masse de probabilités discrètes, voir par exemple Harfouche et al. (2018)[33]. Nous avons comparé les estimateurs récursifs à noyau discrets (WVR, DDU, DT, LR et DOSK) proposés avec les estimateurs non récursifs à noyau discrets (WVR, DDU, DT, LR et DOSK), en utilisant différentes tailles d'échantillon $n = 50, 100, 300, 500$ et 1000 avec un nombre de réplifications $N_{rep} = 500$. Les performances des estimateurs récursifs et non récursifs à noyau discrets (WVR, DDU, DT, LR et DOSK) sont examinées à l'aide du critère de l'erreur quadratique intégrée (ISE) défini comme suit :

$$ISE := \sum_{x \in N} \{\bar{f}_n(x) - f(x)\}^2; \tag{3.23}$$

où : \bar{f}_n est l'estimateur récursif ou non récursif à noyau discret et $f(x)$ est la fonction de masse de probabilité. Le pas est fixé à $\gamma_n = \frac{1}{n}$, et le paramètre de lissage est choisi par $h_n = h_{ucv}n^{-1}$, h_{ucv} est choisi à l'aide de la technique de validation croisée sans biais (UCV). Pour les estimateurs non récursifs à noyau discret, le paramètre de lissage est choisi par la technique de validation croisée sans biais.

Nous considérons six fonctions de masse de probabilité définies comme suit :

(a) **D1** une densité de Poisson avec le paramètre, $\lambda = 8$.

$$f(x) = \frac{8^x}{x!} e^{-8} \quad x \in \mathbb{N}.$$

(b) **D2** un mélange de trois densités de Poisson avec les paramètres $\lambda = 3$, $\lambda = 12$ et $\lambda = 24$.

$$f(x) = \frac{3^x}{x!}e^{-3} + \frac{(12)^x}{x!}e^{-12} + \frac{(24)^x}{x!}e^{-24} \quad x \in \mathbb{N}.$$

(c) **D3** une densité géométrique avec le paramètre $\rho = 0.1$

$$f(x) = (0.1) e^{(x-1)} \quad x \in \mathbb{N}.$$

(d) **D4** est un mélange de distributions de Poisson et géométriques avec les paramètres $\lambda = 10$ et $\rho = 0.1$.

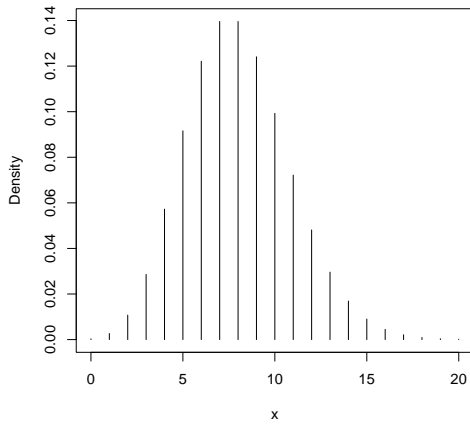
$$f(x) = \frac{10^x}{x!}e^{-10} + (0.1)e^{(x-1)} \quad x \in \mathbb{N}.$$

(e) **D5** une densité binomiale négative avec les paramètres $n = 20$ et $\rho = 2/3$.

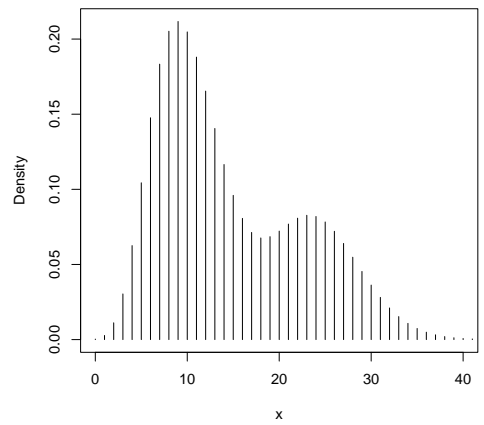
$$f(x) = \frac{(19+x)!}{x!19!} \left(\frac{2}{3}\right)^{20} \left(\frac{1}{3}\right)^x$$

(f) **D6** une densité binomiale avec les paramètres $n = 5$ et $\rho = 0.1$.

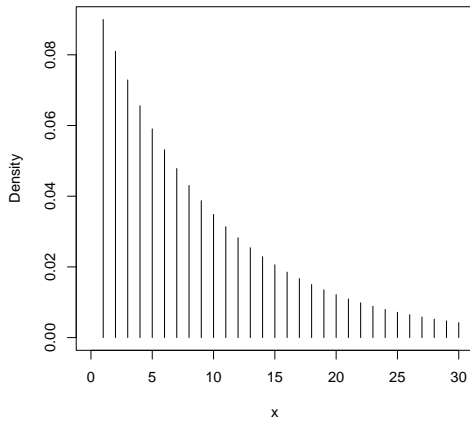
$$f(x) = \frac{5!}{x!(5-x)!} (0.1)^x \cdot (0.9)^{(5-x)} \quad x \in \{0, 1, \dots, 5\}$$



