

UNIVERSITÉ ASSANE SECK DE ZIGUINCHOR



**THÈSE DE DOCTORAT DE L'UNIVERSITÉ ASSANE SECK DE
ZIGUINCHOR**

DOMAINE : SCIENCES ET TECHNOLOGIES

MENTION : MATHÉMATIQUES ET APPLICATIONS

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

OPTION : PROBABILITÉS ET STATISTIQUES

ÉCOLE DOCTORALE SCIENCES, TECHNOLOGIES ET INGÉNIERIE

Présentée par

Marcel Sihintoé BDIANE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ ASSANE SECK DE ZIGUINCHOR

Sujet de la thèse :

**Estimateur de la fonction intensité du processus de Poisson non homogène par
la méthode du noyau**

Sous la direction de **Papa NGOM et Clément MANGA**

Soutenue le **mardi 21 décembre 2021** devant le **jury** composé de :

Président :	M. Marie Salomon SAMBOU	Professeur titulaire, UASZ (Sénégal)
Rapporteurs :	M. Alassane DIEDHIOU	Professeur titulaire, UASZ (Sénégal)
	M. Aliou DIOP	Professeur titulaire, UGB (Sénégal)
	M. Mame Diarra FALL	Maître de Conférences, Université d'Orléans (France)
Examineurs :	M. Emmanuel Nicolas CABRAL	Maître Assistant, UASZ (Sénégal)
Directeurs :	M. Papa NGOM	Professeur titulaire, UCAD (Sénégal)
	M. Clément MANGA	Maître de Conférences, UASZ (Sénégal)



THÈSE EFFECTUÉE AU SEIN DU LABORATOIRE MATHÉMATIQUES ET APPLICATIONS
(LMA) DE L'UFR SCIENCES ET TECHNOLOGIES DE L'UNIVERSITÉ ASSANE SECK DE
ZIGUINCHOR
BP : 523-ZIGUINCHOR-SÉNÉGAL

Dédicaces

Je dédie cette thèse à : Dieu le Père tout puissant qui ne cesse de faire des merveilles pour moi, ma mère Elisabeth MANGA.

Je la dédie aussi à mes frères et sœurs : Albert BADIANE, Paul BADIANE, Alexandre Badiane, Jean Kouhissoré BADIANE, Siré BADIANE, Ginette BADIANE, Denisia BADIANE, Marie Badiane, Alphonse Badiane, pour votre amour, votre soutien sans faille et tout ce que vous avez pu m'apporter pour franchir les obstacles les plus difficiles.

Je la dédie aux docteurs et doctorants du laboratoire de Mathématiques et Applications de notre institution pour leur assistances et conseils, comme Souhaibou Sambou, Abdoulaye DIOUF, Winnie Ossette INGOBA, Alassane AW, Nestor DJINTELBE, Sény DIATTA, Pape Modou SARR, Ibrahima SANE, Bernard MANGA, Agack Alain DIED-HIOU, Guillaume Itbadio SADIO, Ibou GOUDIABY, Alioune COULIBALY, etc. Un merci au département tout entier !

Enfin, à la mémoire de mon défunt cher père Jean Pierre et défunt frère Olivier qui auraient été fiers du fruit de tant d'efforts consentis.

Remerciements

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

Je tiens à exprimer toute ma gratitude à Papa NGOM et Clément MANGA pour m'avoir fait confiance, accepter d'encadrer ce travail. Ils m'ont permis au cours de ces quatre années de comprendre les problématiques liées à l'étude de la fonction intensité du processus de Poisson. Je tiens particulièrement à les remercier pour leur disponibilité et leur convivialité tout au long de ces années. Malgré toutes leurs occupations, ils ont su être à l'écoute de mes préoccupations et de mes questions, ce qui m'a sans cesse aidé à avancer. Leurs conseils et leurs expériences m'ont permis de venir à bout de ce travail. J'espère que nous aurons l'occasion de collaborer dans les prochaines années. Je voudrais enfin les remercier pour m'avoir communiqué leur passion de la recherche en matière de résultats théoriques ou de leur mise en œuvre concrète.

Je remercie également le professeur Marie Salomon SAMBOU, pour la manière dont il a mené l'équipe "des thésards" du laboratoire de Mathématiques et Applications à laquelle j'appartiens et tous les enseignants du département de mathématiques.

Mes remerciements vont aussi à l'endroit de mon directeur de master, monsieur Emmanuel Nicolas CABRAL qui a guidé mes premiers pas en recherche en master 2 et pour tous ces encouragements, ces conseils et ces orientations.

Je tiens à remercier l'ensemble des membres du jury, pour avoir accepté de braver la chaleur et le désert pour participer à ma soutenance et pour l'intérêt porté à mes travaux de recherche. Je remercie tout particulièrement Monsieur Aliou DIOP de l'université Gaston Berger de Saint-Louis, Madame Mame Diarra FALL de l'université d'Orléans (France) et Alassane DIEDHIOU de l'université Assane SECK de Ziguinchor d'avoir accepté d'être rapporteurs de ce travail de recherche. Je remercie aussi tous les lecteurs de ce mémoire ainsi que pour leurs remarques et commentaires qui m'ont permis d'améliorer son contenu. Enfin, j'ai pu trouver dans les questions qui m'ont été formulées des perspectives intéressantes à ce mémoire. Il sera important de les prendre en compte dans les années à venir.

Je remercie également le professeur Marie Salomon SAMBOU d'avoir accepté de présider ce jury et je tiens à exprimer ma reconnaissance à monsieur Emmanuel Nicolas CABRAL, pour le travail d'examinateur. C'est un grand honneur de les compter parmi les membres de mon jury.

Enfin, je remercie du fond du cœur les familles DIATTA et DIEDHIOU, qui m'ont toujours accueilli pendant mes déplacements à Dakar. Merci pour tout ce que vous m'avez apporté, pour votre disponibilité, soutien, gentillesse et amitié, en particulier

Omer DIATTA, Adolphe DIATTA et sa femme, Alexandre DIATTA et sa femme Adriana,
Nasaire SAMBOU, Roger BASSENE, Youssouph MANSALY, Jules NYAFOUNA...

Estimateur de la fonction intensité du processus de Poisson non homogène par la méthode du noyau.

Résumé.

Nous présenterons dans cette thèse, l'approche non paramétrique par noyaux classiques et associés pour les fonctions intensité du processus de Poisson non homogène. Nous commencerons par rappeler d'abord les notions essentielles d'estimations par noyaux continus (classiques) et noyaux associés continus. Nous donnerons la définition et les caractéristiques des estimateurs à noyaux continus (classiques). Nous rappellerons aussi les différentes techniques de choix de paramètres de lissage et nous revisiterons les problèmes de supports ainsi qu'une résolution des effets de bord dans le cas continu.

Les différentes techniques de sélection de la matrice de lissage seront rappelées. Nous proposerons par la suite la forme de l'estimateur à noyau associé continu multivarié. Les propriétés statistiques de cet estimateur seront présentées, ainsi que trois méthodes classiques pour le choix de la matrice de lissage. Par ailleurs, nous présenterons l'alternative bayésienne pour le choix de la matrice de lissage. L'estimateur bayésien de H est obtenu à partir de la loi a posteriori sous la fonction de perte quadratique et entropie. Nous proposerons dans ce cadre bayésien deux approches (locale et adaptative) pour le choix de la matrice de lissage H .

Ensuite, nous détaillerons la nouvelle méthode d'estimation de la fonction intensité par les noyaux associés continus, lesquels englobent les noyaux continus (classiques). Nous définirons les noyaux associés continus et nous rappellerons la méthode mode-dispersion pour leur construction puis nous illustrerons ceci sur un noyau associé non classique de la littérature à savoir le noyau Bêta. Nous examinerons par la suite les propriétés des estimateurs (le biais, la variance et les erreurs quadratiques moyennes ponctuelles et intégrées). Puis, nous proposerons un algorithme de réduction de biais que nous illustrerons sur ce même noyau associé non classique. Des études par simulations seront faites sur ce noyau et comparées avec le noyau gaussien. Par ailleurs, nous étudierons les comportements asymptotiques de cet estimateur Bêta associé continu. Nous montrerons la consistance ainsi que la normalité asymptotique.

Mots-clefs : Fonction intensité, processus de Poisson non homogène, intensité de premier ordre, effet de bord, matrice de lissage, méthode bayésienne locale, méthode bayésienne adaptative, noyau classique, noyau associé, validation croisée.

Kernel intensity function estimation for the inhomogeneous Poisson process for the kernel

Abstract.

We will present in this thesis, the nonparametric approach by classical and associated kernels for the intensity functions of the nonhomogeneous Poisson process. We will begin by recalling first the essential notions of estimations by continuous kernels (classical) and associated continuous kernels. We will give the definition and the characteristics of the continuous kernel estimators (classical). We will also recall the various techniques for choosing smoothing parameters and we will revisit the problems of supports as well as a resolution of edge effects in the continuous case.

The various techniques for selecting the smoothing matrix will be recalled. We will then propose the form of the multivariate continuous associated kernel estimator. The statistical properties of this estimator will be presented, as well as three classical methods for the choice of the smoothing matrix. In addition, we will present the Bayesian alternative for the choice of the smoothing matrix. The Bayesian estimator of H is obtained from the posterior distribution under squared loss and entropy. We will propose in this Bayesian framework two approaches (local and adaptive) for the choice of the smoothing matrix H .

Then, we will detail the new method of estimation of the intensity function by the associated continuous kernels, which include the continuous (classical) kernels. We will define the continuous associated kernels and we will recall the mode-dispersion method for their construction, then we will illustrate this on an associated non-classical nucleus from the literature, namely the Beta nucleus. We will then examine the properties of the estimators (bias, variance and point and integrated mean squared errors). Then, we will propose a bias reduction algorithm that we will illustrate on this same non-classical associated kernel. Simulation studies will be made on this nucleus and compared with the Gaussian kernel. In addition, we will study the asymptotic behaviors of this continuous associated Beta estimator. We will show the consistency as well as the asymptotic normality.

Key words and phrases. Intensity function, inhomogeneous Poisson process, first order intensity, edge effect, smoothing matrix, local Bayesian method, adaptive Bayesian method, classical kernel, associated kernel, cross validation.

Liste des travaux

Publication

M. S. Badiane, P. Ngom and C. Manga. Bayesian selection of local bandwidth in non-homogeneous Poisson process kernel estimators for the intensity function. *Far East Journal of Theoretical Statistics*, 2, 109-144, 2021.

Soumise

M. S. Badiane, P. Ngom and C. Manga. Bayesian selection of adaptive bandwidth in non-homogeneous Poisson process kernel estimators for the intensity function. (Soumise au Biennales)

Table des matières

Introduction générale	3
1 Préliminaires	7
1.1 Introduction	7
1.1.1 Définitions et Notations	7
1.1.2 Loi d'un processus ponctuel	8
1.1.3 Processus ponctuels finis	8
1.1.4 Processus ponctuel de Poisson	9
1.2 L'intensité : (<i>propriété d'ordre 1</i>)	10
1.3 L'intensité : (<i>propriété du second ordre</i>)	10
1.4 Noyaux classiques multivariés	12
1.5 Exemples de noyaux classiques	14
1.6 Choix de la matrice des fenêtres	16
1.6.1 La Méthode de Scott	16
1.6.2 Méthode de la validation croisée par les moindres carrés	17
1.7 Approche bayésienne	17
1.7.1 Lois a priori	18
1.7.2 Loi a posteriori	20
1.7.3 Inférence bayésienne	20
1.8 Noyau associé et problème aux bords	23
1.9 Noyaux asymétriques continus	24
1.10 Estimateur à noyau associé produit	28
1.11 Résolution de problèmes de biais de bordure	28
1.12 Cas continu : réduction du biais de l'estimateur	28
2 Estimateur de la fonction intensité du processus de Poisson non homogène	30
2.1 Introduction	30
2.2 Estimateur consistant	31

2.3	Propriétés de l'estimateur	33
2.3.1	Mesure d'erreur et matrice de lissage optimale.	34
2.3.2	Critère d'erreur	35
2.3.3	Matrice optimale	37
2.4	Conclusion	39
3	Approche bayésienne locale pour le choix de la matrice de lissage de la fonction intensité.	40
3.1	Introduction	40
3.2	Méthodologie	40
3.3	Étude de simulation	44
3.3.1	Données simulées	44
3.3.2	Données réelles	53
3.4	Conclusion	56
4	Approche bayésienne adaptative pour le choix de la matrice de lissage de la fonction intensité.	57
4.1	Introduction	57
4.2	Méthodologie	57
4.2.1	Choix de la loi a priori	58
4.2.2	Choix des paramètres de la loi a priori	58
4.2.3	Estimation de la loi a posteriori	59
4.3	Études de simulation	66
4.3.1	Analyse de sensibilité	66
4.3.2	Approche bayésienne adaptative, locale et la validation croisée.	71
4.3.3	Données réelles	74
4.4	Conclusion	76
5	Estimation de la fonction intensité par le noyau Bêta.	77
5.1	Introduction	77
5.2	Estimateur à noyau Bêta de la fonction intensité	78
5.2.1	Erreur quadratique moyenne (MSE).	84
5.2.2	Erreur quadratique moyenne intégrée (MISE)	84
5.3	L'estimateur de la fonction intensité par le noyau Bêta modifié.	84
5.4	Résultats asymptotiques.	87
5.5	Étude simulation	97
5.6	Application à des données réelles	104
5.7	Conclusion	105

Conclusion et perspectives	106
Références	108

Table des figures

1.1	Paramétrisations de matrice de lissage Duong (2004) : densité cible et forme du noyau associé classique.	14
1.2	Comportements de quelques noyaux classiques de même cible $x = 0$	15
1.3	Comportements de quelques noyaux associés non classiques de la droite réelle positive aux bords pour $x = h = 0.5$	27
3.1	(a) : fonction intensité $\lambda_1(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$	44
3.2	(a) : fonction intensité $\lambda_2(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$	45
3.3	(a) : fonction intensité $\lambda_3(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$	45
3.4	(a) : fonction intensité $\lambda_4(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$	46
3.5	(a) : fonction intensité $\lambda_5(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$	46
3.6	Graphes du \widehat{ISE} avec le noyau gaussien et la loi Gamma inverse des cinq fonctions intensité.	47
3.7	Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_1(x)$ pour $m = 500$	52
3.8	Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_2(x)$ pour $m = 500$	52
3.9	Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_3(x)$ pour $m = 500$	52
3.10	Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_4(x)$ pour $m = 500$	53
3.11	Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_5(x)$ pour $m = 500$	53
3.12	(a) : Distribution de 58 cas de cancer de larynx et (b) : de 978 cas de cancer de poumon de la région de Chorley-Ribble.	54
3.13	Estimation de la fonction intensité par noyau pour le cancer de larynx par la méthode : (a) <i>AMISE</i> , (b) <i>LSCV</i> , (c) Scott, (d) bayésienne locale.	55

3.14	Estimation de la fonction intensité par noyau pour le cancer de poumon par la méthode : (a) <i>AMISE</i> , (b) <i>LSCV</i> , (c) Scott, (d) bayésienne locale.	56
4.1	Grapshe du \widehat{ISE} avec le noyau gaussien et la loi a priori donnée à l'équation (4.2.6) des fonctions intensités sous la fonction perte quadratique.	68
4.2	Grapshe du \widehat{ISE} avec le noyau gaussien et la loi a priori donnée à l'équation (4.2.6) des fonctions intensités pour la fonction perte entropie.	70
4.3	Estimation de la fonction intensité par noyau pour des données du cancer de larynx par la méthode : (a) <i>LSCV</i> , (b) bayésienne locale, (c) bayésienne adaptative avec la fonction de perte entropie, (d) bayésienne adaptative avec la fonction de perte quadratique.	75
4.4	Estimation de la fonction intensité par noyau pour les données du cancer de poumon par la méthode : (a) <i>LSCV</i> , (b) bayésienne locale, (c) bayésienne adaptative avec la fonction de perte entropie, (d) bayésienne adaptative avec la fonction de perte quadratique.	76
5.1	Forme du noyau Bêta pour $x = (x_1, x_2)$ avec paramètres de lissage $h_1 = h_2 = 0, 2$	79
5.2	Grapshe du biais intégré pour les 4 fonctions intensités.	99
5.3	Grapshe de la variance intégrée pour les quatre fonctions intensités.	100
5.4	Grapshe de <i>AMISE</i> des quatre fonctions intensités	103
5.5	(a) processus de Poisson non homogène, (b) contour pour l'estimateur de noyau Gaussien, (c) contour pour l'estimateur de noyau Bêta avec des arbres caryer.	104
5.6	(a) processus de Poisson inhomogène, (b) contour pour l'estimateur de noyau Gaussien, (c) contour pour l'estimateur du noyau Bêta avec des érables.	104
5.7	(a) processus de Poisson inhomogène, (b) contour pour l'estimateur de noyau gaussien, (c) contour pour l'estimateur de noyau Bêta avec des chênes rouges.	105

Liste des tableaux

1.1	Quelques noyaux associés continus univariés.	15
1.2	Quelques noyaux associés continus univariés.	27
3.1	Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$	48
3.2	Résultats de simulation du \widehat{ISE} basé sur 100 répétitions pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$	50
3.3	Comparaison du temps d'exécution (en secondes).	51
3.4	Paramètres de lissages optimaux pour les données de cancer de larynx et du cancer de poumon pour différentes méthodes de sélections.	54
3.5	Comparaison de la valeur du ISE_0 et du temps d'exécution pour les données du cancer de Larynx et du poumon.	55
4.1	Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$ sous la fonction de perte quadratique.	69
4.2	Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$ sous la fonction de perte entropie.	71
4.3	Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$	72
4.4	Comparaison du temps d'exécution (en secondes).	74
4.5	Comparaison du ISE_0 et du temps d'exécution pour les données réelles avec le noyau gaussien produit.	75
5.1	Biais et variance intégrés pour les fonctions intensités $\lambda_{11}(x)$, $\lambda_{22}(x)$, $\lambda_{33}(x)$ et $\lambda_{44}(x)$ respectivement.	101
5.2	Biais intégré, la variance intégrée et $AMISE$ pour les fonctions intensité $\lambda_{11}(x)$, $\lambda_{22}(x)$, $\lambda_{33}(x)$ et $\lambda_{44}(x)$ respectivement.	102

Notations et Abréviations.

AMISE	Erreur quadratique moyenne intégrée asymptotique (Asymptotic Mean Integrate Squared Error)
Cov(.)	Covariance de l'argument
diag(.)	Matrice diagonale
$\mathbb{E}(\cdot)$	Espérance mathématique de l'argument
f	Fonction densité
fidis	Famille des lois en dimensions finie
H	Matrice de lissage
h	Paramètre de lissage
I_2	Matrice identité de \mathbb{R}^2
ISE	Erreur quadratique intégrée (Integrated squared error)
K	Fonction noyau
LSCV	Validation croisée par moindre carrée (Least Squares Cross-Validation)
MISE	Erreur quadratique moyenne intégrée (Mean Integrate Squared Error)
MSE	Erreur quadratique moyenne (Mean Squared Error)
N^{lf}	Ensemble des configurations localement finies
$\sigma(\mathcal{N}^{lf})$	La plus petite σ -algèbre de N^{lf}
\mathbb{P}	Mesure de probabilité
Var(.)	Variance de l'argument
W_1	Support de la fonction densité
vec(H)	Vecteur de H , vecteur de $(k^2 \times 1)$ obtenu en mettant les colonnes de H l'une sous l'autre de gauche à droite
vech(H)	Demi-vecteur de H , vecteur de $(\frac{1}{2}k(k+1) \times 1)$ obtenu de $vec(H)$ par l'élimination des éléments de H situés au-dessus de la diagonale

\mathbf{W}	Sous-ensemble de \mathbb{R}^2
$N(A)$	Nombre d'événements dans A
λ	Fonction intensité du processus
$\mathbb{1}_A$	Fonction indicatrice de A
$ \mathbf{H} $	Déterminant de la matrice H
$\nabla\lambda$	Matrice Hessienne
$\pi(h)$	Distribution a priori
$\pi(h X)$	Distribution a posteriori
\mathcal{H}	Ensemble des matrices diagonales
(\mathbf{W}, \mathbf{d})	Espace métrique complet et séparable
$\text{tr}(H)$	Trace de la matrice H
$\mu_2(K)$	Le moment d'ordre 2 de K ie $\int u^2 K(u) du$
$\mathcal{N}(0, 1)$	Distribution normale ou gaussienne standard
$\mathcal{N}(\mu, \sigma^2)$	Distribution normale de moyenne μ et de variance σ^2
$\mathcal{N}(\mu, \Sigma)$	Distribution normale multivariée de moyenne μ et de matrice de covariance Σ
$\hat{\theta}$	Estimateur du paramètre θ
$\hat{\lambda}_{0,H}$	Estimateur de la fonction intensité
$\hat{\lambda}_{0,H}^{-i}$	Estimateur de la fonction intensité en omettant l'événement i
$o(\cdot)$	$a_n = o(b_n)$ ssi $\frac{a_n}{b_n} \rightarrow 0$ quand $n \rightarrow \infty$ où $(a_n)_{n>0}$ et $(b_n)_{n>0}$ des suites réelles.
$O(\cdot)$	$a_n = O(b_n)$ ssi $\frac{a_n}{b_n} = Cte$ quand $n \rightarrow \infty$
x^T	Transposée du vecteur x .
\mathbb{N}	Ensemble des entiers naturels
\mathbb{R}	Ensemble des nombres réels

- \mathbb{R}^k Espace euclidien réel de dimension k
- Θ L'espace des paramètres
- Ω L'espace des observations

Introduction générale

Dans de nombreux domaines d'application, les jeux de données à analyser se présentent sous la forme de listes d'événements localisés dans l'espace. Ces jeux de données sont appelés semis de points et peuvent correspondre à la disposition de certaines espèces végétales dans une forêt, aux emplacements des épices de secousses sismiques enregistrées, à la localisation de trésors archéologiques retrouvés sur un site, aux adresses de patients affectés d'une certaine maladie dans une région, à la répartition de cellules dans un tissu biologique... Un tel semis de points est noté $\{X_i, i = 1, \dots, N\}$, où $X_i \in X \subseteq \mathbb{R}^k$, $\forall i = 1, \dots, N$, X étant borné.

Cependant, un des problèmes habituellement rencontrés en statistique est celui de l'estimation de fonctions multivariées. Il s'agit d'un problème bien connu dans de nombreux domaines où l'on a souvent à modéliser des phénomènes complexes. On rencontre ce genre de questions en économie, en médecine, en sociologie, en environnement, en marketing, en sport, etc.

Dans le cas multivarié, ces derniers sont souvent décrits par des vecteurs aléatoires à valeurs réelles et à support connu dans $\mathbb{T}_k \subset \mathbb{R}^k$ avec $k \in \{1, 2, \dots\}$. Cet ensemble peut être formé d'axes univariés continus, discrets ou mixtes (à la fois discrets et continus).

Afin d'estimer la fonction intensité inconnue λ , une première approche dite paramétrique consiste à supposer que λ appartient à une famille de fonction intensité comme dans le cas de la fonction densité qui peuvent être décrites par un certain nombre de paramètres réels.

Par exemple, dans le cas continu multidimensionnel, la densité normale multivariée est caractérisée par un vecteur $\nu \in \mathbb{R}^k$ représentant son centre et une matrice semi-définie positive qui est sa matrice de variance covariance. Dans le cas discret multidimensionnel, la distribution de Poisson multivarié est caractérisée par $k + 1$ paramètres réels. Ces deux lois sont considérées généralement comme les distributions standards pour l'analyse des données multidimensionnelles continues et de comptage respectivement. En effet, le statisticien qui opte pour une telle approche possède une bonne connaissance a priori du phénomène aléatoire, en pratique on dit que les valeurs attendues sont régulières. Cependant, lorsqu'aucune information n'est disponible sur le phénomène étudié, l'application d'un modèle paramétrique n'est pas satisfaisante. Pour pallier les insuffisances et les défauts des familles paramétriques, une seconde approche dite non paramétrique est proposée. Les méthodes non paramétriques estiment la fonction intensité directement à partir de l'information disponible sur l'ensemble des données sans aucune hypothèse a priori sur la distribution des données, en pratique on dit que dans cette approche les données parlent d'elles-mêmes.

Récemment, une nouvelle approche dite semi-paramétrique a été introduite par Yue et al. (2011). Cette dernière est un compromis entre la méthode paramétrique et la méthode non paramétrique qui consiste à écrire λ comme la distribution pondérée par une fonction poids appropriée, en pratique on dit que cette approche élimine le bruit.

L'estimateur à noyau a été proposé initialement par Rosenblatt (1956) et Parzen (1962) pour estimer des fonctions de densité à support non borné. Cet estimateur dépend d'une fonction appelée noyau et d'un paramètre appelé fenêtre de lissage. Le choix du noyau dans ce cas est peu important, car il n'influe pas sur l'estimation. Les noyaux les plus utilisés pour les densités à support non borné sont les noyaux symétriques, nous citons par exemple le noyau gaussien et le noyau Epanechnikov. Il existe d'autres noyaux symétriques, comme le noyau triangulaire, le noyau biweight et le noyau uniforme.

Cependant, lorsqu'on veut estimer des densités ou des intensités à support non borné, l'estimateur à noyau classique devient non consistant, à cause des effets du bord. Ce problème est dû à l'utilisation des noyaux symétriques qui assignent un poids en dehors du support lorsque le lissage est pris en compte près du bord. Pour remédier à ce problème, récemment, certains auteurs ont proposé dans le cas des densités à support compact, l'utilisation des noyaux dont le support coïncide avec celui de la densité à estimer. Ceci a efficacement résolu le problème des effets de bord puisque les noyaux utilisés ici sont généralement asymétriques et peuvent changer de forme selon la position du point d'estimation. C'est notamment le cas, en univarié de Chen (1999)(2000a) avec les noyaux Bêta et Gamma pour estimer les densités à support respectivement $[0, 1]$ et $[0, +\infty[$, puis de Scaillet (2004) avec les noyaux gaussiens inverses et sa réciproque pour les densités à support $]0, +\infty[$. Toujours dans le cas continu, Libengué (2013) a proposé une méthode de construction de noyaux associés continus dite "*mode-dispersion*" qui englobe, entre autres, les noyaux de Chen (1999)(2000a) et de Scaillet (2004). En multivarié, les premiers travaux sur l'estimation non paramétrique par la méthode du noyau remontent à Maniya (1961) et Nadaraya (1964), qui ont proposé des estimateurs à noyaux pour la fonction densité bivarié. Cacoullos (1966) a repris l'idée de Nadaraya (1964) et a proposé une extension d-dimensionnelle de l'estimateur à noyau avec un seul paramètre de lissage positif. Epanechnikov (1969) a traité le cas où le paramètre de lissage est un vecteur. Singh (1976) a développé ces idées pour estimer les dérivées partielles de la densité multivariée. Deheuvels (1977) est le premier à avoir introduit la forme générale de l'estimateur à noyau, qui ne dépend pas d'un vecteur, mais d'une matrice symétrique définie positive. Wand et Jones (1995) ont présenté une monographie détaillée sur ce sujet. Le noyau le plus populaire pour l'estimation de la densité de probabilité multidimensionnelle dans \mathbb{R}^k est le noyau gaussien multivarié. Dans ce cas, l'estimateur possède de bonnes propriétés asymptotiques, voir Silverman (1986) et Scott (1992). Cependant, cet estimateur n'est pas adapté lorsque certaines variables sont bornées. Un problème de biais se pose aux bornes, ce qui entraîne la non-consistance de l'estimateur. Ce problème devient plus sévère dans le cas multivarié, car il s'ajoute au problème de dimension de support. Une première solution à ce problème dans le cas multivarié a été donnée par Bouezmarni et Roumbouts (2010), en utilisant des estimateurs à noyaux produits composés des univariés (Bêta et Gamma) de Chen (1999)(2000a).

La qualité de l'estimateur à noyau de la fonction intensité dépend crucialement du choix du paramètre (et/ou) de la matrice de lissage. Pour le problème particulier de

l'estimation de l'intensité du noyau, Diggle (1985) et Diggle (2013) propose une sélection du paramètre de lissage adaptatif aux données basées sur l'hypothèse d'un processus de Cox.

Dans le cas univarié, ceci revient à choisir un paramètre scalaire strictement positif qui contrôle le degré de lissage. Par contre dans le cas multivarié, le paramètre de lissage est une matrice symétrique définie positive. Cette matrice contrôle le degré de lissage et l'orientation par rapport aux axes des coordonnées des contours de la densité cible à estimer. Elle contient $k(k+1)/2$ éléments indépendants à estimer dans \mathbb{R}^k , $k \geq 1$, ce qui rend son choix plus difficile. Pour remédier à cette complexité, certains auteurs supposent que la matrice est diagonale. Mais pour certaines densités cibles, d'autres auteurs ont montré l'importance de la matrice de lissage pleine, Chacón et Duong (2011).

Par ailleurs, l'objectif principal de cette thèse est de proposer l'alternative bayésienne pour le choix de la matrice de lissage et d'introduire l'estimateur à noyau associé continu bivarié pour l'estimation de la fonction intensité du processus de Poisson non homogène. Le formalisme bayésien est caractérisé par le traitement de la matrice de lissage comme une variable aléatoire, en lui associant une loi a priori $\pi(\cdot)$. L'estimation bayésienne est obtenue ainsi via la loi a posteriori $\pi(\cdot | \text{données})$. Tout d'abord, nous allons proposer la définition d'un noyau classique continu multivarié ainsi que l'estimateur basé sur ce noyau pour estimer la fonction intensité du processus de Poisson non homogène. Nous donnerons ensuite ses propriétés statistiques à distance finie et asymptotique. Par la suite, nous allons proposer deux approches bayésiennes (locale et adaptative) pour estimer la matrice de lissage où nous obtenons la forme explicite de l'estimateur de la matrice de lissage. L'approche bayésienne locale consiste à traiter la matrice des fenêtres de lissage localement, en la considérant comme une variable aléatoire de loi a priori $\pi(\cdot)$. Nous allons développer cette approche, en utilisant le noyau gaussien multivarié avec la distribution gamma inverse comme la loi a priori. Cette combinaison permet de dériver explicitement l'estimateur bayésien local. L'approche bayésienne adaptative consiste à associer un vecteur de lissage adaptatif pour chaque vecteur d'observation, et de le traiter comme un vecteur de variables aléatoires de loi a priori $\pi(\cdot)$. Comme dans le cas de l'approche bayésienne locale, nous envisageons d'employer le noyau gaussien avec les deux fonctions pertes : *perte quadratique* et *perte entropie* afin de dériver exactement l'estimateur bayésien adaptatif pour le vecteur de lissage. Ces deux approches bayésiennes seront illustrées par des études de simulations et des applications sur des données réelles. Les performances de ces approches bayésiennes seront aussi examinées et comparées par rapport aux performances de la méthode classique validation croisée (*LSCV*) et par la méthode *AMISE*, en utilisant l'erreur quadratique intégrée. Ensuite, nous utiliserons le noyau Bêta pour estimer la fonction intensité du processus de Poisson non homogène et nous donnerons ses propriétés statistiques et asymptotiques. Enfin, nous comparerons les performances du noyau avec le noyau gaussien en utilisant des données simulées et réelles.

Cette thèse est composée d'une introduction générale, de cinq chapitres et d'une conclusion générale et quelques perspectives.

Le **chapitre 1** rappelle l'état d'art concernant la méthode du noyau classique multivarié et la méthode du noyau associé univarié, pour l'estimation de la fonction intensité du processus de Poisson. Nous soulignerons également les problèmes liés aux différentes structures de supports ainsi qu'aux formes de la matrice de lissage.

Le **chapitre 2** introduit la notion de consistance de l'estimateur non paramétrique de la fonction intensité par noyau classique bivarié. Nous étudierons aussi les propriétés statistiques de cet estimateur. La technique de validation croisée et la méthode de Scott pour choisir la matrice de lissage seront présentées.

Dans le **chapitre 3**, nous présenterons l'estimation de la fonction intensité en utilisant l'approche bayésienne locale pour la sélection de la matrice de lissage.

Le **chapitre 4** propose l'estimateur bayésien adaptatif dans l'estimation de la fonction intensité du processus de Poisson.

Le **chapitre 5** introduit l'estimation non paramétrique de la fonction intensité par noyau associé Bêta bivarié. À travers des simulations et applications sur des données réelles à la fin de chacun de ces trois derniers chapitres, nous montrerons la pertinence et l'efficacité de l'approche bayésienne locale, adaptative et du noyau Bêta.

Enfin, cette thèse se termine par une conclusion générale et quelques perspectives.

1.1 Introduction

Ce chapitre a pour but de présenter quelques éléments de la théorie des processus ponctuels. Ainsi, nous rappelons les méthodes d'estimations par noyaux continus multivariés ainsi qu'associés continus univariés. Nous nous intéressons aux problèmes d'estimation de la fonction intensité. Pour ces derniers, les estimateurs à noyaux continus (classiques) et associés continus seront présentés avec les différentes méthodes du choix de la matrice des fenêtres ainsi que leurs problèmes d'effets de bord.

1.1.1 Définitions et Notations

Nous nous intéressons à des ensembles non ordonnés de points d'un espace donné W , par exemple \mathbb{R}^2 .

Définition 1.1.1. *On appelle configuration un ensemble dénombrable, non ordonné de points de W :*

$$x = \{x_1, \dots, x_n, \dots\}$$

Pour pouvoir considérer des configurations aléatoires, il faut ajouter des contraintes sur W . Nous choisissons de considérer un espace W muni d'une métrique d telle que (W, d) soit complet et séparable. Cette métrique permet de définir une topologie et une σ -algèbre borélienne. En pratique, W sera souvent un sous-ensemble compact de \mathbb{R}^k muni de la distance euclidienne.

Définition 1.1.2. *On dit d'une configuration $x \subseteq W$ qu'elle est localement finie si dans tout borélien borné $A \subseteq W$ elle possède un nombre $N_x(A)$ fini de points. La famille de toutes les configurations localement finies sera notée N^{lf} .*

On définit alors la notion de processus ponctuel :

Définition 1.1.3. *Un processus ponctuel sur W est une application X d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ dans N^{lf} , telle que pour tout borélien $A \subseteq W$, $N(A) = N_X(A)$ est une variable aléatoire (finie).*

Ce qui induit une nouvelle définition :

Définition 1.1.4. *Si l'espace W est borné ou si $N_X(W)$ est fini presque sûrement, le processus ponctuel est dit processus ponctuel fini.*

Les réalisations d'un processus ponctuel X sont donc des configurations aléatoires de points tels que pour tout borélien $A \subseteq W$ le nombre de points dans A soit une variable aléatoire. Cela signifie qu'un processus ponctuel est une variable aléatoire à valeur dans l'espace mesurable $(\mathcal{N}^{lf}, \sigma(\mathcal{N}^{lf}))$, où $\sigma(\mathcal{N}^{lf})$ est le plus petit σ -algèbre telle que pour tout borélien borné $A \subseteq W$ l'application $x \rightarrow N_x(A)$ soit mesurable. La mesure de probabilité induite sur \mathcal{N}^{lf} est appelée la loi de X .

1.1.2 Loi d'un processus ponctuel

La loi d'un processus ponctuel X est la mesure image par l'application X de \mathbb{P} sur $\sigma(\mathcal{N}^{lf})$. Mais comme $\sigma(\mathcal{N}^{lf})$ est définie par la mesurabilité des applications $x \rightarrow N_x(A)$ pour des boréliens $A \subseteq W$, l'analogie de la loi de probabilité pour des variables aléatoires dans un contexte de processus ponctuels est l'ensemble des lois jointes des vecteurs $(N(A_1), \dots, N(A_m))$ où les A_i sont des boréliens bornés.

Définition 1.1.5. *La famille des lois en dimensions finies (**fidis**) d'un processus ponctuel X sur un espace métrique (W, d) complet et séparable est la collection des lois jointes de $(N(A_1), \dots, N(A_m))$ pour tout vecteur fini (A_1, \dots, A_m) de boréliens bornés $A_i \subseteq W$, $i = 1, \dots, m$ de longueur quelconque $m \in \mathbb{N}$.*

L'intérêt de cette définition est justifié par le théorème suivant :

Théorème 1.1.1. *Daley et Jones (1988)*

*La loi d'un processus ponctuel X sur un espace métrique complet et séparable (W, d) est entièrement déterminée par ses **fidis**.*

Donc, deux processus ponctuels partageant les mêmes **fidis** ont la même loi.

1.1.3 Processus ponctuels finis

La plupart des processus ponctuels utilisés en pratique sont observés dans une région bornée. Cette région peut être dictée par l'application tout comme elle peut résulter d'un choix volontaire pour limiter la taille de l'espace d'intérêt. Dans tous les cas, les réalisations du processus ponctuel contiennent presque sûrement un nombre fini de points. Nous verrons plus loin qu'il y a une raison théorique importante qui nous fait considérer cette classe de processus ponctuels. La notion de densité est en effet difficile à mettre en place pour des processus qui ne sont pas finis.

Construction

Pour construire un processus ponctuel fini, nous pouvons utiliser :

1. Une distribution de probabilité discrète (p_m) , $m \in \mathbb{N}$ pour le nombre de points.
2. Une famille de densités de probabilité symétriques $j_m(x_1, \dots, x_m)$, $m \in \mathbb{N}$ sur W^m pour les positions des points. Le deuxième point implique que nous supposons que W soit muni d'une mesure borélienne $\nu(\cdot)$ de manière à pouvoir définir les densités j_m par rapport à la mesure produite $\nu(\cdot)^m$.

Nous pouvons alors construire un processus ponctuel X de la manière suivante : nous considérons $N(W)$ une variable aléatoire de loi $(p_m)_m$, et nous conditionnons ensuite par rapport aux événements $N(W) = m$: on prend $(X_1, \dots, X_m) \in W^m$ un vecteur aléatoire de loi $j_m(\cdot, \dots, \cdot)$. La condition de symétrie est nécessaire puisqu'une configuration est indifférente à l'ordre dans lequel ses points sont recensés. Il convient de vérifier toutefois si nous pouvons passer légalement de vecteurs ordonnés à des vecteurs non ordonnés. Ce point est important puisqu'il servira dans des démonstrations ultérieures où nous basculerons des configurations aux m -uplets, et réciproquement. Définissons d'abord N^f comme l'ensemble des configurations finies. Ensuite, appelons \mathcal{N}^f le plus petit σ -algèbre pour laquelle les applications $x \rightarrow N_x(A)$ (A borélien borné) sont mesurables. Nous définissons ensuite les sous-ensembles de N^f suivants :

$$N_m = \{x \in N^f : N_x(W) = m\}$$

à chacun de ces ensembles, on associe une σ -algèbre $\sigma(\mathcal{N}_m^f)$ en prenant la trace de $\sigma(\mathcal{N}^f)$ sur N_m . Considérons maintenant les fonctions $f_m : W^m \rightarrow N_m$ qui à des vecteurs de taille m associent des configurations de m points. Les f_m sont mesurables par rapport à la σ -algèbre des boréliens, et grâce à leur invariance par permutation, mesurables par rapport à la σ -algèbre $B_s(W^m)$ des boréliens symétriques dans W^m . Par hypothèse, pour tout m , $j_m(\cdot, \dots, \cdot)$ est une densité invariante par permutation et est donc $B_s(W^m)$ -mesurable.

1.1.4 Processus ponctuel de Poisson

Le processus ponctuel de Poisson joue un rôle fondamental dans l'étude des processus ponctuels. En effet, il sert de modèle de référence, mathématiquement étudiable, pour lequel il n'existe aucune interaction entre les points. De nombreux modèles, comme les processus ponctuels de Gibbs, par exemple, sont alors construits à partir du processus ponctuel de Poisson. Dans cette section, nous nous intéressons au processus de Poisson spécifié par une fonction d'intensité.

Définition 1.1.6. *Soit $\nu(\cdot)$ une mesure borélienne sur un espace métrique et séparable (W, d) tel que $\nu(W) > 0$ et $\nu(A) < \infty$ pour tout borélien borné A . (Une telle mesure est dite localement finie). Un processus X sur W est appelé un processus ponctuel de Poisson de mesure d'intensité $\nu(\cdot)$ si :*

- (P_1) $N(A)$ suit une loi de Poisson d'espérance $\nu(A)$ pour tout borélien borné $A \subseteq W$.
- (P_2) Pour k boréliens disjoints A_1, \dots, A_k , les variables aléatoires $N(A_1), \dots, N(A_k)$ sont *i.i.d.*

Si $\nu(\cdot)$ est non atomique, le processus de Poisson est simple. Il est fini si $\nu(W) < \infty$. Lorsque $W = \mathbb{R}^k$, on parle de processus de Poisson homogène lorsque la mesure d'intensité est $\lambda\nu(\cdot)$ où ν est la mesure de Lebesgue et λ un paramètre strictement positif. On appelle alors ce paramètre l'intensité du processus (propriété P_1). La propriété P_2 peut être interprétée comme une propriété de non-corrélation spatiale totale, puisqu'elle affirme que ce qui se passe à l'intérieur d'une fenêtre quelconque est totalement indépendant de ce qui se passe à l'extérieur. Le théorème suivant précise cette idée, en affirmant que les points d'un processus ponctuel de Poisson se comportent aléatoirement et n'interagissent pas les uns avec les autres.

Théorème 1.1.2. *Van Lieshout (2000)*

Soit X un processus ponctuel de Poisson sur un espace métrique complet et séparable (W, d) de mesure d'intensité $\nu(\cdot)$, et $A \subseteq W$ un borélien borné. Alors, conditionnellement à $\{N(A) = m\}$, X restreints à A suit la loi d'un processus binomial de points indépendants, ν -uniformément distribués sur A .

1.2 L'intensité : (propriété d'ordre 1)

Les lois des processus sont très complexes (Moller et al. (2003)), ce qui conduit dans la pratique à utiliser de façon privilégiée des indicateurs qualifiés de propriété d'ordre 1 ou d'ordre 2, comme nous utilisons les moments d'ordre 1 et 2 (espérance et variance) pour appréhender une variable aléatoire de loi inconnue.

Définition 1.2.1. *L'intensité est apparue dans la présentation du processus de Poisson où elle est constante (λ). Il existe d'autres processus X dans lesquels cette hypothèse est relâchée, et où la fonction d'intensité $\lambda(x)$ est variable. Elle est définie par*

$$E[N(B)] = \int_B \lambda(x) dx.$$

En appliquant la définition de l'espérance à une petite région centrée en x et de surface dx , nous pouvons définir l'intensité en ce point x comme le nombre de points attendus dans cette petite surface lorsqu'elle tend vers 0, soit :

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \frac{\mathbb{E}(N(dx))}{|dx|}.$$

Dans le cas où cette intensité n'est pas constante, elle peut être estimée avec les méthodes non paramétriques utilisées pour l'estimation de la densité. Dans sa version la plus simple, sans correction des effets de bord, l'estimateur de l'intensité s'écrit :

$$\lambda(u) = \sum_{i=1}^n K(u - x_i),$$

K désignant le noyau, qui peut être gaussien, ou à support fini (noyau d'Epanechnikov, noyau Biweight de Tukey...). Il doit vérifier $\int_{\mathbb{R}^2} K(u) du = 1$. Comme dans toutes les méthodes non paramétriques, le choix du noyau a un impact limité. En revanche le choix de la bande passante est extrêmement important (voir par exemple Illian et al. 2008)).

1.3 L'intensité : (propriété du second ordre)

Nous allons nous intéresser, pour introduire les propriétés du second ordre d'un processus ponctuel, à la variance et à la covariance des comptages de points, que nous définissons ci-dessous :

$$Var(N(A)) = \mathbb{E}[N(A)^2] - \mathbb{E}[N(A)]^2$$

$$Cov[N(A_1), N(A_2)] = \mathbb{E}[N(A_1)N(A_2)] - \mathbb{E}[N(A_1)]\mathbb{E}[N(A_2)].$$

Définition 1.3.1. *Plutôt que d'utiliser ces indicateurs, on définit le moment d'ordre deux de la façon suivante :*

$$\mathcal{G}(A \times B) = \mathbb{E}[N(A)N(B)] - \mathbb{E}[N(A \cap B)], \quad (1.3.1)$$

qui vaut pour le processus de Poisson : $\lambda^2|A||B|$. Lorsque cette mesure admet une densité, celle-ci, appelée intensité d'ordre 2 et notée λ_2 est définie de telle sorte que $\mathcal{G}(C) = \int_C \lambda_2(u, v) dudv$. Cette intensité du second ordre peut s'interpréter comme :

$$\lambda_2(x, y) = \lim_{|dx| \rightarrow 0 |dy| \rightarrow 0} \frac{\mathbb{E}[N(dx)N(dy)]}{|dx||dy|}.$$

Les intensités du premier et du second ordre permettent de définir une fonction, appelée fonction de corrélation de paire de points de la façon suivante :

$$g_2(u, v) = \frac{\lambda_2(u, v)}{\lambda(u)\lambda(v)}. \quad (1.3.2)$$

Dans le cas d'un processus de Poisson homogène, $\lambda_2(u, v) = \lambda^2$, $g_2(u, v) = 1$.

Lorsqu'un processus est **stationnaire** (au second ordre), l'intensité du second ordre n'est pas affectée par la translation et ne dépend que de la différence entre les points : $\lambda_2(x, y) = \lambda_2(x - y)$.

Lorsqu'il est en plus **isotrope**, le processus n'est pas affecté par la rotation et l'intensité de second ordre ne dépend que de la distance entre x et y . Notons que la **stationnarité** au second ordre et l'**isotropie** sont indispensables pour de nombreux outils de statistique.

Répartition au hasard, agrégation, régularité.

Lorsque l'on étudie une distribution de points, deux grandes questions se posent : les points observés sont-ils distribués au hasard ou y a-t-il une interaction ? S'il y a une interdépendance, est-elle de nature agrégative ou répulsive ? Selon les réponses à ces questions, trois configurations de points sont généralement mises en évidence : une distribution dite **complètement aléatoire**, une **agrégée** et une **régulière**.

La configuration **complètement aléatoire** est centrale pour la théorie. Toutes les configurations de points, en tant que réalisation d'un processus ponctuel, sont aléatoires, mais celle-ci correspond à une distribution "**complètement au hasard**" de points sur une surface : les points sont localisés partout avec la même probabilité et indépendamment les uns des autres. Cette configuration correspond à un tirage d'un processus de Poisson homogène. Il n'y a pas dans ce cas aucune interaction entre les points, mais seule l'utilisation d'indicateurs permet de juger si la distribution observée s'écarte significativement d'une distribution **complètement aléatoire**. En effet, il est délicat à l'œil nu d'identifier une telle configuration.

La seconde répartition de points est dite **régulière (ou répulsive)** : nous pouvons penser à la répartition spatiale des arbres dans un verger ou le long des rues en ville, à celle des transats sur une plage, etc. Dans une telle configuration, les points sont plus régulièrement

espacés qu'ils ne le seraient sous une distribution complètement aléatoire. Les points se repoussent et créent une distribution de points dispersée.

Enfin, la dernière configuration possible est qualifiée d'**agrégée**. Dans ce cas, une interaction entre les points est mise en évidence, ils s'attirent, créant des agrégats : une concentration géographique sera alors détectée.

1.4 Noyaux classiques multivariés

Cette section présente la méthode d'estimation de la fonction intensité multidimensionnelle par noyaux continus classiques.

Définition 1.4.1. Une fonction K de support continu $W_1 \subset \mathbb{R}^2$ est dite *noyau classique* si elle est une densité de probabilité symétrique ($K(-x) = K(x)$) de vecteur moyen μ_K nul ($\mu_K := \int_{W_1} xK(x)dx = 0$) de matrice de variance-covariance Σ_K ($\Sigma_K := \int_{W_1} xx^T K(x)dx$) de carré intégrable ($\int_{W_1} K^2(x)dx < +\infty$).

Le noyau classique K est positif de masse totale égale à l'unité (i.e. pour tout élément x de W_1 , $K(x) \geq 0$ et $\int_{W_1} K(x)dx = 1$).

Définition 1.4.2. Soit $H = H_m \rightarrow 0$ quand $m \rightarrow \infty$ une matrice des fenêtres symétrique, définie positive et K la fonction noyau vérifiant la définition 1.4.1. L'estimateur à noyau classique multivarié de la fonction intensité λ selon Cucala (2006) peut-être défini par :

$$\hat{\lambda}_H(x) = \frac{1}{p_H(x)|H_N|} \sum_{i=1}^N K(H_N^{-1}(x - X_i)) \mathbb{1}_{N \neq 0}, \quad (1.4.1)$$

$$\hat{\lambda}_H(x) = \frac{1}{p_H(x)|H_N|^{\frac{1}{2}}} \sum_{i=1}^N K(H_N^{-\frac{1}{2}}(x - X_i)) \mathbb{1}_{N \neq 0}, \quad (1.4.2)$$

Nous pouvons unifier ces écritures par :

$$\hat{\lambda}_H(x) = \frac{1}{p_H(x)} \sum_{i=1}^N K_{H_N}(x - X_i) \mathbb{1}_{N \neq 0}, \quad (1.4.3)$$

où

- $K_{H_N}(\cdot) = \frac{1}{|H_N|} K(H_N^{-1} \cdot)$ pour l'estimateur donné en (1.4.1),
- $K_{H_N}(\cdot) = \frac{1}{|H_N|^{\frac{1}{2}}} K(H_N^{-\frac{1}{2}} \cdot)$ pour celui de (1.4.2),
- $p_H(\cdot)$ est le terme de correction de bord,
- $|H_N|$ est le déterminant de la matrice de lissage H_N ,
- N représente le nombre d'observations.

Notons que la formulation de l'estimateur en (1.4.1) est la plus connue alors que la forme en (1.4.2) a été introduite par Wand et Jones (1994) pour faciliter une technique de choix de la matrice des fenêtres. Dans la forme de l'estimateur (1.4.3), la matrice des fenêtres

H_N a une influence certaine. Lorsqu'elle tend vers la matrice nulle 0_2 , cela conduit au phénomène de *sous lissage*, tandis que, lorsque H_N devient grand, l'estimateur est *sur lissé*. Il est bien connu que le choix de H_N est d'une importance capitale, contrairement aux choix du noyau. Ainsi, l'obtention d'une estimation plus proche de la réalité implique donc un choix optimal de H_m . Dans la suite de ce chapitre, la matrice des fenêtres H_N sera remplacée par H .

Le type d'orientation de la fonction noyau classique K est contrôlé par la paramétrisation de la matrice de lissage. Wand et Jones (1994) ont étudié les différentes formes de paramétrisation de la matrice de lissage bivariée H . Les auteurs ont dénombré trois formes principales et trois autres hybrides :

- (i) la classe des matrices "pleines" symétriques définies positives

$$H = \begin{pmatrix} h_1 & h_{12} \\ h_{12} & h_2 \end{pmatrix},$$

- (ii) la classe des matrices diagonales $H = \text{diag}(h_1, h_2)$,

- (iii) la classe formée du produit d'une constante positive et de la matrice identité $H = h\mathbf{I}_2$,

- (iv) la classe formée du produit d'une constante positive et de la matrice de variance-covariance empirique S ,

$$H = hS = \begin{pmatrix} hS_1^2 & hS_{12} \\ hS_{21} & hS_2^2 \end{pmatrix},$$

- (v) la classe formée d'une constante positive et de la diagonale de S ,

$$H = \begin{pmatrix} hS_1^2 & 0 \\ 0 & hS_2^2 \end{pmatrix},$$

- (vi) la classe des matrices obtenues en utilisant le coefficient de corrélation ρ_{12} pour déterminer la rotation

$$H = \begin{pmatrix} h_1^2 & \rho_{12}h_1h_2 \\ \rho_{12}h_1h_2 & h_2^2 \end{pmatrix},$$

Ces différentes paramétrisations sont décrites et illustrées par la figure 1.1. Pour une utilisation générale, la forme (iii) est la moins intéressante, parce que trop restrictive. Comme le montre la Partie (b) de la figure 1.1, les contours de la densité cible sont des ellipses pendant que celles du noyau sont circulaires. Les formes (iv) et (v) imposent au noyau de s'aligner suivant la matrice de variance-covariance de la cible f et ne sont donc pas conseillées de manière générale (voir Partie (c) de la figure 1.1). La troisième forme hybride (vi) dépend du choix judicieux du paramètre de corrélation du noyau. La partie (d) de la figure 1.1, montre bien que le noyau classique est orienté suivant la matrice de corrélation et n'est donc pas généralement utilisé. La forme diagonale (ii) est la plus utilisée, mais reste inappropriée pour des densités cibles à forte structure de corrélation comme la partie (a) de la figure 1.1. Le noyau classique utilisant cette forme garde la même orientation selon les axes quand bien même la densité cible f est fortement corrélée (i.e. orientée de manière oblique). Puisqu'on s'intéresse à toutes les formes de corrélation de f , la plus générale est (i) qui a une matrice de lissage pleine.

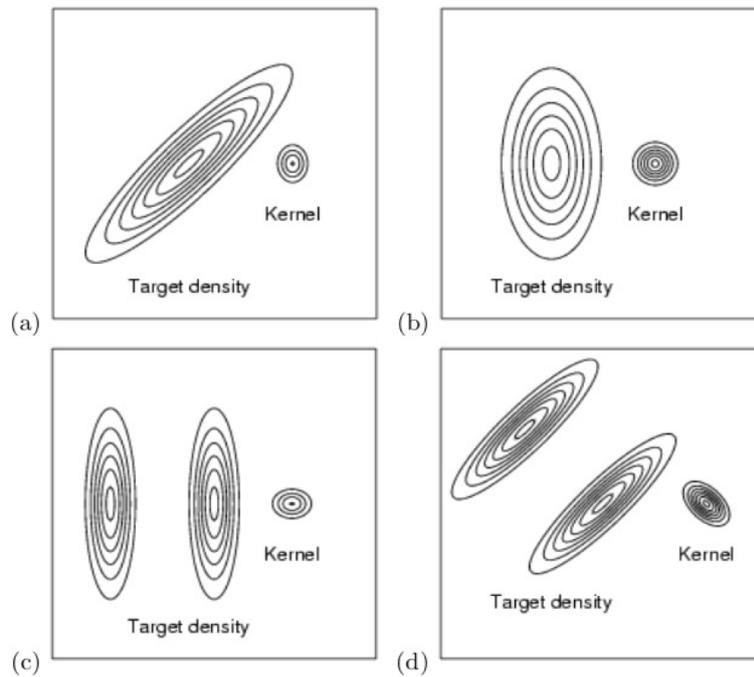


FIGURE 1.1 – Paramétrisations de matrice de lissage Duong (2004) : densité cible et forme du noyau associé classique.

1.5 Exemples de noyaux classiques

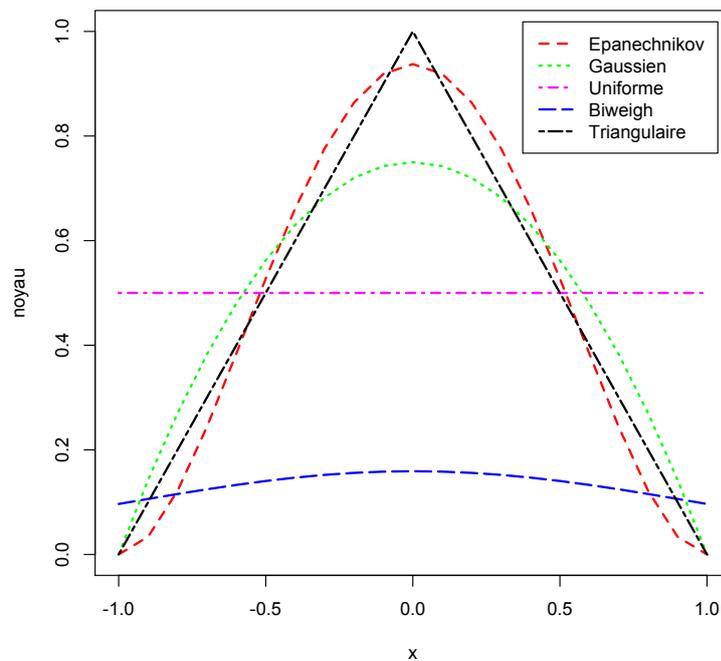
Le plus utilisé des noyaux classiques multivariés est la loi normale multivariée, donnée par $\mathcal{N}(\mu, \Sigma)$ où μ est le vecteur moyenne et Σ est la matrice de variance covariance :

$$f(x) = \frac{1}{(2\pi)^{\frac{k}{2}} (\det \Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad \forall x \in \mathbb{R}^k. \quad (1.5.1)$$

En plus de la loi normale multidimensionnelle, nous pouvons construire plusieurs autres noyaux classiques multivariés à partir des noyaux univariés. Quelques-uns sont rappelés dans le tableau 1.1. Tous ces noyaux univariés vérifient la définition 1.4.1 et ont pour support $W_1 = \mathbb{R}$ ou $[-1, 1]$.

TABLE 1.1 – Quelques noyaux associés continus univariés.

Noyaux	Support	Densité
Epanechnikov	$[-1, 1]$	$\frac{3}{4}(1 - x^2)$
Cosinus	$[-1, 1]$	$\frac{\pi}{4} \cos(\frac{\pi}{2}x)$
Biweight	$[-1, 1]$	$\frac{15}{16}(1 - x^2)^2$
Triweight	$[-1, 1]$	$\frac{35}{32}(1 - x^2)^3$
Triangulaire	$[-1, 1]$	$1 - x $
Gaussien	\mathbb{R}	$\frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$
Uniforme	$[-1, 1]$	$\frac{1}{2}$
Double Epanechnikov	$[-1, 1]$	$3 x (1 - x)$
Double Exponentielle	\mathbb{R}	$\frac{1}{2} \exp(-\frac{1}{2} x)$

FIGURE 1.2 – Comportements de quelques noyaux classiques de même cible $x = 0$.

1.6 Choix de la matrice des fenêtres

Dans cette section, nous rappelons trois méthodes pratiques du choix de la matrice des fenêtres H , à savoir la méthode de Scott, la méthode de validation croisée ("*Cross-validation*") ainsi que l'approche bayésienne avec deux de ces variantes (locale et adaptative).

1.6.1 La Méthode de Scott

La sélection empirique de la bande passante donne une formule résultante de la matrice optimale pour une distribution de référence. Évidemment, la fonction densité de probabilité d'une distribution normale multivariée $\mathcal{N}(\mu, \Sigma)$ est un bon candidat pour une distribution de référence dans le cas multivarié. Supposons que le noyau K est gaussien multivarié, c'est-à-dire K a pour fonction densité de probabilité $\mathcal{N}(0, \mathbb{I})$. Notons que $\mu_2(K) = 1$ et $\int K(u)^2 du = 2^k \pi^{-k/2}$ dans ce cas. Avec l'équation du *AMISE* du chapitre 2 et le fait que

$$\int \left[\text{tr}(H^T D^2 H) \right]^2 dx = \frac{1}{2^{k+2} \pi^{k/2} |H|^{1/2}} \left[2 \text{tr}(H^T \Sigma^{-1} H)^2 + (\text{tr}(H^T \Sigma^{-1} H))^2 \right], \quad (1.6.1)$$

cf. Wand & Jones (1994), nous pouvons facilement dériver des formules empiriques pour différentes hypothèses sur H et Σ . Dans le cas le plus simple, c'est-à-dire que l'on considère H et Σ comme des matrices diagonales $H = \text{diag}(h_1, \dots, h_d)$ et $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_k^2)$, cela conduit à

$$\tilde{h}_j = \left(\frac{4}{k+2} \right)^{\frac{1}{k+2}} n^{-1/(k+4)} \sigma_j. \quad (1.6.2)$$

Notez que cette formule coïncide avec la règle empirique de Silverman dans le cas $d = 1$ et Silverman (1986). Comme la dimension k varie, la constante dans l'équation (1.6.2) s'étend sur l'intervalle $[0,924, 1,059]$, avec une limite égale à 1. La constante est exactement 1 dans le cas bivarié, on arrive à la règle de Scott :

$$\hat{h}_j = n^{-1/(k+4)} \hat{\sigma}_j. \quad (1.6.3)$$

(voir Scott (1992), p. 152). Il n'est pas possible de dériver la règle empirique pour le général H et Σ . Cependant, (1.6.2) montre qu'il peut être judicieux de choisir la matrice de bande passante H proportionnel à la règle Σ :

$$\hat{H} = n^{-1/(k+4)} \hat{\Sigma}^{1/2}. \quad (1.6.4)$$

Nous remarquons que cette règle équivaut à appliquer une transformation de Mahalanobis aux données (pour transformer la matrice de covariance estimée en identité), puis à calculer l'estimation du noyau avec la règle de Scott (1.6.3) et enfin à retransformer la fonction densité de probabilité estimée en échelle d'origine.

Principalement, toutes les méthodes de plug-in pour l'estimation de la densité du noyau unidimensionnel peuvent être étendues au cas multivarié. Cependant, en pratique, cela est

fastidieux, car la dérivation des asymptotiques implique des dérivées multivariées et des développements de Taylor d'ordre supérieur.

1.6.2 Méthode de la validation croisée par les moindres carrés

L'idée de base des méthodes de validation croisée consiste à trouver une fonction de score $LSCV(H)$ dont le calcul est plus simple que $MISE(H)$. Pour plus de détails, on peut se référer à de nombreux auteurs tels que Bowman (1984), Rudemo (1982) dans le cas univarié. L'approche la plus utilisée est celle de validation croisée par les moindres carrés. Duong et Hazelton (2005) ont généralisé ces résultats pour une matrice de lissage H symétrique définie positive quelconque. Le principe de base de cette méthode est de minimiser par rapport à H le critère d'erreur quadratique intégrée ($ISE(H)$) défini par

$$ISE = \int_W (\hat{\lambda}_{0,H}(x) - \lambda_0(x))^2 dx \quad (1.6.5)$$

$$= \int_W \hat{\lambda}_{0,H}(x)^2 dx - 2 \int_W \hat{\lambda}_{0,H}(x) \lambda_0(x) dx + \int_W \lambda_0(x)^2 dx. \quad (1.6.6)$$

Compte tenu du fait que la quantité $\int_W \lambda_0(x)^2 dx$ ne dépend pas de H , on choisit la matrice des paramètres de lissage de façon à ce qu'elle minimise un estimateur de $LSCV(H)$ défini par

$$LSCV(H) = \int_W \hat{\lambda}_{0,H}(x)^2 dx - \frac{2}{N} \sum_{i=1}^N \hat{\lambda}_{0,H}^{-i}(X_i), \quad (1.6.7)$$

où $\hat{\lambda}_{0,H}^{-i}(x) = \frac{1}{N-1} \sum_{j \neq i} K_H(x - X_j)$ est l'estimateur calculé à partir de l'échantillon privé de l'événement X_i . La matrice optimale notée H_{LSCV} est donnée par

$$H_{LSCV} = \arg \min_{\mathcal{H}} LSCV(H), \quad (1.6.8)$$

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

1.7 Approche bayésienne

L'objectif de cette partie est d'une part, rappeler quelques notions de base du formalisme bayésien et les estimateurs paramétriques bayésiens, utiles pour des considérations majeures dans le cadre de cette étude et d'autre part, passer en revue les principales méthodes de calcul bayésien qui peuvent être employées lors de l'implémentation pratique de l'approche bayésienne. L'inférence bayésienne mobilise deux sources d'information : d'une part les données, via la vraisemblance et d'autre part le savoir de l'expert via la distribution a priori sur les paramètres et via les hypothèses structurelles sur lesquelles repose le modèle utilisé. Le concept fondamental de l'approche bayésienne est la distribution a posteriori $\pi(\theta|x)$. En effet cette distribution opère de façon conditionnelle sur les observations et met donc en œuvre automatiquement l'inversion des probabilités, tout en incluant le principe de vraisemblance. La distribution a posteriori représente l'actualisation de l'information disponible sur les paramètres θ , au vu de l'information contenue dans la vraisemblance $\pi(x|\theta)$, tandis que la distribution a priori $\pi(\theta)$ représente l'information disponible préalablement à l'observation de x .

Définition 1.7.1. *Un modèle statistique bayésien est constitué d'un modèle statistique paramétrique, $f(x|\theta)$, et d'une distribution a priori pour les paramètres, $\pi(\theta)$.*

L'impact du théorème de Bayes provient de la décision audacieuse de mettre les causes et effets sur le même niveau conceptuel puisque les deux sont aléatoires.

1.7.1 Lois a priori

Le point le plus critiquable et le plus critiqué de l'analyse bayésienne est le choix de la loi a priori. Cependant, une fois cette dernière connue, l'inférence peut être conduite d'une façon quasi mécanique. Dans une certaine mesure, c'est aussi la plus difficile à déterminer. Dans la pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori, au sens où plusieurs lois de probabilité peuvent être compatibles avec cette information. Il y a plusieurs raisons pour cela : le décideur, le client ou le statisticien n'ont pas forcément le temps, les ressources ou souvent la volonté de chercher à construire un a priori exact et doivent compléter l'information partielle qu'ils ont rassemblée à l'aide de données subjectives afin d'obtenir une loi a priori. Dans le cadre de cette étude, on se contente de ne présenter que les types de densités a priori les plus courants : les densités a priori conjuguées et les densités a priori non informatives.

A priori conjugué

Quand l'information a priori sur le modèle est trop vague ou peu fiable, une construction subjective complète de la distribution a priori est évidemment impossible. Mais d'autres considérations plus techniques peuvent être prises en compte. En particulier, Raiffa et Schlaifer (1961) ont proposé l'utilisation des densités a priori dites conjuguées afin de faciliter l'analyse bayésienne.

Définition 1.7.2. *Une famille \mathcal{F} de distribution de probabilité sur Θ (Θ est l'espace des paramètres) est dite conjuguée (ou fermée par échantillonnage) par une fonction de vraisemblance $f(x|\theta)$ si, pour tout $\pi \in \mathcal{F}$, la distribution a posteriori $\pi(\cdot|x)$ appartient également à \mathcal{F} .*

Cette définition est formellement vague puisque si l'on choisit \mathcal{F} comme la famille de toutes les densités de probabilité, la famille de la densité a priori \mathcal{F} est toujours conjuguée par la fonction de vraisemblance. Il est plus intéressant d'étudier des familles de densités a priori conjuguées tel que celles-ci aient la même forme fonctionnelle que la fonction de vraisemblance. Dans ce cas, le passage de la fonction de vraisemblance à la densité a posteriori se réduit à un changement de paramètres et non à une modification de la forme fonctionnelle de la famille correspondante. Les densités a posteriori sont donc toujours calculables et l'évaluation des statistiques a posteriori est grandement simplifiée.

A priori non informatif

A priori de Laplace

Historiquement, Laplace fut le premier à utiliser des techniques non informatives. Il munit les paramètres d'une loi a priori qui prend en compte son ignorance en donnant la

même vraisemblance à chaque valeur du paramètre, donc en utilisant une loi uniforme. Ainsi, la densité a priori d'un paramètre θ est défini par

$$\pi(\theta) = k,$$

où k est une constante.

Trois critiques ont été avancées sur ce choix :

- Les lois résultantes sont impropres quand l'espace des paramètres n'est pas compact, ces dernières mènent à des difficultés comme le paradoxe de marginalisation.
- Le principe des événements équiprobables de Laplace n'est pas cohérent en termes de partitionnement.
- Une autre critique concerne le problème de l'invariance par reparamétrisation, si on passe de $\theta \in \Theta$ à $\eta = g(\theta)$ par une transformation bijective g l'information a priori reste totalement inexistante et ne devrait pas être modifiée.

Cependant, si $\pi(\theta) = 1$, la loi a priori sur η est $\pi(\eta) = \left| \frac{d}{d\eta} g^{-1}(\eta) \right|$ par le changement de variable. Donc $\pi(\eta)$ est le plus souvent non constante.

Exemple 1.7.1. Si p , une proportion, suit une loi uniforme sur $[0, 1]$, le paramètre de rapport des chances $\gamma = \frac{p}{1-p}$ suit une loi a priori $\pi(\gamma) = \frac{1}{(1-\gamma)^2}$, qui est donc non constant.

A priori de Jeffreys

Afin d'éviter le besoin de prendre en compte une structure d'invariance potentielle, Jeffreys (1946) propose des lois a priori non informatives fondées sur l'information de Fisher donnée par

$$I(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 \right].$$

Dans le cas unidimensionnel, sous certaines conditions de régularité, cette information est aussi égale à

$$I(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right].$$

La loi a priori de Jeffreys est

$$\pi(\theta) = \sqrt{I(\theta)},$$

elle définit un coefficient de normalisation près quand π est propre. Ce choix, dépendant de l'information de Fisher se justifie par le fait que $I(\theta)$ est largement acceptée comme un indicateur de la quantité d'information apportée par le modèle sur θ (Fisher (1956)).

A priori impropre

La loi a priori peut être impropre c'est-à-dire $\int_{\Theta} \pi(\theta) d\theta = \infty$. Ce choix de type de loi n'a donc plus d'intérêt que calculatoire et s'interprète difficilement. La construction de lois non informatives peut conduire à des lois a priori de ce type.

1.7.2 Loi a posteriori

L'analyse statistique se ramène fondamentalement à une inversion, car elle doit déterminer les causes réduites aux paramètres du mécanisme probabiliste générateur, à partir des effets résumés par les observations. Une description générale de l'inversion des probabilités est donnée par le théorème de Bayes :

Théorème 1.7.1. *Si A et E sont deux événements tels que $P(E) \neq 0$, $P(A|E)$ et $P(E|A)$ sont reliés par*

$$\begin{aligned} P(A|E) &= \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} \\ &= \frac{P(E|A)P(A)}{P(E)}. \end{aligned}$$

Bayes donne en réalité une version continue de ces résultats, à savoir, pour deux variables aléatoires X et Y donc la distribution conditionnelle $f(x|y)$ et la distribution marginale $g(y)$, la distribution conditionnelle de y sachant x est

$$g(y|x) = \frac{f(x|y)g(y)}{\int f(x|y)g(y)dy}.$$

Bayes et Laplace ont considéré que l'incertitude sur le paramètre θ d'un modèle peut être décrite par une distribution de probabilité π sur Θ , appelée distribution a priori. L'inférence est alors fondée sur la distribution de θ conditionnel à X , $\pi(\theta|x)$, appelée distribution a posteriori et définie par

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta}.$$

1.7.3 Inférence bayésienne

Dans cette section, nous présentons les estimateurs ponctuels bayésiens, et leurs propriétés. Nous la débutons par une discussion sur le concept de fonction de risque bayésien dont la minimisation conduit aux différents estimateurs paramétriques bayésiens. En particulier, nous présentons, en détail, les estimateurs par maximum a posteriori et par moyenne a posteriori.

Fonctions de coût

Étant donné $\hat{\theta}$ l'estimateur du paramètre θ , on définit une fonction de coût non négative $C(\epsilon)$ telle que $\epsilon = \theta - \hat{\theta}$ est l'erreur d'estimation pour un vecteur d'observation x donné. L'objectif est de déterminer l'estimateur θ qui minimise le coût moyen $\mathbb{E}[C(\epsilon)]$ nommé le risque de Bayes "R" défini par :

$$R = \mathbb{E}[C(\epsilon)] = \int_{\Theta} \left[\int_{\Omega} C(\epsilon)\pi(x|\theta)d\theta \right] \pi(\theta)dx, \quad (1.7.1)$$

où Ω est l'espace des observations. Sous certaines conditions de régularité, l'équation (1.7.1) peut s'écrire sous la forme :

$$R = \mathbb{E}[C(\epsilon)] = \int_{\Omega} \left[\int_{\Theta} C(\epsilon) \pi(\theta|x) dx \right] \pi(\theta) m(x) d\theta. \quad (1.7.2)$$

Pour x fixé, l'estimateur qui minimise R est l'estimateur qui minimise le coût moyen sur θ , définit par :

$$\mathbb{E} [C(\theta - \hat{\theta}(x))|x] = \int_{\Omega} C(\theta - \hat{\theta}(x)) \pi(\theta|x) d\theta. \quad (1.7.3)$$

On présente les estimateurs bayésiens correspondant aux fonctions de coût classiques.

Coût quadratique

Considérons le cas où $C(\epsilon)$ est quadratique

$$C(\theta - \hat{\theta}(x)) = (\theta - \hat{\theta}(x))^2. \quad (1.7.4)$$

L'estimateur de Bayes qui minimise le risque de Bayes étant donné cette fonction de coût doit minimiser :

$$\begin{aligned} \mathbb{E} [(\theta - \hat{\theta}(x))^2|x] &= \hat{\theta}(x)^2 - 2\hat{\theta}(x)\mathbb{E}(\theta|x) + \mathbb{E}(\theta^2|x) \\ &= \hat{\theta}(x)^2 - 2\hat{\theta}(x)\mathbb{E}(\theta|x) - \mathbb{E}^2(\theta|x) + \mathbb{E}^2(\theta|x) + \mathbb{E}(\theta^2|x) \\ &= \mathbb{V}(\theta|x) + (\hat{\theta}(x) - \mathbb{E}(\theta|x))^2. \end{aligned}$$

Une variante de cette fonction de perte est une fonction de perte quadratique pondérée (fonction de perte quadratique généralisée) de la forme

$$C(\theta - \hat{\theta}(x)) = w(\theta)(\theta - \hat{\theta}(x))^2. \quad (1.7.5)$$

L'égalité qui correspond au minimum du risque quadratique est atteinte pour

$$\hat{\theta}_{MMSE}(x) = \mathbb{E}(\theta|x). \quad (1.7.6)$$

Par conséquent, pour une fonction de coût quadratique, l'estimateur de risque de Bayes minimum est la moyenne a posteriori ou en notation anglo-saxonne *Minimum Mean Squared Error (MMSE)*.

Coût entropique

Calabria et Pulcini (1994) ont proposé une fonction de perte qui découle de la fonction de perte Linex appelée la fonction de perte entropie définie par

$$C(\theta - \hat{\theta}(x)) = \left(\frac{\hat{\theta}(x)}{\theta} \right)^p - p \ln \frac{\hat{\theta}(x)}{\theta} - 1. \quad (1.7.7)$$

L'estimateur de Bayes de paramètre θ sous cette fonction de perte est

$$\hat{\theta}_{MME}(x) = \left(\mathbb{E}(\theta^{-p}|x) \right)^{\frac{-1}{p}}. \quad (1.7.8)$$

- Lorsque $p = 1$, l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique pondéré $\frac{\hat{\theta}(x) - \theta}{\theta}$.
- Lorsque $p = -1$, l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique.

Estimateurs MMSE

L'estimateur *MMSE* est déterminé par la moyenne de la densité a posteriori considérée. Étant donné, un vecteur de paramètres θ et un vecteur d'observations x , l'estimateur de $\hat{\theta}_{MMSE}$ est définie par

$$\hat{\theta}_{MMSE} = \int_{\Theta} \theta \pi(\theta|x) d\theta. \quad (1.7.9)$$

- Une intégration est requise pour estimer la moyenne de la densité a posteriori $\pi(\theta|x)$.
- Le principe d'invariance n'est pas vérifié pour l'estimateur *MMSE*.

A l'exception des cas académiques, l'inférence bayésienne se heurte très rapidement à des problèmes d'intégration et/ou d'optimisation. En effet, l'évaluation des intégrales suivantes intervient souvent dans l'analyse bayésienne :

- calcul de la constante de normalisation

$$m(y) = \int_{\Theta} f(y|\theta) \pi(\theta) d\theta, \quad (1.7.10)$$

- évaluation d'une densité marginale a posteriori
Soit $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2$

$$\pi(\theta_1|y) = \int_{\Theta_2} \pi(\theta_1, \theta_2|y) d\theta_2, \quad (1.7.11)$$

- calcul de moment a posteriori

$$\mathbb{E}(h(\theta)|y) = \int_{\Theta} \pi(\theta|y) h(\theta) d\theta. \quad (1.7.12)$$

Approche bayésienne locale

L'approche bayésienne locale a été introduite par De Lima et Atuncar (2011). Cette méthode est une généralisation au cas multidimensionnel de la méthode décrite par Gangopadhyay et Cheung (2002) en utilisant le noyau gaussien multivarié. Elle consiste à estimer H en chaque vecteur cible x et traite donc H comme une quantité aléatoire de loi a priori $\pi(\cdot)$. À partir de la formule de Bayes, la loi a posteriori prend la forme suivante :

$$\pi(H|x) = \lambda_0(x) \pi(H) \left(\int_{\mathcal{H}} \hat{\lambda}_0(x) \pi(H) dH \right)^{-1}. \quad (1.7.13)$$

Comme le modèle λ_0 est inconnu, on le remplace par son estimateur à noyau $\hat{\lambda}_{0,H}$. Ainsi, l'estimateur de la loi a posteriori de $\pi(H|x)$ devient

$$\hat{\pi}(H|x, X_1, \dots, X_N) = \hat{\lambda}_{0,H}(x) \pi(H) \left(\int_{\mathcal{H}} \hat{\lambda}_{0,H}(x) \pi(H) dH \right)^{-1}. \quad (1.7.14)$$

En fait, $\hat{\lambda}_{0,H}$ représente l'information nécessaire sur $H = H(x)$ pour compléter la loi a priori $\pi(\cdot)$. L'utilisation de l'estimateur de Bayes sous la perte quadratique conduit au meilleur estimateur \hat{H} de H au point x par

$$\hat{H}(x) = \int_{\mathcal{H}} H \hat{\pi}(H|x, X_1, \dots, X_N) dH. \quad (1.7.15)$$

Approche bayésienne adaptative

L'approche bayésienne adaptative a été introduite par Zougab et al. (2014) en utilisant le noyau gaussien multivarié. Cette approche consiste à sélectionner un vecteur de fenêtres h_i correspondant à l'observation X_i en lieu et place du vecteur $h = (h_1, h_2)^T$ initial et traiter donc h_i comme une quantité aléatoire de loi a priori $\pi(\cdot)$. L'approche que nous décrivons ici consiste dans un premier temps à l'estimateur à noyau adaptatif donné par :

$$\hat{\lambda}_{0,h_i}(x) = \frac{1}{N p_{h_i}} \sum_{i=1}^N \prod_{j=1}^2 K_{h_i}(x - X_{ij}) \mathbb{1}_{N \neq 0} \quad (1.7.16)$$

où $X_i = (X_{i1}, X_{i2})^T$, $i = 1, \dots, N$, $h_i = (h_{i1}, h_{i2})^T$. L'information apportée par les observations pour h_i est obtenue par une estimation par validation croisée de $\lambda_0(X_i)$ donnée par

$$\hat{\lambda}_{0,H_i}^{-i}(X_i) = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \prod_{l=1}^2 K_{h_{il}}(X_{il} - X_{jl}), \quad (1.7.17)$$

où $\hat{\lambda}_{0,h_i}^{-i}$ est l'estimateur "leave-one-out" c'est-à-dire l'estimateur de λ_0 sans l'événement X_i et K_{h_i} est le noyau classique. À partir de la formule de Bayes et de l'estimateur (1.7.17), l'estimateur de la loi a posteriori pour chaque H_i prend la forme suivante

$$\hat{\pi}(h_i|X_i) = \hat{\lambda}_{0,h_i}^{-i}(X_i|h_i)\pi(h_i) \left(\int_{\mathcal{H}} \hat{\lambda}_{0,h_i}^{-i}(X_i|h_i)\pi(h_i) dh_i \right)^{-1}. \quad (1.7.18)$$

Par l'estimateur de Bayes sous les pertes quadratique et entropique, le meilleur estimateur des h_i est la moyenne de $\pi(h_i|X_i)$ donnée respectivement par

$$\hat{h}_i^{qua} = \mathbb{E}(h_i|X_i) \quad (1.7.19)$$

et

$$\hat{h}_i^{ent} = \left[\mathbb{E}(h_i^{-1}|X_i) \right]^{-1}. \quad (1.7.20)$$

Remarque 1.7.1. *L'estimateur à noyau gaussien multivarié n'est pas adapté lorsque certaines variables sont bornées. Un problème de biais se pose aux bornes, ce qui entraîne la non-consistance de l'estimateur. Ce problème devient plus sévère dans le cas multivarié, car il s'ajoute au problème de dimension de support. Une première solution à ce problème dans le cas multivarié a été donnée par Bouezmarni et Roumbouts (2010), ils proposent d'utiliser l'estimateur à noyau produit composé des noyaux associés uni-variés.*

Dans ce qui suit, nous allons rappeler la notion unifiée d'un noyau associé univarié continu de cible x et de paramètre de lissage h ainsi que quelques exemples illustratifs.