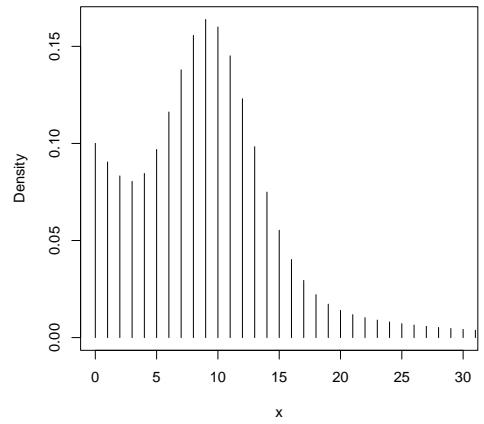
(a) D1



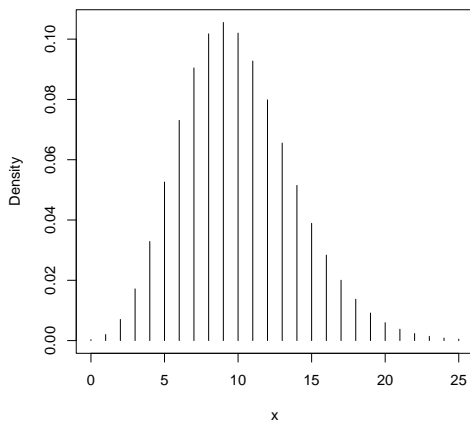
(b) D2



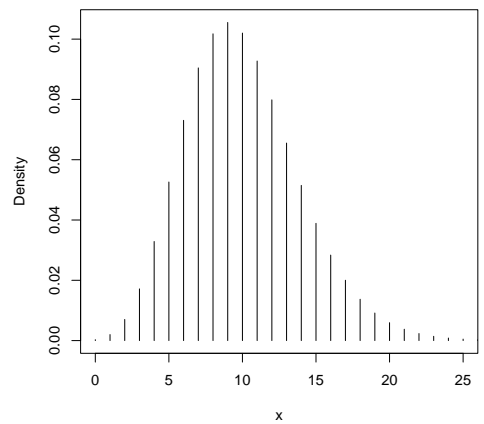
(c) D3



(d) D4



(e) D5



(f) D6

FIGURE 3.6 – La fonction de masse de probabilité (pmf).

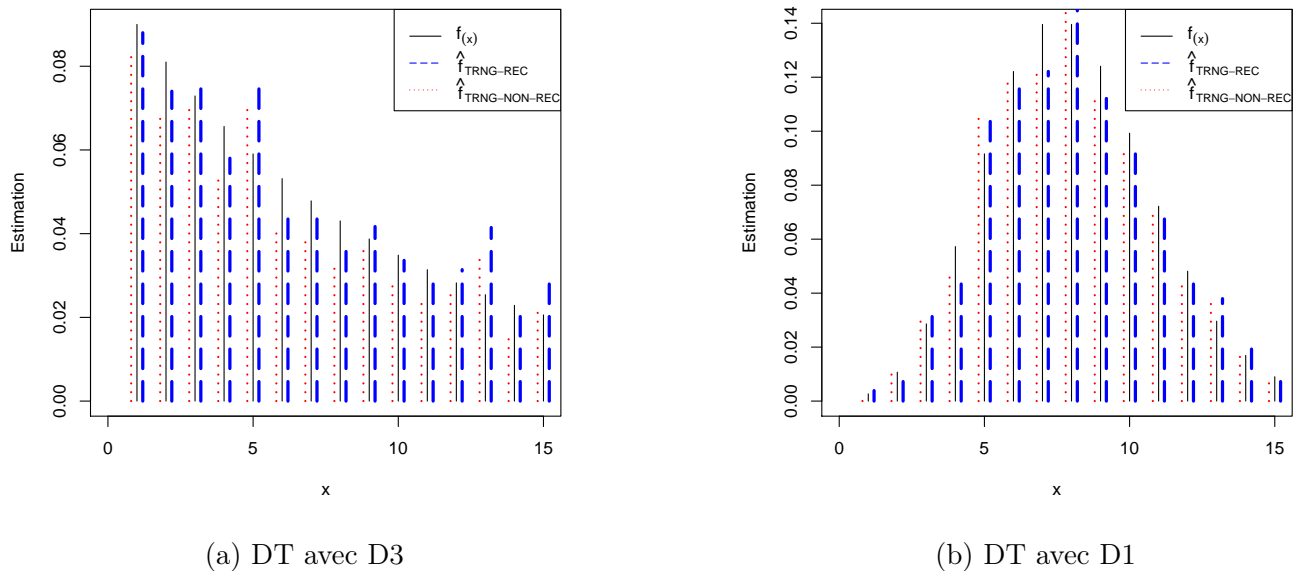


FIGURE 3.7 – Estimation pmf de données poissonniennes et géométriques avec $n = 500$ en utilisant l’estimateur récursif et non récursif à noyau discret DT($a = 2$).

La figure 3.7 présente les graphiques des estimations pour la fonction de masse de probabilité discrète (D1, D3) en utilisant l’estimateur récursif et non récursif à noyau triangulaire (DT avec $a = 2$). Les résultats sont donnés pour une taille d’échantillon $n = 500$. La figure montre que la qualité du lissage obtenue par l’estimateur récursif pour la fonction de masse de probabilité discrète (D1, D3) est meilleure que celle de l’estimateur non récursif. L’estimateur récursif est meilleur que l’estimateur non récursif dans le cas des données de comptage.

Les résultats sont présentés dans les tableaux 3.11 et 3.12. Nous pouvons constater que :

- * Les valeurs du ISE moyen diminuent à mesure que la taille de l’échantillon augmente pour tous les estimateurs.
- * Pour n grand, nous pouvons observer que l’estimateur récursif est meilleur que l’estimateur non récursif.
- * Il est évident que l’estimateur récursif donne de meilleurs résultats que l’estimateur non récursif pour les noyaux LR, DT et WVR, à l’exception de la densité D6.
- * Nous pouvons observer que l’estimateur non récursif est meilleur que l’estimateur récursif pour les noyaux DOSK et DDU.

TABLEAU 3.9 – Statistiques sommaires des données sur le choix du mode de transport

Catégories	Air	Train	Bus	Voiture	Total
Numéro choisissant	58	63	30	59	210
p	0.28	0.30	0.14	0.28	1

TABLEAU 3.10 – Données de longévité des insectes adultes observés en jours.

Jours	1	2	3	4	5	6	7	8	9	Total
Fréquences observées	29	16	22	8	2	4	0	0	1	82

TABLEAU 3.11 – ISE×1000 moyenne des noyaux discrets, avec $(\gamma_n) = (1/n)$.

Modèle	Noyau	n=50	n=100	n=300	n=500	n=1000
D1	DOSK-REC	1.070885	0.782153	0.254062	0.161280	0.075439
	DOSK-NON-REC	0.115462	0.226666	0.028851	0.014642	0.007868
	WVR-REC	0.106533	0.028247	0.006829	0.001586	0.000018
	WVR-NON-REC	0.904893	0.052116	0.026920	0.026230	0.006858
	LR-REC	1.291807	0.375850	0.030452	0.001368	0.000425
	LR-NON-REC	4.531648	1.565053	0.319136	0.154435	0.008018
D2	DOSK-REC	0.791484	0.284462	0.090633	0.052015	0.029146
	DOSK-NON-REC	0.308141	0.047789	0.006557	0.002644	0.001176
	WVR-REC	2.028765	1.028473	0.307973	0.178956	0.084453
	WVR-NON-REC	0.174374	0.101851	0.029736	0.019704	0.009892
	LR-REC	2.668544	1.190547	0.350131	0.187839	0.088816
	LR-NON-REC	6.382475	2.688518	0.827380	0.451326	0.214763
D3	DOSK-REC	0.272924	0.258900	0.083007	0.046970	0.029015
	DOSK-NON-REC	0.264000	0.069168	0.012925	0.004189	0.001264
	WVR-REC	0.418818	0.73784	0.111160	0.087513	0.002297
	WVR-NON-REC	0.299320	0.054167	0.004032	0.010205	0.004153
	LR-REC	0.492650	0.128700	0.022970	0.008742	0.003962
	LR-NON-REC	1.650800	0.609960	0.151430	0.054292	0.041507
D4	DOSK-REC	1.698286	0.344394	0.142203	0.061901	0.037935
	DOSK-NON-REC	0.798637	0.095777	0.030411	0.004858	0.003980
	WVR-REC	0.711365	0.180390	0.020010	0.016547	0.009437
	WVR-NON-REC	0.474203	0.116573	0.016700	0.0146490	0.008540
	LR-REC	0.783367	0.271730	0.028728	0.019490	0.003988
	LR-NON-REC	2.501586	0.970332	0.182951	0.113086	0.044337
D5	DOSK-REC	3.022705	0.409805	0.173936	0.121183	0.006051
	DOSK-NON-REC	1.340444	0.010231	0.011167	0.003874	0.000595
	WVR-REC	0.437006	0.444562	0.033828	0.009901	0.004472
	WVR-NON-REC	0.302034	0.423796	0.036711	0.008199	0.008783
	LR-REC	1.228516	0.103028	0.021926	0.001110	0.000601
	LR-NON-REC	3.938572	0.850126	0.263450	0.126916	0.061486
D6	DOSK-REC	0.151812	0.013052	0.004709	0.003706	0.002180
	DOSK-NON-REC	2.623111	0.963111	0.420123	0.196984	0.099080
	WVR-REC	2.500506	1.38014	0.519008	0.291521	0.122194
	WVR-NON-REC	1.185435	1.004197	0.296462	0.156256	0.064864
	LR-REC	2.716215	0.826873	0.488865	0.261704	0.145554
	LR-NON-REC	3.043221	0.113932	0.136929	0.124570	0.041506

TABLEAU 3.12 – ISE×1000 moyenne des noyaux discrets, avec $(\gamma_n) = (1/n)$.

Modèle	Noyau	n=50	n=100	n=300	n=500	n=1000
D1	DDU-REC	80.38538	74.41650	69.16628	54.54546	42.70712
	DDU-NON-REC	1.106441	0.929253	0.506661	0.356700	0.189654
	DT-REC	5.294644	2.706687	1.338549	0.225223	0.000383
	DT-NON-REC	3.462416	2.512117	2.161361	0.2381364	0.002150
D2	DODU-REC	80.07992	74.134498	58.72824	53.65726	41.65353
	DDU-NON-REC	1.356065	0.599272	0.233446	0.161883	0.009798
	DT-REC	3.901649	1.101240	0.359552	0.194347	0.094931
	DT-NON-REC	2.902801	1.011110	0.364273	0.210888	0.117054
D3	DDU-REC	80.08322	74.15399	68.80260	53.88910	41.96592
	DDU-NON-REC	0.556604	0.255273	0.109009	0.070625	0.046299
	DT-REC	1.431746	0.145778	0.016559	0.008249	0.004074
	DT-NON-REC	1.032827	0.146866	0.096719	0.071004	0.041650
D4	DDU-REC	80.71639	74.99684	69.85007	55.71490	44.73805
	DDU-NON-REC	6.919494	3.751645	1.701841	1.431081	1.253994
	DT-REC	1.529044	0.379873	0.454098	0.219860	0.068774
	DT-NON-REC	0.807493	0.344176	0.104122	0.094338	0.012989
D5	DDU-REC	80.30123	74.35315	69.05373	54.30937	42.26751
	DDU-NON-REC	0.838461	0.797002	0.352401	0.271131	0.327579
	DT-REC	1.51872	0.375587	0.199411	0.054222	0.015444
	DT-NON-REC	0.990168	0.344030	0.256319	0.045285	0.011029
D6	DDU-REC	80.04697	74.09574	68.71693	53.65734	41.77817
	DDU-NON-REC	0.833805	0.218997	0.078697	0.049532	0.032473
	DT-REC	2.206054	1.081984	0.538374	0.250160	0.141519
	DT-NON-REC	2.339445	1.085930	0.538970	0.250773	0.141729