1.8 Noyau associé et problème aux bords

Les noyaux classiques, de par leurs propriétés, présentent de nombreux avantages. Cependant, du fait de leurs symétries, ils deviennent inappropriés lorsqu'il s'agit d'estimer

des densités à support partiellement ou totalement borné. En effet, ils assignent des poids à l'extérieur du support de la densité à estimer lorsque le lissage se fait aux points de bord. Cela crée ainsi les biais dits de bord qui rendent l'estimateur non consistant. En multivarié, les conséquences des problèmes de bord sont bien plus sévères en ce sens que les régions de bord augmentent avec la dimension k . Les solutions aux problèmes de bord ont été moins bien étudiées que dans le cas univarié. Par exemple, Muller & Stadtmüller (1999) ont proposé des noyaux de bord définis sur des supports arbitraires où les noyaux multivariés sont obtenus en minimisant un problème variationnel. Nous pouvons aussi nous référer à Epanechnikov (1969). Bien que ces estimateurs aient des propriétés intéressantes, ils ne sont pas aisés à mettre en œuvre. Pour le cas univarié, nous pourrions nous référer, entre autres, aux méthodes dites des données reflétées ("*reflection data method*" en anglais) de Silverman (1986) et Cline & Hart (1991) et à celui des noyaux de bord ("*boundary kernels method*" en anglais) de Gasser et al. (1985) et Zhang & Karunamuni (2010).

Récemment, certains auteurs ont proposé dans le cas des densités à support compact, l'utilisation de noyaux dont le support coïncide avec celui de la densité à estimer. Ceci a efficacement résolu le problème des effets de bord puisque les noyaux utilisés ici sont généralement asymétriques et peuvent changer de forme selon la position du point d'estimation. C'est notamment le cas, en univarié, de Chen (1999)(2000a) avec les noyaux Bêta et Gamma pour estimer les densités à support respectivement $[0, 1]$ et $[0, +\infty[$, puis de Scaillet (2004) avec les noyaux gaussiens inverses et sa réciproque pour les densités à support $]0, +\infty[$. Toujours dans le cas continu, Libengué (2013) a proposé une méthode de construction de noyaux associés continus qui englobe, entre autres, les noyaux de Chen (1999)(2000a) et de Scaillet (2004). En multivarié continu, Bouezmarni & Rombouts (2010) ont proposé des estimateurs à noyaux produits composés des univariés (Bêta et Gamma) de Chen (1999)(2000a). Li & Racine (2007) ont utilisé ce produit de noyaux univariés auquel ils ont ajouté des noyaux univariés gaussiens pour les données continues. Tous ces noyaux multivariés ont une matrice de lissage diagonale alors que le rôle des matrices de lissages pleines a été démontré dans certaines situations.

1.9 Noyaux asymétriques continus

Cette section présente l'essentiel des travaux sur les noyaux univariés continus de Chen (1999)(2000a)(2000b) et Libengué (2013). Nous désignerons par W_1 le support de la fonction intensité à estimer. L'ensemble W_1 sera continu $[a, b]$ ou $[0, +\infty]$ pour tous réels $a < b$.

Définition 1.9.1. *On appelle type de noyau K_θ continu toute fonction de densité paramétrée par $\theta \in \Theta \subseteq \mathbb{R}^2$, de support $\mathbf{S}_\theta \subseteq \mathbb{R}$ et de carré intégrable.*

Nous ne considérerons dans la suite que les types de noyaux continus pour la construction des noyaux associés continus.

Définition 1.9.2. *Soit $W_1 \subseteq \mathbb{R}$ le support de la fonction densité de probabilité à estimer, une cible $x \in W_1$ et un paramètre de lissage h . Une fonction densité de probabilité $K_{x,h}(\cdot)$ de support $\mathbf{S}_{x,h} \subseteq \mathbb{R}$ est appelée "noyau associé" lorsque les conditions suivantes sont*

satisfaites :

$$x \in \mathbf{S}_{x,h}, \quad (1.9.1)$$

$$\mathbb{E}(\mathcal{Z}_{x,h}) = x + a(x, h), \quad (1.9.2)$$

$$\text{Var}(\mathcal{Z}_{x,h}) = b(x, h), \quad (1.9.3)$$

où $\mathcal{Z}_{x,h}$ est la variable aléatoire de fonction de densité $K_{x,h}$ et les quantités $a(x, h)$ et $b(x, h)$ tendent vers 0 quand h tend vers 0.

Nous notons que si l'on dispose d'un noyau associé continu, l'estimateur correspondant peut être facilement déduit. Si tel n'est pas le cas, il est possible de construire un noyau associé en utilisant le type de noyau approprié (continu). Le type de noyau utilisé doit avoir au moins le même nombre de paramètres que le nombre de composantes dans le couple (x, h) comme paramètres du noyau associé à construire. Dans le cas discret, il n'y a pas encore de méthode et dans le cas continu, Libengué (2013) a proposé une méthode dite "*mode-dispersion*" qui est rappelée ci-dessous.

Méthode de construction dans le cas continu

Méthodologie

Soit $K_{\theta(a,b)}$ un type de noyau uni modal sur $\mathbf{S}_{\theta(a,b)}$, de mode $M(a, b)$ et de paramètre de dispersion $D(a, b)$. La méthode "*mode-dispersion*" permet la construction d'une fonction $K_{\theta(x,h)}$, dépendant de x et h en résolvant en a et b le système.

$$\begin{cases} M(a, b) = x \\ D(a, b) = h. \end{cases} \quad (1.9.4)$$

Alors $\theta(x, h) = \theta(a(x, h), b(x, h))$ où $a(x, h)$ et $b(x, h)$ sont solutions du système précédent, pour $h > 0$ et $x \in W_1$ (support de la densité f à estimer).

Nous pourrions se référer à Jørgensen (1997)(2013) et Jørgensen & Kokonendji (2011)(2016) pour de plus amples informations sur les notions de dispersion unie et multivariée. La proposition suivante assure que la fonction noyau $K_{\theta(x,h)}$ issu de la méthode "*mode-dispersion*" est bien un noyau associé continu. En d'autres termes, il est montré que $K_{\theta(x,h)}$ satisfait les conditions de la définition 1.9.2.

Proposition 1.9.1. *Soit W_1 le support de la densité f , à estimer. Pour tout $x \in W_1$ et $h > 0$, la fonction noyau $K_{\theta(x,h)}$ construit par la méthode "*mode-dispersion*", de support $\mathbf{S}_{\theta(x,h)} = \mathbf{S}_{\theta(a(x,h), b(x,h))}$, vérifie*

$$x \in \mathbf{S}_{\theta(x,h)}, \quad (1.9.5)$$

$$\mathbb{E}(\mathcal{Z}_{\theta(x,h)}) - x = A_{\theta}(x, h), \quad (1.9.6)$$

$$\text{Var}(\mathcal{Z}_{\theta(x,h)}) = B_{\theta}(x, h), \quad (1.9.7)$$

où $\mathcal{Z}_{\theta(x,h)}$ est la variable aléatoire de fonction de densité $K_{\theta(x,h)}$ et les quantités $A_{\theta}(x,h)$ et $B_{\theta}(x,h)$ tendent vers 0 quand h tend vers 0.

La définition suivante présente l'estimateur à noyau associé continu obtenu soit de la définition 1.9.2, soit de la proposition 1.9.1 ; c'est-à-dire $K_{\theta(x,h)} = K_{x,h}$.

Définition 1.9.3. Soit X_1, \dots, X_N une suite de variables aléatoires i.i.d. de fonction intensité $\hat{\lambda}_{0,h}$ sur W_1 . L'estimateur à noyau associé continu

$$\hat{\lambda}_{0,h}(x) = \frac{1}{N} \sum_{i=1}^N K_{x,h}(X_i) \mathbf{1}_{N \neq 0}, \quad \forall x \in W_1, \quad (1.9.8)$$

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

Nous rappelons dans la suite quelques propriétés de l'estimateur (1.9.8) à noyau associé continu. Les propriétés asymptotiques ont, quant à elles, été étudiées par Libengué [91] pour le cas continu.

Proposition 1.9.2. Soit X_1, X_2, \dots, X_n un échantillon i.i.d. de fonction intensité λ_0 sur W_1 . Soit $\hat{\lambda}_{0,H} = \hat{\lambda}_{0,h,K}$ l'estimateur (1.9.8) de λ_0 utilisant un noyau associé. Alors, pour tout $x \in W_1$ et $h > 0$, on

$$\mathbb{E}[\hat{\lambda}_{0,H}(x)] = \mathbb{E}[\lambda_0(\mathcal{Z}_{x,h})], \quad (1.9.9)$$

où $\mathcal{Z}_{x,h}$ est la variable aléatoire liée à la fonction densité de probabilité $K_{x,h}$ sur $\mathcal{S}_{x,h}$. De plus, pour une fonction intensité, on a $\hat{\lambda}_{0,H}(x) > 0$ pour tout $x \in W_1$ et

$$\int_{W_1} \hat{\lambda}_{0,h}(x) \nu(dx) = \wedge_n, \quad (1.9.10)$$

où $\wedge_n = \wedge(n; h, K)$ est une constante positive et finie $\int_{W_1} K_{x,h}(t) \nu(dt)$ pour tout $t \in W_1$, et ν est une mesure de Lebesgue sur W_1 .

Exemples de noyaux associés

Le tableau 1.2 rappelle quelques noyaux associés univariés continus. Nous trouverons les détails sur ces noyaux ainsi que leurs applications dans les travaux de Libengué (2013) et Ziane et al. (2015).

TABLE 1.2 – Quelques noyaux associés continus univariés.

Noyaux	Support	Densité
Bêta étendu	$[a, b]$	$\frac{(y-a)^{\frac{(x-a)}{h(b-a)}} (b-y)^{\frac{(b-x)}{h(b-a)}}}{(b-a)^{1+h-1} B\left(1+\frac{(x-a)}{h(b-a)}, 1+\frac{(b-x)}{h(b-a)}\right)}$
Gamma	$[0, \infty[$	$\frac{1}{\Gamma(\frac{x}{h}+1)} \frac{y^{\frac{x}{h}} e^{-\frac{y}{h}}}{h^{\frac{x}{h}+1}}$
Gaussien inverse réciproque	$]0, \infty[$	$\frac{1}{\sqrt{2\pi h y}} \exp\left(-\frac{x-h}{2h} \left(\frac{y}{x-h} - 2 + \frac{x-h}{y}\right)\right)$
Lognormal	$[0, \infty[$	$\frac{1}{hy\sqrt{2\pi}} \exp\left(-\frac{1}{2h} \left(\log(y) - \log(x) + \sigma^2\right)^2\right)$
Gaussien inverse	$]0, \infty[$	$\frac{1}{\sqrt{2\pi h y^3}} \exp\left(-\frac{1}{2hx} \left(\frac{y}{x} - 2 + \frac{x}{y}\right)\right)$
Birnbaum-Saunders	$]0, \infty[$	$\frac{\sqrt{2\pi}}{2\sqrt{h}} \left(\sqrt{\frac{1}{xy}} + \sqrt{\frac{x}{y^3}}\right) \exp\left(-\frac{1}{2\alpha^2} \left(\frac{y}{x} - 2 + \frac{x}{y}\right)\right)$

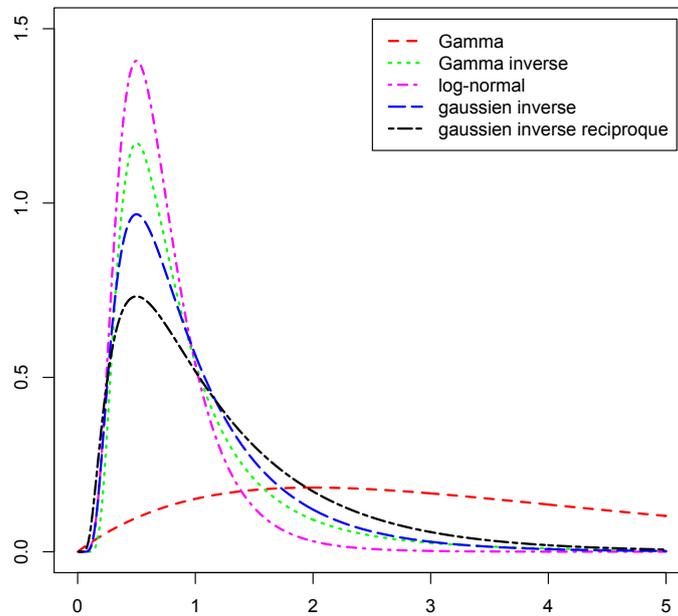


FIGURE 1.3 – Comportements de quelques noyaux associés non classiques de la droite réelle positive aux bords pour $x = h = 0.5$.

1.10 Estimateur à noyau associé produit

L'estimateur à noyau associé produit a été introduit par Bouezmarni et Roumbouts (2010), et vu le lien entre la densité et la fonction intensité, l'estimateur de la fonction intensité est de la forme suivante :

$$\hat{\lambda}_{0,h}(x_1, \dots, x_k) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^k K^j(h_j, X_{ij}) \mathbb{1}_{N \neq 0}(x_j), \quad (1.10.1)$$

où $(h_1, \dots, h_k)^T$ est le vecteur des paramètres de lissage et K^j est un noyau associé univarié pour la variable x_j . Ces deux auteurs ont étudié deux cas de variables à support bornées : le cas où le support de la variable est non-négatif, ils utilisent le noyau "local linear", noyau Gamma ou le noyau Gamma modifié. Si la variable est à support compact (par exemple $[0, 1]$), ils utilisent le noyau Bêta ou bien le noyau Bêta modifié.

1.11 Résolution de problèmes de biais de bordure

Tous les noyaux asymétriques du tableau 1.2 changent de forme en fonction de la cible x et n'ont apparemment pas de problèmes aux bords. Cependant, le biais dans le cas continu est différent de celui du noyau classique gaussien. Nous pourrions nous référer dans le cas continu à Malec & Schienle (2014), Hirukawa & Sakudo (2014) et Igarashi & Kakizawa (2015) qui ont traité ce problème de biais de bord pour certains noyaux asymétriques. Dans ce qui suit, nous rappelons l'algorithme de réduction du biais pour tout noyau associé univarié continu de Libengué (2013).

1.12 Cas continu : réduction du biais de l'estimateur

Dans le biais de l'estimateur $\hat{\lambda}_{0,h}(x)$ à noyau associé, on remarquera la présence du terme $\lambda'_0(x)$ (où $\lambda'_0(x)$ est la dérivée première de la fonction intensité λ_0) qui augmente ce biais (chapitre 5). S'inspirant des travaux de Zhang (2010) et Zhang & Karunamuni (2010), Libengué (2013) a proposé un algorithme pour éliminer le terme $\lambda'_0(x)$ dans la majeure partie du support continu W_1 . La première étape consiste à définir les régions intérieures et de bord ; tandis que la seconde traite le noyau associé modifié qui fournit un biais réduit.

1. Nous divisons le support $W_1 = [t_1, t_2]$ en deux régions d'ordre $\alpha(h) > 0$ avec $\alpha(h) \rightarrow 0$ quand $h \rightarrow 0$;
 - région intérieure (la plus grande région pouvant contenir 95% des observations) notée $W_{\alpha(h),0}$ et définie par l'intervalle

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

- régions de bord (pouvant être vide) représentées par les deux intervalles $W_{\alpha(h),-1}$ et $W_{\alpha(h),+1}$ respectivement définis par

$$W_{\alpha(h),-1} = [t_1, t_1 + \alpha(h)] \quad (\text{à gauche})$$

et

$$W_{\alpha(h),+1} = [t_2 - \alpha(h), t_2] \quad (\text{\`a droite}).$$

Les régions de bord constituent le complémentaire de la région intérieure et nous notons :

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

2. Nous modifions le noyau associé $K_\theta(x, h)$ correspondant à $A(x, h)$ et $B(x, h)$ en une nouvelle fonction noyau $K_{\tilde{\theta}}(x, h)$ correspondant à

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

et

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

de sorte que, pour un h fixé

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

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

soit continue sur W_1 et constant sur $W_{\alpha(h),0}^c$.

Estimateur de la fonction intensité du processus de Poisson non homogène

2.1 Introduction

Soit X un processus ponctuel spatial défini sur \mathbb{R}^2 et X_1, \dots, X_N une réalisation de X observés sur une région bornée $W \in \mathbb{R}^2$ et N est la variable aléatoire comptant le nombre d'événements. Pour faciliter la notation tout au long de ce travail, les événements des configurations de points (semis de points) sont écrits en majuscules tandis que les minuscules désignent tout point de \mathbb{R}^2 . Si dx est une région infinitésimale contenant le point $x \in \mathbb{R}^2$, la fonction d'intensité de premier ordre de X (Diggle (2003)) est défini comme suit :

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(dx)]}{|dx|} \right\} \quad (2.1.1)$$

où $|dx|$ et $N(dx)$ désignent la zone et le nombre d'événements de X dans dx , respectivement. Intuitivement, $\lambda(x)|dx|$ est la probabilité pour que dx contienne exactement un événement de X . Un processus ponctuel est homogène si son intensité de premier ordre est constante (i.e $\lambda(x) = \lambda > 0$) et sinon non-homogène. Bien que les processus de points spatiaux soient définis sur \mathbb{R}^2 , la région d'observation définit un support borné pour $\lambda(x)$.

Un problème principal lors de la modélisation des semis de points spatiaux non homogènes est l'estimation de leur intensité de premier ordre. En effet, nous pouvons supposer une structure paramétrique pour la fonction d'intensité et estimer les paramètres inconnus, par exemple, par pseudo vraisemblance maximale Diggle (2003) ; Waagepetersen (2007). Cependant, des estimations peu fiables peuvent être obtenues si le modèle paramétrique supposé s'écarte de la vraie fonction intensité.

Ainsi, nous pouvons obtenir des estimateurs de la fonction intensité λ en utilisant les techniques non paramétriques d'estimation de densité bivariée. Diggle (1985) introduit un estimateur à noyau de $\lambda(x)$ donné par :

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^N K_h(x - X_i) = \frac{1}{p_h(x)h^2} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right) \quad (2.1.2)$$

où la fonction du noyau, $K(\cdot)$, est une fonction de densité de probabilité bivariée radialement symétrique, $h > 0$ est le paramètre de lissage ou la bande passante, K_h est le

noyau lissé et $p_h(x) = \int_W h^{-2} K\left(\frac{x-y}{h}\right) dy$ est le terme de correction de bord nécessaire lorsque le domaine d'observation est limité.

2.2 Estimateur consistant

Soit X_1, \dots, X_N une réalisation du processus ponctuel spatial non homogène X observés sur W . Supposons que le processus ponctuel est un processus de Poisson (Diggle (2003)). L'hypothèse de Poisson est nécessaire pour développer la théorie asymptotique et garantir la consistance de l'estimateur du noyau de la densité des lieux d'événements. Cependant, l'estimateur du noyau et le paramètre de lissage proposés dans ce travail peuvent être appliqués à d'autres processus ponctuels. Dans l'étude de simulation, nous analysons leurs performances pour les processus de Poisson ponctuels.

Notez que nous ne pouvons pas distinguer **l'hétérogénéité** et **l'interaction** dans une configuration de point observé, sauf si nous avons des informations supplémentaires, telles que des covariables ou un modèle paramétrique (Diggle (2003)). Ainsi, la pratique courante dans l'analyse de semis de point spatial suppose que le processus ponctuel est de Poisson, en estimant l'intensité du premier ordre, et puis estimer les propriétés de second ordre pour tester l'hypothèse de Poisson. Par conséquent, l'hypothèse de Poisson n'est pas très restrictive en pratique. Étant donné que le nombre d'événements d'un processus ponctuel de Poisson non homogène, N a la distribution de

$$Poisson\left(\int_W \lambda(x) dx\right) = Poisson(m),$$

nous pouvons établir la relation suivante entre une densité bivariée et l'intensité de premier ordre d'un processus ponctuel spatial donné par :

$$f(x) = \frac{\lambda(x)}{\int_W \lambda(x) dx} = \frac{\lambda(x)}{m}. \quad (2.2.1)$$

Compte tenu de cette relation, Cucala (2006) a défini la densité des emplacements des événements comme

$$\lambda_0(x) = \frac{\lambda(x)}{m}.$$

Soit H une matrice symétrique et définie positive, l'estimateur du noyau de $\lambda_0(\cdot)$ est donné par :

$$\hat{\lambda}_{0,H}(x) = \frac{\hat{\lambda}_H(x)}{N} \mathbb{1}_{[N \neq 0]} = \frac{|H|^{-1/2}}{p_H(x)N} \sum_{i=1}^N K\left(H^{-1/2}(x - X_i) \mathbb{1}_{[N \neq 0]}\right). \quad (2.2.2)$$

Un critère d'erreur globale commune pour mesurer les performances d'un estimateur à noyau est le *MISE*. Il s'agit de **l'erreur quadratique moyenne intégrée** entre une estimation et la fonction cible. Dans notre cas

$$MISE(H) = \mathbb{E} \left[\int_W (\hat{\lambda}_{0,H}(x) - \lambda_0(x))^2 dx \right] = \int_W B(x, H)^2 dx + \int_W Var(x, H) dx \quad (2.2.3)$$

où \mathbb{E} est l'espérance du caractère aléatoire à la fois de l'emplacement et du nombre d'événements et $B(x, H)$, $Var(x, H)$ sont le biais et la variance de $\hat{\lambda}_{0,H}(x)$ respectivement. Pour obtenir l'expression asymptotique du $MISE(H)$, nous devons supposer un cadre asymptotique. Pour l'estimation de la densité du noyau dans \mathbb{R}^k , le cadre asymptotique suppose que la taille de l'échantillon déterministe, n , tend vers l'infini et tous les éléments de la matrice de lissage H tendent vers 0 telles que $n|H|^{1/2} \rightarrow \infty$. Dans l'analyse des semis de points spatiaux, deux travaux dans le cadre asymptotique peuvent être supposés :

- l'asymptotique **extensive** (increasing domain asymptotics) : c'est la situation où le nombre d'observations disponible croît avec le domaine d'observation W ;
- l'asymptotique **intensive** (infill asymptotics) : c'est la situation où les observations augmentent dans un domaine W fixé et borné (nous avons de plus en plus de sites d'observation tandis que la région reste fixe).

Le premier (asymptotique **extensive**) (Cressie (1993) ; Guan (2008b)) dicte que le nombre prévu d'événements tend vers l'infini avec la taille de la région d'observation. Ce cadre, qui augmente le nombre d'événements en gardant la distance entre eux, peut être supposé analyser la structure du second ordre des semis de points spatiaux (Guan (2007)) et d'estimer l'intensité de premier ordre si nous avons des informations supplémentaires tel qu'un modèle paramétrique (Guan (2008a)) ou des covariables (Guan (2008b)).

Cependant, dans cette hypothèse, nous ajoutons uniquement des informations dans la limite, plutôt qu'ailleurs, et lorsque tous les éléments de H tendent vers zéro, l'intensité estimée à chaque point dépend d'un nombre attendu d'événements tendant vers zéro.

Pour surmonter ce problème, Diggle et Marron (1988) ont adopté le cadre asymptotique **intensive**, qui dicte que le nombre attendu d'événements tend vers l'infini, à savoir $\int_{\mathbb{R}^2} \lambda(x) dx \rightarrow \infty$, où $m = \int_W \lambda(x) \rightarrow \infty$ pour un domaine d'observation borné. Ainsi, à chaque emplacement, l'intensité de premier ordre est estimée à l'aide d'un certain nombre d'événements qui tendent vers 1. Par conséquent, nous développons la théorie asymptotique de ce travail sous l'asymptotique **intensive**. Ce cadre a également été supposé par Cowling et al. (1996) pour l'estimation de l'intensité du noyau dans \mathbb{R} et par Cucala (2006), qui a montré la consistance de (2.2.1) avec une bande passante scalaire et a fourni une expression asymptotique pour son $MISE$ sur \mathbb{R}^2 . L'asymptotique **extensive** peut être utile en épidémiologie, lorsque le nombre d'événements et la zone d'étude augmentent à mesure que la maladie se propage. Cependant, l'asymptotique **intensive** est plus appropriée lorsque, comme dans le modèle de point des incendies de forêt analysé, la région est fixée à l'avance et le nombre d'événements augmente avec le temps.

2.3 Propriétés de l'estimateur

Conditions de régularité

Nos résultats reposent sur les hypothèses de régularité portées sur la matrice des paramètres de lissages et sur l'intensité de premier ordre suivant :

H.2.1 La matrice H est symétrique et définie positive et que $H \rightarrow 0$ et $m^{-1}|H|^{-1/2} \rightarrow 0$, quand $m \rightarrow \infty$.

H.2.2 $\lambda(\cdot)$ admet des dérivées partielles jusqu'à l'ordre quatre, toutes ses dérivées partielles sont continues et de carrées intégrables.

H.2.3 $K(\cdot)$ est une densité continue, symétrique et de carrée intégrale tel que

$$\int_{\mathbb{R}^2} uu^T K(u) du = \mu_2(K) \mathbb{I}_2,$$

avec $\mu_2(K) < \infty$.

Avant d'analyser la performance de (2.2.2), une notation sera introduite.

- Soit A une matrice carrée. La trace de A , notée $tr(A)$, est la somme des éléments diagonaux de A et nous avons

$$tr(AB) = tr(BA),$$

pour toute matrice carrée B de même ordre que A .

- Soit A une matrice carrée d'ordre k . Le vecteur de A , noté par $vecA$, est le $(k^2 \times 1)$ vecteur obtenu en mettant les colonnes de A l'une sous l'autre dans l'ordre, de gauche à droite.
- Le demi-vecteur de A , noté par $vechA$, est le $(\frac{1}{2}k(k+1) \times 1)$ vecteur obtenu de $vecA$ par l'élimination des éléments de A situés au-dessus de la diagonale principale.

Exemple 2.3.1. Si $A = \begin{pmatrix} 1 & 4 \\ 7 & 3 \end{pmatrix}$, alors $vecA = \begin{pmatrix} 1 \\ 7 \\ 4 \\ 3 \end{pmatrix}$ et $vechA = \begin{pmatrix} 1 \\ 7 \\ 3 \end{pmatrix}$.

- Pour toute fonction de carrée intégrable g , $R(g) = \int_{\mathbb{R}^2} g(x)^2 dx$. Soit $r = (r_1, r_2)^T$, où r_1 et r_2 sont des entiers non négatifs et $|r| = r_1 + r_2$, alors la dérivée partielle r de λ_0 peut être écrite comme :

$$\lambda_0^{(r)}(x) = \frac{\partial^{|r|} \lambda_0(x)}{\partial x_1^{r_1} \partial x_2^{r_2}}.$$

- Notons $D^r \lambda_0$ la $r^{ième}$ dérivée de λ_0 . En utilisant cette notation $D\lambda_0 = \nabla \lambda_0$ et $D^2 \lambda_0 = \nabla^2 \lambda_0$ sont respectivement le vecteur gradient et la matrice hessienne de λ_0 .

Ψ_4 est la matrice 3×3 donnée par :

$$\Psi_4 = \int_{\mathbb{R}^2} vech[2\nabla^2 \lambda_0(x) - dg \nabla^2 \lambda_0(x)] vech^T [2\nabla^2 \lambda_0(x) - dg \nabla^2 \lambda_0(x)] dx \quad (2.3.1)$$

où pour tout $j = |r|$, chaque entrée dans ψ_j est la fonction dérivée de la fonction intensité intégrée

$$\psi_r = \int_{\mathbb{R}^2} \lambda_0^{(r)}(x) \lambda_0(x) dx = \mathbb{E}[\lambda_0^{(r)}(x)] \quad (2.3.2)$$

ici \mathbb{E} représente l'espérance sur la densité bivariée $\lambda_0(x)$. En utilisant l'intégration par parties, nous pouvons voir que $\psi_r \neq 0$ seulement si $|r|$ est pair. Pour $j = 4$

$$\Psi_4 = \begin{pmatrix} \psi_{40} & 2\psi_{31} & \psi_{22} \\ 2\psi_{31} & 4\psi_{22} & 2\psi_{13} \\ \psi_{22} & 2\psi_{13} & \psi_{04} \end{pmatrix}. \quad (2.3.3)$$

2.3.1 Mesure d'erreur et matrice de lissage optimale.

L'estimateur du noyau pour la densité des emplacements d'événements d'une configuration de points observés sur une région bornée W est donné par :

$$\hat{\lambda}_{0,H}(x) = \frac{\hat{\lambda}_H(x)}{N} \mathbb{1}_{[N \neq 0]} = \frac{|H|^{-1/2}}{p_H(x)N} \sum_{i=1}^N K(H^{-1/2}(x - X_i) \mathbb{1}_{[N \neq 0]}). \quad (2.3.4)$$

En considérant $g(X_i) = K_H(x - X_i)$ dans (2.2.2) et en appliquant un développement de Taylor à l'ordre 2, la moyenne de $\hat{\lambda}_{0,H}(x)$ peut s'écrire comme suit :

$$\begin{aligned} p_H(x)(1 - e^{-m})^{-1} \mathbb{E}(\hat{\lambda}_{0,H}(x)) &= \int_W K_H(x - y) \lambda_0(y) dy = \int_{B_{x,H}} \lambda_0(x - u^T H^{1/2}) K(u) du \\ &= \lambda_0(x) \int_{B_{x,H}} K(u) du - \int_{B_{x,H}} u^T H^{1/2} D\lambda_0(x) K(u) du + \frac{1}{2} \int_{B_{x,H}} u^T H D^2 \lambda_0(x) K(u) du + o(\text{tr}(H)) \end{aligned}$$

avec

$$B_{x,H} = \{H^{-1}(x - y); y \in W\} = H^{-1}(x - W)$$

et

$$p_H(x) = \int_W K_H(x - y) dy = \int_{B_{x,H}} K(u) du.$$

Alors, le biais de $\hat{\lambda}_{0,H}(x)$ est

$$\begin{aligned} B(x, H) &= -e^{-m} \lambda_0(x) - \frac{(1 - e^{-m})}{p_H(x)} \int_{B_{x,H}} u^T H^{1/2} D\lambda_0(x) K(u) du \\ &\quad + \frac{(1 - e^{-m})}{p_H(x)} \left(\frac{1}{2} \int_{B_{x,H}} u^T H D^2 \lambda_0(x) K(u) du + o(\text{tr}(H)) \right). \end{aligned}$$

D'après les hypothèses **H.2.1** – **H.2.3** $B(x, H) \rightarrow 0$ quand $m \rightarrow \infty$ et les éléments de $H \rightarrow 0$, alors $\hat{\lambda}_{0,H}(x)$ est asymptotiquement sans biais. Cependant, comme $\lim_{|H| \rightarrow 0} B_{x,H}$ dépend de W , nous pouvons donc obtenir une expression simplifiée de $B(x, H)$.

Quand $W = \mathbb{R}^2$, $p_H(x) = 1$ et $B_{x,H} = \mathbb{R}^2$. En ajoutant les conditions de l'hypothèse **H.2.3** c'est-à-dire que K est un noyau symétrique et $\int_{\mathbb{R}^2} uu^T K(u) du = \mu_2(K) \mathbf{I}_2$, avec $\mu_2(K) < \infty$, alors

$$B(x, H) = -e^{-m} \lambda_0(x) - \frac{(1 - e^{-m})}{2} \mu_2(K) \text{tr}(HD^2 \lambda_0(x)) + o((1 - e^{-m} \text{tr}(H))) \quad (2.3.5)$$

est d'ordre $O(\text{tr}(H))$.

Considérant $g(X_i) = K_H(x - X_i)$, la variance de $\hat{\lambda}_{0,H}(x)$ est

$$\text{Var}(x, H) = \frac{A(m)}{p_H(x)} \int_W (K_H(x - y))^2 \lambda_0(y) dy - \frac{A(m) - e^{-m} + e^{-2m}}{p_H(x)} \left[\int_W (K_H(x - y)) \lambda_0(y) dy \right]^2$$

où

$$A(m) = \mathbb{E} \left[\frac{1}{N} \mathbf{1}_{[N>0]} \right] = e^{-m} \sum_{k=1}^{\infty} \frac{m^k}{kk!}.$$

Par le développement de Taylor, le premier terme donne :

$$\begin{aligned} \int_W (K_H(x - y))^2 \lambda_0(y) dy &= |H|^{-1/2} \int_W K(u)^2 \lambda_0(x - u^T H^{1/2} du) \\ &= |H|^{-1/2} \lambda_0(x) \int_W K(u)^2 du + o(|H|^{-1/2}). \end{aligned}$$

En considérant que $\lambda_0(x)$ borné et K une fonction densité bivariée, l'intégrale du second terme de l'expression de la variance est d'ordre 1 et

$$\text{Var}(x, H) = \frac{A(m)|H|^{-1/2}}{p_H(x)} \lambda_0(x) \int_{B_{x,H}} K(u)^2 du + o(A(m)|H|^{-1/2}) \quad (2.3.6)$$

tend vers 0 comme $m^{-1}|H|^{-1/2} \rightarrow 0$ quand $m \rightarrow \infty$. Cependant, le terme de correction de bord n'affecte pas la vitesse de convergence de la variance.

2.3.2 Critère d'erreur

L'évaluation de la similarité entre l'estimateur $\hat{\lambda}_{0,H}(x)$ et la vraie intensité λ_0 à estimer nécessite des critères d'erreur. La mesure la plus naturelle utilisée est l'erreur quadratique moyenne. Nous définissons d'abord l'erreur quadratique moyenne ("Mean Squared Error" en anglais)(MSE) par :

$$\begin{aligned}
MSE(x, H) &= B^2(x, H) + Var(x, H) & (2.3.7) \\
&= \frac{(1 - e^{-m})^2}{4} \mu_2(K)^2 tr^2(HD^2\lambda_0(x)) + e^{-2m} \lambda_0^2(x) \\
&\quad - e^{-m}(1 - e^{-m}) \lambda_0(x) \mu_2(K) tr(HD^2\lambda_0(x)) + A(m) |H|^{-1/2} \lambda_0(x) R(K) \\
&\quad + o((1 - e^{-m})^2 tr^2(H)) + o(e^{-m}(1 - e^{-m}) tr(H)) + o(A(m) |H|^{-1/2}) & (2.3.8)
\end{aligned}$$

où $R(K) = \int K^2(x) dx$.

Une mesure globale de l'efficacité de l'estimateur $\hat{\lambda}_{0,H}(x)$ est obtenue en intégrant le MSE sur tout le support W de λ_0 . Il s'agit de " l'Erreur Quadratique Moyenne Intégrée", en anglais "*Mean Integrated Squared Error*" ($MISE$), elle est donnée par :

$$\begin{aligned}
MISE(H) &= \int_{\mathbb{R}^2} (B^2(x, H) + Var(x, H)) dx \\
&= \frac{(1 - e^{-m})^2}{4} \mu_2(K)^2 \int_{\mathbb{R}^2} tr^2(HD^2\lambda_0(x)) dx + e^{-2m} \int_{\mathbb{R}^2} \lambda_0^2(x) dx \\
&\quad - e^{-m}(1 - e^{-m}) \mu_2(K) \int_{\mathbb{R}^2} \lambda_0(x) tr(HD^2\lambda_0(x)) dx + A(m) |H|^{-1/2} R(K) \\
&\quad + o((1 - e^{-m})^2 tr^2(H)) + o(e^{-m}(1 - e^{-m}) tr(H)) + o(A(m) |H|^{-1/2}). & (2.3.9)
\end{aligned}$$

Comme indiqué par Marron et Wand (1992) pour l'estimateur de densité du noyau, le terme de rappel dans $MISE$ peut être considérablement grand pour les processus ponctuels à faible intensité moyenne. Cependant, nous avons observé que même lorsque le $AMISE$ est un mauvais estimateur du $MISE$, il fonctionne raisonnablement bien comme critère de sélection de la bande passante. La minimisation de $AMISE(H)$ ne peut être effectuée que numériquement. Pour ce faire, nous pouvons appliquer un algorithme de Newton-Raphson, comme proposé par Wand (1992), pour les estimateurs de densité de noyau multivarié.

Afin de choisir la matrice de lissage optimale, nous chercherons une matrice qui minimise le MSE asymptotique (d'où le $AMSE$) ou le $MISE$ asymptotique ($AMISE$).

Sous les mêmes hypothèses **H.2.1**, **H.2.2** et **H.2.3**, l'expression du $AMSE$ et du $AMISE$ deviennent

$$\begin{aligned}
AMSE(x, H) &= \left(-e^{-m} \lambda_0(x) - \frac{(1 - e^{-m})}{2} \mu_2(K) tr(HD^2\lambda_0(x)) \right)^2 \\
&\quad + \frac{A(m) |H|^{-1/2}}{p_H(x)} \lambda_0(x) \int_W K(u)^2 du & (2.3.10)
\end{aligned}$$

et

$$\begin{aligned}
AMISE(H) &= \int_W \left(-e^{-m} \lambda_0(x) - \frac{(1 - e^{-m})}{2} \mu_2(K) \text{tr}(HD^2 \lambda_0(x)) \right)^2 dx \\
&+ \int_W \frac{A(m) |H|^{-1/2}}{p_H(x)} \lambda_0(x) \int_{B_{x,H}} K(u)^2 du dx. \tag{2.3.11}
\end{aligned}$$

Comme indiqué par Marron et Wand (1992) pour l'estimateur de densité du noyau, le terme de rappel dans $MISE$ peut être considérablement grand pour les processus ponctuels à faible intensité moyenne. Cependant, nous avons observé que même lorsque le $AMISE$ est un mauvais estimateur du $MISE$, il fonctionne raisonnablement bien comme critère de sélection de la bande passante. La minimisation de $AMISE(H)$ ne peut être effectuée que numériquement. Pour ce faire, nous pouvons appliquer un algorithme de Newton-Raphson, comme proposé par Wand (1992), pour les estimateurs de densité de noyau multivarié.

2.3.3 Matrice optimale

Pour avoir une bonne estimation par la méthode des noyaux, il faut bien choisir les paramètres de la matrice de lissage puisque cette matrice joue un rôle crucial dans le processus. Lorsque les paramètres de lissage sont très petits, l'estimateur est très volatile et on parle de sous-lissage (“*under-smoothing*” en anglais). En revanche, lorsque les paramètres grandissent, l'estimateur est alors de moins en moins influencé par les données, on parle d'un effet de sur lissage (“*over-smoothing*” en anglais). En pratique, il est primordial de trouver la meilleure matrice de lissage qui permet d'éviter le sous-lissage et le sur lissage. Les méthodes existantes pour le choix de la matrice de lissage peuvent être classées en deux catégories. La première catégorie est constituée des méthodes purement théoriques qui sont basées sur la minimisation de l'erreur quadratique moyenne intégrée ($MISE$). En effet, la valeur idéale théorique de H s'obtient en minimisant le $MISE$ asymptotique donné en (2.3.11). Ainsi, pour un échantillon de taille donnée et pour un noyau (classique) K fixé, cette valeur idéale de H est donné par minimisation du $AMISE$. Si nous restreignons H à la famille des matrices diagonales définies positives, c'est-à-dire $H = \text{diag}(h_1^2, h_2^2)$, nous obtenons une expression simplifiée de (2.3.11). Le $AMISE$ de $\hat{\lambda}_{0,H}(x)$ pour $h = (h_1, h_2)^T$ est

$$AMISE(H) = \frac{1}{4} \mu_2(K)^2 (h_1^4 \psi_{40} + 2h_1^2 h_2^2 \psi_{22} + h_2^4 \psi_{04}) + \frac{A(m)}{\sqrt{h_1^2 h_2^2}} R(K). \tag{2.3.12}$$

Nous minimisons le $AMISE$ en annulant le gradient donné par :

$$DAMISE(h) = \begin{pmatrix} \frac{1}{4} \mu_2(K)^2 (2h_1^2 \psi_{40} + 2h_2^2 \psi_{22}) + \frac{A(m)}{2\sqrt{h_1^4 h_2^2}} R(K) \\ \frac{1}{4} \mu_2(K)^2 (2h_2^2 \psi_{04} + 2h_1^2 \psi_{22}) + \frac{A(m)}{2\sqrt{h_1^2 h_2^4}} R(K) \end{pmatrix} = 0 \tag{2.3.13}$$

les expressions de h_1^2 et h_2^2 sont obtenues en résolvant le $DAMISE$. Nous obtenons donc :

$$h_{2,AMISE}^2 = h_{1,AMISE}^2 \left(\frac{\psi_{40}}{\psi_{04}} \right)^{1/2} \tag{2.3.14}$$

et

$$\mu_2(K)^2 \left(h_1^6 \frac{\psi_{40}^{5/4} \psi_{04}^{1/2}}{\psi_{40}^{3/4}} + h_1^6 \frac{\psi_{40}^{3/4} \psi_{22}}{\psi_{40}^{3/4}} \right) + \frac{A(m)}{\sqrt{h_1^2 h_2^4}} R(K) = 0$$

par conséquent,

$$h_{1,AMISE}^2 = \left[\frac{\psi_{04} A(m) R(K)}{\mu_2(K)^2 \psi_{40}^{3/4} (\psi_{40}^{1/2} \psi_{04}^{1/2} + \psi_{22})} \right]^{1/3}. \quad (2.3.15)$$

Finalement, pour $H = h^2 \mathbf{I}_2$, l'expression du $AMISE$ est donnée par :

$$AMISE(h) = \frac{h^4}{4} \mu_2(K)^2 (\psi_{40} + 2\psi_{22} + \psi_{04}) + \frac{A(m)}{h^2} R(K)$$

et le paramètre de lissage correspondant est donné par :

$$h_{AMISE}^2 = \left[\frac{2A(m)R(K)}{\mu_2(K)^2 (\psi_{40} + 2\psi_{22} + \psi_{04})} \right]^{1/3}. \quad (2.3.16)$$

Théorème 2.3.1. *Lo (2018) Sous les hypothèses **H.2.1**, **H.2.2** et **H.2.3**,*

a) *Les paramètres de lissages $h_{1,AMISE}$ et $h_{2,AMISE}$ minimisent $AMISE(H)$ donné par l'équation (2.3.11) si*

$$\psi_{40}, \psi_{04}, (\psi_{40}^{1/2} \psi_{04}^{1/2} + \psi_{22}) \neq 0.$$

b) *Les paramètres de lissages $h_{1,AMISE}$ et $h_{2,AMISE}$ minimisent $AMSE(H)$ donné par l'équation (2.3.11) si*

$$\frac{\partial^2 \lambda_0(x)}{\partial x_1^2} \frac{\partial^2 \lambda_0(x)}{\partial x_2^2} > 0.$$

Corollaire 2.3.1. *Lo (2018) Sous les mêmes hypothèses que le théorème 2.3.1,*

(i) $MSE(x, H) = AMSE(x, H) + o\left(A(m)|H|^{-\frac{1}{2}} + tr^2(H)\right).$

(ii) $MISE(H) = AMISE(H) + o\left(A(m)|H|^{-\frac{1}{2}} + tr^2(H)\right).$

Le $AMISE$ de $\hat{\lambda}_{0,H}(x)$ dépend de m et Ψ_4 qui, ainsi que l'intensité de premier ordre, sont inconnus. Ainsi, afin d'utiliser le $MISE$ comme critère d'erreur pour sélectionner la matrice de lissage optimale, nous devons obtenir un estimateur de $AMISE(H)$.

Deux catégories de méthodes classiques ont été proposées pour le choix de ces paramètres de lissage : les méthodes plug-in et la méthode validation croisée.

Pour les méthodes plug-in, plusieurs points peuvent être soulignés. D'abord la technique de plug-in est satisfaisante théoriquement dans le cas où la matrice de lissage optimal obtenu en minimisant le critère $AMISE$ s'exprime d'une manière simple.

Cependant, plusieurs difficultés se posent en pratique pour cette technique. En effet, pour calculer la matrice optimale, il faut estimer une ou plusieurs quantités inconnues. Ceci s'avère techniquement très difficile selon le type du noyau utilisé et des restrictions sont aussi imposées sur la fonction intensité à estimer λ souvent très difficiles à vérifier dans la pratique. De plus, la matrice de lissage obtenu par plug-in n'est valable qu'asymptotiquement.

Le principal intérêt des méthodes validation croisées est qu'elles n'exigent pas des hypothèses sur la fonction intensité à estimer λ_0 comme c'est le cas des méthodes plug-in. Pour être applicables en pratique, ces méthodes sont guidées entièrement par les événements. Cependant, cette approche présente quelques inconvénients majeurs : la matrice de lissage estimée par la technique de validation croisée présente une grande variabilité, c'est-à-dire que les valeurs des paramètres de la matrice de lissage peuvent être très différentes après plusieurs applications de la méthode à un même échantillon. Ceci relève du problème d'automatisation de la méthode validation croisée.

2.4 Conclusion

Ce chapitre est consacré à l'introduction de l'estimateur à noyau dans le cas bivarié. Ainsi, nous avons rappelé la forme de l'estimateur à noyau dans le cas multidimensionnel continu cité dans la littérature et étudié ses propriétés statistiques (espérance, biais, variance et erreur quadratique moyenne intégrée). Nous sommes intéressés aussi à la sélection de la matrice de lissage minimisant l'erreur quadratique moyenne intégrée à partir d'un noyau unidimensionnel produit.

Approche bayésienne locale pour le choix de la matrice de lissage de la fonction intensité.

3.1 Introduction

L'approche bayésienne locale a été proposée respectivement par Gangopadhyay, Cheung (2002) et De Lima et Atuncar (2011) dans le cas uni et multidimensionnel continu et par Zougab et al. (2012) et Belaid et al. (2016) dans le cas uni et multidimensionnel discret. Nous proposons dans ce chapitre de l'appliquer dans le cas de la fonction intensité du processus de Poisson de premier ordre. Cette approche consiste à sélectionner la matrice des fenêtres de lissage localement, c'est à dire en chaque vecteur cible x . La matrice de lissage H est considérée comme une matrice de variables aléatoires de loi a priori $\pi(\cdot)$. Nous utilisons le noyau gaussien univarié produit avec comme loi a priori la loi Gamma inverse. En exploitant la conjugabilité entre le noyau gaussien utilisé et la loi a priori choisie, nous obtenons la forme exacte de l'estimateur de la matrice H . Une étude de simulation ainsi que des applications sur des données réelles seront présentées.

3.2 Méthodologie

Dans cette approche, la matrice de lissage H est considérée comme une matrice de variables aléatoires de loi a priori $\pi(H)$. À partir de la formule de Bayes, la loi a posteriori de H est de la forme :

$$\pi(H|x) = \lambda_0(x)\pi(H) \left(\int_{\mathcal{H}} \lambda_0(x)\pi(H)dH \right)^{-1} \quad (3.2.1)$$

où \mathcal{H} est l'ensemble des matrices diagonales symétriques définies positives. Comme le modèle $\lambda_0(x)$ est inconnu, nous le remplaçons par son estimateur à noyau $\hat{\lambda}_{0,H}$. Ainsi, une estimation ou approximation de la loi a posteriori $\pi(H|x)$ est :

$$\hat{\pi}(H|x, X_1, \dots, X_N) = \hat{\lambda}_{0,H}(x)\pi(H) \left(\int_{\mathcal{M}} \hat{\lambda}_{0,H}(x)\pi(H)dH \right)^{-1}. \quad (3.2.2)$$

En fait, $\hat{\lambda}_{0,H}(x)$ représente l'information nécessaire sur $H = H(x)$ pour compléter la loi a priori $\pi(\cdot)$. L'utilisation de l'estimateur de Bayes sous la perte quadratique conduit au meilleur estimateur $\hat{H}(x)$ de H en x par :

$$\hat{H}(x) = \int_{\mathcal{M}} H \hat{\pi}(H|x, X_1, \dots, X_N) dH. \quad (3.2.3)$$

Dans certains cas, les expressions de la loi a posteriori (3.2.2) et l'estimateur bayésien de H (3.2.3) peuvent être obtenues explicitement en utilisant un prior conjugué. Dans le cas où nous ne pouvons pas les obtenir explicitement, nous faisons appel aux méthodes d'approximations de Monté Carlo.

Approche bayésienne avec le noyau classique pour l'estimation de la fonction intensité.

L'estimateur de la fonction intensité associé au noyau classique produit est :

$$\hat{\lambda}_{0,H}(x) = \frac{1}{Np_H(x)} \sum_{i=1}^N \prod_{j=1}^2 K_{x_j, h_j}(X_{ij}) \mathbb{1}_{N \neq 0} \quad (3.2.4)$$

$$= \frac{1}{Np_H(x)} \sum_{i=1}^N \frac{1}{\sqrt{h_1^2}} \frac{1}{\sqrt{h_2^2}} K\left(\frac{x_1 - X_{i1}}{\sqrt{h_1^2}}\right) K\left(\frac{x_2 - X_{i2}}{\sqrt{h_2^2}}\right) \mathbb{1}_{N \neq 0} \quad (3.2.5)$$

où $X_i = (X_{i1}, X_{i2})^T$, $i = 1, 2, \dots, N$, $x = (x_1, x_2)^T$ et h_1, h_2 sont des paramètres de lissages univariés.

En utilisant le noyau gaussien multivarié produit donné par

$$K(u) = \frac{1}{2\pi} \exp\left(-\frac{u_1^2}{2}\right) \exp\left(-\frac{u_2^2}{2}\right), u = (u_1, u_2)^T \quad (3.2.6)$$

l'estimateur devient :

$$\hat{\lambda}_{0,H}(x) = \frac{1}{2\pi Np_H(x)} \sum_{i=1}^N \frac{1}{\sqrt{h_1^2}} \frac{1}{\sqrt{h_2^2}} \exp\left(-\frac{1}{2} \frac{(x_1 - X_{i1})^2}{h_1^2}\right) \exp\left(-\frac{1}{2} \frac{(x_2 - X_{i2})^2}{h_2^2}\right) \mathbb{1}_{N \neq 0}. \quad (3.2.7)$$

Notons que la loi a posteriori de l'estimateur bayésien de H par la fonction perte quadratique peut s'obtenir d'une manière explicite en utilisant un certain choix de la loi a priori. Nous exploitons notre approche dans ce sens pour trouver la forme explicite de la loi a posteriori et de l'estimateur bayésien de H .

Choix de la loi a priori

Dans ce qui suit, nous supposons que chaque h_j^2 a une distribution a priori gamma inverse $IG(\alpha, \beta)$ avec le paramètre de forme $\alpha > 0$ et le paramètre d'échelle $\beta > 0$. Rappelons que la densité de $IG(\alpha, \beta)$ avec $a, b > 0$ est défini comme

$$\Phi_{a,b}(y) = \frac{b^a}{\Gamma(a)} \frac{1}{y^{a+1}} \exp\left(-\frac{b}{y}\right) \quad (3.2.8)$$

et donc la loi priori est donnée par :

$$\pi(H) = \prod_{j=1}^2 \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(h_j^2)^{\alpha+1}} \exp\left(-\frac{\beta}{h_j^2}\right). \quad (3.2.9)$$

Estimateur de la loi a posteriori et de la matrice de lissage

De ces considérations, on obtient une forme explicite de l'estimateur de la loi a posteriori et de l'estimateur bayésien de $\hat{H}(x) = (\hat{h}_1(x_1), \hat{h}_2(x_2))$ par le résultat suivant :

Théorème 3.2.1. (i) d'après l'équation (3.2.2), l'estimateur de la densité a posteriori $\hat{\pi}(H|x, X_1, \dots, X_N)$ est donné par :

$$\hat{\pi}(H|x, X_1, X_2, \dots, X_N) = \frac{1}{D_{ij}} \sum_{i=1}^N \prod_{j=1}^2 (A_{ij})^{-\alpha-\frac{1}{2}} \Phi_{\alpha+\frac{1}{2}, A_{ij}}(h_j^2) \quad (3.2.10)$$

avec

$$D_{ij} = \sum_{i=1}^N (A_{i1}A_{i2})^{-\alpha-\frac{1}{2}}, \quad i = 1, \dots, N, \quad j = 1, 2.$$

$$A_{i1} = \beta + \frac{(x_1 - X_{i1})^2}{2}, \quad A_{i2} = \beta + \frac{(x_2 - X_{i2})^2}{2}$$

(ii) Sous la fonction perte quadratique, les éléments diagonaux de l'estimateur bayésien de la matrice H pour chaque $h_j(x_j)$, $i = 1, \dots, N$ et $j = 1, 2$ sont :

$$\hat{h}_1^2(x_1) = \frac{1}{(\alpha - \frac{1}{2})D_{ij}} \sum_{i=1}^N A_{i1}^{-\alpha+\frac{1}{2}} A_{i2}^{-\alpha-\frac{1}{2}} \quad (3.2.11)$$

$$\hat{h}_2^2(x_2) = \frac{1}{(\alpha - \frac{1}{2})D_{ij}} \sum_{i=1}^N A_{i1}^{-\alpha-\frac{1}{2}} A_{i2}^{-\alpha+\frac{1}{2}} \quad (3.2.12)$$

avec $\alpha > \frac{1}{2}$. \square

Preuve :

i) On réécrit $\hat{\pi}(H|x, X_1, \dots, X_N)$ de (3.2.2) comme quotient de $M(H|x, X_1, \dots, X_N) = \hat{\lambda}_{0,H}(x)\pi(H)$ et $\int_{]0,+\infty[^2} M(H|x, X_1, \dots, X_N)dH$

$$\begin{aligned} M(H|x, X_1, \dots, X_N) &= \frac{1}{\Gamma(\alpha)^2 N p_H(x)} \frac{\beta^{2\alpha}}{2\pi} \sum_{i=1}^N (h_1^2 h_2^2)^{-\alpha-\frac{3}{2}} \exp\left(-\frac{1}{h_1^2} \left(\beta + \frac{(x_1 - X_{i1})^2}{2}\right)\right) \\ &\quad \times \exp\left(-\frac{1}{h_2^2} \left(\beta + \frac{(x_2 - X_{i2})^2}{2}\right)\right) \\ &= \frac{1}{\Gamma(\alpha)^2 N p_H(x)} \frac{\beta^{2\alpha}}{2\pi} \sum_{i=1}^N (h_1^2)^{-\alpha-3/2} (h_2^2)^{-\alpha-3/2} \exp\left(-\frac{A_{i1}}{h_1^2}\right) \\ &\quad \times \exp\left(-\frac{A_{i2}}{h_2^2}\right) \end{aligned}$$

avec $A_{i1} = \left(\beta + \frac{(x_1 - X_{i1})^2}{2}\right)$ et $A_{i2} = \left(\beta + \frac{(x_2 - X_{i2})^2}{2}\right)$.

Par un calcul direct du terme $\int_{]0,+\infty[^2} M(H|x, X_1, \dots, X_N) dH$, nous obtenons :

$$\int_{\mathcal{H}} \hat{\lambda}_{0,H}(x) \pi(H) dH = \int_{]0,+\infty[^2} \frac{1}{\Gamma(\alpha)^2 N p_H(x)} \frac{\beta^{2\alpha}}{2\pi} \sum_{i=1}^N (h_1^2 h_2^2)^{-\alpha - \frac{3}{2}} \exp\left(-\frac{A_{i1}}{h_1^2}\right) \times \exp\left(-\frac{A_{i2}}{h_2^2}\right) dh_1^2 dh_2^2 \quad (3.2.13)$$

$$= \frac{\beta^{2\alpha} \Gamma(\alpha + \frac{1}{2})^2}{2\pi \Gamma(\alpha)^2 N p_H(x)} \sum_{i=1}^N (A_{i1} A_{i2})^{-\alpha - \frac{1}{2}} \quad (3.2.14)$$

$$= \frac{\beta^{2\alpha} \Gamma(\alpha + \frac{1}{2})^2}{2\pi \Gamma(\alpha)^2 N p_H(x)} D_{ij} \quad (3.2.15)$$

avec

$$D_{ij} = \sum_{i=1}^N (A_{i1} A_{i2})^{-\alpha - \frac{1}{2}}, \quad i = 1, \dots, N, \quad j = 1, 2.$$

Finalement, le quotient formé conduit à :

$$\hat{\pi}(H|x, X_1, X_2, \dots, X_N) = \frac{1}{D_{ij}} \sum_{i=1}^N \prod_{j=1}^2 (A_{ij})^{-\alpha - \frac{1}{2}} \Phi_{\alpha + \frac{1}{2}, A_{ij}}(h_j^2). \quad (3.2.16)$$

- ii) Nous rappelons que la moyenne de la loi gamma inverse $IG(\alpha, \beta)$ est $\frac{\beta}{\alpha-1}$ et que $\mathbb{E}((H|x, X_1, \dots, X_N)) = \int_{]0,+\infty[^2} H \hat{\pi}(H|x, X_1, \dots, X_N) dH$ avec $\hat{\pi}(H|x, X_1, \dots, X_N)$ l'estimateur de la loi a postériori. Pour la matrice des fenêtres, nous avons

$$\hat{H}(x) = \int_{]0,+\infty[^2} H \hat{\pi}(H|x, X_1, \dots, X_N) dH$$

que nous pouvons écrire comme :

$$\begin{aligned} \hat{H}(x) &= \int_{]0,+\infty[^2} (h_1^2, h_2^2) \hat{\pi}(H|x, X_1, \dots, X_N) dh_1^2 dh_2^2 \\ &= \int_{]0,+\infty[^2} h_1^2 \hat{\pi}(H|x, X_1, \dots, X_N) dh_1^2 dh_2^2, \int_{]0,+\infty[^2} h_2^2 \hat{\pi}(H|x, X_1, \dots, X_N) dh_1^2 dh_2^2. \end{aligned}$$

Par conséquent, la forme exacte des éléments diagonaux de l'estimateur bayésien de la matrice H est :

$$\hat{h}_1^2(x_1) = \frac{1}{(\alpha - \frac{1}{2}) D_{ij}} \sum_{i=1}^N A_{i1}^{-\alpha + \frac{1}{2}} A_{i2}^{-\alpha - \frac{1}{2}}$$

$$\hat{h}_2^2(x_2) = \frac{1}{(\alpha - \frac{1}{2}) D_{ij}} \sum_{i=1}^N A_{i1}^{-\alpha - \frac{1}{2}} A_{i2}^{-\alpha + \frac{1}{2}}$$

avec $\alpha > \frac{1}{2}$. \square

Notez que la meilleure matrice de lissage locale dépend des paramètres de la loi a priori. Ainsi, il est important d'explorer la sensibilité de l'estimateur d'intensité $\hat{\lambda}_{0,H}$ au choix des paramètres de la loi a priori. Cette question a été explorée dans notre étude de simulation, qui est discutée dans la section suivante. Si l'ensemble de données présente une grande

variabilité et que la taille de l'échantillon est modérée, il a été démontré que *AMISE*, Scott ou *LSCV* sélectionne souvent une matrice de lissage sensiblement différente de l'optimum. Étant donné que l'estimateur d'intensité du noyau est extrêmement sensible au choix de la matrice de lissage, cela peut avoir un effet négatif sur l'estimation. Notre étude de simulation suggère que ce type de problème est moins susceptible de se produire dans notre approche.

3.3 Étude de simulation

3.3.1 Données simulées

Dans cette section, nous comparons les performances des estimateurs à noyaux de la fonction intensité du processus de Poisson non homogène, en utilisant l'approche bayésienne locale, la méthode de Scott, le *AMISE* et la méthode *LSCV* pour le choix de la matrice des fenêtres de lissage. Le critère de comparaison est l'erreur quadratique intégrée (*ISE*) donné par :

$$ISE = \int_W (\hat{\lambda}_{0,H}(x) - \lambda_0(x))^2 dx. \quad (3.3.1)$$

Nous considérons les fonctions intensités suivantes :

1. Le premier processus a pour fonction intensité :

$$\lambda_1(x) = b \exp(-3x_1), \quad x = (x_1, x_2) \in [0, 1]^2 \quad (3.3.2)$$

utilisant plusieurs valeurs de b (c'est-à-dire $b = 66, 165, 330, 660, 1320, 1700$) pour obtenir $m = \int_W \lambda(x) dx = 20, 50, 100, 200, 400, 500$ respectivement.

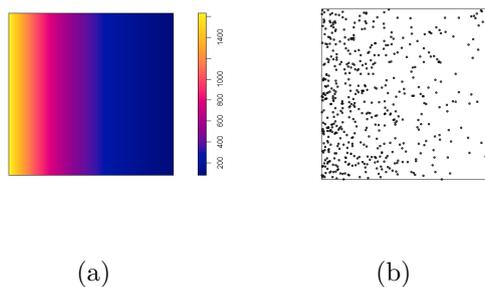


FIGURE 3.1 – (a) : fonction intensité $\lambda_1(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$.

2. Nous générons un deuxième et un troisième processus en considérant la fonction intensité donnée par :

$$\lambda_k(x) = b\Phi_{(0.3-0.2x_2, \sigma)}(x_1) + 25, \quad x = (x_1, x_2) \in [0, 1]^2 \quad (3.3.3)$$

où Φ est la densité univariée de la loi normale de moyenne $\mu = 0.3 - 0.2x_2$ et de variance $\sigma = 0.1$ pour $k = 2$ et $\sigma = 0.02$ pour $k = 3$, utilisant plusieurs valeurs b (

c'est-à-dire $b = 20, 50, 100, 200, 400, 500$) pour obtenir une fonction moyenne $m = \int_W \lambda(x) dx = 20, 50, 100, 200, 400, 500$ respectivement. Ce modèle a été utilisé par Barr et Schoenberg [5] pour analyser les performances de l'estimateur de Voronoi d'intensité du premier ordre.

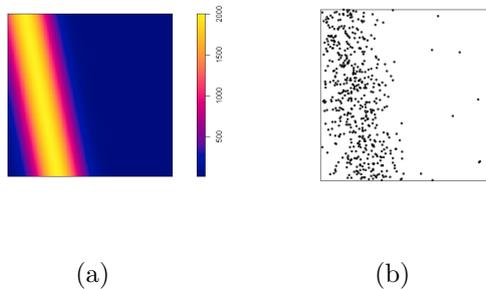


FIGURE 3.2 – (a) : fonction intensité $\lambda_2(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$.

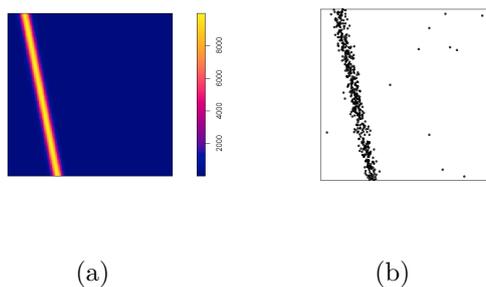


FIGURE 3.3 – (a) : fonction intensité $\lambda_3(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$.

- Enfin, nous avons simulé des processus ponctuels spatiaux avec des intensités de premier ordre données par :

$$\lambda_4(x) = b(\Phi_{x_2,0.1}(x_1) + \Phi_{1-x_2,0.2}(x_1) - \frac{1}{5}\Phi_{2,(\mu,\Sigma)}(x)) + 30, \quad x = (x_1, x_2) \in [0, 1]^2 \quad (3.3.4)$$

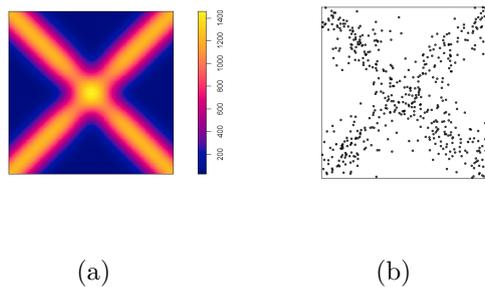


FIGURE 3.4 – (a) : fonction intensité $\lambda_4(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$.

et

$$\lambda_5(x) = b(\Phi_{x_2,0.1}(x_1) + \Phi_{1-x_2,0.2}(x_1) - \frac{1}{20}\Phi_{2,(\mu,\Sigma)}(x)) + 50, \quad x = (x_1, x_2) \in [0, 1]^2 \quad (3.3.5)$$

où $\Phi_{2,(\mu,\Sigma)}$ est la densité bivariée de la loi normale de moyenne $\mu = (0.5, 0.5)$, et de matrice variance $\Sigma = 0.01\mathbb{I}_2$, enfin d'obtenir $m = 20, 50, 100, 200, 400, 500$ pour $b = 12, 30, 60, 120, 240, 275$ respectivement.

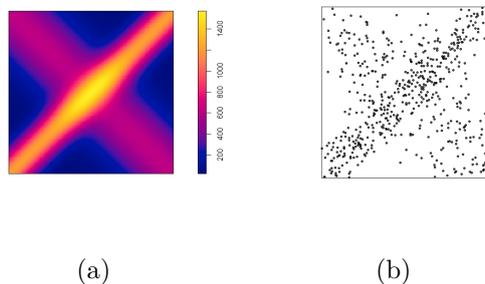


FIGURE 3.5 – (a) : fonction intensité $\lambda_5(x)$. (b) : réalisation d'un processus de Poisson non homogène pour $m = 500$.

Étude de sensibilité

La performance de l'approche bayésienne locale dépend du choix des paramètres de la loi a priori. L'analyse de sensibilité des paramètres (α, β) sera traitée comme dans Zougab et al. (2012) et Belaid et al. (2016). Nous examinons la sensibilité du modèle $\hat{\lambda}_{0,H}(x)$ où la matrice diagonale des fenêtres et le paramètre du lissage dans le cas unidimensionnel sont obtenus par l'approche bayésienne locale. Nous avons calculé le \widehat{ISE} où \widehat{ISE} est l'estimation du ISE en combinant quatre valeurs de $\alpha \in \{0.6, 2, 5, 10\}$ et quatre valeurs de $\beta \in \{1, 5, 10, 15\}$ de la loi a priori Gamma inverse, pour un nombre de répétitions égales à 100. Nous utilisons tous les modèles avec $m = 500$. Les figures 3.6 rapportent l'allure de $\beta \mapsto \widehat{ISE}(\beta)$ pour les quatre valeurs de α fixes. Nous avons tracé les cinq modèles $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 respectivement.

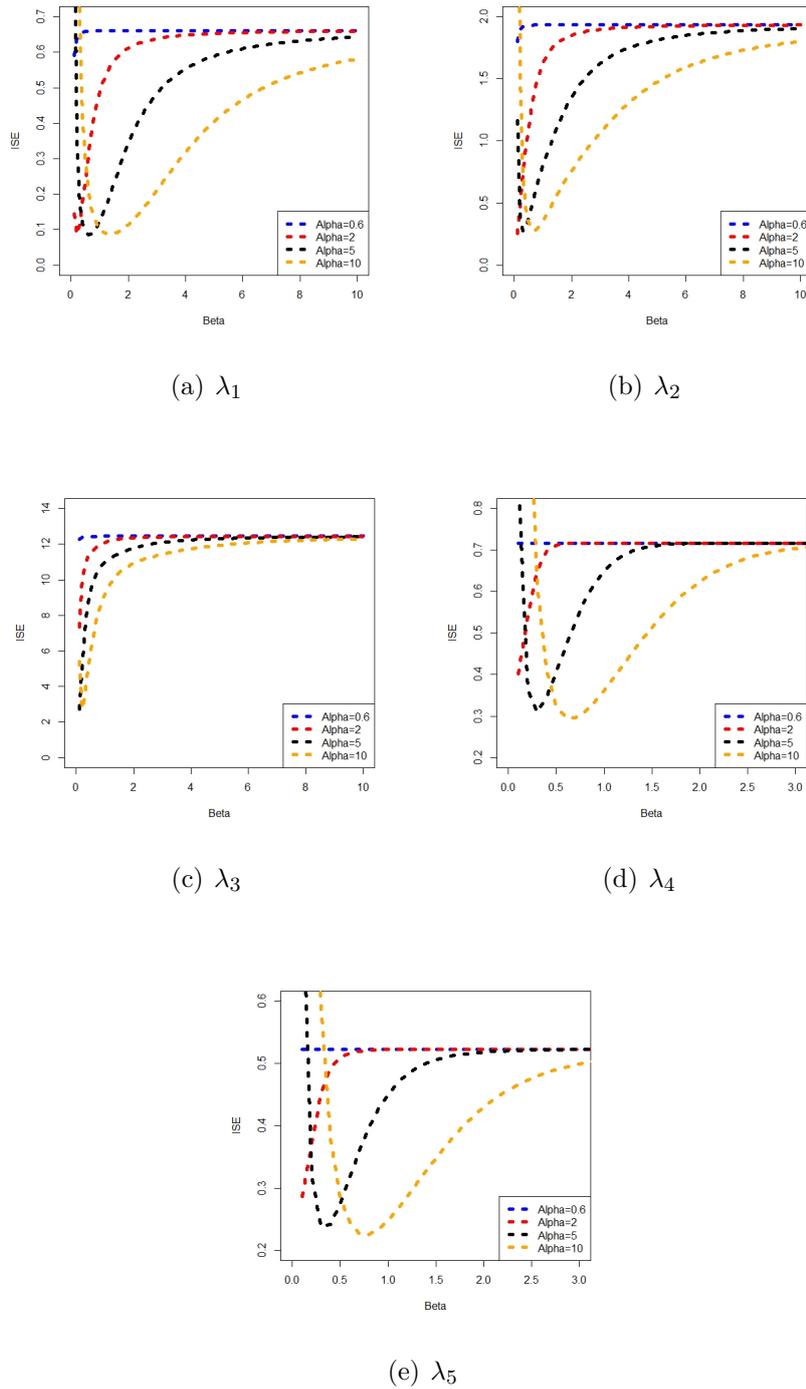


FIGURE 3.6 – Graphes du \widehat{ISE} avec le noyau gaussien et la loi Gamma inverse des cinq fonctions intensité.

Les résultats de l'approche bayésienne locale ont été obtenus en utilisant la fonction intensité et la loi Gamma inverse comme loi a priori et les valeurs des paramètres de la loi Gamma inverse ont été choisies comme suit :

- pour λ_1 : $(\alpha, \beta) = (5, 0.7)$;
- pour λ_2 : $(\alpha, \beta) = (10, 0.7)$;
- pour λ_3 : $(\alpha, \beta) = (10, 0.7)$;

- pour λ_4 : $(\alpha, \beta) = (10, 0.7)$;
- pour λ_5 : $(\alpha, \beta) = (10, 0.7)$.

TABLE 3.1 – Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$.

		λ_1	λ_2	λ_3	λ_4	λ_5
α	β	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}
0.6	0.7	0.66901	1.92787	12.56351	0.70829	0.53084
	5	0.67068	1.93068	12.56657	0.70829	0.53084
	10	0.67071	1.93073	12.56662	0.70829	0.53084
	15	0.67072	1.93074	12.56663	0.70829	0.53084
2	0.7	0.39029	1.42762	12.0035	0.70932	0.52717
	5	0.66231	1.91678	12.55144	0.70830	0.53084
	10	0.66859	1.92721	12.56279	0.70829	0.53084
	15	0.66977	1.92917	12.56492	0.70829	0.53084
5	0.7	0.08463	0.59219	10.58585	0.51142	0.35444
	5	0.59899	1.81004	12.43466	0.70845	0.53074
	10	0.65175	1.89925	12.53234	0.70831	0.53083
	15	0.66219	1.91662	12.55126	0.70830	0.53084
10	0.7	0.16017	0.28304	7.95844	0.22878	0.22723
	5	0.41321	1.47769	12.06175	0.70939	0.52818
	10	0.59059	1.79587	12.41909	0.70848	0.53071
	15	0.63359	1.86887	12.49917	0.70835	0.53081

NB : En gras le \widehat{ISE} minimum.

À partir du tableau 3.1, nous avons les valeurs minimales du \widehat{ISE} (en gras) pour chaque valeur de α et β . À partir de là, nous choisissons $(\alpha, \beta) = (5, 0.7)$ comme paramètres de la loi a priori pour la fonction intensité λ_1 . Pour les fonctions intensités λ_2 , λ_3 , λ_4 et λ_5 , nous avons obtenu les valeurs des paramètres (α, β) qui donnent les valeurs minimales du \widehat{ISE} données par $(\alpha, \beta) = (10, 0.7)$.

Après avoir choisi les paramètres de la loi a priori Gamma inverse, nous comparons maintenant les performances des méthodes validation croisée, *AMISE*, la méthode de Scott et la méthode bayésienne locale. Le tableau 3.2 présente les valeurs de l'erreur quadratique intégrée \widehat{ISE} pour les quatre différentes approches de sélections de la matrice de lissage en utilisant les fonctions intensités λ_1 , λ_2 , λ_3 , λ_4 et λ_5 pour différentes tailles d'échantillon $m = \{20, 50, 100, 200, 400, 500\}$. Les résultats sont obtenus pour un nombre de répétitions égal à 100.

Les résultats obtenus montrent l'intérêt de l'approche bayésienne locale au sens du *ISE* donnée par l'équation 3.3.1. Pour les données simulées à partir de λ_1 , λ_2 , λ_3 , λ_4 et λ_5 , l'approche bayésienne locale fonctionne mieux pour les échantillons de petites tailles

($m = 20$ et 50), moyennes ($m = 100, 200$ et 400) et relativement grandes ($m = 500$) que les autres estimateurs.

Cependant, la valeur du \widehat{ISE} obtenue par les approches diminue à mesure que la taille de l'échantillon m augmente. Nous avons aussi comparé le temps d'exécution de chaque méthode pour les fonctions d'intensités $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 avec une taille d'échantillon $m \in \{20, 50, 100, 200, 400, 500\}$. Les calculs ont été effectués par un ordinateur personnel standard à $4,0\text{ GHz}$ avec 4G RAM et par le logiciel R pour 100 réplifications. Les résultats obtenus sont présentés dans le tableau 3.3. Ils montrent l'avantage en temps d'exécution de la méthode de Scott pour les modèles, le temps d'exécution de l'approche bayésienne locale est meilleur que celui de la méthode validation croisée, mais aussi celui du *AMISE*. Les figures 3.1, 3.2, 3.3, 3.4 et 3.5 représentent les estimateurs par noyau des fonctions intensités $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 respectivement, et pour $m = 500$.

TABLE 3.2 – Résultats de simulation du \widehat{ISE} basé sur 100 répétitions pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$.

Intensités	m	\widehat{ISE}			
		AMISE	LSCV	Scott	Bayes-loc
λ_1	20	1.77828	2.05799	1.67838	1.06629
	50	0.76831	1.67185	0.64775	0.41261
	100	0.44069	1.66827	0.30389	0.19864
	200	0.29453	1.61114	0.17838	0.13907
	400	0.21920	1.60389	0.10400	0.08923
	500	0.21158	1.52249	0.08746	0.08463
λ_2	20	2.67632	3.44664	1.85426	1.51401
	50	1.15848	3.15100	0.99096	0.94609
	100	0.74519	2.99807	0.78143	0.71966
	200	0.66218	2.94038	0.52222	0.49979
	400	0.39482	2.92211	0.44543	0.31276
	500	0.34916	2.87909	0.42991	0.28304
λ_3	20	19.20312	14.91222	12.17067	11.75374
	50	13.01852	13.95714	10.71801	9.53085
	100	11.0800	13.76103	10.41801	8.85307
	200	9.55484	13.37586	9.926011	8.21080
	400	9.04308	13.29754	9.883853	8.01764
	500	8.88654	13.18517	9.740522	7.95844
λ_4	20	2.27746	2.03730	1.93589	1.48579
	50	1.88108	1.79842	1.00109	0.90025
	100	0.97251	1.74923	0.67914	0.66764
	200	0.72077	1.72217	0.49941	0.41640
	400	0.62006	1.68298	0.40942	0.32410
	500	0.49610	1.67720	0.39563	0.22878
λ_5	20	1.81585	1.85584	1.77502	1.27327
	50	0.85063	1.62308	0.80204	0.66453
	100	0.61517	1.59969	0.53284	0.51778
	200	0.45408	1.57946	0.40414	0.33645
	400	0.37612	1.53382	0.27341	0.24863
	500	0.36428	1.52249	0.27197	0.22723

NB : En gras la valeur du \widehat{ISE} minimum.

TABLE 3.3 – Comparaison du temps d'exécution (en secondes).

Intensités	m	Temps			
		<i>AMISE</i>	<i>LSCV</i>	<i>Scott</i>	Bayes-loc
λ_1	20	7.87	7.19	3.96	4.37
	50	7.89	8.02	3.99	4.71
	100	9.97	12.51	3.99	5.32
	200	11.35	29.88	4.01	6.48
	400	14.38	106.16	4.07	8.75
	500	16.49	170.31	4.12	9.86
λ_2	20	6.73	7.14	3.87	4.35
	50	8.01	8.00	3.88	4.65
	100	10.48	10.74	3.92	5.22
	200	13.85	23.35	3.97	6.38
	400	15.43	76.35	4.99	8.48
	500	18.52	120.12	5.02	9.40
λ_3	20	7.41	7.89	3.87	4.38
	50	8.54	9.33	3.88	4.79
	100	10.87	14.44	3.91	5.27
	200	12.14	40.41	4.68	7.71
	400	15.23	120.66	4.73	8.66
	500	20.12	191.87	4.92	9.78
λ_4	20	7.40	7.39	3.88	4.36
	50	8.63	8.91	3.90	4.76
	100	10.23	12.95	3.95	5.42
	200	11.89	14.03	3.98	6.47
	400	14.15	32.10	4.12	8.87
	500	19.62	163.56	4.34	9.69
λ_5	20	6.16	7.07	3.86	4.40
	50	7.65	8.19	3.96	4.72
	100	9.02	12.30	3.98	5.31
	200	13.45	28.72	4.00	6.52
	400	16.68	101.82	4.12	8.98
	500	21.16	169.60	4.33	10.04

NB : En gras la valeur du \widehat{ISE} minimum.