3.4.4 Temps de calcul

Nous avons effectué une analyse comparative des temps de calcul nécessaires pour évaluer à la fois les estimateurs récursifs et non récursifs utilisant l’approche UCV pour la sélection des paramètres de lissage. Dans un premier temps, nous avons fixé un échantillon $\{X_1, \dots, X_{100}\}$ du noyau DT avec innovation de poisson (D1) pour dériver les paramètres basés sur l’approche UCV ; plus précisément, $h_n = h_{CV,n}$ dans le cas non récursif et $h_i = h_{CV,n}i^{-1}$ dans le cas récursif. Ensuite, en générant des observations supplémentaires X_{n+1}, \dots, X_{n+N} , nous avons calculé les estimateurs à noyau discret DT basés sur le sous-échantillon $\{X_1, \dots, X_{n+k}\}$ pour chaque $k = 1, \dots, N$. Dans le contexte de $(n, N) = (100, 250)$, la figure 3.8 illustre les temps de calcul (en secondes) nécessaires pour calculer les estimateurs non récursifs (ligne continue) et les estimateurs récursifs (ligne pointillée) en utilisant $h_{n+k} = h_{CV,n+k}$ et $h_i = h_{CV,n}i^{-1}$, respectivement. Ces calculs ont été exécutés sur un ordinateur personnel standard fonctionnant à 3.0 GHz avec 8GB RAM en utilisant le logiciel R.

Remarque 3.4.1. Il est évident que la vitesse de l’estimateur récursif à noyau DT est très élevée.

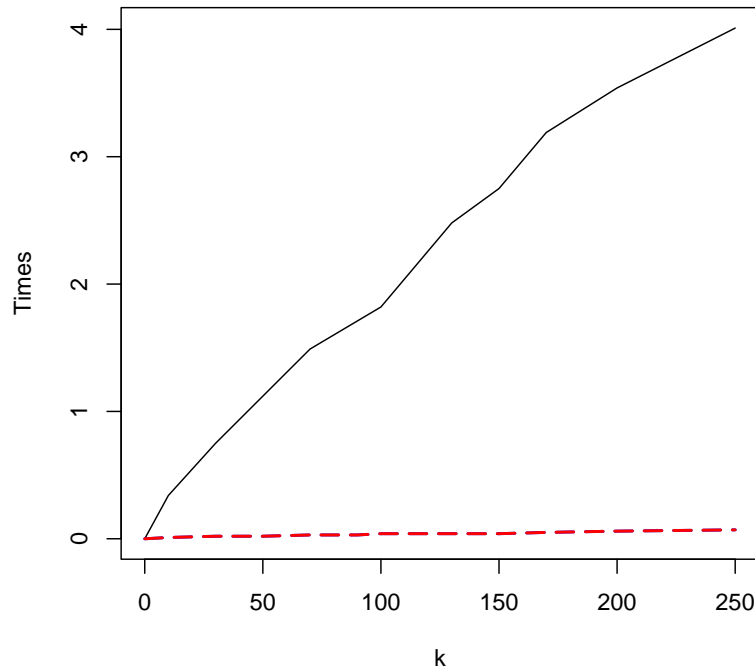


FIGURE 3.8 – Temps de calcul (en secondes) de l’estimateur à noyau DT non récursif (en trait plein)/récursif (en pointillé) pour le modèle **D1** à partir de $\{X_1, \dots, X_{100+k}\}$.

3.4.5 Illustrations à partir des données réelles

Pour enrichir nos simulations de Monte-Carlo, cette section examine deux applications de données réelles. Tout d’abord, nous présentons les performances des estimateurs récursifs et non récursifs cas discrets, basés sur le noyau $DT(a = 2)$. La première application concerne le choix du mode de transport entre Sydney et Melbourne en Australie, tel que décrit par Baker et Greene (2011) [4]. Les données comprennent 210 observations réparties en 4 catégories : avion, train, bus et voiture, avec des proportions relatives de 0,28, 0,30, 0,14 et 0,28 respectivement. La figure 3.9 montre une comparaison de ces observations.

La deuxième application concerne le développement d’un insecte parasite appelé "mouche blanche spiralée", observé en République tchèque et décrit par Senga Kiessé and Mizère (2013)[42] en République du Congo. Cet insecte ravageur provoque des dégâts tels que la succion de la sève, la réduction de l’activité de photosynthèse et le dessèchement des feuilles des plantes. Les biologistes congolais recherchent un modèle approprié en étudiant les données de comptage caractérisant la croissance en spirale des aleurodes ainsi que la longévité des insectes adultes, comme le montre le tableau 3.10.

La figure 3.9 illustre l'estimateur récursif et non récursif à noyau DT pour les deux applications réelle. Nous pouvons voir que les estimateurs récursifs et non récursifs sont capables de reproduire l'uni-modalité et la bi-modalité des données. Cependant, le paramètre de lissage fourni par les estimateurs récursifs à noyau DT est plus adapté. Ces applications à des ensembles de données réelles montrent les bonnes performances d'estimateur récursif DT.