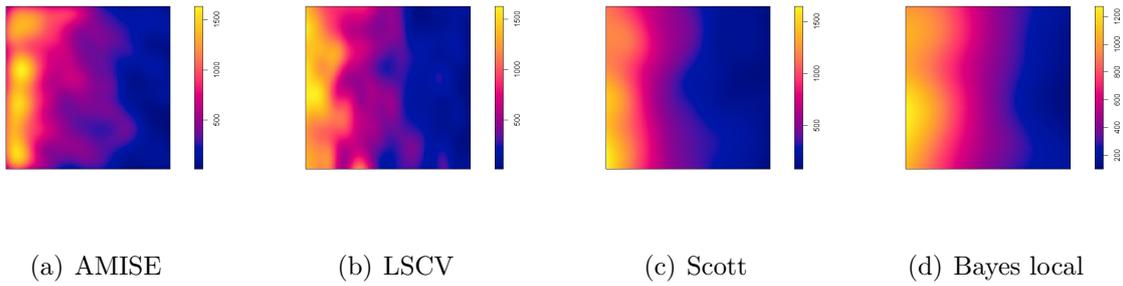


FIGURE 3.7 – Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_1(x)$ pour $m = 500$.

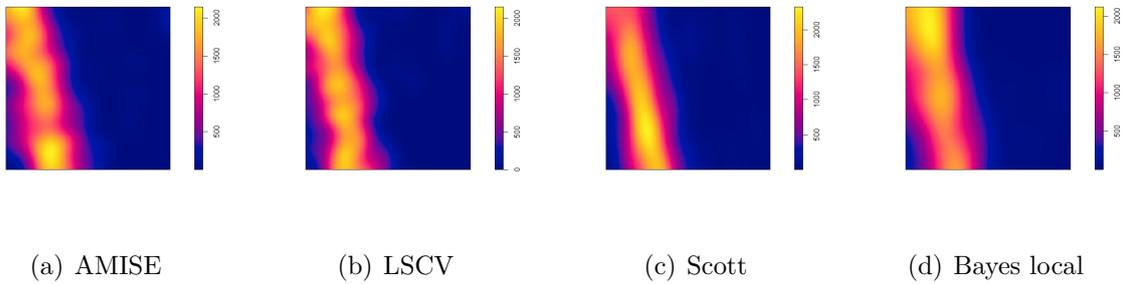


FIGURE 3.8 – Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_2(x)$ pour $m = 500$.

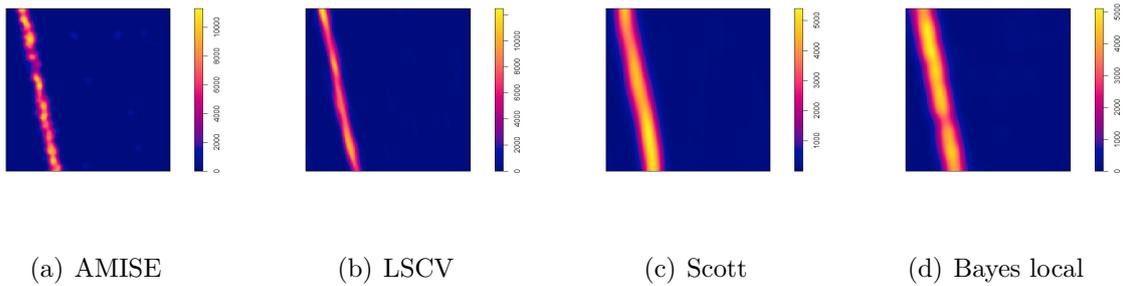


FIGURE 3.9 – Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_3(x)$ pour $m = 500$.

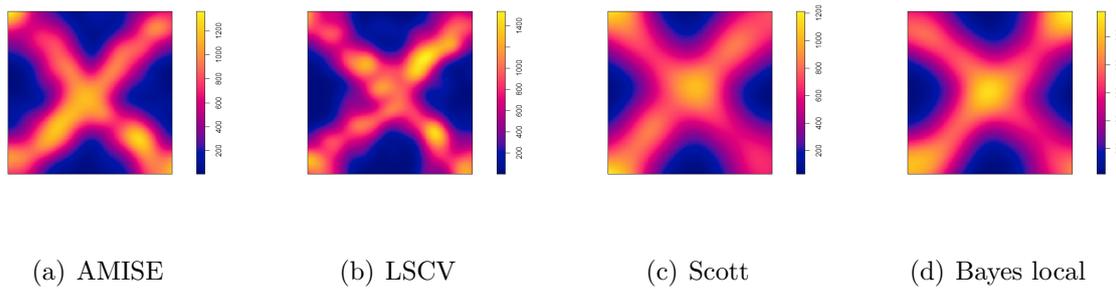


FIGURE 3.10 – Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_4(x)$ pour $m = 500$.

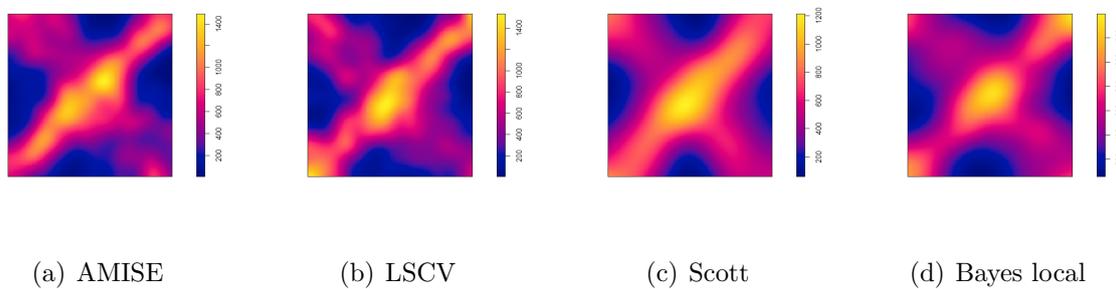


FIGURE 3.11 – Estimateur par noyau de la fonction intensité du processus de Poisson non homogène $\lambda_5(x)$ pour $m = 500$.

3.3.2 Données réelles

Nous utilisons ici l'ensemble de données de localisations des 58 cas de cancer du larynx et 978 cas de cancer du poumon enregistrés dans la région de Chorley-Ribble au cours des années 1974 – 1983. Les données sont représentées sous la forme d'un modèle de configuration de points, les points indiquant l'emplacement spatial de l'adresse du domicile de chaque individu et les marques identifiant si chaque point est un cas de cancer du larynx ou du poumon. Dans cette section, nous analysons les modèles de configurations de points pour les 58 cas de cancers de larynx et les 978 cas de cancers du poumon dans la région de Chorley et South Ribble Health Authority du Lancashire (Angleterre) entre 1974 et 1983. Afin de comparer leur structure spatiale, quatre méthodes ont été appliquées pour sélectionner la meilleure matrice des paramètres de lissage pour estimer l'intensité de premier ordre des deux semis de points.

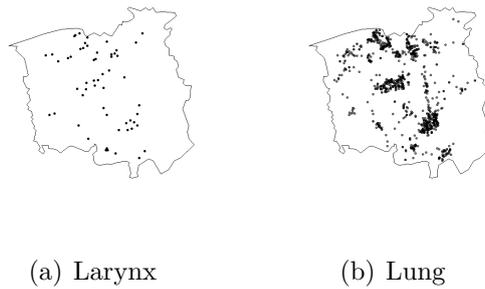


FIGURE 3.12 – (a) : Distribution de 58 cas de cancer de larynx et (b) : de 978 cas de cancer de poumon de la région de Chorley-Ribble.

Pour évaluer la performance des estimateurs, nous utilisons l’erreur quadratique intégrée empirique (ISE_0) définie par :

$$ISE_0 = \int_W (\hat{\lambda}_{0,H}(x) - \hat{\lambda}_h(x))^2 dx \quad (3.3.6)$$

où $\hat{\lambda}_h$ est l’estimateur empirique de Diggle de la fonction intensité donné par l’équation (2.1.2). Les résultats sont donnés dans le tableau 3.5 avec le noyau gaussien produit. Ces résultats montrent que l’approche bayésienne locale est meilleure que la méthode classique *LSCV*, la méthode *AMISE* et la méthode de Scott.

TABLE 3.4 – Paramètres de lissages optimaux pour les données de cancer de larynx et du cancer de poumon pour différentes méthodes de sélections.

Données	Larynx $m = 58$		Lung $m = 978$	
Paramètres	h_1	h_2	h_1	h_2
AMISE	2.950758	4.944416	0.9546852	1.807898
LSCV	0.8291104	0.6829351	1.404456	$1.578721 \cdot 10^{-10}$
Scott	1.870306	2.421046	1.111556	1.529619
Bayes-local	0.29000	3.64000	2.162632	1.844211

TABLE 3.5 – Comparaison de la valeur du ISE_0 et du temps d'exécution pour les données du cancer de Larynx et du poumon.

	Méthodes	$AMISE$	$LSCV$	Scott	Bayes Local
Larynx	ISE_0	0.03266	0.03266	0.00973	0.00626
	Temps d'exécution	0.19163	0.19134	0.11436	0.11851
Poumon	ISE_0	0.01981	0.01991	0.00688	0.00515
	Temps d'exécution	0.29663	9.57779	0.12256	0.23938

NB : En gras la valeur du ISE_0 minimum.

Pour illustrer ces résultats, nous avons tracé les estimations obtenues par les approches (bayésienne locale, $AMISE$, $LSCV$ et Scott) pour les données de cancer de Larynx et de cancer du poumon.

Les figures 3.13 et 3.14 montrent les estimateurs des fonctions intensités par noyau fourni par les différents sélecteurs de la matrice de lissage. Pour le cancer du larynx, la méthode de Scott conduit à un sur lissage des données, tandis que le $LSCV$ et la méthode $AMISE$ ont fourni des estimateurs moins lisses que celui obtenu par la méthode bayésienne locale. Pour le cancer du poumon, les méthode $AMISE$ et bayésienne locale fournissent des estimateurs similaires, ainsi que pour le cancer du larynx, les méthodes $LSCV$ et la méthode locale bayésienne fournissent des estimateurs plus lisses.

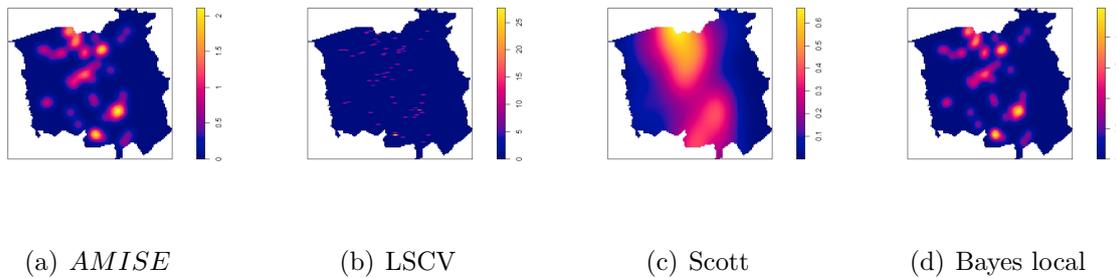


FIGURE 3.13 – Estimation de la fonction intensité par noyau pour le cancer de larynx par la méthode : (a) $AMISE$, (b) $LSCV$, (c) Scott, (d) bayésienne locale.

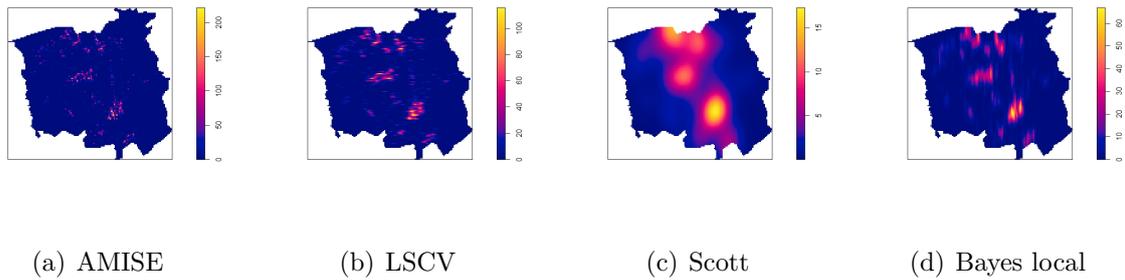


FIGURE 3.14 – Estimation de la fonction intensité par noyau pour le cancer de poumon par la méthode : (a) *AMISE*, (b) *LSCV*, (c) Scott, (d) bayésienne locale.

3.4 Conclusion

Ce chapitre aborde le problème de la sélection de la matrice de lissage des estimateurs à noyau de la fonction intensité du processus de Poisson non homogène. Nous avons proposé l'approche bayésienne locale dans le cas bidimensionnel pour le choix de la matrice des fenêtres de lissage. L'avantage de cette méthode est qu'elle fournit la forme exacte de l'estimateur du paramètre et de la matrice de lissage. Notre étude montre que l'approche bayésienne locale donne de meilleurs résultats que la méthode classique *LSCV*, la méthode *AMISE* et la méthode de Scott au sens du *ISE*.

Approche bayésienne adaptative pour le choix de la matrice de lissage de la fonction intensité.

4.1 Introduction

Dans ce chapitre, nous proposons l'estimateur bayésien adaptatif pour la sélection de la matrice des fenêtres de lissage dans l'estimation de la fonction intensité du processus de Poisson non homogène. L'approche bayésienne adaptative a été proposée respectivement par Brewer (2000) et Zougab et al. (2014) dans le cas uni et multidimensionnel continu dans le cas de la densité et par Zougab et al. (2013) dans le cas unidimensionnel discret. Cette approche consiste à sélectionner le vecteur de lissage h_i en chaque événement X_i . Nous utilisons le noyau gaussien produit avec une loi a priori. Nous estimons h_i en utilisant deux fonctions pertes : perte quadratique et perte entropie. Pour étudier la qualité de cette approche bayésienne, nous présentons une partie application sur des données simulées et sur des données réelles.

4.2 Méthodologie

L'approche que nous proposons cette approche consiste à sélectionner un vecteur de fenêtres h_i correspondant à l'observation X_i en lieu et place du vecteur $h = (h_1, h_2)^T$ initial. Elle consiste dans un premier temps à considérer l'estimateur à noyau multivarié de la fonction intensité avec vecteurs de fenêtres variables $h_i = (h_{i1}, h_{i2})^T$ est donné par

$$\hat{\lambda}_{0, h_i}(x) = \frac{1}{N p_{h_i}} \sum_{i=1}^N \prod_{j=1}^2 K_{x_j, h_{ij}}(X_{ij}) \mathbb{1}_{N \neq 0} \quad (4.2.1)$$

où $K(\cdot)$ est un noyau symétrique et h_{ij} (composante du vecteur h_i) est le paramètre de lissage adaptatif pour chaque événement X_{ij} (composante du vecteur X_i c'est-à-dire $X_i = (X_{i1}, X_{i2})^T$), $i = 1, \dots, N$ et $j = 1, 2$ et $p_{h_i}(\cdot)$ est le terme de correction de bord. Nous proposons d'utiliser la technique de validation croisée pour l'estimation de $\hat{\lambda}_{0, h_i}(X_i)$ par l'estimateur à noyau adaptatif basé sur l'ensemble des événements sauf l'événement X_i . Le résultat est donné par :

$$\hat{\lambda}_{0, h_i}^{-i}(X_i | h_i) = \frac{1}{p_{h_i}(N-1)} \sum_{\substack{j=1 \\ j \neq i}}^N \prod_{l=1}^2 K_{X_{il}, h_{il}}(X_{jl}) \mathbb{1}_{N \neq 0}. \quad (4.2.2)$$

où $-i$ désigne les événements sans le $i^{\text{ème}}$ événement. C'est ce $\hat{\lambda}_{0, h_i}^{-i}(X_i | h_i)$ qui apporte

l'information dont on a besoin pour avoir un estimateur de la loi a posteriori de h_i . Soit $\pi(h_i)$ la loi a priori de h_i , l'estimateur de la loi a posteriori de chaque variable du vecteur h_i pour chaque vecteur d'observation X_i fournie à partir de la règle de Bayes est :

$$\hat{\pi}(h_i|X_i) = \hat{\lambda}_{0,h_i}^{-i}(X_i|h_i)\pi(h_i) \left(\int_{\mathcal{H}} \hat{\lambda}_{0,h_i}^{-i}(X_i|h_i)\pi(h_i)dh_i \right)^{-1}, \quad (4.2.3)$$

où \mathcal{H} est l'espace des vecteurs positives.

Par l'estimateur de Bayes sous la perte quadratique et la perte entropie, le meilleur estimateur pour \hat{h}_i de h_i $i = 1, \dots, N$ est la moyenne de $\hat{\pi}(h_i|X_i)$ donné respectivement par :

$$\hat{h}_i^{qua} = \mathbb{E}(h_i|X_i) = (\mathbb{E}(h_{i1}|X_i), \mathbb{E}(h_{i2}|X_i))^T \quad (4.2.4)$$

$$\hat{h}_i^{ent} = [\mathbb{E}(h_i^{-1}|X_i)]^{-1} = ((\mathbb{E}(h_{i1}^{-1}|X_i), \mathbb{E}(h_{i2}^{-1}|X_i))^{-1})^T. \quad (4.2.5)$$

4.2.1 Choix de la loi a priori

L'inférence bayésienne nécessite le choix de la loi a priori sur le(s) paramètre(s) à estimer. Dans le cas du paramètre de lissage, ce choix n'est pas unique. Zheng (2010) a utilisé la loi a priori définie par :

$$\pi(h_{ij}) = \frac{2}{\Gamma(\alpha)\beta^\alpha} \frac{1}{h_{ij}^{2\alpha+1}} \exp\left(-\frac{1}{\beta h_{ij}^2}\right), \quad h_{ij} > 0, \quad \alpha > 0, \quad \beta > 0. \quad (4.2.6)$$

La moyenne et la variance de la loi a priori pour chaque variable h_{ij} sont données par :

$$\mathbb{E}(h_{ij}) = \frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma^2(\alpha)\beta^{\frac{1}{2}}}, \quad \alpha > \frac{1}{2} \quad (4.2.7)$$

et

$$Var(h_{ij}) = \frac{\Gamma(\alpha - 1)\Gamma(\alpha) - \Gamma^2(\alpha - \frac{1}{2})}{\Gamma^2(\alpha)\beta}. \quad (4.2.8)$$

Notons que pour $\alpha > 1$ fixé et pour $\beta \rightarrow \infty$, la distribution de $h_{ij} \rightarrow 0$. Le choix de la loi a priori peut se faire dans le but d'avoir la conjugabilité avec le noyau gaussien.

4.2.2 Choix des paramètres de la loi a priori

Notez que l'implémentation et les performances de l'approche bayésienne adaptative dépendent du choix des paramètres de la loi a priori α et β ; voir, par exemple, De Lima et Atunca (2011) et Zougab (2014) pour le choix des paramètres de la distribution a priori dans le contexte de l'estimation bayésienne des matrices de lissage dans l'estimation de densité multivariée du noyau. Basé sur la même idée de De Lima (2011) et Zougab (2014), le choix des paramètres de la loi a priori (α, β) peut être réalisé comme suit :

- le paramètre α est choisi tel que :

$$\text{Var}(h_{ij}) = \frac{\Gamma(\alpha - 1)\Gamma(\alpha) - \Gamma^2(\alpha - \frac{1}{2})}{\Gamma^2(\alpha)\beta} > 0,$$

donc α est choisi pour satisfaire $\Gamma(\alpha - 1)\Gamma(\alpha) - \Gamma^2(\alpha - \frac{1}{2}) > 0$. Par conséquent, la valeur de α est prise supérieure à 1,

- pour le choix de β :

$$\mathbb{E}(h_{ij}) = \frac{\Gamma(\alpha - \frac{1}{2})}{\Gamma^2(\alpha)\beta^{\frac{1}{2}}} > 0.$$

Nous pouvons observer que pour une grande valeur de β , la loi a priori de h_i est concentrée à 0. Ensuite, le choix de β en fonction de la taille de l'échantillon m est nécessaire pour la consistance de l'estimateur adaptatif.

4.2.3 Estimation de la loi a posteriori

Nous proposons d'utiliser le noyau gaussien produit et la loi a priori donnée par l'équation (4.2.6) pour chaque composante de la matrice h_i . Cette combinaison permet de dériver explicitement la loi a posteriori pour chaque h_i . L'estimateur à noyau gaussien produit adaptatif est alors de la forme :

$$\hat{\lambda}_{0,h_i}(x) = \frac{1}{p_{h_i}(x)N} \sum_{i=1}^N \prod_{j=1}^2 \frac{1}{h_{ij}\sqrt{2\pi}} \exp\left(-\frac{(x - X_{ij})^2}{2h_{ij}^2}\right) \mathbf{1}_{N \neq 0}. \quad (4.2.9)$$

Par la technique de validation croisée, nous estimons $\lambda_0(X_i)$ sur l'ensemble des événements sauf l'événement X_i , nous obtenons :

$$\hat{\lambda}_{0,h_i}^{-1}(X_i|h_i) = \hat{\lambda}_{0,h_i}(X_i|X_{-i}, h_i) = \frac{1}{p_{h_i}(X_i)(N-1)} \sum_{\substack{j=1 \\ j \neq i}}^N \prod_{l=1}^2 \frac{1}{h_{il}\sqrt{2\pi}} \exp\left(-\frac{(X_{il} - X_{jl})^2}{2h_{ij}^2}\right) \mathbf{1}_{N \neq 0}. \quad (4.2.10)$$

où X_{-i} est l'ensemble de toutes les événements sans l'événement X_i . C'est ce $\hat{\lambda}_{0,h_i}^{-1}$ qui apporte l'information dont on a besoin pour avoir une loi a posteriori de h_i donnée par (4.2.3).

Nous supposons maintenant que chaque composante h_{il} , $i = 1, \dots, N$ et $l = 1, 2$ de h_i suit une loi a priori donnée en (4.2.6) conjuguée avec le noyau gaussien. Le théorème suivant donne la forme explicite de la loi a posteriori et de l'estimateur bayésien de h_i sous les deux fonctions perte quadratique et entropie.

Théorème 4.2.1. (i) L'estimateur de la densité a posteriori $\hat{\pi}(h_i|X_i)$ donnée en (4.2.3) est :

$$\begin{aligned} & \hat{\pi}(h_i|X_i) \\ &= \frac{\sum_{\substack{j=1 \\ j \neq i}}^N (h_{i1}h_{i2})^{-2\alpha-2} \exp\left(-\frac{1}{h_{i1}^2} \left(\frac{(X_{i1}-X_{j1})^2}{2} + \frac{1}{\beta}\right)\right) \exp\left(-\frac{1}{h_{i2}^2} \left(\frac{(X_{i2}-X_{j2})^2}{2} + \frac{1}{\beta}\right)\right)}{\sum_{\substack{j=1 \\ j \neq i}}^N \frac{\Gamma(\alpha+1/2)^2}{4} \left(\left(\frac{(X_{i1}-X_{j1})^2}{2} + \frac{1}{\beta}\right) \left(\frac{(X_{i2}-X_{j2})^2}{2} + \frac{1}{\beta}\right)\right)^{-\alpha-\frac{1}{2}}}. \end{aligned} \quad (4.2.11)$$

(ii) Sous la fonction perte quadratique, l'estimateur bayésien de chaque élément h_{il} $i = 1, \dots, N$ et $l = 1, 2$ est donné par :

$$\hat{h}_{i1}^{qua} = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha+\frac{1}{2})} \sum_{\substack{j=1 \\ j \neq i}}^N A_{ij}^{\alpha+\frac{1}{2}} B_{ij}^{\alpha} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}B_{ij})^{\alpha+\frac{1}{2}} \right]^{-1} \quad (4.2.12)$$

$$\hat{h}_{i2}^{qua} = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha+\frac{1}{2})} \sum_{\substack{j=1 \\ j \neq i}}^N A_{ij}^{\alpha} B_{ij}^{\alpha+\frac{1}{2}} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}B_{ij})^{\alpha+\frac{1}{2}} \right]^{-1}. \quad (4.2.13)$$

(iii) Sous la fonction perte entropie, l'estimateur bayésien de chaque élément h_{il} $i = 1, \dots, N$ et $l = 1, 2$ devient :

$$\hat{h}_{i1}^{ent} = \frac{\Gamma(\alpha+\frac{1}{2})}{\sqrt{2\beta}\Gamma(\alpha+1)} \sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}B_{ij})^{\alpha+1} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}^{\alpha+1}B_{ij}^{\alpha+\frac{1}{2}}) \right]^{-1} \quad (4.2.14)$$

$$\hat{h}_{i2}^{ent} = \frac{\Gamma(\alpha+\frac{1}{2})}{\sqrt{2\beta}\Gamma(\alpha+1)} \sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}B_{ij})^{\alpha+1} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}^{\alpha+\frac{1}{2}}B_{ij}^{\alpha+1}) \right]^{-1}. \quad (4.2.15)$$

où

$$A_{ij} = \frac{1}{\beta(X_{i1}-X_{j1})^2+2}, \quad B_{ij} = \frac{1}{\beta(X_{i2}-X_{j2})^2+2}.$$

Démonstration :

i) $L = \hat{\lambda}_{0,h_i}^{-i}(X_i|h_i)\pi(h_i)$ s'exprime successivement par :

$$\begin{aligned} L &= \frac{1}{(N-1)p_{h_i}(X_i|h_i)} \frac{4}{\Gamma^2(\alpha)\beta^{2\alpha}2\pi} \sum_{\substack{j=1 \\ j \neq i}}^N (h_{i1})^{-2\alpha-1} (h_{i2})^{-2\alpha-1} \exp\left(-\frac{1}{h_{i1}^2} \left(\frac{(X_{i1}-X_{j1})^2}{2} + \frac{1}{\beta}\right)\right) \\ &\quad \times \exp\left(-\frac{1}{h_{i2}^2} \left(\frac{(X_{i2}-X_{j2})^2}{2} + \frac{1}{\beta}\right)\right) \end{aligned}$$

Par un calcul direct, le second terme $\int_{\mathcal{H}} \hat{\lambda}_{0,h_i}^{-i} \pi(h_i) dh_i$ de l'équation (4.2.3) devient

$$\int_{[0,+\infty]^2} \hat{\lambda}_{0,h_i}^{-i} \pi(h_i) dh_i = \frac{\Gamma(\alpha + \frac{1}{2})^2}{2\pi\Gamma(\alpha)^2(N-1)\beta^{2\alpha}p_{h_i}(X_i|h_i)} \sum_{\substack{j=1 \\ j \neq i}}^N \left(\beta + \frac{(X_{i1} - X_{j1})^2}{2} \right)^{-\alpha - \frac{1}{2}} \\ \times \left(\frac{(X_{i2} - X_{j2})^2}{2} + \frac{1}{\beta} \right)^{-\alpha - \frac{1}{2}}$$

Ensuite, nous obtenons l'expression explicite de la loi a posteriori donnée par :

$$\hat{\pi}(h_i|X_i) = \frac{\sum_{\substack{j=1 \\ j \neq i}}^N (h_{i1}h_{i2})^{-2\alpha-2} \exp\left(-\frac{1}{h_{i1}^2} \left(\frac{(X_{i1} - X_{j1})^2}{2} + \frac{1}{\beta}\right)\right) \exp\left(-\frac{1}{h_{i1}^2} \left(\frac{(X_{i2} - X_{j2})^2}{2} + \frac{1}{\beta}\right)\right)}{\sum_{\substack{j=1 \\ j \neq i}}^N \frac{\Gamma(\alpha + 1/2)^2}{4} \left(\left(\frac{(X_{i1} - X_{j1})^2}{2} + \frac{1}{\beta}\right) \left(\frac{(X_{i2} - X_{j2})^2}{2} + \frac{1}{\beta}\right)\right)^{-\alpha - \frac{1}{2}}} \\ (4.2.16)$$

$$A_{ij} = \frac{(X_{i1} - X_{j1})^2}{2} + \frac{1}{\beta}, \quad B_{ij} = \frac{(X_{i2} - X_{j2})^2}{2} + \frac{1}{\beta}.$$

ii) Soit,

$$\hat{h}_i^{qua} = \mathbb{E}(h_i|X_i) = \int_{\mathcal{H}} h_i \pi(h_i|X_i) dh_i \\ = \int_{[0,+\infty]^2} (h_{i1}, h_{i2}) \pi(h_i|X_i) dh_{i1} dh_{i2} \\ = \int_{[0,+\infty]^2} h_{i1} \pi(h_i|X_i) dh_{i1} dh_{i2}, \int_{[0,+\infty]^2} h_{i2} \pi(h_i|X_i) dh_{i1} dh_{i2}$$

Nous obtenons successivement, pour chaque h_{il} , $i = 1, \dots, N$ and $l = 1, 2$,

$$\hat{h}_{i1}^{qua} = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha + \frac{1}{2})} \sum_{\substack{j=1 \\ j \neq i}}^N A_{ij}^{\alpha + \frac{1}{2}} B_{ij}^{\alpha} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij} B_{ij})^{\alpha + \frac{1}{2}} \right]^{-1} \quad (4.2.17)$$

$$\hat{h}_{i2}^{qua} = \frac{\Gamma(\alpha)}{\sqrt{2\beta}\Gamma(\alpha + \frac{1}{2})} \sum_{\substack{j=1 \\ j \neq i}}^N A_{ij}^{\alpha} B_{ij}^{\alpha + \frac{1}{2}} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij} B_{ij})^{\alpha + \frac{1}{2}} \right]^{-1}. \quad (4.2.18)$$

iii) En remplaçant h_i par h_i^{-1} et en utilisant les mêmes calculs que précédemment, nous obtenons facilement les composantes de \hat{h}_i^{ent} données par :

$$\hat{h}_{i1}^{ent} = \frac{\Gamma(\alpha + \frac{1}{2})}{\sqrt{2\beta}\Gamma(\alpha + \frac{1}{2})} \sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij} B_{ij})^{\alpha + 1} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}^{\alpha + 1} B_{ij}^{\alpha + \frac{1}{2}}) \right]^{-1} \quad (4.2.19)$$

$$\hat{h}_{i2}^{ent} = \frac{\Gamma(\alpha + \frac{1}{2})}{\sqrt{2\beta}\Gamma(\alpha + 1)} \sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij} B_{ij})^{\alpha + \frac{1}{2}} \left[\sum_{\substack{j=1 \\ j \neq i}}^N (A_{ij}^{\alpha + \frac{1}{2}} B_{ij}^{\alpha + 1}) \right]^{-1}. \quad (4.2.20)$$

Le théorème suivant montre que si l'on choisit les paramètres du vecteur de lissage de manière appropriée (fixer α et laisser β diverger avec un taux approprié lorsque la taille de l'échantillon augmente), on peut obtenir une séquence de paramètres de lissage qui converge vers zéro presque sûrement, pour chaque X_i .

Théorème 4.2.2. *Soit la fonction λ_0 continu et de support borné. Si $\alpha > 0$ est fixé, $\beta \rightarrow \infty$ et $m\beta^{-\frac{1}{2}} \rightarrow \infty$, quand $m \rightarrow \infty$ alors*

$$\hat{h}_i^{qua} \rightarrow 0, \quad m\hat{h}_i^{qua} \rightarrow \infty$$

et

$$\hat{h}_i^{ent} \rightarrow 0, \quad m\hat{h}_i^{ent} \rightarrow \infty$$

avec une probabilité égale à 1, quand $m \rightarrow \infty$.

Démonstration :

Ici, nous montrons que $h_{ij} \rightarrow 0$ et $mh_{ij} \rightarrow \infty$, quand $m \rightarrow \infty$.

Puisque β diverge quand la taille de l'échantillon augmente, nous écrivons β comme β_m . Choissant ϵ_m , tel que $\epsilon_m \rightarrow \infty$, et $\epsilon_m \sqrt{\beta_m} \rightarrow 0$, quand $m \rightarrow \infty$. Puisque $\Gamma(\alpha)$ est une constante pour α fixé, l'équation (4.2.17) peut être réécrite comme suit :

$$\hat{h}_{i1}^{qua} = C_\alpha \frac{\sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^\alpha \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha + \frac{1}{2}}}{\sqrt{\beta} \sum_{\substack{j=1 \\ j \neq i}}^N \left(\left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right) \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right) \right)^{\alpha + \frac{1}{2}}} \quad (4.2.21)$$

avec $C_\alpha = \frac{\Gamma(\alpha)}{\sqrt{2}\Gamma(\alpha+1/2)}$. Considérons le numérateur de l'équation (4.2.21),

$$\begin{aligned}
 & \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha \left(\frac{1}{\beta_m(X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2} \\
 & \leq \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha \\
 & = \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\left[|X_{i1} - X_{j1}| \geq \frac{\epsilon_m}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha \\
 & \quad + \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\epsilon_m}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha \\
 & < \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\epsilon_m}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha + m \left(\frac{1}{\epsilon_m^2 + 2} \right)^\alpha \\
 & = \frac{m\epsilon_m}{\sqrt{\beta_m}} \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\frac{m\epsilon_m}{\sqrt{\beta_m}}} \mathbf{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\epsilon_m}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m(X_{i1} - X_{j1})^2 + 2} \right)^\alpha + m \left(\frac{1}{\epsilon_m^2 + 2} \right)^\alpha \\
 & \leq \frac{m\epsilon_m}{\sqrt{\beta_m}} \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\frac{m\epsilon_m}{\sqrt{\beta_m}}} \mathbf{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\epsilon_m}{\sqrt{\beta_m}}\right]} \left(\frac{1}{2} \right)^\alpha + m \left(\frac{1}{\epsilon_m^2 + 2} \right)^\alpha
 \end{aligned}$$

Considérons le dénominateur de l'équation (4.2.21)

$$\begin{aligned}
 & \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+1/2} \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+1/2} \\
 & \geq \sum_{\substack{j=1 \\ j \neq i}}^N \sqrt{\beta_m} \mathbb{1}_{\left[|X_{i1} - X_{j1}| \geq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+\frac{1}{2}} \\
 & \times \mathbb{1}_{\left[|X_{i2} - X_{j2}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+\frac{1}{2}} \\
 & = m\sqrt{\epsilon_m} \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\frac{m\sqrt{\epsilon_m}}{\sqrt{\beta_m}}} \mathbb{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+\frac{1}{2}} \\
 & \times \mathbb{1}_{\left[|X_{i2} - X_{j2}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2} \left(\frac{1}{\epsilon_m + 2} \right)^{2(\alpha+\frac{1}{2})} \\
 & \geq m\sqrt{\epsilon_m} \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{\frac{m\sqrt{\epsilon_m}}{\sqrt{\beta_m}}} \mathbb{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+\frac{1}{2}} \\
 & \times \mathbb{1}_{\left[|X_{i2} - X_{j2}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+\frac{1}{2}} \left(\frac{1}{\epsilon_m + 2} \right)^{(\alpha+\frac{1}{2})}
 \end{aligned}$$

Soit g_{1m} défini par

$$\sum_{\substack{j=1 \\ j \neq i}}^N \frac{\sqrt{\beta_m}}{m\epsilon_m} \mathbb{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]},$$

et g_{2m} défini par

$$\sum_{\substack{j=1 \\ j \neq i}}^N \frac{\sqrt{\beta_m}}{m\sqrt{\epsilon_m}} \mathbb{1}_{\left[|X_{i1} - X_{j1}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]} \mathbb{1}_{\left[|X_{i2} - X_{j2}| \leq \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}}\right]}.$$

Combinant les deux inégalités précédent, nous obtenons :

$$\begin{aligned}
 & \frac{\sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^\alpha \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2}}{\sqrt{\beta} \sum_{\substack{j=1 \\ j \neq i}}^N \left(\left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right) \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right) \right)^{\alpha+1/2}} \\
 & \leq \frac{m \frac{\epsilon_m}{\sqrt{\beta_m}} g_{1m} \left(\frac{1}{2}\right)^\alpha + m \left(\frac{1}{\epsilon_m^2 + 2}\right)^\alpha}{m \sqrt{\epsilon_m} g_{2m} \left(\frac{1}{\epsilon_m + 2}\right)^{\alpha+1/2}} \\
 & = \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}} \frac{g_{1m} \left(\frac{1}{2}\right)^\alpha}{g_{2m} \left(\frac{1}{\epsilon_m + 2}\right)^{\alpha+1/2}} + \frac{\left(\frac{1}{\epsilon_m^2 + 2}\right)^\alpha}{\sqrt{\epsilon_m} g_{2m} \left(\frac{1}{\epsilon_m + 2}\right)^{\alpha+1/2}}
 \end{aligned}$$

Les deux expressions suivantes $\sum_{m=1}^{\infty} \exp(-\gamma m (\frac{\epsilon_m}{\sqrt{\beta_m}})^2)$ et $\sum_{m=1}^{\infty} \exp(-\gamma m (\frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}})^2)$ converge pour $\gamma > 0$ avec nos choix de ϵ_m, β_m , alors

$$\sup_{X_j} |g_{1m} - g| \xrightarrow{a.s} 0, \quad \sup_{X_j} |g_{2m} - g| \xrightarrow{a.s} 0$$

comme $m \rightarrow \infty$, par le théorème 2.1.3 de Rao (1983).

Notons que $\sqrt{\epsilon_m} \rightarrow \infty$ et $\frac{\sqrt{\epsilon_m}}{\beta_m} \rightarrow 0$, comme $m \rightarrow \infty$, alors

$$\begin{aligned}
 & \frac{\sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^\alpha \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2}}{\sqrt{\beta} \sum_{\substack{j=1 \\ j \neq i}}^N \left(\left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right) \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right) \right)^{\alpha+1/2}} \\
 & \leq \left(\frac{1}{2}\right)^\alpha \frac{\sqrt{\epsilon_m}}{\sqrt{\beta_m}} \frac{g}{g \left(\frac{1}{\epsilon_m + 2}\right)^{\alpha+1/2}} + \frac{\left(\frac{1}{\epsilon_m^2 + 2}\right)^\alpha}{\sqrt{\epsilon_m} g \left(\frac{1}{\epsilon_m + 2}\right)^{\alpha+1/2}} \\
 & \leq \left(\frac{1}{2}\right)^\alpha C_1 \frac{\sqrt{\epsilon_m} (\epsilon_m + 2)^{\alpha+1/2}}{\sqrt{\beta_m}} + C_2 \frac{(\epsilon_m + 2)^{\alpha+1/2}}{\sqrt{\epsilon_m} g (\epsilon_m^2 + 2)^\alpha} \\
 & \leq C_3 \left(\frac{\epsilon_m^{\alpha+1}}{\sqrt{\beta_m}} + \frac{1}{\epsilon_m^\alpha} \right)
 \end{aligned}$$

comme $m \rightarrow \infty$, C_1, C_2, C_3 sont des constantes. Par conséquent, m assez grand,

$$\hat{h}_{i1}^{qua} \leq C \beta_m^{-\frac{\alpha}{2(2\alpha+1)}}$$

$$\begin{aligned} & \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^\alpha \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2} \\ & \geq \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{1}{\beta_m (X_{i1} - X_{j1})^2 + 2} \right)^{\alpha+1/2} \left(\frac{1}{\beta_m (X_{i2} - X_{j2})^2 + 2} \right)^{\alpha+1/2} \end{aligned}$$

$$\hat{h}_{i1}^{qua} \geq C_\alpha \beta_m^{-1/2}.$$

Par conséquent,

$$C_\alpha \beta_m^{-1/2} \leq \hat{h}_{i1}^{qua} \leq C \beta_m^{-\frac{\alpha}{2(2\alpha+1)}}.$$

Par les hypothèses, $\beta_m \rightarrow \infty$ et $m\beta_m^{-1/2} \rightarrow \infty$, nous obtenons $\hat{h}_{i1}^{qua} \rightarrow 0$ et $mh_i \rightarrow \infty$, quand $m \rightarrow \infty$. Cependant, \hat{h}_{i1}^{qua} est un bon paramètre de lissage de l'estimateur de la fonction intensité.