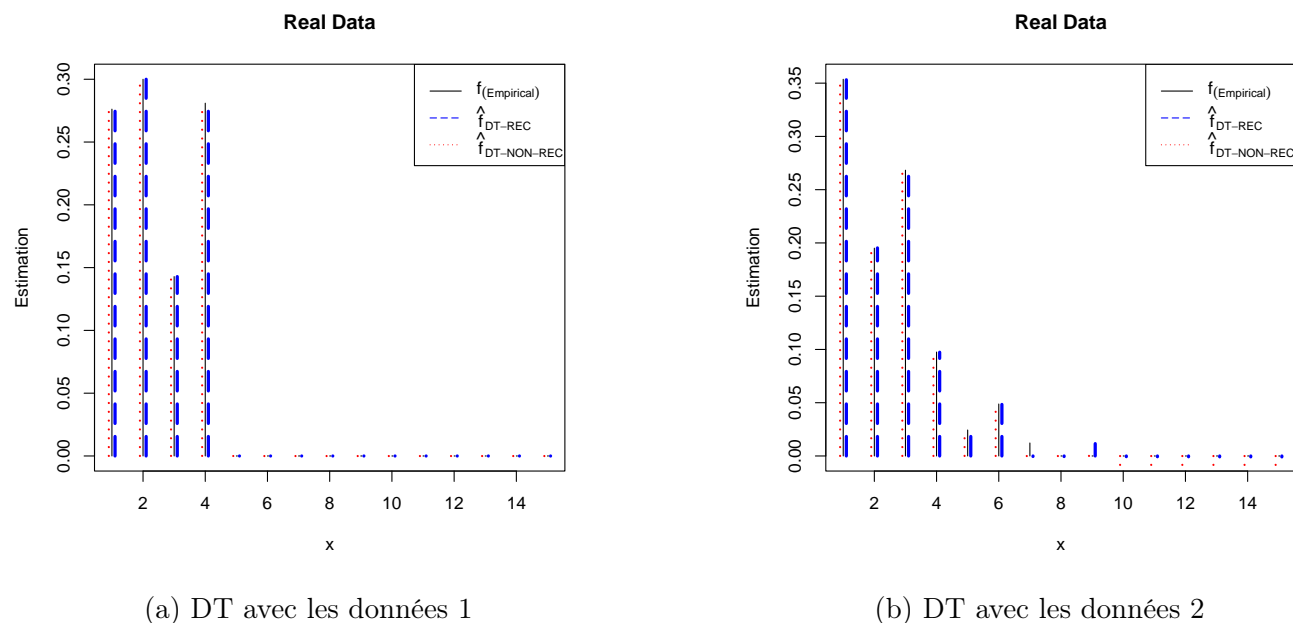


FIGURE 3.9 – Estimateurs récursifs et non récursifs à noyau discret (DT) pour les données 1 du tableau 3.9 avec $n = 210$, et les données 2 du tableau 3.10 avec $n = 82$.

3.5 Conclusion

Dans ce chapitre, nous avons d'abord proposé l'estimateur récursif à noyau asymétrique dans les cas des données dépendantes. Nous avons établi les propriétés asymptotiques biais, variance et MISE. On a réalisé une étude de simulation de monte Carlo basée sur des données dépendantes d'un modèle de durée conditionnelle autorégressive (DCA) et d'un modèle de volatilité stochastique (VS) en utilisant les noyaux asymétriques MG, W et NK. On a réalisé deux applications sur des données réelles basées sur les données de la SNCF et indice des prix. L'estimateur récursif est meilleur que l'estimateur non récursif dans le cas de données dépendantes. Nous avons présenté l'estimateur récursif à noyau discret pour des données de comptage. On a établi les propriétés asymptotiques biais, variance et MISE. On a réalisé une étude de simulation sur des données réelles en utilisant des noyaux discrets tels que WVR, DDU, DT, LR et DOSK. L'estimateur récursif est meilleur que l'estimateur non récursif dans le cas de données de comptage.

Conclusion générale et perspective

Dans cette thèse, nous avons proposé des estimateurs récursifs basés sur des noyaux continus asymétriques et discrets, ainsi que la méthode UCV pour la sélection du paramètre de lissage. Tout d'abord, nous avons réalisé un état de l'art sur les propriétés des estimateurs à noyaux standards. Ensuite, nous avons présenté l'estimateur à noyau asymétrique (non récursif) dans le contexte des données dépendantes positives, en définissant le coefficient de α -mélange.

Nous avons ensuite proposé l'estimateur récursif à noyau asymétrique (GG) pour les données dépendantes. Les propriétés asymptotiques de biais, de variance et de MISE ont été établies. Une étude de simulation de Monte Carlo a été réalisée sur des données dépendantes, utilisant un modèle de durée conditionnelle autorégressive (DCA) et un modèle de volatilité stochastique (VS) avec les noyaux asymétriques MG, W et NK. De plus, deux applications pratiques ont été menées sur des données réelles de la SNCF et de l'indice des prix.

Dans le cas discret, nous avons présenté l'estimateur récursif à noyau discret pour les données de comptage. Nous avons établi les propriétés asymptotiques de biais, de variance et de MISE. Une étude de simulation a été réalisée sur des données générées suivant plusieurs fonctions de masse de probabilités, en utilisant des noyaux discrets tels que WVR, DDU, DT, LR et DOSK, ainsi que sur des données réelles concernant la longévité des insectes adultes et le choix du mode de transport.

Les résultats obtenus démontrent en général l'avantage des estimateurs récursifs par rapport aux estimateurs non récursifs pour les échantillons de moyenne et grande taille dans le cas continu. En revanche, pour les estimateurs à noyau discret, les estimateurs non récursifs sont plus performants pour les échantillons de petite et moyenne taille, tandis que les estimateurs récursifs surpassent les non récursifs pour les échantillons de grande taille.

Les estimateurs récursifs à noyau jouent un rôle crucial dans la mise à jour d'un échantillon de taille n à un échantillon de taille $n + 1$. Il est évident que le temps de calcul de l'estimateur non récursif à noyau est considérablement plus élevé que celui de l'estimateur récursif à noyau, conférant aux estimateurs récursifs un avantage décisif en termes d'efficacité computationnelle.

Cette thèse met en lumière les bénéfices des estimateurs récursifs à noyau, tant pour les données continues que discrètes, en soulignant leur pertinence et efficacité dans divers contextes d'application.