Remplaçant \hat{h}_{i1}^{qua} par \hat{h}_{i2}^{qua} , \hat{h}_{i1}^{ent} et \hat{h}_{i2}^{ent} et utilisant les mêmes calculs, nous obtenons les mêmes résultats \square .

4.3 Études de simulation

Cette section présente une comparaison des performances de l'approche bayésienne adaptative avec la méthode classique globale *LSCV* pour la sélection de la matrice de lissage h_i en utilisant comme mesure de comparaison l'erreur quadratique intégrée (*ISE*) donnée par

$$ISE = \int_W \left(\hat{\lambda}_{0, h_i}(x) - \lambda_0(x) \right)^2 dx \quad (4.3.1)$$

Dans cette étude de simulation, nous considérons les mêmes fonctions intensités du chapitre 3 données par les équations (3.3.2), (3.3.3), (3.3.4) et (3.3.2).

4.3.1 Analyse de sensibilité

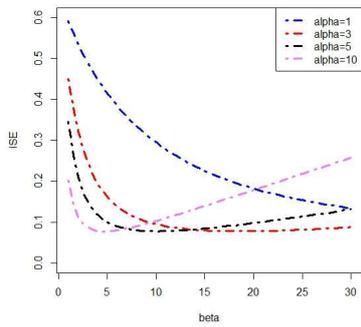
L'analyse de sensibilité de la loi a priori de paramètres (α, β) sera traitée comme dans Zougab et al. (2012)(2014). Nous examinons la sensibilité du modèle $\hat{\lambda}_{0, h_i}$ où le vecteur h_i est obtenu par l'approche bayésienne adaptative, via le *ISE* donné en (4.3.1). Dans le tableau 4.1, nous calculons le \widehat{ISE} (l'estimateur du *ISE*) en utilisant l'approche bayésienne adaptative avec le noyau gaussien produit sous la fonction perte quadratique, et en combinant quatre valeurs de $\alpha \in \{1, 3, 5, 10\}$ et quatre valeurs de $\beta \in \{1, 5, 10, 15\}$ sur la base de 100 répétitions et nous utilisons tous les modèles pour des échantillons de taille $m = 500$.

Pour la fonction de perte quadratique.

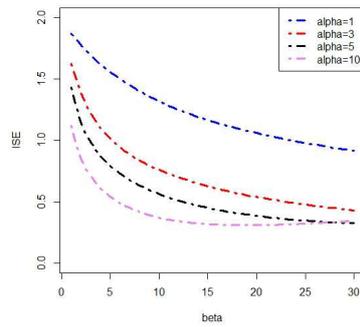
La figure 4.1 rapporte l'allure de $\beta \mapsto \widehat{ISE}(\beta)$ pour les quatre valeurs de α fixées. Sous la fonction de perte quadratique, les paramètres α et β sont choisis comme suit :

- pour λ_1 : $(\alpha, \beta) = (10, 5)$;
- pour λ_2 : $(\alpha, \beta) = (10, 10)$;
- pour λ_3 : $(\alpha, \beta) = (10, 10)$;
- pour λ_4 : $(\alpha, \beta) = (10, 10)$;
- pour λ_5 : $(\alpha, \beta) = (10, 10)$.

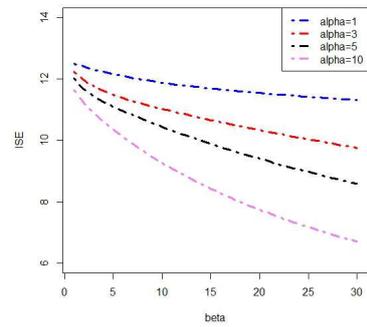
À partir de la figure 4.1, nous observons que pour $\alpha \in \{1, 3, 5, 10\}$ pour la loi a priori donnée par (4.2.6), les valeurs du \widehat{ISE} diminuent pour une certaine valeur de β puis augmentent selon que β croît. Cependant, $\alpha = 10$ donne les plus petites valeurs de \widehat{ISE} pour les fonctions intensités $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ et λ_5 comparé avec $\alpha \in \{1, 3, 5\}$ pour toutes les valeurs de β .



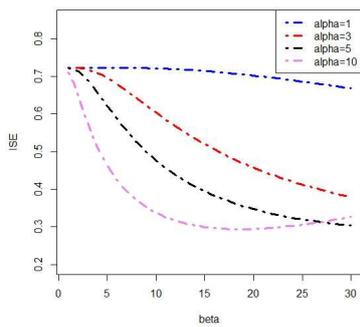
(a) λ_1



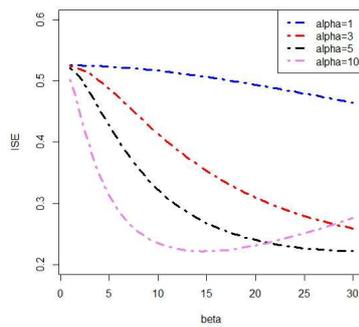
(b) λ_2



(c) λ_3



(d) λ_4



(e) λ_5

FIGURE 4.1 – Graphe du \widehat{ISE} avec le noyau gaussien et la loi a priori donnée à l'équation (4.2.6) des fonctions intensités sous la fonction perte quadratique.

TABLE 4.1 – Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$ sous la fonction de perte quadratique.

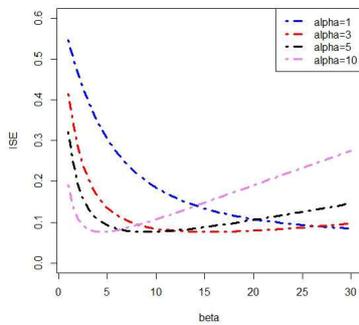
		λ_1	λ_2	λ_3	λ_4	λ_5
α	β	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}
1	1	0.59764	1.82785	12.52033	0.71837	0.52014
	5	0.42546	1.51624	12.17444	0.71916	0.51845
	10	0.30727	1.28191	11.89736	0.71742	0.51227
	15	0.23851	1.13150	11.70381	0.71063	0.50246
3	1	0.45927	1.58260	12.24831	0.71870	0.51919
	5	0.17480	0.98051	11.49379	0.71392	0.48367
	10	0.10645	0.73158	11.03024	0.67627	0.41239
	15	0.08951	0.59902	10.66389	0.62149	0.35287
5	1	0.35779	1.39086	12.02785	0.71893	0.51615
	5	0.11205	0.76260	11.10282	0.61980	0.42494
	10	0.08615	0.53511	10.44153	0.47745	0.32297
	15	0.09053	0.41969	9.88265	0.39600	0.26935
10	1	0.21505	1.08686	11.64761	0.70749	0.49859
	5	0.08372	0.51292	10.35633	0.46365	0.31302
	10	0.10678	0.33794	9.25831	0.33975	0.23746
	15	0.14194	0.26370	7.407172	0.20122	0.22601

NB : En gras le \widehat{ISE} minimum.

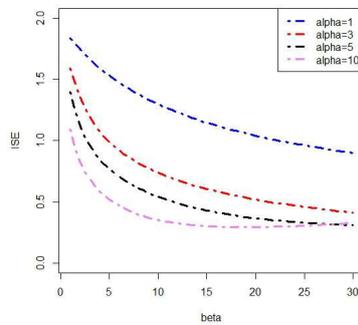
Pour la fonction de perte entropie.

La figure 4.2 représente le graphe de $\beta \mapsto \widehat{ISE}(\beta)$ pour quatre valeurs fixées de $\alpha \in \{1, 3, 5, 10\}$ où h_i est obtenu sous la fonction de perte entropie et les paramètres α et β sont choisis comme suit :

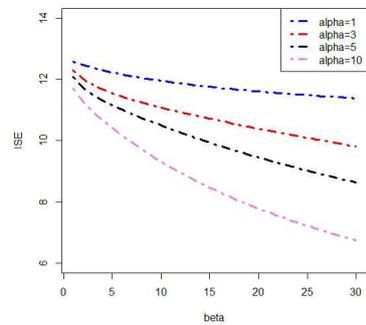
- pour λ_1 : $(\alpha, \beta) = (1, 10)$,
- pour λ_2 : $(\alpha, \beta) = (3, 10)$
- pour λ_3 : $(\alpha, \beta) = (10, 15)$
- pour λ_4 : $(\alpha, \beta) = (3, 10)$ et
- pour λ_5 : $(\alpha, \beta) = (3, 10)$



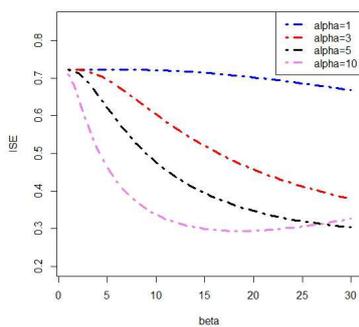
(a) λ_1



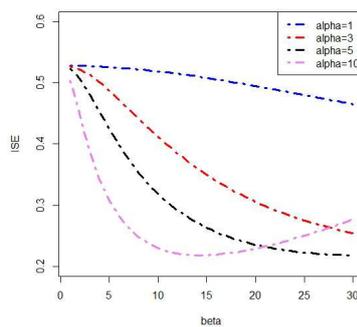
(b) λ_2



(c) λ_3



(d) λ_4



(e) λ_5

FIGURE 4.2 – Graphe du \widehat{ISE} avec le noyau gaussien et la loi a priori donnée à l'équation (4.2.6) des fonctions intensités pour la fonction perte entropie.

TABLE 4.2 – Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$ sous la fonction de perte entropie.

		λ_1	λ_2	λ_3	λ_4	λ_5
α	β	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}	\widehat{ISE}
1	1	0.32019	1.30349	12.02590	0.71044	0.52349
	5	0.08995	0.64463	10.91175	0.54736	0.38399
	10	0.08258	0.41597	9.99137	0.38731	0.27237
	15	0.10823	0.31854	9.22765	0.31736	0.23225
3	1	0.15473	0.92354	11.50698	0.67499	0.48262
	5	0.09125	0.38668	9.79109	0.36394	0.25727
	10	0.14671	0.26527	8.37281	0.28109	0.22285
	15	0.21031	0.27653	7.35893	0.28306	0.24850
5	1	0.10802	0.75291	11.18454	0.61093	0.43035
	5	0.11939	0.30291	8.95289	0.30073	0.22529
	10	0.21503	0.28043	7.29440	0.28487	0.25159
	15	0.31362	0.34263	6.20742	0.34026	0.32346
10	1	0.08300	0.53793	10.55882	0.47319	0.32947
	5	0.20303	0.27754	7.45946	0.28115	0.24476
	10	0.38670	0.40532	5.61196	0.39666	0.38568
	15	0.56830	0.57649	4.58638	0.55360	0.55019

NB : En gras le \widehat{ISE} minimum.

À partir de la figure 4.1, nous observons que pour $\alpha \in \{1, 3, 5, 10\}$ pour la loi a priori donnée par l'équation (4.2.9), les valeurs du \widehat{ISE} diminuent pour une certaine valeur de β puis augment selon que β croit. Cependant, $\alpha = 10$ donne les plus petites valeurs de \widehat{ISE} pour les fonctions intensité λ_1 , λ_2 , λ_3 , λ_4 et λ_5 pour toutes valeurs de β .

4.3.2 Approche bayésienne adaptative, locale et la validation croisée.

Nous comparons les performances de l'approche bayésienne adaptative avec la méthode bayésienne locale et la méthode classique globale *LSCV* pour la sélection de la matrice de lissage adaptatif h_i dans le contexte de l'estimation de la fonction intensité du processus de poisson non homogène bivarié. Par la méthode du noyau. Cette évaluation de performance sera basée sur le critère \widehat{ISE} donné en (4.3.1). Le tableau 4.2 présente les résultats de simulation du \widehat{ISE} pour le noyau gaussien, par l'approche bayésienne adaptative avec la fonction perte quadratique et entropie notée respectivement par *Bayes – quad* et *Bayes – ent*, la méthode bayésienne locale *Bayes – local* et par la méthode classique notée par *LSCV*.

TABLE 4.3 – Valeurs du \widehat{ISE} moyen pour les fonctions intensités $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$.

Intensité	m	\widehat{ISE}			
		$LSCV$	Bayes local	Bayes-ent	Bayes-qua
λ_1	20	2.05799	1.46629	1.42883	1.39620
	50	1.67185	0.44261	0.42622	0.41725
	100	1.66827	0.29864	0.23694	0.23507
	200	1.61114	0.14907	0.13983	0.13924
	400	1.60389	0.08953	0.08908	0.08932
	500	1.52249	0.08463	0.08258	0.08372
λ_2	20	3.44664	1.51401	1.33669	1.29262
	50	3.15100	0.94609	0.90627	0.93242
	100	2.99807	0.71966	0.65435	0.69796
	200	2.94038	0.69979	0.57686	0.62545
	400	2.92211	0.33276	0.31102	0.32578
	500	2.87909	0.28304	0.26527	0.26370
λ_3	20	14.91222	12.75374	12.00605	12.19158
	50	13.95714	9.53085	9.08855	9.41468
	100	13.76103	8.85307	6.85120	8.76765
	200	13.37586	8.21080	5.62105	8.11495
	400	13.29754	8.01764	4.96485	7.73272
	500	13.18517	7.95844	4.58638	7.407172
λ_4	20	2.03730	1.48579	1.36749	1.01470
	50	1.79842	0.90025	0.87859	0.81900
	100	1.74923	0.69864	0.69797	0.69499
	200	1.72217	0.54640	0.49321	0.52948
	400	1.68298	0.35410	0.32936	0.32641
	500	1.67720	0.29689	0.28109	0.20122
λ_5	20	1.85584	1.27327	1.03410	1.16707
	50	1.62308	0.66453	0.61474	0.55712
	100	1.59969	0.56778	0.5290	0.52770
	200	1.57946	0.43645	0.40364	0.40363
	400	1.53382	0.39486	0.36100	0.27398
	500	1.52249	0.22878	0.22285	0.22601

NB : En gras le \widehat{ISE} minimum.

D'après le tableau 4.3, nous observons que :

- la valeur \widehat{ISE} (erreur quadratique intégrée) diminue avec l'augmentation du nombre d'événements pour tous les modèles considérés ;
- l'estimateur de la fonction d'intensité utilisant l'approche bayésienne adaptative est meilleur que celui utilisant la méthode $LSCV$ et l'approche bayésienne locale,

- quelle que soit la fonction intensité cible et pour toutes les tailles d'échantillon ;
- L'approche bayésienne adaptative avec la fonction de perte entropie est aussi performante que l'approche bayésienne adaptative avec la fonction de perte quadratique.

Nous avons également comparé le temps d'exécution de chaque méthode (bayésienne adaptative, bayésienne locale et *LSCV*). Les résultats sont donnés pour les $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$, $\lambda_4(x)$ et $\lambda_5(x)$ avec des tailles d'échantillon $m \in \{20, 50, 100, 200, 400, 500\}$ et pour 100 simulations. Les résultats sont présentés dans le tableau 4.4. Ils indiquent l'avantage de l'approche bayésienne adaptative.

TABLE 4.4 – Comparaison du temps d'exécution (en secondes).

Intensité	m	temps			
		LSCV	Bayes local	Bayes-ent	Bayes-qua
λ_1	20	7.19	4.37	4.37	4.27
	50	8.02	4.71	4.57	4.61
	100	12.51	5.32	5.29	5.31
	200	29.88	6.48	6.43	6.34
	400	106.16	8.75	8.64	8.57
	500	170.31	9.86	9.85	9.69
λ_2	20	7.14	4.35	4.31	4.32
	50	8.00	4.65	4.69	4.70
	100	10.74	5.22	5.17	5.20
	200	23.35	6.38	6.24	6.35
	400	76.35	8.48	8.30	8.26
	500	120.12	9.40	9.23	9.41
λ_3	20	7.89	4.38	4.29	4.31
	50	9.33	4.79	4.72	4.74
	100	14.44	5.27	5.26	5.23
	200	40.41	7.71	6.38	6.56
	400	120.66	8.66	8.44	8.64
	500	191.87	9.78	9.53	9.77
λ_4	20	7.39	4.36	4.30	4.27
	50	8.91	4.76	4.69	4.70
	100	12.95	5.42	5.24	5.26
	200	14.03	6.47	6.30	6.39
	400	32.10	8.87	8.41	8.55
	500	163.56	9.69	9.55	9.63
λ_5	20	7.07	4.40	4.29	4.27
	50	8.19	4.72	4.69	4.72
	100	12.30	5.31	5.23	5.26
	200	28.72	6.52	6.32	6.43
	400	101.82	8.98	8.65	8.61
	500	169.60	10.04	9.61	9.58

NB : En gras le temps minimum.

4.3.3 Données réelles

Nous utilisons ici l'ensemble de données de localisations des 58 cas de cancer du larynx et 978 cas de cancer du poumon enregistrés dans la région de Chorley-Ribble au cours des années 1974 – 1983 (voir chapitre 3, partie 3.3.2).

Pour évaluer les performances des estimateurs, nous utilisons le critère d'erreur quadra-

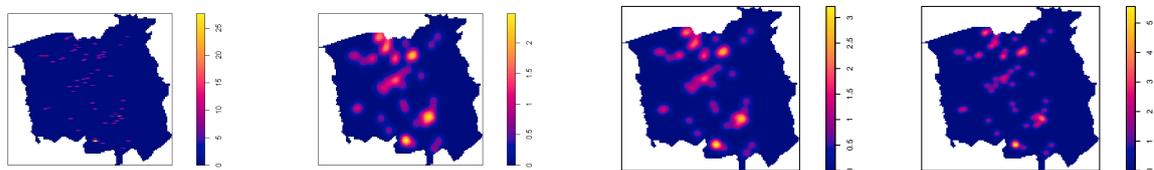
tique intégrée ISE_0 défini par (3.3.6). Nous choisissons $(\alpha, \beta) = (10, \sqrt{m})$ pour le cancer du larynx et pour le cancer de poumon, nous utilisons les paramètres $(\alpha, \beta) = (3, \sqrt{m})$ sous les fonctions de pertes quadratique et entropie respectivement.

Les résultats sont donnés dans le tableau 4.4 avec l'utilisation du noyau gaussien produit. Ces résultats montrent que l'approche bayésienne adaptative est meilleure que l'approche bayésienne locale et la méthode classique $LSCV$.

TABLE 4.5 – Comparaison du ISE_0 et du temps d'exécution pour les données réelles avec le noyau gaussien produit.

	Methods	LSCV	Bayes local	Bayes-ent	Bayes-quad
Larynx	ISE_0	0.03266	0.00626	0.00432	0.00361
	Temps	0.19134	0.11851	0.11487	0.11716
Poumon	ISE_0	0.04021	0.00515	0.00340	0.00431
	Temps	9.57779	0.23938	0.13124	0.12187

NB : En gras le ISE_0 minimum.



(a) LSCV

(b) Bayes local

(c) Bayes adap-ent

(d) Bayes adap-qua

FIGURE 4.3 – Estimation de la fonction intensité par noyau pour des données du cancer de larynx par la méthode : (a) $LSCV$, (b) bayésienne locale, (c) bayésienne adaptative avec la fonction de perte entropie, (d) bayésienne adaptative avec la fonction de perte quadratique.

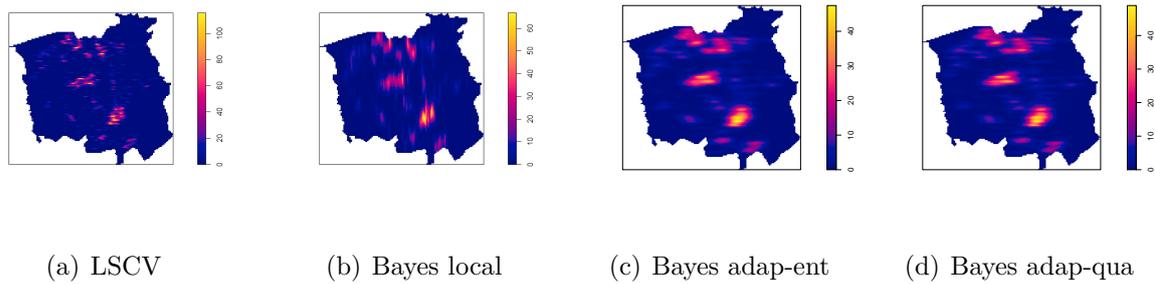


FIGURE 4.4 – Estimation de la fonction intensité par noyau pour les données du cancer de poumon par la méthode : (a) *LSCV*, (b) bayésienne locale, (c) bayésienne adaptative avec la fonction de perte entropie, (d) bayésienne adaptative avec la fonction de perte quadratique.

Les figures 4.3 et 4.4 montrent les estimateurs d'intensités du noyau fourni par les différentes méthodes de sélection de la matrice de lissage. Pour le cancer du larynx et du poumon, la méthode bayésienne adaptative a fourni un estimateur plus lisse que ceux obtenus par la méthode bayésienne locale et la méthode *LSCV*.

Nous fournissons l'erreur quadratique intégrée et les temps d'exécution de ces méthodes dans le tableau 4.4. L'estimateur bayésien adaptatif proposé a un ISE_0 plus petit que l'estimateur utilisant la technique *LSCV* et l'estimateur bayésien local pour ces données réelles.

4.4 Conclusion

Le présent chapitre aborde le problème de la sélection de la matrice de lissage de l'estimateur à noyau de la fonction intensité du processus de Poisson non homogène. L'approche bayésienne adaptative pour sélectionner la matrice de lissage de l'estimation de la fonction intensité à noyau multivarié est une meilleure alternative par rapport à la méthode *LSCV* et à l'approche bayésienne locale. L'avantage de la méthode bayésienne adaptative est que nous avons obtenu la formule explicite des éléments de la matrice de lissage en exploitant la conjugabilité entre le noyau gaussien et la loi a priori. Les études de simulation et les données réelles montrent que l'approche bayésienne adaptative est plus performante que l'approche bayésienne locale et la méthode *LSCV* classique selon le critère *ISE* et en temps d'exécution.

Estimation de la fonction intensité par le noyau Bêta.

5.1 Introduction

Malgré leur robustesse sur les modèles les plus généraux, les estimateurs à noyaux (classiques) connaissent quelques inconvénients parmi lesquels se situent les problèmes de support. Le premier cas est celui des effets de bord dans l'estimation des densités à support borné au moins d'un côté. Ceci est dû au fait que le noyau classique étant symétrique assigne des poids à l'extérieur du support de la densité à estimer lorsque le lissage se fait aux points de bord. Cela crée ainsi les biais dits de bord qui rendent l'estimateur non consistant. Plusieurs solutions à ce problème ont été proposées dans la littérature. Schuster (1985), Silverman (1986) et Cline et Hart (1991) ont fait usage de la méthode dite des données reflétées ("*reflection data method*" en anglais). Cette méthode est spécialement conçue pour le cas des densités vérifiant $f'(0) = 0$ et $f'(t_1) = 0$, où f' désigne la dérivée première de f définie sur un intervalle du type $[0, t_1]$. Ensuite, Gasser et Müller (1979), Gasser et al. (1985), Jones (1993), Zhang et Karunamuni (1999) et Karunamuni et Albert (2005) ont suggéré l'utilisation de la méthode des noyaux de bord ("*boundary kernels method*" en anglais). Cette méthode est plus générale que celle des données reflétées dans le sens où elle peut s'adapter à n'importe quelle forme de la densité. Cependant, un inconvénient de cette méthode est que les estimations faites pourraient être négatives à proximité des extrémités, surtout quand $f(0) \approx 0$ ou $f(t_1) \approx 0$. Pour corriger cette déficience de la méthode des noyaux de bord, certains remèdes ont été proposés ; se référer à Jones (1993), Jones et Foster (1996), Gasser et Müller (1979), Karunamuni et al. (2006) ainsi que Karunamuni et Zhang (2008) pour un approfondissement.

Toujours dans l'optique de résoudre le problème des effets de bord, Cowling et Hall (1996) ont proposé la méthode des pseudo-données ("*pseudo-data method*" en anglais). L'idée de cette méthode consiste à générer des observations supplémentaires $X_{(i)}$ à l'aide de ce qu'ils appellent les "*trois points pour mieux régner*", puis les combiner avec les observations d'origine X_i pour former un type d'estimateur à noyau. Dans le même ordre d'idée, Marron et Ruppert (1994) ont proposé la méthode de transformation ("*transformation method*" en anglais). Elle est constituée de trois étapes. Tout d'abord, une transformation g est choisie dans une famille paramétrique de telle sorte que la densité des variables $Y_i = g(X_i)$ ait une dérivée première qui est approximativement égale à 0 aux bords. Ensuite, un estimateur à noyau de la réflexion est appliquée à Y_i . Enfin, cet estimateur est converti par la formule de changement de variables pour obtenir une estimation de f . Notons aussi que d'autres auteurs comme Bouezmarni et al. (2005) ainsi que Gustafson

et al. (2006) ont abordé le même sujet en combinant parfois quelques-unes de ces méthodes.

En multivarié, les conséquences des problèmes de bords sont bien plus sévères en ce sens que les régions de bords augmentent avec la dimension k . Les solutions aux problèmes de bords ont été moins bien étudiées que dans le cas univarié. Par exemple, Muller et Stadmüller (1999) ont proposé des noyaux de bord définies sur des supports arbitraires où les noyaux multivariés sont obtenus en minimisant un problème variationnel. On peut aussi se référer à Epanechnikov (1969). Bien que ces estimateurs aient des propriétés intéressantes, ils ne sont pas aisés à mettre en œuvre.

Récemment, certains auteurs ont proposé, dans le cas des densités à support compact, l'utilisation de noyaux dont le support coïncide avec celui de la densité à estimer. Ceci a efficacement résolu le problème des effets de bord puisque les noyaux utilisés ici sont généralement asymétriques et peuvent changer de forme selon la position du point d'estimation. C'est notamment le cas de Chen (1999) (2000a) avec les noyaux bêta et gamma pour estimer les densités à support respectivement $[0, 1]$ et $[0, +\infty[$ puis Scaillet (2004) avec les noyaux inverses gaussien et sa réciproque pour les densités à support $]0, \infty[$. Dans cette même optique, Senga Kiessé (2008) et Kokonendji & Senga Kiessé (2011) ont proposé l'utilisation des noyaux associés discrets pour le lissage des données catégorielles ou de dénombrement ; puisque jusqu'alors, certains auteurs comme Simonof (1996) ainsi que Simonof & Tutz (2000) dans leur tentative d'estimation des données discrètes font usage des noyaux continus.

En multivarié continu, Bouezmarni et Rombouts (2010) ont proposé des estimateurs à noyaux produits composés des univariés (bêta et gamma) de Chen (1999)(2000a). Li et Racine (2007) ont utilisé ce produit de noyaux univariés auquel il a ajouter des noyaux univariés gaussiens pour les données continues. Tous ces noyaux multivariés ont une matrice de lissage diagonale alors que la rôle des matrices de lissages pleines a été démontré dans certaines situations (voir, par exemple, figure 1.1).

Ce chapitre se propose d'homogénéiser la théorie concernant l'estimateur à noyau associé continu Bêta pour la fonction intensité du processus de Poisson non homogène. Il s'organise de la manière suivante : nous présentons d'abord une définition du noyau associé continu Bêta incluant les cas classiques. Par la suite, nous utilisons ce noyau associé pour estimer des fonctions intensité. Différentes propriétés sont présentées, en particulier la convergence au sens du risque quadratique moyen et intégré asymptotique (" *Asymptotic Mean Integrated Squared Error* " (*AMISE*) en anglais). Enfin, nous présentons les résultats des études par simulation sur quatre différents types de fonctions d'intensité et sur des données réelles.

5.2 Estimateur à noyau Bêta de la fonction intensité

Soit X un processus ponctuel spatial défini dans \mathbb{R}^2 et X_1, \dots, X_N une réalisation de X observée sur une région bornée $W \in \mathbb{R}^2$. L'estimateur à noyau continu bivarié associé $\hat{\lambda}_{0,h}$ de λ_0 est défini comme

$$\hat{\lambda}_{0,h}^{(1)}(x) = \frac{1}{N} \sum_{i=1}^N K_{x,h}(X_i) \mathbf{1}_{[N \neq 0]}. \quad (5.2.1)$$

où N est une variable aléatoire discrète suivant une loi de Poisson de paramètre $\int_W \lambda(x)dx$, $x = (x_1, x_2)^T \in W$, $X_i = (X_{i1}, X_{i2})^T$, $i = 1, \dots, N$ et $h = (h_1, h_2)^T$ est le vecteur de bandes passantes. Soit

$$K_{x,h}(\cdot) = \prod_{j=1}^2 K_{x_j, h_j}(\cdot)$$

un noyau associé continu bivarié paramétré par x et h , qui vérifie les conditions suivantes :

$$x \in \mathcal{S}_{x,h}, \mathbb{E}(\mathcal{Z}_{x,h}) = x + a(x, h) \tag{5.2.2}$$

et

$$Cov(\mathcal{Z}_{x,h}) = B(x, h) \tag{5.2.3}$$

L'intensité du premier ordre est définie à la suite de Diggle (2003) comme :

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{\mathbb{E}[N(dx)]}{|dx|} \right\} \tag{5.2.4}$$

où $|dx|$ désigne l'aire d'une région infinitésimale contenant le point $x \in W$.

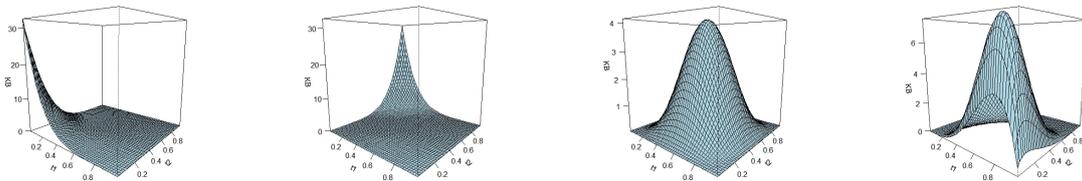
Nous supposons que le support compact W est connu et, sans perte de généralité, $W = [0, 1]^2$ et λ_0 a une dérivée seconde continue. Soit $K_{p,q}$ la fonction de densité d'une variable aléatoire $Beta(p, q)$. L'estimateur à noyau produit Bêta de la fonction intensité est donné par :

$$\hat{\lambda}_{0,h}^{(1)}(x) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1}(X_{ij}) \mathbb{1}_{[N \neq 0]} \tag{5.2.5}$$

avec

$$K_{\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1}(y) = \frac{y^{\frac{x_j}{h_j}} (1-y)^{\frac{1-x_j}{h_j}}}{B\left(\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1\right)} \tag{5.2.6}$$

est un noyau et $x_j \in [0, 1]$ Chen (1999).



(a) $(x_1, x_2) = (0, 0)$ (b) $(x_1, x_2) = (0, 1)$ (c) $(x_1, x_2) = (0.5, 0.5)$ (d) $(x_1, x_2) = (0.8, 0.1)$

FIGURE 5.1 – Forme du noyau Bêta pour $x = (x_1, x_2)$ avec paramètres de lissage $h_1 = h_2 = 0, 2$.

Propriétés de l'estimateur

Dans cette partie, nous examinons les différentes propriétés statistiques à distance finie et asymptotique de l'estimateur de la fonction intensité par l'utilisation du noyau associé Bêta donné par l'équation (5.2.5).

Soit (X_1, \dots, X_N) une réalisation du processus de Poisson spatial non homogène X avec intensité du premier ordre $\lambda(x)$ observée sur une région $[0, 1]^2 = W \subset \mathbb{R}^2$ Étant donné que le nombre d'événements N a une distribution de *Poisson* $\left(\int_W \lambda(x) dx\right) = \text{Poisson}(m)$.

Espérance

L'espérance mathématique de l'estimateur est donnée par :

$$\begin{aligned} \mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x)] &= \mathbb{E}\left(\frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_{ij}) \mathbb{1}_{N \neq 0}\right) \\ &= \mathbb{E}[\mathbb{E}[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_j) | N > 0]] \\ &= \sum_{i=1}^{\infty} \mathbb{E}\left(\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_{ij})\right) P(N = k) \\ &= (1 - e^{-m}) \mathbb{E}(\lambda_0(Y)). \end{aligned}$$

où les $Y = (Y_1, Y_2)$ et les Y_j sont i.i.d et suivent une loi Bêta de moyenne

$$\mu_{Y_j} = x_j + h_j(1 - 2x_j) + \Delta_1(x_j) \quad (5.2.7)$$

et de variance

$$\sigma_{Y_j}^2 = h_j x_j (1 - x_j) + \Delta_2(x_j) \quad (5.2.8)$$

où $\Delta_j(x_j) \leq M h_j^2$ pour $j = 1, 2$. Par conséquent, les autres termes des extensions ci-dessus pour μ_{Y_j} et $\sigma_{Y_j}^2$ sont uniformément $O(h^2)$.

Le noyau produit $\prod_{j=1}^2 K_{x_j, h_j}^*(t)$ est utilisé dans (5.2.5) comme fonction d'intensité. Pour trouver l'espérance, faisons le développement de Taylor de $\lambda_0(y)$ au point μ_j

$$\begin{aligned} \lambda_0(y) &= \lambda_0(\mu) + \sum_{j=1}^2 (y_j - \mu_j) \frac{\partial \lambda_0(y)}{\partial x_j} + \sum_{j=1}^2 (y_j - \mu_j)^2 \frac{\partial^2 \lambda_0(y)}{\partial x_j^2} \\ &\quad + \sum_{j \neq l} (y_j - \mu_j)(y_l - \mu_l) \frac{\partial \lambda_0(y)}{\partial x_j \partial x_l} + o\left(\sum_{j=1}^2 h_j\right). \end{aligned}$$

$$\begin{aligned} \mathbb{E}(\hat{\lambda}_{0,h}^{(1)}(x)) &= (1 - e^{-m})\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 h_j(1 - 2x_j)\lambda_0^j(x) \\ &\quad + \frac{(1 - e^{-m})}{2} \sum_{j=1}^2 x_j h_j(1 - x_j)\lambda_0^{jj}(x) + o((1 - e^{-m}) \sum_{j=1}^2 h_j^2) \end{aligned}$$

Biais

Le biais de l'estimateur de la fonction intensité est donné par :

$$Bias(\hat{\lambda}_{0,h}^{(1)}(x)) = -e^{-m}\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 h_j u_j(x) + o\left((1 - e^{-m}) \sum_{j=1}^2 h_j^2\right) \quad (5.2.9)$$

où

$$u_j(x) = \left((1 - 2x_j)\lambda_0^j(x) + \frac{x_j}{2} h_j(1 - x_j)\lambda_0^{jj}(x) \right), \quad \lambda_0^j(x) = \frac{\partial \lambda_0(x)}{\partial x_j}, \quad \lambda_0^{jj}(x) = \frac{\partial^2 \lambda_0(x)}{\partial x_j^2}.$$

Variance

Par définition la variance de l'estimateur de la fonction intensité est donnée par :

$$\begin{aligned} Var(\hat{\lambda}_{0,h}^{(1)}) &= \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_{ij}) \mathbf{1}_{N \neq 0} \right)^2 \right] \\ &\quad - \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_{ij}) \mathbf{1}_{N \neq 0} \right) \right]^2 \end{aligned}$$

et

$$\begin{aligned}
& \mathbb{E} \left[\left(\frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_{ij}) \mathbf{1}_{N \neq 0} \right)^2 \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\left(\frac{\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(X_j) \prod_{l=1}^2 K_{\frac{x_l}{h_l}+1, \frac{1-x_l}{h_l}+1}(X_l)}{N^2} \right) \middle| N > 0 \right] \right] \\
&= \sum_{k=1}^{\infty} \left[\frac{k-1}{k} \int_W \int_W \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \prod_{l=1}^2 K_{\frac{x_l}{h_l}+1, \frac{1-x_l}{h_l}+1}(y_l) \lambda_0(x) \lambda_0(y) dx dy \right. \\
&\quad \left. + \frac{1}{k} \int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right]^2 \lambda_0(x) dx \right] \frac{e^{-m} m^k}{k!} \\
&= \mathbb{E} \left[\frac{N-1}{N} \mathbf{1}_{N>0} \right] \left(\int_W \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \lambda_0(x) dx \right)^2 \\
&\quad + \mathbb{E} \left(\frac{1}{N} \mathbf{1}_{N>0} \right) \int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right]^2 \lambda_0(x) dx
\end{aligned}$$

En posant

$$A(m) = \mathbb{E} \left(\frac{1}{N} \mathbf{1}_{N>0} \right),$$

nous avons :

$$\begin{aligned}
\text{Var}(\hat{\lambda}_{0,h}^{(1)}(x)) &= A(m) \int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right]^2 \lambda_0(x) dx \\
&\quad + \left[\mathbb{E} \left(\frac{N-1}{N} \mathbf{1}_{N>0} \right) + (1 - e^{-m})^2 \right] \left[\int_W \prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \lambda_0(x) dx \right]^2 \\
&= A(m) \int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right]^2 \lambda_0(x) dx \\
&\quad + (A(m) - e^{-m} + e^{-2m}) \left[\int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right] \lambda_0(x) dx \right]^2
\end{aligned}$$

Par un développement de Taylor, l'intégrale du premier terme dans la partie droite de cette expression est donnée par :

$$\begin{aligned}
A(m) \int_W \left[\prod_{j=1}^2 K_{\frac{x_j}{h_j}+1, \frac{1-x_j}{h_j}+1}(x_j) \right]^2 \lambda_0(x) dx &= A(m) \left(\prod_{j=1}^2 A_{h_j}(x_j) \right) \mathbb{E}(\lambda_0(\gamma_x)) \\
&= A(m) \prod_{j=1}^2 (A_{h_j}(x_j)) \lambda_0(x) + o \left(A(m) \sum_{j=1}^2 h_j^2 \right)
\end{aligned}$$

où γ_{x_j} est une variable aléatoire Bêta $B(\frac{2x_j}{h_j} + 1, \frac{2(1-x_j)}{h_j} + 1)$.

$$A_{h_j}(x_j) = \frac{B(\frac{2x_j}{h_j} + 1, \frac{2(1-x_j)}{h_j} + 1)}{B(\frac{x_j}{h_j} + 1, \frac{(1-x_j)}{h_j} + 1)}.$$

Dans chen (1999),

$$A_{h_j}(x_j) = \begin{cases} \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}, & \text{si } \frac{x_j}{h_j} \text{ et } \frac{1-x_j}{h_j} \rightarrow \infty \\ \frac{h_j^{-1}\Gamma(2k_j+1)}{2^{2k_j+1}\Gamma^2(k_j+1)}, & \text{si } \frac{x_j}{h_j} \rightarrow k_j \text{ ou } \frac{1-x_j}{h_j} \rightarrow k_j. \end{cases}$$

$$Var(\hat{\lambda}_{0,h}^{(1)}(x)) = A(m) \prod_{j=1}^2 A_{h_j}(x_j) \mathbb{E}(\lambda_0(Z(x))) \quad (5.2.10)$$

où $Z = (Z_1, Z_2)$ et Z_j sont des variables aléatoires *i.i.d* de loi Bêta.