Au cours de ce travail de recherche, nous avons constaté que de nombreux problèmes restent encore à explorer. Parmi les aspects à approfondir concernant les estimateurs récursifs de la densité de probabilité par la méthode du noyau, les plus intéressants semblent être les suivants :

- Estimation récursive de la densité à noyau pour les données dépendantes positives dans le cas de données multivariées.
 - Développement d'un estimateur récursif à noyau discret pour les données de comptage dans le cas de données dépendantes et multivariées.
 - Proposition des estimateurs récursifs dans l'estimation de la fonction de régression par noyaux associés.
 - Création d'un package sous R, pour l'implémentation des différents estimateurs récursifs.
-

Bibliographie

- [1] A. Aboubacar and C.C. Kokonendji. Asymptotic results for recursive multivariate associated kernel estimators of the probability density mass function of a data stream. *Communications in Statistics – Theory and Methods*, Accepted for publication (DOI :...), pp.1–16, pages 1–16, 2024.
- [2] J. Aitchison and C. G. G. Aitken. Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3) :413, December 1976.
- [3] A. Amiri. Sur une famille paramétrique d’estimateurs séquentiels de la densité pour un processus fortement mélangeant. *Comptes Rendus. Mathématique*, 347(5-6) :309–314, 2009.
- [4] A. Baker and K. Greene. The latin american left’s mandate : free-market policies and issue voting in new democracies. *World Politics*, 63(1) :43–77, 2011.
- [5] O. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 64(2) :253–280, 2002.
- [6] L. Bauwens and P. Giot. *Econometric modelling of stock market intraday activity*, volume 38. Journal of springer Science and Business Media, 2001.
- [7] L. Bauwens and D. Veredas. The stochastic conditional duration model : a latent variable model for the analysis of financial durations. *Journal of econometrics*, 119(2) :381–412, 2004.
- [8] D. Bosq. Regression estimation and prediction for discrete time processes. *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction*, pages 65–93, 1996.
- [9] T. Bouezmarni and J. VK. Rombouts. Nonparametric density estimation for positive time series. *Computational Statistics and Data Analysis*, 54(2) :245–261, 2010.
- [10] A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360, 1984.
- [11] M. J. Brewer. A modelling approach for bandwidth selection in kernel density estimation. In *Compstat : Proceedings in Computational Statistics 13th Symposium held in Bristol, Great Britain, 1998*, pages 203–208. Springer, 1998.

-
- [12] M. J. Brewer. A bayesian model for local smoothing in kernel density estimation. *Statistics and Computing.*, 10 :299–309, 2000.
- [13] M. Carrasco and X. Chen. Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory.*, 18 :17–39, 2002.
- [14] N. Cencov. Estimation of an unknown distribution density from observations. *Soviet Math.*, 3 :1559–1566, 1962.
- [15] S. X. Chen. Beta kernel estimators for density functions. *Computational Statistics and Data Analysis*, 31(2) :131–145, 1999.
- [16] S. X. Chen. Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics*, 52(3) :471–480, 2000.
- [17] H. I. Davies. Strong consistency of a sequential estimator of a probability density function. *Bulletin of Mathematical Statistics*, 15(4) :49–54, 1973.
- [18] P. Deheuvels. Sur l’estimation séquentielle de la densité. *Thèse de Doctorat de l’université Paris 6.*, 1973.
- [19] P. Deheuvels. Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité. *CR Acad. Sci. Paris*, 278 :1217–1220, 1974.
- [20] M. Diaz, J. Geffroy, and M. Courvoisier. On-set realization of fail-safe sequential machines. *IEEE Transactions on Computers*, 100(2) :133–138, 1974.
- [21] L. Dobele-Kpoka and F. G. Baudin. Méthode non-paramétrique des noyaux associés mixtes et applications. *Thèse de Doctorat de l’université de Besançon.*, 2013.
- [22] P. Doukhan. Mixing : Properties and examples. *Springer Science and Business Media.*, 85 :17–39, 1994.
- [23] F. C. Drost and Bas J M. Werker. Semiparametric duration models. *Journal of Business and Economic Statistics*, 22(1) :40–50, 2004.
- [24] R. Engle. The econometrics of ultra-high-frequency data. *Econometrica*, 68(1) :1–22, 2000.
- [25] R. Engle and J. R. Russell. Autoregressive conditional duration : a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162, 1998.
- [26] M. Fernandes and J. Grammig. A family of autoregressive conditional duration models. *Econometrics*, 130(1) :1–23, 2006.
- [27] J. Galambos and E. Seneta. Regularly varying sequences. *Proceedings of the American Mathematical Society*, 41(1) :110–116, 1973.
- [28] A K. Gangopadhyay and K N. Cheung. Bayesian approach to the choice of smoothing parameter in kernel density estimation. *Journal of Nonparametric Statistics.*, 14 :655–664, 2002.
-

-
- [29] C. Gouriéroux and A. Monfort. *Séries temporelles et modèles dynamiques*. Livre de Economica Paris, 1990.
- [30] P. Hall, J.S. Marron, and B. Park. Smoothed cross-validation. *Probability theory and related fields*, 92(1) :1–20, 1992.
- [31] P. Hall and P. Patil. On the efficiency of on-line density estimators. *Journal of IEEE Transactions on Information Theory*, 40(5) :1504–1512, 1994.
- [32] J. Hamilton and O. Jorda. A model of the federal funds rate target. *Journal of Political Economy*, 110(5) :1135–1167, 2002.
- [33] L. Harfouche, S. Adjabi, N. Zougab, and B. Funke. Multiplicative bias correction for discrete kernels. *Statistical Methods and Applications*, 27 :253–276, 2018.
- [34] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Journal of Economic Studies*, 61(2) :247–264, 1994.
- [35] M. Hirukawa and M. Sakudo. Family of the generalised gamma kernels : a generator of asymmetric kernels for nonnegative data. *Nonparametric Statistics*, 27(1) :41–63, 2015.
- [36] E. Isogai. Joint asymptotic normality of nonparametric recursive density estimators at a finite number of distinct points. *Journal of the Japan Statistical Society, Japanese Issue*, 14(2) :125–135, 1984.
- [37] A. Jemai. Estimation fonctionnelle non paramétrique au voisinage du bord. *Thèse de Doctorat de l'Université de Poitiers.*, 2018.
- [38] X. Jin, J. Kawczak, et al. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Annals of Economics and Finance*, 4 :103–124, 2003.
- [39] Y. Kakizawa. Recursive asymmetric kernel density estimation for nonnegative data. *Nonparametric Statistics*, 33(2) :197–224, 2021.
- [40] G. Kerkycharian and D. Picard. Density estimation by kernel and wavelets methods : optimality of besov spaces. *Statistics and Probability Letters*, 18(4) :327–336, 1993.
- [41] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, pages 462–466, 1952.
- [42] T. Senga Kiessé and D. Mizère. Weighted poisson and semiparametric kernel models applied to parasite growth. *Australian and New Zealand Journal of Statistics*, 55(1) :1–13, 2013.
- [43] C.C. Kokonendji and T. Senga Kiessé. Discrete associated kernels method and extensions. *Journal of Statistical Methodology*, 8(6) :497–516, 2011.
- [44] C.C. Kokonendji and FG Libengué. Méthode des noyaux associés continus et estimation de densité. *Statistique de la SFdS*, 6, 2011.
-

-
- [45] C.C. Kokonendji, T. Senga Kiessé, and S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. *Nonparametric Statistics*, 19(6-8) :241–254, 2007.
- [46] R. Kronmal and M. Tarter. The estimation of probability densities and cumulatives by fourier series methods. *Journal of the American Statistical Association*, 63(323) :925–952, 1968.
- [47] S. Makhloufi, N. Zougab, Y. Ziane, and S. Adjabi. A family of asymmetric kernels based on log-symmetric distributions. *Communications in Statistics-Simulation and Computation*, pages 1–18, 2021.
- [48] P. Malec and M. Schienle. Nonparametric kernel density estimation near the boundary. *computational Statistics and Data Analysis*, 72 :57–76, 2014.
- [49] S. M. O. Maouloud. Quelques aspects fonctionnels et non fonctionnels des grandes déviations et des déviations modérées en estimation non-paramétrique. *Thèse de Doctorat de l'Université Reims-Champagne Ardenne*, 2007.
- [50] C. Marchant, K. Bertin, V. Leiva, and H. Saulo. Generalized birnbaum–saunders kernel density estimators and an analysis of financial data. *computational Statistics and Data Analysis*, 63 :1–15, 2013.
- [51] K. A. Mezhoud, Z. Mohdeb, and S. Louhichi. Recursive kernel estimation of the density under η -weak dependence. *Journal of the Korean Statistical Society*, 43(3) :403–414, 2014.
- [52] A. Mokkadem, M. Pelletier, and Y. Slaoui. The stochastic approximation method for the estimation of a multivariate probability density. *Journal of Statistical Planning and Inference*, 139(7) :2459–2478, 2009.
- [53] B. Park and J. Marron. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85(409) :66–72, 1990.
- [54] E. Parzen. On estimation of a probability density function and mode. *Annals of mathematical statistics*, 33(3) :1065–1076, 1962.
- [55] E. Parzen. On estimation of a probability density function and mode. *Journal of the annals of mathematical statistics*, 33(3) :1065–1076, 1962.
- [56] B. T. Polyak and A. B. Tsybakov. Optimal order of accuracy of search algorithms in stochastic optimization. *Problemy Peredachi Informatsii*, 26(2) :45–53, 1990.
- [57] J. Racine and Q. Li. Nonparametric estimation of regression functions with both categorical and continuous data. *Econometrics*, 119(1) :99–130, 2004.
- [58] P. Révész. Robbins-monro procedure in a hilbert space and its application in the theory of learning processes i. *I. Studia Sci. Math. Hung*, 8 :391–398, 1973.
-

-
- [59] P. Révész. How to apply the method of stochastic approximation in the non-parametric estimation of a regression function. *Journal of Theoretical and Applied Statistics*, 8(1) :119–126, 1977.
- [60] H. Robbins and S. Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.
- [61] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the national Academy of Sciences*, 42(1) :43–47, 1956.
- [62] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The annals of mathematical statistics*, pages 832–837, 1956.
- [63] G. Roussas. Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation. *Nonparametric Statistics*, 1(3) :171–195, 1992.
- [64] M. Rudemo. Empirical choice of histograms and kernel density estimators. *Journal of Statistics*, 9(2) :65–78, 1982.
- [65] A. Saab. Sur une condition nécessaire et suffisante de l_1 -convergence presque complète de l'estimateur de la partition fixe pour une densité. *Journal of Comptes Rendus*, 286(16) :1107–1110, 1976.
- [66] N. Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. *Communications in Statistics Theory and Methods*, 38(19) :3583–3595, 2009.
- [67] O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Nonparametric statistics*, 16(1-2) :217–226, 2004.
- [68] S. Schwartz. Estimation of probability density by an orthogonal series. *The Annals of Mathematical Statistics*, pages 1261–1265, 1967.
- [69] D. Scott and G. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400) :1131–1146, 1987.
- [70] T. Senga Kiessé. Approche non-paramétrique par noyaux associés discrets des données de dénombrement. *Thèse de Doctorat de l'Université de Pau.*, 2008.
- [71] J. Sheather, C. Michael, and M. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Methodological.*, 53 :683–690, 1991.
- [72] N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility model. *Stochastic volatility : selected readings*, 6(2) :480–490, 2005.
- [73] R. Singh. Applications of estimators of a density and its derivatives to certain statistical problems. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(3) :357–363, 1977.
-