Par le développement de Taylor,

$$\mathbb{E}(\lambda_0(Z)) = \lambda_0(x) + O(h)$$

Par conséquent,

$$Var(\hat{\lambda}_{0,h}^{(1)}(x)) = \begin{cases} A(m) \prod_{j=1}^2 \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}} (\lambda_0(x) + o(1)), & \text{si } \frac{x_j}{h_j} \text{ et } \frac{1-x_j}{h_j} \rightarrow \infty \\ A(m) \prod_{j=1}^2 \frac{h_j^{-1}\Gamma(2c_j+1)}{2^{2c_j+1}\Gamma^2(c_j+1)} (\lambda_0(x) + o(1)), & \text{si } \frac{x_j}{h_j} \rightarrow c_j \text{ ou } \frac{1-x_j}{h_j} \rightarrow c_j. \end{cases} \quad (5.2.11)$$

où

$$\lim_{\substack{\frac{x_j}{h_j} \rightarrow \infty, \\ \frac{(1-x_j)}{h_j} \rightarrow \infty}} \frac{B(\frac{2x_j}{h_j} + 1, \frac{2(1-x_j)}{h_j} + 1)}{B^2(\frac{x_j}{h_j} + 1, \frac{(1-x_j)}{h_j} + 1)} = \frac{1}{\sqrt{h_j}} \frac{1}{2\sqrt{\pi x_j(1-x_j)}} \quad (5.2.12)$$

$$\lim_{\frac{(1-x_j)}{h_j} \rightarrow \infty} \frac{B(2c_j + 1, \frac{2(1-x_j)}{h_j} + 1)}{B^2(c_j + 1, \frac{(1-x_j)}{h_j} + 1)} = \frac{1}{h_j} \frac{\Gamma(2c_j + 1)}{2^{2c_j+1}\Gamma^2(c_j + 1)}, \quad \text{pour } c_j \text{ fixé}, \quad (5.2.13)$$

$$\lim_{\frac{x_j}{h_j} \rightarrow \infty} \frac{B(\frac{2x_j}{h_j} + 1, 2c_j + 1)}{B^2(\frac{x_j}{h_j} + 1, c_j + 1)} = \frac{1}{h_j} \frac{\Gamma(2c_j + 1)}{2^{2c_j+1}\Gamma^2(c_j + 1)}, \quad \text{pour } c_j \text{ fixé}. \quad (5.2.14)$$

5.2.1 Erreur quadratique moyenne (MSE).

L'erreur quadratique moyenne est donnée par :

$$\begin{aligned} MSE(x) &= \mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x) - \lambda_0(x)] = Bias^2(\hat{\lambda}_{0,h}^{(1)}(x)) + Var(\hat{\lambda}_{0,h}^{(1)}(x)) \\ &= \left[-e^{-m}\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 h_j u_j(x) \right]^2 + A(m) \prod_{j=1}^2 \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}} \lambda_0(x). \end{aligned}$$

5.2.2 Erreur quadratique moyenne intégrée (MISE)

L'erreur quadratique moyenne intégrée est donnée par :

$$\begin{aligned} MISE(\hat{\lambda}_{0,h}^{(1)}(x)) &= \int_W \mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x) - \lambda_0(x)] dx = \int_W Bias(\hat{\lambda}_{0,h}^{(1)}(x)) dx + \int_W Var(\hat{\lambda}_{0,h}^{(1)}(x)) dx \\ &= \int_W \left[-e^{-m}\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 h_j u_j(x) \right]^2 dx \\ &\quad + \int_W A(m) \prod_{j=1}^2 \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}} \lambda_0(x) dx. \end{aligned}$$

Il est facile de voir que l'estimateur du noyau Bêta $\hat{\lambda}_{0,h}^{(1)}$ est positif. Cependant, l'équation (5.2.11) montre que la variance de $\hat{\lambda}_{0,h}^{(1)}(x)$ a un taux de convergence plus lent (ou une variance plus grande) pour les points x satisfaisant $\frac{x_j}{h_j} \rightarrow c_j$ ou $\frac{1-x_j}{h_j} \rightarrow c_j$. Ces points sont étiquetés comme des points limites dans Chen (2000a).

Une propriété indésirable de l'estimateur du noyau Bêta $\hat{\lambda}_{0,h}^{(1)}(x)$ est l'implication de $\lambda_0(x)^j$ dans l'expression du biais de l'équation (5.4.1), ce qui rend la comparaison difficile entre l'estimateur du noyau bêta et d'autres méthodes existantes comme l'estimateur du noyau gaussien. Pour faciliter la comparaison entre l'estimateur du noyau Bêta et d'autres estimateurs, Chen (2000a) a proposé une version modifiée de la fonction densité et vu le lien entre la fonction densité et la fonction intensité, l'estimateur $\hat{\lambda}_{0,h}^{(1)}(x)$ est modifié en reparamétrant soigneusement le noyau Bêta.

5.3 L'estimateur de la fonction intensité par le noyau Bêta modifié.

A la suite de Chen (1999) et Libengué (2013), un autre estimateur est donné par :

$$\hat{\lambda}_{0,h}^{(2)}(x) = \begin{cases} \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}, \frac{1-x_j}{h_j}}(X_{ij}) \mathbb{1}_{[N \neq 0]}, & \text{si } x_j \in [2h_j, 1 - 2h_j], \\ \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\rho(x_j), \frac{1-x_j}{h_j}}(X_{ij}) \mathbb{1}_{[N \neq 0]}, & \text{si } x_j \in [0, 2h_j] \\ \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^2 K_{\frac{x_j}{h_j}, \rho(1-x_j)}(X_{ij}) \mathbb{1}_{[N \neq 0]}, & \text{si } x_j \in (1 - 2h_j, 1]. \end{cases} \quad (5.3.1)$$

où $K_{x,h}$ sont des noyaux Bêta bornés.

$$\text{ici } \rho_{h_j}(x_j) = 2h_j^2 + 2,5 - \sqrt{4h_j^4 + 6h_j^2 + 2,25 - x_j^2 - \frac{x_j}{h_j}}.$$

Notez que pour chaque point fixe h_j , $\rho(x_j, h_j)$ est une fonction croissante monotone de x_j entre $[0, 2h_j]$. En particulier, il a $y = x_j/h_j$ comme sa ligne tangente à $x_j = 2h_j$ et $\rho(0, h_j) = 1$, où $h = (h_1, h_2)$ est le vecteur des paramètres de lissage tels que $h_j \rightarrow 0$ quand $m \rightarrow \infty$. Soit $K_{\alpha,\beta}$ la fonction de densité d'une variable aléatoire $B(\alpha, \beta)$.

Propriétés de l'estimateur modifié.

Biais

Le biais de l'estimateur est donné par :

$$\text{Bias}(\hat{\lambda}_{0,h}^{(2)}(x)) \sim -e^{-m} \lambda_0(x) + \begin{cases} \frac{(1-e^{-m})}{2} \sum_{j=1}^2 h_j x_j (1-x_j) \lambda_0^{jj}(x), & \text{si } x_j \in [2h_j, 1 - 2h_j] \\ (1-e^{-m}) \sum_{j=1}^2 \xi(x_j) h_j \lambda_0^j(x), & \text{si } x_j \in [0, 2h_j], \\ -(1-e^{-m}) \sum_{j=1}^2 \xi(1-x_j) h_j \lambda_0^j(x), & \text{si } x_j \in (1 - 2h_j, 1] \end{cases}$$

avec $\xi(x_j) = (1-x_j)(\rho(x_j) - \frac{x_j}{h_j}) / (1 + h_j \rho(x_j) - x_j)$.

Variance

La variance de $\hat{\lambda}_{0,h_j}^{(2)}(x_j)$ est donnée par :

$$\text{Var}(\hat{\lambda}_{0,h_j}^{(2)}(x_j)) = A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x) + o(A(m) \sum_{j=1}^2 h_j) \quad (5.3.2)$$

où

$$A_{h_j}^*(x_j) = \begin{cases} \frac{B\left(\frac{2x_j}{h_j}-1, \frac{2(1-x_j)}{h}-1\right)}{B^2\left(\frac{x_j}{h_j}, \frac{(1-x_j)}{h_j}\right)}, & \text{si } x_j \in [2h_j, 1-2h_j], \\ \frac{B\left(2\rho(x_j)-1, \frac{2(1-x_j)}{h_j}-1\right)}{B^2\left(\rho(x_j), \frac{(1-x_j)}{h_j}\right)}, & \text{si } x_j \in [0, 2h_j] \\ \frac{B\left(\frac{2x_j}{h_j}-1, 2\rho(1-x_j)-1\right)}{B^2\left(\frac{x_j}{h_j}, \rho(1-x_j)\right)}, & \text{si } x_j \in (1-2h_j, 1]. \end{cases} \quad (5.3.3)$$

Il est évident d'après l'expression du biais de $\hat{\lambda}_{0,h}^{(2)}(x)$ que son biais ne dépend pas de λ_0^j dans l'intervalle $[2h_j, 1-2h_j]$. En combinant les l'expression du biais et de la variance ci-dessus, nous obtenons les expressions du *MSE* et du *MISE*.

MSE

L'erreur quadratique moyenne de l'estimateur Bêta modifié est donnée par :

$$MSE(\hat{\lambda}_{0,h}^{(2)}(x)) \sim \begin{cases} \left[-e^{-m}\lambda_0(x) + \frac{(1-e^{-m})}{2} \sum_{j=1}^{2h_j} x_j(1-x_j)\lambda_0^{jj}(x) \right]^2 + A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x), \\ \text{si } x_j \in [2h_j, 1-2h_j], \\ \left[-e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^2 \xi(x_j)h_j\lambda_0^j(x) \right]^2 + A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x), \\ \text{si } x_j \in [0, 2h_j], \\ \left[-e^{-m}\lambda_0(x) - (1-e^{-m}) \sum_{j=1}^2 \xi(1-x_j)h_j\lambda_0^j(x) \right]^2 + A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x), \\ \text{si } x_j \in (1-2h_j, 1] \end{cases}$$

MISE

Il est affirmé dans Chen (2000a) que $A_{h_j}^*(x_j)$ a une expression asymptotique similaire à celle de $A_{h_j}(x)$, et donc $Var(\hat{\lambda}_{0,h}^{(1)}(x)) \approx Var(\hat{\lambda}_{0,h}^{(2)}(x))$. Sur cette base, le *MISE* asymptotique simplifié de $\hat{\lambda}_{0,h}^{(2)}(x)$ est donné par :

$$\begin{aligned}
MISE(\hat{\lambda}_{0,h}^{(2)}(x)) &\sim \prod_{j=1}^2 \int_0^{2h_j} \left[-e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^2 \xi(x_j)h_j\lambda_0^{jj}(x) \right]^2 dx_j \\
&+ \prod_{j=1}^2 \int_{1-2h_j}^1 \left[-e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^2 \xi(1-x_j)h_j\lambda_0^j(x) \right]^2 dx_j \\
&+ \prod_{j=1}^2 \int_{2h_j}^{1-2h_j} \left[-e^{-m}\lambda_0(x) + \frac{(1-e^{-m})}{2} \sum_{j=1}^2 h_jx_j(1-x_j)\lambda_0^j(x) \right]^2 dx_j \\
&+ \int_0^1 \int_0^1 A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x) dx
\end{aligned}$$

5.4 Résultats asymptotiques.

Dans cette section, nous donnons les propriétés asymptotiques des estimateurs (5.2.5) et (5.3.1). À cette fin, nous donnons les expressions des biais et des variances dans les lemmes suivants. Dans les lemmes suivants consacrés au biais et aux variances, nous supposons que $\lambda_0(x)$ est une fonction deux fois différentiables. Le lemme suivant énonce les biais des estimateurs d'intensité non paramétrique.

Lemme 5.4.1.

$$Bias(\hat{\lambda}_{0,h}^{(1)}(x)) = -e^{-m}\lambda_0(x) + (1-e^{-m}) \sum_{j=1}^2 h_j u_j(x) + o\left((1-e^{-m}) \sum_{j=1}^2 h_j\right) \quad (5.4.1)$$

$$Bias(\hat{\lambda}_{0,h}^{(2)}(x)) = -e^{-m}\lambda_0(x) + \frac{(1-e^{-m})}{2} \sum_{j=1}^2 h_j v_j(x) + o\left((1-e^{-m}) \sum_{j=1}^2 h_j\right) \quad (5.4.2)$$

avec

$$v_j(x) = \frac{x_j(1-x_j)}{2} \frac{\partial^2 \lambda_0(x)}{\partial x_j^2}$$

Remarque 5.4.1. Si $h_1 = h_2 = h$ et $h \rightarrow 0$, $mh \rightarrow 0$ quand $m \rightarrow \infty$, alors les biais des estimateurs de la fonction d'intensité sont donnés par :

$$Bias(\hat{\lambda}_{0,h}^{(1)}(x)) = -e^{-m}\lambda_0(x) + \frac{(1-e^{-m})h}{2} \sum_{j=1}^2 u_j(x) + o((1-e^{-m})h) \quad (5.4.3)$$

$$Bias(\hat{\lambda}_{0,h}^{(2)}(x)) = -e^{-m}\lambda_0(x) + \frac{(1-e^{-m})h}{2} \sum_{j=1}^2 v_j(x) + o((1-e^{-m})h) \quad (5.4.4)$$

Démonstration :

Rappelons que N n'est pas une constante, mais une variable aléatoire, nous avons donc un double scénario stochastique, d'un côté le caractère aléatoire fourni par N et de l'autre le caractère aléatoire du processus ponctuel. Pour faire face à cela, nous utilisons la moyenne conditionnelle et nous considérons également certains outils liés aux séries de nombres réels. Considérons d'abord la moyenne,

$$\begin{aligned}\mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x)] &= \mathbb{E}[\mathbb{E}[\hat{\lambda}_{0,h}^{(1)}|N > 0]] = \sum_{n=1}^{\infty} \mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x)|N > 0]\mathbb{P}(N = n) \\ &= \sum_{n=1}^{\infty} \mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x)|N = n] \frac{e^{-m} m^n}{n!} = (1 - e^{-m})\mathbb{E}[\hat{\lambda}_{0,h}^1(x)|N = n].\end{aligned}$$

Donc nous avons :

$$\mathbb{E}[\hat{\lambda}_{0,h}^{(1)}(x)|N = n] = \int \lambda_0(x) \prod_{j=1}^2 K_{x_j, h_j}^*(t) dt = \mathbb{E}_\eta(\lambda_0(Y))$$

où les Y_j sont i.i.d de loi Bêta ($B(\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1)$)

Soient μ_j et σ_j^2 soit la moyenne et la variance de Y_j respectivement. D'après Johnson, Kotz et Balakrishnan (1994), nous pouvons montrer qu'il existe une constante M telle que :

$$\mu_{Y_j} = x_j + h_j(1 - 2x_j) + \Delta_1(x_j) \quad (5.4.5)$$

$$\sigma_{Y_j}^2 = b_j x_j(1 - x_j) + \Delta_2(x_j) \quad (5.4.6)$$

où $\Delta_j(x_j) \leq Mh_j^2$ pour $j = 1, 2$. Par conséquent, les autres termes des expressions ci-dessus pour μ_{Y_j} et $\sigma_{Y_j}^2$ sont uniformément $O(h^2)$.

Le noyau produit $\prod_{j=1}^2 K_{x_j, h_j}^*(t)$ est utilisé dans (5.2.5) comme fonction d'intensité. Pour cela, nous faisons le développement de Taylor de $\lambda_0(y)$ au point μ_j

$$\begin{aligned}\lambda_0(y) &= \lambda_0(\mu) + \sum_{j=1}^2 (y_j - \mu_j) \frac{\partial \lambda_0(y)}{\partial x_j} + \sum_{j=1}^2 (y_j - \mu_j)^2 \frac{\partial^2 \lambda_0(y)}{\partial x_j^2} \\ &\quad + \sum_{j \neq l} (y_j - \mu_j)(y_l - \mu_l) \frac{\partial \lambda_0(y)}{\partial x_j \partial x_l} + o\left(\sum_{j=1}^2 h_j\right).\end{aligned}$$

Nous pouvons écrire cette dernière équation par :

$$\begin{aligned}\mathbb{E}(\hat{\lambda}_{0,h}^{(1)}(x)) &= (1 - e^{-m})\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 h_j(1 - 2x_j)\lambda_0^j(x) + \frac{(1 - e^{-m})}{2} \sum_{j=1}^2 x_j h_j(1 - x_j)\lambda_0^{jj}(x) \\ &\quad + o((1 - e^{-m}) \sum_{j=1}^2 h_j^2)\end{aligned}$$

Par conséquent, le biais de l'estimation de la fonction d'intensité bivariée est donnée par :

$$\text{Bias}(\hat{\lambda}_{0,h}^{(1)}(x)) = -e^{-m}\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 u_j(x) + o((1 - e^{-m}) \sum_{j=1}^2 h^2) \quad (5.4.7)$$

où

$$u_j(x) = h_j \left((1 - 2x_j) \frac{\partial \lambda_0(x)}{\partial x_j} + \frac{x_j}{2} (1 - x_j) \frac{\partial^2 \lambda_0(x)}{\partial x_j^2} \right),$$

Quand $h_1 = h_2 = h$, alors nous avons

$$\int (\mathbb{E}(\hat{\lambda}_{0,h}^{(1)}(x)) - \lambda_0(x))^2 dx = \int \left(-e^{-m}\lambda_0(x) + (1 - e^{-m}) \sum_{j=1}^2 u_j(x) \right)^2 dx + o\left((1 - e^{-m}) \sum_{j=1}^2 h^2 \right). \quad (5.4.8)$$

□

Les variances sont données par :

Lemme 5.4.2.

$$\text{Var}(\hat{\lambda}_{0,h}^{(1)}(x)) = A(m) \left(\prod_{j=1}^2 A_{h_j}(x_j) \right) \lambda_0(x) + o(A(m) \sum_{j=1}^2 h_j) \quad (5.4.9)$$

et

$$\text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) = A(m) \left(\prod_{j=1}^2 A_{h_j}^*(x_j) \right) \lambda_0(x) + o(A(m) \sum_{j=1}^2 h_j) \quad (5.4.10)$$

Remarque 5.4.2. Si $h_1 = h_2 = h$ et $h \rightarrow 0$, $mh \rightarrow 0$ quand $m \rightarrow \infty$, alors la variance des estimateurs de la fonction intensité est égal :

$$\text{Var}(\hat{\lambda}_{0,h}^{(1)}(x)) \sim \text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) = \frac{A(m)}{h} \left(\prod_{j=1}^2 \frac{(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}} \right) \lambda_0(x) + o\left(A(m) \prod_{i=1}^2 h^{-1/2} \right) \quad (5.4.11)$$

Hypothèses de régularités

H.5.1 $h_j \rightarrow 0$ et $m^{-1} \prod_{j=1}^2 h_j^{-1/2} \rightarrow 0$, quand $m \rightarrow 0$;

H.5.2 $h_j \rightarrow 0$ et $\log(m)A(m) \prod_{j=1}^2 h_j^{-2} \rightarrow 0$, quand $m \rightarrow 0$.

Le résultat suivant indique l'erreur quadratique moyenne intégrée (*MISE*) de l'estimateur non paramétrique.

Théorème 5.4.1. *Supposons que $\lambda_0(x)$ soit deux fois différentiables. Soient $\hat{\lambda}_{0,h}^{(1)}(x)$ et $\hat{\lambda}_{0,h}^{(2)}(x)$ les estimateurs Bêta 1 et Bêta 2 de la fonction intensité. Sous l'hypothèse **H.5.1***

$$AMISE(\hat{\lambda}_{0,h}^{(1)}(x)) = (1 - e^{-m})^2 \int_W \left(\sum_{j=1}^2 h_j u_j(x) \right)^2 dx + A(m) \int_W \prod_{j=1}^2 h_j^{-1/2} \times (x_j(1-x_j))^{-1/2} \lambda_0(x) dx \quad (5.4.12)$$

$$AMISE(\hat{\lambda}_{0,h}^{(2)}(x)) = (1 - e^{-m})^2 \frac{1}{4} \int_W \left(\sum_{j=1}^2 x_j h_j (1-x_j) \lambda_0(x) \right)^2 dx + A(m) \int_W \prod_{j=1}^2 h_j^{-\frac{1}{2}} \times (x_j(1-x_j))^{-1/2} \lambda_0(x) dx \quad (5.4.13)$$

Démonstration :

Comme

$$A(m) = \mathbb{E} \left(\frac{1}{N} \mathbb{1}_{N \neq 0} \right)$$

et nous avons alors :

$$\mathbb{E} \left(\prod_{j=1}^2 (K_{x_j, h_j}^*)^2(X_j) \right) = \prod_{j=1}^2 A_{h_j}(x_j) \mathbb{E}(\lambda_0(Z)),$$

où γ_{x_j} est une variable aléatoire de loi $B\left(\frac{2x_j}{h_j} + 1, \frac{2(1-x_j)}{h_j} + 1\right)$.

$$A_{h_j}(x_j) = \frac{B\left(\frac{2x_j}{h_j} + 1, \frac{2(1-x_j)}{h_j} + 1\right)}{B\left(\frac{x_j}{h_j} + 1, \frac{(1-x_j)}{h_j} + 1\right)}.$$

Pour Chen (1999),

$$A_{h_j}(x_j) = \begin{cases} \frac{h_j^{-\frac{1}{2}} (x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}, & \text{si } \frac{x_j}{h_j} \text{ et } \frac{1-x_j}{h_j} \rightarrow \infty \\ \frac{h_j^{-1} \Gamma(2k_j+1)}{2^{2k_j+1} \Gamma^2(k_j+1)}, & \text{si } \frac{x_j}{h_j} \rightarrow k_j \text{ ou } \frac{1-x_j}{h_j} \rightarrow k_j. \end{cases}$$

$$Var(\hat{\lambda}_{0,h}^{(2)}(x)) = A(m) \prod_{j=1}^2 A_{h_j}(x_j) \mathbb{E}(\lambda_0(Z(x))) \quad (5.4.14)$$

où $Z = (Z_1, Z_2)$ et les variables aléatoires Z_j sont indépendantes et distribuées selon une loi Bêta.

Par le développement de Taylor,

$$\mathbb{E}(\lambda_0(Z)) = \lambda_0(x) + O(h).$$

Par conséquent,

$$\frac{1}{A(m)} \text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) = \lambda_0(x) \left(\prod_{j \in I} \frac{h^{-\frac{1}{2}}(x(1-x))^{-\frac{1}{2}}}{2\sqrt{\pi}} \right) \left(\prod_{j \in I^c} \frac{\Gamma(2k_j + 1)}{2^{2k_j+1}\Gamma^2(k_j + 1)} h_j^{-1} \right) \quad (5.4.15)$$

où $I = \{j, \frac{x_j}{h_j} \rightarrow \infty\}$ et I^c est le complémentaire de I .

Le deuxième produit disparaît dans la variance intégrée. Soit $\delta_j = h_j^{1-\epsilon}$, $0 < \epsilon < 1/2$ et $\delta = (\delta_1, \delta_2)$, alors

$$\begin{aligned} \int \text{Var}(\hat{\lambda}_{0,h}^{(1)}(x)) dx &= A(m) \int_0^\delta \text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) dx + \int_\delta^1 \text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) dx \\ &= O\left(\prod_{j \in I} h_j^{1/2-\epsilon} \prod_{j \in I^c} h_j^{-\epsilon}\right) + \int \left(\prod_{j \in I} \frac{h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}}\right) \lambda_0(x) dx \\ &= o\left(\prod_{j=1}^2 h_j^{-1/2}\right) + \frac{1}{2\sqrt{\pi}} \int \left(\prod_{j=1}^2 h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}\right) \lambda_0(x) dx \\ \int \text{Var}(\hat{\lambda}_{0,h}^{(2)}(x)) dx &= \frac{1}{2\sqrt{\pi}} \int \left(\prod_{j=1}^2 h_j^{-\frac{1}{2}}(x_j(1-x_j))^{-\frac{1}{2}}\right) \lambda_0(x) dx + o\left(\prod_{j=1}^2 h_j^{-1/2}\right) \end{aligned} \quad (5.4.16)$$

En combinant (5.4.8) et (5.4.16), nous obtenons l'erreur quadratique moyenne intégrée de l'estimateur non paramétrique avec l'intensité du noyau Bêta. \square

Le théorème 5.4.1 prouve que le taux de convergence du biais de l'estimateur non paramétrique est uniforme, donc exempt de biais aux limites. Le taux de convergence de l'erreur quadratique moyenne intégrée devient plus lent lorsque la dimension de la variable aléatoire augmente. Ceci est connu comme la malédiction de la dimensionnalité. Nous voyons que dans la région limite, la variance de l'estimateur produit par noyau est plus grande que la variance dans la région intérieure. Cependant, l'augmentation de la variance est compensée par un biais plus faible dans cette région. Loin des limites, nous avons l'effet inverse, c'est-à-dire une variance plus faible et un biais légèrement plus élevé. Heureusement, la deuxième dérivée de la fonction de densité est négligeable à partir de zéro.

Remarque 5.4.3. *Sous les conditions des lemmes 5.4.1 et 5.4.2, l'erreur quadratique moyenne intégrée MISE est donnée par :*

$$\begin{aligned} AMISE(\hat{\lambda}_{0,h}^{(1)}(x)) &= (1 - e^{-m})^2 \frac{h^2}{4} \int_W \left(\sum_{j=1}^2 u_j(x) \right)^2 dx + A(m) h^{-1} \\ &\quad \times \int_W \prod_{j=1}^2 (x_j(1-x_j))^{-1/2} \lambda_0(x) dx \end{aligned} \quad (5.4.17)$$

$$AMISE(\hat{\lambda}_{0,h}^{(2)}(x)) = (1 - e^{-m})^2 \frac{h^2}{4} \int_W \left(\sum_{j=1}^2 v_j(x) \right)^2 dx + A(m)h^{-1} \int_W \prod_{j=1}^2 (x_j(1-x_j))^{-1/2} \lambda_0(x) dx \quad (5.4.18)$$

En utilisant les résultats du (5.4.1), nous pouvons obtenir la borne supérieure du $MISE$ et trouver les expressions des paramètres de lissage optimales $\hat{h}^1 = (\hat{h}_1^1, \hat{h}_2^1)$ pour les estimateurs $\hat{\lambda}_{0,h}^{(1)}(x)$ et $\hat{h}^{(2)} = (\hat{h}_1^{(2)}, \hat{h}_2^{(2)})$ pour $\hat{\lambda}_{0,h}^{(2)}(x)$. Par conséquent, le théorème suivant :

Théorème 5.4.2. *Sous l'hypothèse H.5.1, les paramètres de lissage optimal sont donnés par :*

$$\hat{h}_1^{(1)} = \left[\frac{K^{\frac{3}{2}} \int U(x) dx}{4a[K \int u_1^2(x) dx + \int u_1(x)u_2(x) dx]} \right]^{\frac{1}{3}}, \quad \hat{h}_2^{(1)} = \left[\frac{K^{-\frac{3}{2}} \int U(x) dx}{4a[K \int u_1^2(x) dx + \int u_1(x)u_2(x) dx]} \right]^{\frac{1}{3}} \quad (5.4.19)$$

$$\hat{h}_1^{(2)} = \left[\frac{C^{\frac{3}{2}} \int U(x) dx}{4a[C \int v_1^2(x) dx + \int v_1(x)v_2(x) dx]} \right]^{\frac{1}{3}}, \quad \hat{h}_2^{(2)} = \left[\frac{C^{-\frac{3}{2}} \int U(x) dx}{4a[C \int v_1^2(x) dx + \int v_1(x)v_2(x) dx]} \right]^{\frac{1}{3}} \quad (5.4.20)$$

où

$$u_j(x) = (1 - 2x_j)\lambda_0^j(x) + x_j(1 - x_j)\lambda_0^{jj}, \quad v_j(x) = x_j(1 - x_j)\lambda_0^{jj}, \quad a = (1 - e^{-m})^2$$

$$U(x) = A(m) \prod_{j=1}^2 \frac{(x_j(1-x_j))^{-1/2}}{2\sqrt{\pi}}, \quad K = \frac{\int u_2(x) dx}{\int u_1(x) dx}, \quad C = \frac{\int v_2(x) dx}{\int v_1(x) dx}.$$

Démonstration :

Pour déterminer les éléments de la matrice diagonale, nous minimisons la fonction $AMISE$ donnée par :

$$AMISE(\hat{\lambda}_{0,h}^{(1)}(x)) = (1 - e^{-m})^2 \int_W \left(\sum_{j=1}^2 h_j u_j(x) \right)^2 dx + A(m)h_1^{-1/2}h_2^{-1/2} \int_W U(x) dx \quad (5.4.21)$$

En dérivant la fonction $AMISE$ par rapport à h_1 et h_2 , nous obtenons les équations suivantes :

$$c \left(2h_1 \int u_1^2(x) dx + 2h_2 \int u_1(x)u_2(x) dx \right) - \frac{1}{2h_1^{-\frac{3}{2}}h_2^{-\frac{1}{2}}} \int U(x) dx = 0 \quad (5.4.22)$$

$$c \left(2h_2 \int u_1^2(x) dx + 2h_1 \int u_1(x)u_2(x) dx \right) - \frac{1}{2h_1^{-\frac{1}{2}}h_2^{-\frac{3}{2}}} \int U(x) dx = 0 \quad (5.4.23)$$

En multipliant ces deux équations par h_1 et h_2 respectivement, nous obtenons :

$$c \left(2h_1^2 \int u_1^2(x) dx + 2h_1h_2 \int u_1(x)u_2(x) dx \right) - \frac{1}{2h_1^{-\frac{1}{2}}h_2^{-\frac{1}{2}}} \int U(x) dx = 0 \quad (5.4.24)$$

$$c \left(2h_2^2 \int u_1^2(x) dx + 2h_1h_2 \int u_1(x)u_2(x) dx \right) - \frac{1}{2h_1^{-\frac{1}{2}}h_2^{-\frac{1}{2}}} \int U(x) dx = 0 \quad (5.4.25)$$

Par soustraction des deux dernières équations, nous obtenons l'équation donnée par :

$$h_1 = h_2 \left[\frac{\int u_2(x) dx}{\int u_1(x) dx} \right]^{\frac{1}{2}} \quad (5.4.26)$$

Posons $K = \left[\frac{\int u_2(x) dx}{\int u_1(x) dx} \right]^{\frac{1}{2}}$.

En remplaçant l'équation (5.4.26) dans (5.4.24), nous avons

$$\hat{h}_1^1 = \left[\frac{\int U(x) dx}{4aK^{-\frac{3}{2}} [K \int u_1^2(x) dx + \int u_1(x)u_2(x) dx]} \right]^{\frac{1}{3}}, \quad \hat{h}_2^1 = \left[\frac{\int U(x) dx}{4aK^{\frac{3}{2}} [K \int u_1^2(x) dx + \int u_1(x)u_2(x) dx]} \right]^{\frac{1}{3}} \quad (5.4.27)$$

Une condition nécessaire pour que h_1 et h_2 minimise le *AMISE* est que

$$\frac{\partial^2 \lambda_0(x)}{\partial x_1^2} \frac{\partial^2 \lambda_0(x)}{\partial x_2^2} \geq 0 \quad (5.4.28)$$

Par les mêmes méthodes, nous calculons les paramètres de la matrice de lissage de l'estimateur $\hat{\lambda}_{0,h}^{(1)}(x)$. \square

Remarque 5.4.4. Sous les conditions des lemmes 5.4.1 et 5.4.2, les paramètres de lissage minimisant le MISE de $\hat{\lambda}_{0,h}^{(1)}$ et MISE de $\hat{\lambda}_{0,h}^{(2)}$ sont donnés par

$$\hat{h}^1 = \left[\frac{2A(m) \int_W U(x) \lambda_0(x) dx}{(1 - e^{-m})^2 \int_W \left(\sum_{j=1}^2 u_j(x) \right)^2 dx} \right]^{1/3}, \quad \hat{h}^2 = \left[\frac{A(m) \int_W U(x) \lambda_0(x) dx}{(1 - e^{-m})^2 \int_W \left(\sum_{j=1}^2 v_j(x) \right)^2 dx} \right]^{1/3}. \quad (5.4.29)$$

Le théorème suivant établit la consistance uniforme de l'estimateur d'intensité non paramétrique

Théorème 5.4.3. Soit λ_0 une fonction d'intensité sur $[0, 1]^2$ de $\hat{\lambda}_{0,h}^{(2)}$. Sous l'hypothèse **H.5.2**, on a

$$\sup_{x \in [0,1]^2} |\hat{\lambda}_{0,h}^{(2)}(x) - \lambda_0(x)| \xrightarrow{a.s} 0 \quad \text{quand } m \longrightarrow \infty \quad (5.4.30)$$

Démonstration :

Nous désignons par $\mu_x = (\mu_{x_1}, \mu_{x_2})$ où μ_{x_j} est la moyenne de la loi Bêta de paramètre $(\frac{x_j}{h_j} + 1, \frac{1-x_j}{h_j} + 1)$.

Comme λ_0 est continue,

$$\mu_{x_j} = x_j + h_j(1 - 2x_j)$$

et $h_j \longrightarrow 0$, nous avons, pour tout ϵ il existe δ tel que :

$$|\lambda_0(y) - \lambda_0(x)| < \epsilon, \quad \|y - x\| < \delta. \quad (5.4.31)$$

Nous commençons par :

$$\begin{aligned} \left| \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) - \lambda_0(x) \right| &\leq \int_{\|y-x\| \leq \delta} (K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) |\lambda_0(y) - \lambda_0(x)| dy \\ &\quad + \int_{\|y-x\| \geq \delta} (K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) |\lambda_0(y) - \lambda_0(x)| dy \\ &= D_1 + D_2. \end{aligned}$$

À partir de l'équation (5.4.31)

$$D_1 < \epsilon \int_{\|y-x\| \leq \delta} (K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) |\lambda_0(y) - \lambda_0(x)| dy \quad (5.4.32)$$

D'autre part, en utilisant l'inégalité de Chebyshev et que

$$\text{Var}(K_{h_j, y_j}(x_j)) = h_j x_j (1 - x_j),$$

nous obtenons :

$$\begin{aligned}
D_2 &\leq 2 \sup_{x \in [0,1]^2} |\lambda_0(x)| \int_{\|y-x\| \geq \delta} (K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) dy \\
&= 2 \sup_{x \in [0,1]^2} |\lambda_0(x)| \prod_{j=1}^2 (x_j h_j (1 - x_j)) \\
&= o(1). \\
D_2 &= o(1). \tag{5.4.33}
\end{aligned}$$

Par conséquent, d'après (5.4.32) et (5.4.33)

$$\left| \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) - \lambda_0(x) \right| \longrightarrow 0.$$

Il reste maintenant à prouver que le terme $\left| \mathbb{E} \left(\lambda_0(x) - \hat{\lambda}_{0,h}^{(2)}(x) \right) \right|$ converge presque sûrement vers zéro. Utilisant l'intégration par parties,

$$\begin{aligned}
|\hat{\lambda}_{0,h}^{(2)}(x) - \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x))| &= \left| \int (K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) d(F_N(t) - F(t)) \right| \\
&\leq \left| \int (F_N(y) - F(y)) d(K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) \right| \\
&\leq \sup_{x \in [0,1]^2} |(F_N(y) - F(y))| \left| \int d(K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) \right|.
\end{aligned}$$

Nous pouvons voir que :

$$\left| \int d(K_{h_1, y_1}(x_1) K_{h_2, y_2}(x_2)) \right| \leq 4(h_1 h_2)^{-1}.$$

Par conséquent, et d'après Kiefer (1961),

$$\begin{aligned}
\mathbb{P} \left(\left| \hat{\lambda}_{0,h}^{(2)}(x) - \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) \right| > \epsilon \right) &\leq \mathbb{P} \left(\sup_{x \in [0,1]^2} |(F_m(x) - F(x))| > \epsilon 2^{-2} h_1 h_2 \right) \\
&= C(2) \exp \left(-c 2^{-4} \epsilon^2 A(m)^{-1} (h_1 h_2)^2 \right).
\end{aligned}$$

Prenons $\epsilon_m = \frac{\alpha}{\sqrt{c}} \sqrt{A(m) \log(m)} (h_1 h_2)^{-2}$ avec $\alpha \geq 2^2$, $c < 2$.

Ceci implique que :

$$\sum_m \exp \left(-c 2^{-4} A(m)^{-1} (h_1 h_2)^2 \right) < \infty \tag{5.4.34}$$

Par conséquent, $\left| \mathbb{E} \left(\lambda_0(x) - \hat{\lambda}_{0,h}^{(2)}(x) \right) \right|$ converge presque sûrement. \square

Le théorème suivant traite de la normalité asymptotique de l'estimateur d'intensité non paramétrique de l'estimateur à noyau Bêta.

Théorème 5.4.4. *Supposons que λ_0 soit deux fois différentiables par rapport à x . Supposons également que les paramètres de lissage satisfont*

$$h_j = k_j A(m)^{1/2},$$

nous avons

$$\sigma^{*-1} A(m)^{-\frac{1}{2}} \prod_{j=1}^2 h_j^{\frac{1}{4}} (\hat{\lambda}_{0,h}^{(2)}(x) - \lambda_0 - \mu^*) \longrightarrow \mathcal{N}(0, 1) \quad (5.4.35)$$

où

$$\sigma^*(x) = \begin{cases} \left(\frac{h_j^{-\frac{1}{2}} (x_j(1-x_j))^{-\frac{1}{2}}}{2\sqrt{\pi}} \right)^{\frac{1}{2}}, & \text{si } \frac{x_j}{h_j} \text{ et } \frac{1-x_j}{h_j} \rightarrow \infty \\ \left(\frac{h_j^{-1} \Gamma(2c_j+1)}{2^{2c_j+1} \Gamma^2(c_j+1)} \right)^{\frac{1}{2}}, & \text{si } \frac{x_j}{h_j} \rightarrow c_j \text{ ou } \frac{1-x_j}{h_j} \rightarrow c_j. \end{cases} \quad (5.4.36)$$

et

$$\mu^* = \sum_{j=1}^2 h_j B_j(x), \quad B_j = \begin{cases} x_j(1-x_j)\lambda_0^{jj}(x), & \text{si } x_j \in [2h_j, 1-2h_j] \\ \xi_{h_j}(x_j)\lambda_0^j(x), & \text{si } x_j \in [0, 2h_j), \\ -\xi_{h_j}(1-x_j)\lambda_0^j(x), & \text{si } x_j \in (1-2h_j, 1] \end{cases} \quad (5.4.37)$$

Démonstration :

Nous commençons par la décomposition classique en utilisant l'expression du biais asymptotique :

$$\begin{aligned} \left(\hat{\lambda}_{0,h}^{(2)}(x) - \lambda_0(x) \right) &= \left(\hat{\lambda}_{0,h}^{(2)}(x) - \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) \right) + \left(\mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) - \lambda_0(x) \right) \\ &= \left(\hat{\lambda}_{0,h}^{(2)}(x) - \mathbb{E}(\hat{\lambda}_{0,h}^{(2)}(x)) \right) + \sum_{j=1}^2 h_j B_j(x) + O\left(\sum_{i=1}^2 h_j^2 \right). \end{aligned}$$

Par conséquent, nous obtenons $h_j = O(A(m)^{\frac{1}{2}})$

$$\sigma^{*-1} A(m)^{-\frac{1}{2}} \prod_{j=1}^2 h_j^{\frac{1}{4}} (\hat{\lambda}_{0,h}^{(2)}(x) - \lambda_0(x) - \mu^*) = \sum_{i=1}^N Y_i + O\left(A(m)^{\frac{1}{2}} \right),$$

où

$$Y_i = \sigma^{*-1} A(m)^{\frac{1}{2}} \prod_{j=1}^2 h_j^{\frac{1}{4}} (K(h_1, X_{i1})(x_1)K(h_2, X_{i2})(x_2) - \mathbb{E}(K(h_1, X_{i1})(x_1)K(h_2, X_{i2})(x_2))).$$

Maintenant, nous appliquons le théorème central limite de Liapunov pour prouver la normalité asymptotique de

$$S_m = \sum_{i=1}^m Y_i.$$

À partir de l'équation (5.4.16),

$$\text{Var}(S_m) = 1 + o(1).$$

Utilisant (5.4.16) et que $h_j = O(A(m)^{\frac{1}{2}})$,

$$\mathbb{E}(|Y_i|^3) = o(A(m)).$$

Par conséquent,

$$S_m = \sum_{i=1}^m Y_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \square$$

5.5 Étude simulation

Dans cette section, nous étudions les propriétés des estimateurs d'intensité non paramétrique pour les données bivariées avec supports bornés. Nous comparons le biais, la variance et l'erreur quadratique moyenne intégrée des estimateurs en utilisant les noyaux produits Bêta 1, Bêta 2 (Bêta modifié), gaussien. Nous considérons quatre fonctions intensités données par :

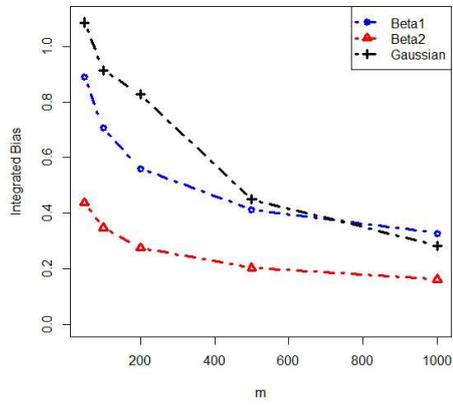
1. $\lambda_{11}(x) = b \exp(-2x_1 - x_2)$, $x = (x_1, x_2) \in [0, 1]^2$, utilisant plusieurs valeurs de b ($b = 183$ ou 365 ou 730 ou 1825 ou 3650) afin d'obtenir $m = \int_W \lambda(x) dx \approx 50$ ou 100 ou 200 ou 500 ou 1000 respectivement.
2. $\lambda_{22}(x) = b \Phi_{(0.3-0.2x_2, 0.02)}(x_1) + 25$, $x = (x_1, x_2) \in [0, 1]^2$, où Φ est la densité de la loi normale de moyenne $\mu = 0.3 - 0.2x_2$ et écart-type $\sigma = 0.02$, utilisant plusieurs valeurs de b ($b = 50$ ou 100 ou 200 ou 500 ou 1000) afin d'obtenir $m = \int_W \lambda(x) dx \approx 50$ ou 100 ou 200 ou 500 ou 1000 respectivement.
3. $\lambda_{33}(x) = b(\cos(\sqrt{(x_1 - 3)^2(x_2 - 3)^2}))^2$, $x = (x_1, x_2) \in [0, 1]^2$, utilisant plusieurs valeurs de b ($b = 127$ ou 330 ou 254 ou 634 ou 1267) afin d'obtenir $m = \int_W \lambda(x) dx \approx 50$ ou 100 ou 200 ou 500 ou 1000 respectivement.
4. $\lambda_{44}(x) = b(x_1^2 + x_2^3)$, $x = (x_1, x_2) \in [0, 1]^2$, utilisant plusieurs valeurs de b ($b = 86$ ou 172 ou 344 ou 860 ou 1720) afin d'obtenir $m = \int_W \lambda(x) dx \approx 50$ ou 100 ou 200 ou 500 ou 1000 respectivement.