-
- [74] Y. Slaoui. Bandwidth selection for recursive kernel density estimators defined by stochastic approximation method. *Journal of Probability and Statistics*, 2014, 2014.
- [75] C. Stone. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, pages 1285–1297, 1984.
- [76] S. Taylor. Modelling financial time series. *Journal of world scientific*, 61(5) :247–264, 2008.
- [77] L. Tat Tran. Density estimation for time series by histograms. *Journal of statistical planning and inference*, 40(1) :61–79, 1994.
- [78] A.B. Tsybakov. Recurrent estimation of the mode of a multidimensional distribution. *Journal of problems Inform. Transmission*, 26(1) :119–128, 1990.
- [79] G. Wahba. Interpolating spline methods for density estimation i. equi-spaced knots. *The Annals of Statistics*, pages 30–48, 1975.
- [80] G. Wahba. Spline interpolation and smoothing on the sphere. *Journal of Scientific and Statistical Computing*, 2(1) :5–16, 1981.
- [81] J. Walter and R. Sidarous. Comparison of direct bladder and sacral nerve stimulation in spinal cats. *Journal of J Rehabil Res Dev*, 29(2) :13–22, 1992.
- [82] M. Wang and J. Van Ryzin. A class of smooth estimators for discrete distributions. *Biometrika*, 68(1) :301–309, 1981.
- [83] G. Watson. Smooth regression analysis. *Sankhyā : The Indian Journal of Statistics, Series A*, pages 359–372, 1964.
- [84] E. Wegman and H. Davies. Remarks on some recursive estimators of a probability density. *The Annals of Statistics*, pages 316–327, 1979.
- [85] C. Wolverton and T. Wagner. Recursive estimates of probability densities. *Journal of Transactions on Systems Science and Cybernetics*, 5(3) :246–247, 1969.
- [86] H. Yamato. Sequential estimation of a continuous probability density function and mode. *Journal of research Association of Statistical Sciences*, 14(3) :1–12, 1971.
- [87] X. Zhang, M. King, and R. Hyndman. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 50(11) :3009–3031, 2006.
- [88] N. Zougab. Approche bayésienne dans l’estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne. *Thèse de Doctorat de l’Université Béjaia.*, 2013.
- [89] N. Zougab, S. Adjabi, and C. C. Kokonendji. Adaptive smoothing in associated kernel discrete functions estimation using bayesian approach. *Journal of Statistical Computation and Simulation.*, 83 :2219–2231, 2013.
-

RÉSUMÉ : Dans cette thèse, nous avons procédé à l'estimation de la densité de probabilité pour des données dépendantes à support non négatif en utilisant la méthode du noyau. Fondées sur des estimateurs récursifs et non récursifs, ainsi que la méthode UCV pour la sélection du paramètre de lissage. Plus spécifiquement, nous avons développé deux estimateurs récursifs. Le premier est un estimateur récursif à noyau gamma généralisé destiné à l'estimation de données dépendantes positives issues d'un processus stationnaire α mélange. Le second est un estimateur récursif à noyau discret utilisé pour estimer des données de comptage. Nous avons établi les propriétés asymptotiques : biais, variance, ainsi que l'erreur quadratique intégrée moyenne (MISE) en tant que mesure globale. Une étude de simulation a comparé les performances des estimateurs récursifs et non récursifs en utilisant le critère ISE. Les résultats obtenus suggèrent que les estimateurs récursifs surpassent leurs homologues non récursifs en termes de performance.

Mots clés : erreur quadratique intégrée moyenne ; α mélange ; données de comptage ; méthode du noyau ; paramètre de lissage ; estimateur récursif ; erreur quadratique intégrée (ISE).

ABSTRACT : In this thesis, we carried out probability density estimation for dependent data with non-negative support using the kernel method. Based on recursive and non-recursive estimators, as well as the UCV method for smoothing parameter selection. More specifically, we have developed two recursive estimators. The first is a recursive estimator with a generalized gamma kernel for estimating positive dependent data from a stationary α -mixture process. The second is a discrete kernel recursive estimator used to estimate count data. We have established the asymptotic properties, including bias and variance, as well as the mean integrated squared error (MISE) as a global measure. A simulation study compared the performance of these estimators using the ISE criterion. Results suggest that recursive estimators outperform their non-recursive counterparts.

Key words : mean integrated squared error ; α mixture ; count data ; kernel method ; smoothing parameter ; recursive estimator ; Integrated Squared Error (ISE).

المخلص : في هذه الأطروحة، قمنا بتقدير كثافة الاحتمال للبيانات المعتمدة مترابطة وذات دعم غير سالب باستخدام طريقة النواة. استكشفنا مقاربات غير متماثلة مختلفة مستمرة، استنادًا إلى المقدرات التكرارية وغير التكرارية، بالإضافة إلى طريقة UCV لاختيار معلمة التنعيم. بشكل أكثر تحديدًا، قمنا بتطوير مقدرين تكراريين. الأول هو مقدر تكراري بنواة جاما عامة لتقدير البيانات المترابطة الإيجابية من عملية مزيج ألفا ثابتة. الثاني هو مقدر تكراري للبنية المتعددة القطاعات يستخدم لتقدير بيانات العد. لقد قمنا بتحديد الخصائص الحدية، بما في ذلك الانحياز والتباين، بالإضافة إلى الخطأ المتوسط المربع المتكامل كقياس عالمي. قامت دراسة المحاكاة بمقارنة أداء هذه المقدرات باستخدام معيار ISE. تشير النتائج إلى أن المقدرات التكرارية تفوق نظرائها غير التكرارية.

الكلمات الرئيسية: الخطأ المتوسط المربع المتكامل؛ مزيج ألفا؛ بيانات العد؛ طريقة النواة؛ معلمة التنعيم؛ مقدر تكراري؛ الخطأ المربع المتكامل للعدد