Dans les simulations, nous considérons les tailles d'échantillon 50, 100, 200, 500 et 1000 et nous effectuons pour chaque modèle. Dans chaque réplcation, la matrice de lissage est choisie de telle sorte que l'erreur quadratique intégrée soit minimisée.

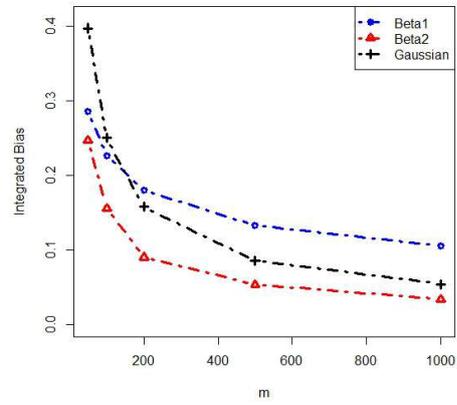
Le biais intégré, la variance intégrée et l'erreur quadratique moyenne intégrée ont été

calculés comme des mesures de performance pour chacun des estimateurs. Les résultats sont résumés dans les figures 5.2, 5.3 et 5.4. La simulation a montré que l'estimateur du noyau Bêta 1 $\hat{\lambda}_{0,h}^{(1)}$ a toujours un biais et une variance plus grands que $\hat{\lambda}_{0,h}^{(2)}$. Le biais, la variance et l'erreur quadratique moyenne intégrée diminuent tous lorsque la taille de l'échantillon augmente. Et les trois mesures semblaient converger à mesure que les m augmentent. Nous observons sur la figure 5.2 que l'estimateur du noyau Bêta 2 $\hat{\lambda}_{0,h}^{(2)}$ a la plus petite variance intégrée pour tous les cas considérés.

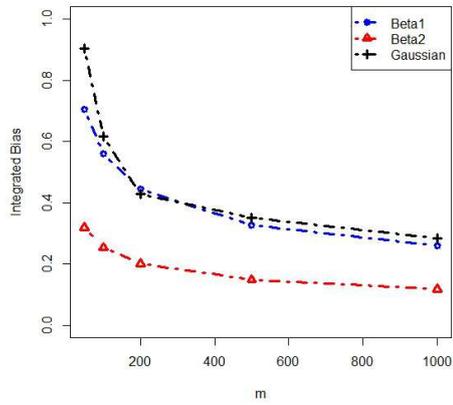
En combinant le biais au carré intégré et la variance intégrée donnée dans les tableaux 5.1 et 5.2, nous avons obtenu les erreurs intégrées moyennes minimisées illustrées à la figure 5.4. La figure 5.4 montre que l'estimateur à noyau Bêta 2 a la plus petite erreur quadratique moyenne intégrée pour toutes les fonctions intensités considérées.



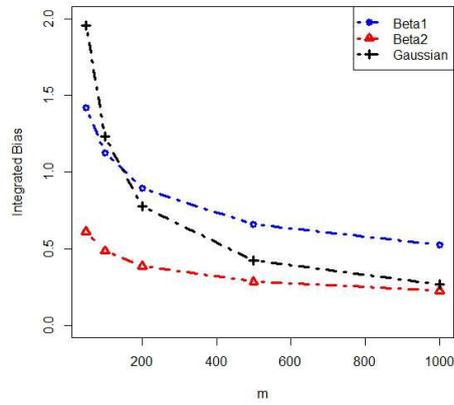
(a) λ_{11}



(b) λ_{22}

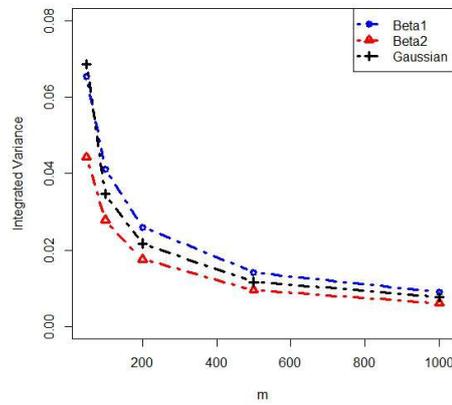


(c) λ_{33}

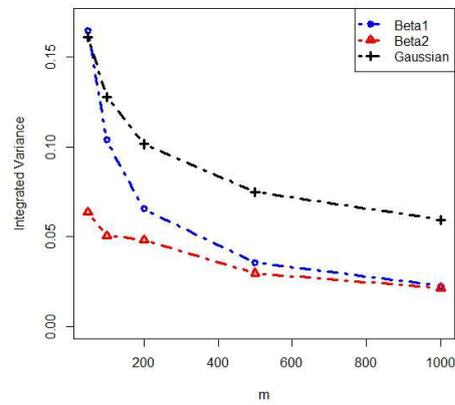


(d) λ_{44}

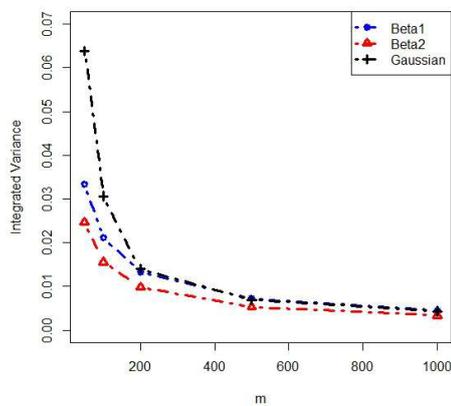
FIGURE 5.2 – Graphe du biais intégré pour les 4 fonctions intensités.



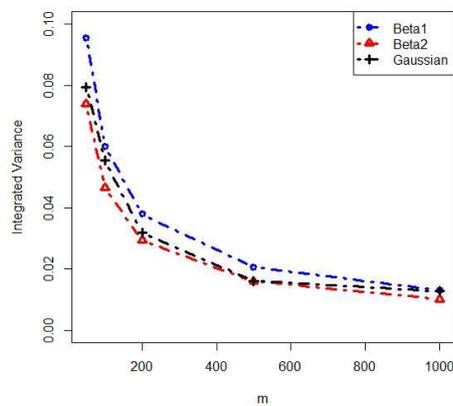
(a) λ_{11}



(b) λ_{22}



(c) λ_{33}



(d) λ_{44}

FIGURE 5.3 – Graphe de la variance intégrée pour les quatre fonctions intensités.

Pour la figure 5.3, nous remarquons que le biais des estimateurs est décroissant pour tous les modèles et toutes les valeurs de m . Nous remarquons que le noyau Bêta modifié a le plus petit biais que le noyau Bêta et le noyau gaussien. Mais aussi le biais de l'estimateur d'intensité du noyau Bêta est plus petit que l'estimateur du noyau gaussien pour les petites valeurs m et plus grand pour les grandes valeurs m .

TABLE 5.1 – Biais et variance intégrés pour les fonctions intensités $\lambda_{11}(x)$, $\lambda_{22}(x)$, $\lambda_{33}(x)$ et $\lambda_{44}(x)$ respectivement.
$$\lambda_{11}(x)$$

Kernel	Bêta 1		Beta2		Gaussian	
	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>
m						
50	0.890175	0.065398	0.438080	0.044296	1.086967	0.068601
100	0.706532	0.041198	0.347705	0.027904	0.914707	0.034763
200	0.560775	0.025953	0.275973	0.017579	0.828213	0.021718
500	0.413182	0.014089	0.203339	0.009543	0.449623	0.011633
1000	0.327943	0.008875	0.161390	0.006011	0.283245	0.007685

$$\lambda_{22}(x)$$

Kernel	Bêta 1		Beta2		Gaussian	
	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>
m						
50	0.285453	0.164689	0.246807	0.063492	0.397468	0.160992
100	0.226564	0.103747	0.155479	0.050393	0.250389	0.127779
200	0.179824	0.065356	0.089997	0.047945	0.157735	0.101418
500	0.132495	0.035481	0.053173	0.029470	0.085631	0.074726
1000	0.105162	0.022351	0.033496	0.021390	0.053944	0.059310

$$\lambda_{33}(x)$$

Kernel	Bêta 1		Beta2		Gaussian	
	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>
m						
50	0.704867	0.033462	0.318732	0.024667	0.902631	0.063814
100	0.559453	0.021080	0.252978	0.015539	0.616783	0.030649
200	0.444038	0.013279	0.200789	0.009789	0.429521	0.014020
500	0.327170	0.007209	0.147942	0.005314	0.350333	0.006961
1000	0.259675	0.004541	0.117422	0.003347	0.283692	0.004350

$$\lambda_{44}(x)$$

Kernel	Bêta 1		Bêta 2		Gaussian	
	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>	<i>Biais</i>	<i>Variance</i>
m						
50	1.419621	0.095508	0.611610	0.073895	1.959342	0.079393
100	1.126754	0.060166	0.485435	0.046551	1.234308	0.055387
200	0.894305	0.037902	0.385290	0.029325	0.777565	0.031857
500	0.658929	0.020576	0.283884	0.015920	0.422127	0.016104
1000	0.522992	0.012962	0.225319	0.010029	0.265923	0.012782

TABLE 5.2 – Biais intégré, la variance intégrée et $AMISE$ pour les fonctions intensité $\lambda_{11}(x)$, $\lambda_{22}(x)$, $\lambda_{33}(x)$ et $\lambda_{44}(x)$ respectivement.
$$\lambda_{11}(x)$$

Noyau	Bêta 1		Bêta 2		Gaussian	
	$Biais^2$	$MISE$	$Biais^2$	$MISE$	$Biais^2$	$MISE$
50	11.368660	11.43058	0.256379	0.300675	6.187647	6.256248
100	7.161809	7.203007	0.161509	0.189413	2.455569	2.490332
200	4.511657	4.537610	0.101744	0.119323	0.974493	0.996211
500	2.449304	2.463393	0.055235	0.064778	0.287205	0.298838
1000	1.542965	1.551840	0.034796	0.040807	0.113977	0.121662

$$\lambda_{22}(x)$$

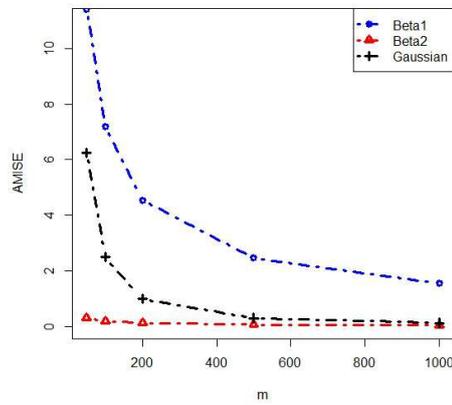
Noyau	Bêta 1		Bêta 2		Gaussian	
	$Biais^2$	$MISE$	$Biais^2$	$MISE$	$Biais^2$	$MISE$
50	13.96110	14.125790	1.110869	1.174361	22.17056	22.33155
100	8.794944	8.898691	0.699803	0.750196	8.798394	8.926173
200	5.540467	5.605823	0.440848	0.488793	3.491645	3.593063
500	3.007828	3.043309	0.239329	0.268799	1.029066	1.103792
1000	1.894813	1.917164	0.150768	0.172158	0.408385	0.467695

$$\lambda_{33}(x)$$

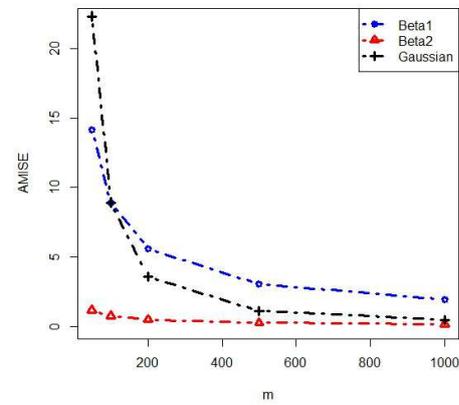
Noyau	Bêta 1		Bêta 2		Gaussian	
	$Biais^2$	$MISE$	$Biais^2$	$MISE$	$Biais^2$	$MISE$
50	2.842427	2.875889	0.137356	0.162023	5.564476	5.62829
100	1.790617	1.811697	0.086529	0.102068	2.208264	2.238913
200	1.128018	1.141297	0.054509	0.064298	0.876350	0.890370
500	0.612382	0.619591	0.029592	0.034906	0.258280	0.265241
1000	0.385776	0.390317	0.018642	0.021989	0.102498	0.106848

$$\lambda_{44}(x)$$

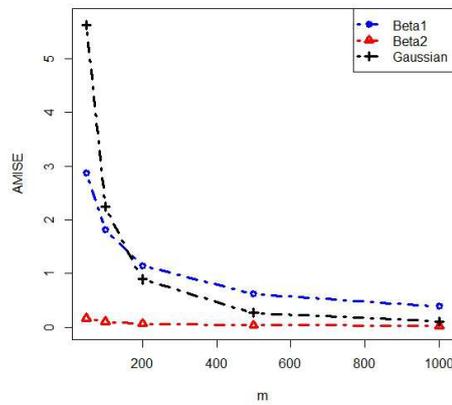
Noyau	Beta1		Beta2		gaussien	
	$Biais^2$	$MISE$	$Biais^2$	$MISE$	$Biais^2$	$MISE$
50	8.099679	8.195187	0.427505	0.501400	4.043768	4.123161
100	5.102478	5.162644	0.269311	0.315862	1.604770	1.660157
200	3.214360	3.252262	0.169655	0.198980	0.636853	0.668710
500	1.745023	1.765599	0.092103	0.108023	0.187695	0.203799
1000	1.099296	1.112258	0.058021	0.068050	0.074486	0.087268



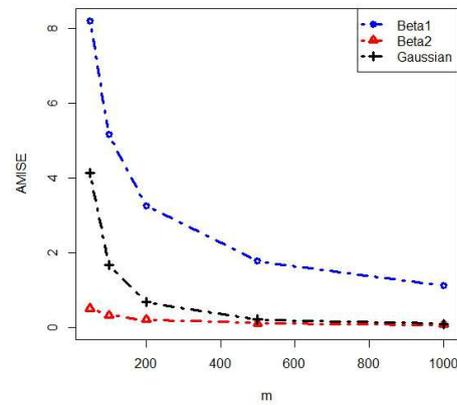
(a) λ_{11}



(b) λ_{22}



(c) λ_{33}



(d) λ_{44}

FIGURE 5.4 – Graphe de *AMISE* des quatre fonctions intensités

En termes de biais intégré, l'estimateur du noyau Bêta modifié est plus performant que les estimateurs du noyau Bêta et gaussien, et le noyau Bêta est meilleur que l'estimateur Gaussien pour les grandes valeurs m et plus que pour les petites valeurs m . Par exemple, pour $m = 500$, le biais pour le noyau Bêta 1, Bêta 2 et gaussien est de 0, 203339, 0, 413182 et 0, 449623 respectivement pour λ_{11} .

En termes de variance, les estimateurs ont également presque les mêmes performances. Par

exemple, pour $m = 500$, la variance intégrée pour les noyaux Bêta 1, Bêta 2 et gaussien est de 0,005314, 0,007209 et 0,006961, respectivement pour la fonction intensité λ_{33} .

5.6 Application à des données réelles

Nous utilisons ici l'ensemble de données d'une enquête sur une parcelle de $924ft \times 924ft$ (19,6 acres) à Lansing Woods, comté de Clinton, Michigan USA par D.J. Gerrard. Les données donnent l'emplacement de 2251 arbres et leur classification botanique (en caryer, érable, chêne rouge, chêne blanc, chêne noir et arbres divers). La taille originale de la parcelle (924×924 pieds) a été redimensionnée au carré de l'unité. Nous considérons trois espèces particulières d'arbres : les caryers (703 arbres), les érables (514 arbres) et les chênes rouges (346 arbres).

Nous avons obtenu des estimations d'intensité pour ces trois espèces d'arbres en utilisant le noyau gaussien et le noyau Bêta, ce qui montre nos résultats pour les essences de caryer, d'érable et de chêne rouge, respectivement.

Les figures 5.5, 5.6 et 5.7 affichent les tracés des emplacements pour les arbres et les contours pour les estimations d'intensité non paramétriques. Nous avons obtenu des estimations d'intensité pour ces trois espèces en utilisant les estimateurs d'intensité du noyau, le noyau Bêta modifié et le noyau Gaussien. Les figures 5.5 à 5.7 montrent nos résultats pour les essences de chêne de caryer, érable et chênes rouge, respectivement.

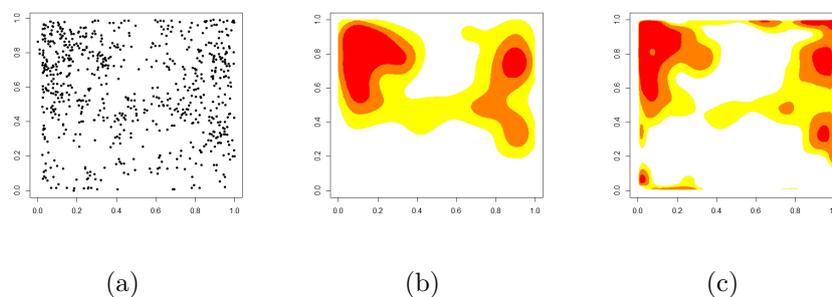


FIGURE 5.5 – (a) processus de Poisson non homogène, (b) contour pour l'estimateur de noyau Gaussien, (c) contour pour l'estimateur de noyau Bêta avec des arbres caryer.

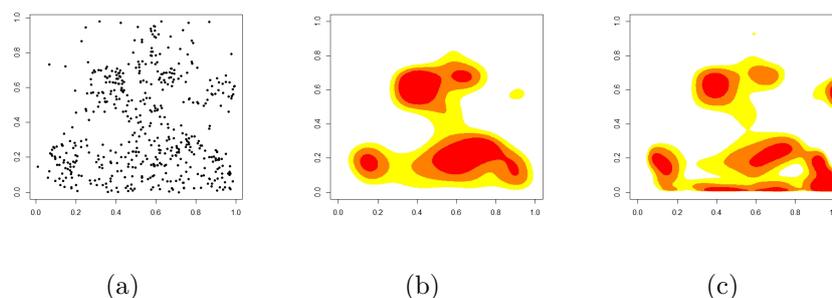


FIGURE 5.6 – (a) processus de Poisson inhomogène, (b) contour pour l'estimateur de noyau Gaussien, (c) contour pour l'estimateur du noyau Bêta avec des érables.

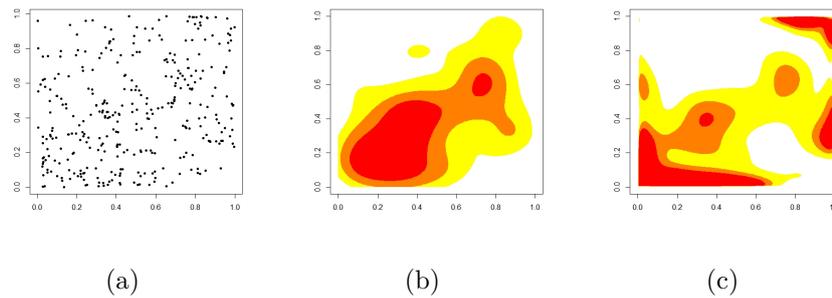


FIGURE 5.7 – (a) processus de Poisson inhomogène, (b) contour pour l’estimateur de noyau gaussien, (c) contour pour l’estimateur de noyau Bêta avec des chênes rouges.

5.7 Conclusion

Ce chapitre propose un estimateur non paramétrique pour les fonctions intensités du processus de Poisson non homogène bivarié utilisant le noyau Bêta. L’estimateur est basé sur le noyau Bêta lorsque le support est un ensemble compact. En utilisant des noyaux symétriques, les estimateurs de la fonction d’intensité sont robustes au problème des limites. Nous fournissons les propriétés asymptotiques de l’estimateur et montrons que le taux optimal de convergence de l’erreur quadratique moyenne intégrée est obtenu. Nous examinons la performance de l’échantillon fini dans quatre simulations. En fin, nous constatons que les estimateurs que nous proposons fonctionnent presque aussi bien que l’estimateur Gaussien lorsqu’il n’y a pas de problèmes de limites.

Conclusion et perspectives

Cette thèse porte sur l'estimation non paramétrique de la fonction intensité du processus de Poisson non homogène, en utilisant la méthode du noyau classique et associé. Nous avons présenté d'abord l'état de l'art sur l'estimateur à noyau classique de la fonction intensité bivariée continue et ses propriétés statistiques.

Par la suite, l'estimateur standard à noyau associé Bêta multivarié dans le cas continu a été défini. Il a été montré que cet estimateur est sans effet de bord, mais possède un biais plus grand. En effet, le nombre de termes du biais de cet estimateur à noyau Bêta est supérieur à celui à noyau classique. Pour corriger cela, un algorithme de réduction du biais consiste à partitionner le support W en deux régions (de bord et la plus grande partie d'intérieur) puis à modifier le noyau associé de départ de sorte que le biais de l'estimateur issu de ce noyau soit de même ordre que celui de l'estimateur à noyau classique multivarié.

Les différentes techniques de sélection de la matrice de lissage ont été rappelées. Nous avons proposé par la suite la forme de l'estimateur à noyau associé continu multivarié. Les propriétés statistiques de cet estimateur ont été présentées, ainsi que deux méthodes classiques pour le choix de la matrice de lissage. Par ailleurs, nous avons présenté l'alternative bayésienne pour le choix de la matrice de lissage. Cette matrice est considérée comme une variable aléatoire de loi a priori $\pi(\cdot)$. L'estimateur bayésien de H est obtenu à partir de la loi a posteriori sous les pertes quadratique et entropie. Nous avons proposé dans ce cadre bayésien deux approches pour le choix de la matrice de lissage H .

Cette approche a été comparée avec les méthodes classiques validation croisée et *AMISE* en utilisant le critère *ISE*. Les résultats obtenus sur des données simulées et sur des données réelles montrent l'avantage de l'approche bayésienne locale par rapport à la méthode validation croisée et *AMISE*. Nous avons aussi estimé la matrice de lissage par l'approche bayésienne adaptative. Nous avons aussi pu dériver explicitement les formes exactes de la loi a posteriori et de l'estimateur de la matrice de lissage H grâce à la conjugativité entre le noyau utilisé et la loi a priori choisie. Les résultats numériques obtenus sur des données simulées et sur des données réelles montrent que l'approche bayésienne adaptative en utilisant les deux fonctions (perte quadratique et perte entropie) sont meilleures que la méthode classique *LSCV*. Enfin une comparaison entre les deux approches bayésiennes a été illustrée sur des données simulées et réelles. Les résultats obtenus montrent que l'approche bayésienne adaptative est meilleure que l'approche bayésienne locale. L'approche bayésienne adaptative est meilleure que l'approche bayésienne locale en temps d'exécution.

Enfin, nous avons fait une comparaison entre le noyau gaussien et les noyaux Bêta et Bêta modifié en utilisant le biais, la variance et le *MISE* comme critères d'erreurs de

comparaison. Ces critères montrent que le noyau Bêta modifié est meilleur que le noyau gaussien pour toutes les fonctions intensités considérées.

Les travaux réalisés offrent des perspectives très intéressantes. On peut citer :

- Refaire les travaux présentés dans le cas d'une matrice H symétrique définie positive pleine,
- On peut envisager estimer la fonction intensité en utilisant d'autres noyaux différents des noyaux gaussiens et Bêta,
- Introduire l'approche semi-paramétrique pour l'estimation de la fonction intensité du premier ordre,
- Estimer l'intensité du premier ordre dans le cas de données fonctionnelles,
- L'idée de construction d'une version récursive de l'estimateur à noyau de la fonction intensité est de même souhaitable.
- Puisque les matrices des fenêtres jouent un rôle crucial dans la qualité des différents lissages, il serait intéressant d'étudier la performance du choix bayésien local et adaptatif en utilisant des matrices pleines en utilisant des lois à priori non conjuguées.
- Proposer une méthode d'estimation de la fonction intensité qui se base sur deux noyaux. L'un des noyaux prend en compte la structure spatiale par la mesure de la proximité entre les événements et l'autre contrôle la distance entre les événements.
- Deux types d'asymptotiques ont été considérés dans le Chapitre 2. Il est naturel de regarder le cas où l'un des domaines est fixé (par exemple la région est fixée) et l'autre évolue.
- Refaire les travaux réalisés dans le cas des fonctions intensités discontinues.

Bibliographie

- [1] Abdous, B. and Kokonendji, C. C. (2009). Consistency and asymptotic normality for discrete associated-kernel estimator. *African Diaspora Journal of Mathematics* 8, 63–70.
- [2] Baddeley, A. J., Moller, J. & Waagepetersen, R. P. (2000). Non and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statist. Neer.* 54, 329–50.
- [3] Baddeley, A. and Turner, R. (2005). Spatstat : an R package for analyzing spatial point patterns. *J. Stat. Soft.* 12, (6), 1-42.
- [4] Badiane, M. S, Ngom, P and Manga. C. (2021). Bayesian selection of local bandwidth in non-homogeneous poisson process kernel estimators for the intensity function. *Far East Journal of Theoretical Statistics*, 61(2) :109–144.
- [5] Barr, C. D. and Schoenberg, F. P. (2010). On the voronoi estimator for the intensity of an inhomogeneous planar poisson process. *Biometrika*, 97(4) :977-984.
- [6] Belaid, N., Adjabi, S., Kokonendji, C. C. and Zougab, N. (2016). Bayesian local bandwidth selector in multivariate associated kernel estimators for joint probability mass functions. *Journal of Statistical Computation and Simulation.* 86, 3667–3681.
- [7] Belaid, N.S. Adjabi, N.S., Kokonendji, C. C. and Zougab, N. (2018). Bayesian adaptive bandwidth selector for multivariate discrete kernel estimator. *Communications in Statistics-Theory and Methods*, 47(12) :2988–3001.
- [8] Berman, M. & Diggle, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* 51, (1), 81–92.
- [9] Bernardeau, F. & Van Deweygaert, R. (1996). A new method for accurate estimation of velocity field statistics. *Mon. Not. R. Astron. Soc.* 279, 693–711.
- [10] Bouezmarni, T. and Rolin, J. M. (2003). Consistency of the beta kernel density function estimator. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 89–98.
- [11] Bouezmarni, T. and Roumbouts, J. V. K. (2010). Non-parametric density estimation for multivariate bounded data. *Journal of Statistical Planning and Inference* 140, 139–152.
- [12] Bouezmarni, T., Rombouts, J., (2006). Nonparametric density estimation for positive time series. Technical report, Université catholique de Louvain, Center for Operations Research.
- [13] Bouezemarni, T., Karunamuni, R.J. & Alberts, T. (2007). Nonparametric density estimation for multivariate bounded data. Technical report, HEC Montréal, Institut d'économie appliquée.

- [14] Bouezemarni, T. & Scaillet, O. (2005). Consistency of Asymmetric kernel density estimators and smoothed histograms with application to income data. *Econometric Theory* 21, 390-412.
- [15] Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2) :353–360.
- [16] Brewer, M. J. (1998) . A modelling approach for bandwidth selection in kernel density estimation. In : *Proceedings of COMPSTAT Physica Verlag, Heidelberg*, pp :203–208.
- [17] Brewer, M.J (2000). A Bayesian model for local smoothing in kernel density estimation. *Statistics and Computing*, 10 :299–309.
- [18] Brooks, M. M. and Marron, J. S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Process. Appl.* 38, (1), 157-165.
- [19] Cacoullos, T. (1966). Estimation of a multivariate density. *Annals of the institute of statistical Mathematics*, 18, 179–189.
- [20] Calabria, R. and Pulcini, G. (1994). An engineering approach to bayes estimation for the weibull distribution. *Microelectronics Reliability*, 34(5) :789–802.
- [21] Cao, R. (1993). Bootstrapping the mean integrated squared error. *J. Multivariate Anal.* 45, (1), 137–160.
- [22] Cao, R., Cuevas, A. & González Manteiga,W. (1994). A comparative study of several smoothing methods in density estimation. *Comput. Statist. Data Anal.* 17, (2), 153–176.
- [23] Chacón,J. E.& Duong, T.(2011). Unconstrained pilot selectors for smoothed cross validation, *Australian and New Zealand Journal of Statistics* 53, 331–351.
- [24] Chen, S. (1999) : A Beta Kernel Estimation for Density Functions, *Computational Statistics and Data analysis*, 31, 131-145.
- [25] Chen, S.X. (2000). Beta kernels smoothers for regression curves, *Statistica Sinica* 52, 73–91.
- [26] Chen, S. (2000) : Probability Density Functions Estimation Using Gamma Kernels, *Annals of the Institute of Statistical Mathematics*, 52, 471–480.
- [27] Cheng, M., Fan, J., Marron, J., (1997). On automatic boundary corrections. *Annals of Statistics* 25, 1691–1708.
- [28] Cline, D. B. H. & Hart, J. D. (1991). Kernel estimation of densities with discontinuities or discontinuous derivatives. *Statistics : A Journal of Theoretical and Applied Statistics*, 22(1), 69-84.
- [29] Cowling, A., Hall, P. & Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a poisson point process. *J. Amer. Statist. Assoc.* 91, (436), 1516–1524.
- [30] Cressie, Noel A. C. (1993). *Statistics for spatial data*, (Revised)., Wiley-Interscience, New York.
- [31] Cronie, O. & Van Lieshout, M. N. M. (2016). Summary statistics for inhomogeneous marked point processes. *Ann. Inst. Statist. Math.* 68, 905–28.
- [32] Cronie, O., Van Lieshout, M .N. M. (2018) A non-model based approach to bandwidth selection for kernel estimators of spatial intensity functions. *Biometrika* 105 :455–462.

- [33] Cucala, L. (2006). *Espacements bidimensionnels et données entachées d'erreurs dans l'analyse des processus ponctuels spatiaux*, Ph.D. Thesis, Université des Sciences Sociales, Toulouse I.
- [34] Cucala, L. (2008). Intensity estimation for spatial point processes observed with noise. *Scand. J. Stat.* 35, (2), 322–334.
- [35] Daley, D. J. and Vere-Jones, D. (1988). An introduction to the theory of point processes. Springer Verlag.
- [36] De Lima, M. S. and Atuncar, G. S. (2011). Unconstrained pilot selectors for smoothed crossvalidation. *Journal of Nonparametric Statistics*, 23, 137–148.
- [37] Deheuvels, P. (1977). estimation non paramétrique de la densité par histogrammes généralisées. *The Annual Review of Statistics and Its Application*, 25, 5–42.
- [38] Diggle, P. (1985). A kernel method for smoothing point process data. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 34, (2), 138–147.
- [39] Diggle, P. and Marron, J. S. (1988). Equivalence of smoothing parameter selectors in density and intensity estimation. *J. Amer. Statist. Assoc.* 83, 793–800.
- [40] Diggle, P. J. (1990). A point process modelling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 153(3), 349–362.
- [41] Diggle, P. J. (2003). *Statistical analysis of spatial point patterns*, (2nd ed.), A Hodder Arnold Publication, Oxford, UK.
- [42] Diggle, P. J. (2014). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Boca Raton, Florida : Taylor & Francis/CRC Press, 3rd ed.
- [43] Diggle, P. J., Besag, J., and Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics*, pages 659–667.
- [44] Duong, T. (2013). ks : *Kernel smoothing*. R package version 1.8.13.
- [45] Duong, T. (2004). *Bandwidth selectors for multivariate kernel density estimation*. University of Western Australia.
- [46] Duong, T. and Hazelton, M. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *J. Nonparametr. Stat.* 15, (1), 17–30.
- [47] Duong, T. and Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scand. J. Stat.* 32, (3), 485–506.
- [48] Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density, *Theory of Probability and Its Applications* 14, 153–158.
- [49] Fisher, R. A. (1956). *Statistical methods and scientific inference*. Hafner Publishing Co.
- [50] Fuentes-Santos, I., González-Manteiga, W., and Mateu, J. (2016). Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial poisson point processes. *Scandinavian Journal of Statistics*, 43(2), 416–435.
- [51] Gangopadhyay, A.K., Cheung, K.N. (2002). Bayesian approach to the choice of smoothing parameter in kernel density estimation. *J Nonparametric Stat*, 14, 655–664.
- [52] Gasser, T., and Müller, H.G. (1979). Kernel estimation of regression functions. *In Smoothing techniques for curve estimation*, pages 23–68. Springer.
- [53] Gasser, T., Müller, H.G. & Mammitzsch, V. (1985). Kernels for nonparametric curve estimation, *Journal of the Royal Statistical Society* 2, 238–252.

- [54] Gatrell, A. C., Bailey, T. C., Diggle, P. J., and Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, 21(1), 256-274.
- [55] Gelfand, A. E., Diggle, P. J., Fuentes, M. & Guttorp, P. E. (2010). *Handbook of Spatial Statistics*. Boca Raton, Florida : Taylor & Francis/CRC Press.
- [56] González-Manteiga, W., Cao, R. & Marron, J. S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *J. Amer. Statist. Assoc.* 91, (435), 1130–1140.
- [57] Gourieroux, C., Monfort, A. (2006). Pricing with splines. *Annales d'Economie et de Statistique*, pages 3–33.
- [58] Guan, Y. (2008). A goodness-of-fit test for inhomogeneous spatial poisson processes. *Biometrika* 95, (4), 831-845.
- [59] Guan, Y. (2008). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *J. Amer. Statist. Assoc.* 103, (483), 1238-1247.
- [60] Guan, Y. and Loh, J.M.(2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *J. Amer. Statist. Assoc.* 102, (480), 1377-1386.
- [61] Gustafsson, J., Hagmann, M., Nielsen, J. P., and Scaillet, O. (2006). Local transformation kernel density estimation of loss. *Journal of Business and Economic Statistics*.
- [62] Hall, P., 1983. Large sample optimality of least squares cross-validation in density estimation. *Annals of Statistics* 11, 1156–1174.
- [63] Hirukawa, M. and Sakudo, M. (2014). Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics & Data Analysis*, 75 :112–123.
- [64] Hu, S., Poskitt, D. and Zhang, X. (2012). Bayesian adaptive bandwidth kernel density estimation of irregular multivariate distributions. *Computational Statistics and Data Analysis* 56, 732–740.
- [65] Igarashi, G. and Kakizawa, Y. (2015). Bias corrections for some asymmetric kernel estimators. *Journal of Statistical Planning and Inference*, 159 :37–63.
- [66] Illian, D. J., Penttinen, P. A., Stoyan, D. H. and Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*, John Wiley & Sons, Chichester, UK.
- [67] Illian, J. B., Moller, J., and Waagepetersen, R. P. (2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics*, 16(3), 389-405.
- [68] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007) :453–461.
- [69] Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Wiley, New York.
- [70] Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* 3, 135-146.
- [71] Jones, M.C. & Foster, P.J. (1996). A simple non negative boundary correction method for kernel density estimation. *Statistics Sinica* 4, 1005-1013.
- [72] Jørgensen, B. (1997). *The Theory of Dispersion Models*, Chapman & Hall, London.

- [73] Jørgensen, B. (2013). Construction of multivariate dispersion models, *Brazilian Journal of Probability and Statistics* 27, 285–309.
- [74] Jørgensen, B. & Kokonendji, C.C. (2016). Discrete dispersion models and their tweedie asymptotics. *AStA Advances in Statistical Analysis*, 100(1) :43–78
- [75] Jørgensen, B. & Kokonendji, C.C. (2011). Dispersion models for geometric sums, *Brazilian Journal of Probability and Statistics* 25, 263–293.
- [76] Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Statistical Methodology*, 2(3) :191–212.
- [77] Karunamuni, R. J. and Alberts, T. (2006). A locally adaptive transformation method of boundary correction in kernel density estimation. *Journal of Statistical Planning and Inference*, 136(9) :2936–2960.
- [78] Karunamuni, R. J. and Zhang, S. (2008). Some improvements on a boundary corrected kernel density estimator. *Statistics & Probability Letters*, 78(5) :499–507.
- [79] Kerscher, M. (2000). Statistical analysis of large-scale structure in the universe. *In Statistical physics and spatial statistics, volume 554*, pages 36-71. Springer.
- [80] Kiefer, J. (1961). On large deviations of the empiric df of vector chance variables and a law of the iterated logarithm. *Pacific journal of mathematics*, 11(2) :649–660.
- [81] Kokonendji, C. C., Senga Kiessé, T. and Zocchi, S. S. (2007). Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Non-parametric Statistics* 19, 241–257.
- [82] Kokonendji, C. C. and Senga Kiessé, T. (2011). Discrete associated kernels method and extensions. *Statistical Methodology* 8, 497–516.
- [83] Kokonendji, C. C. and Somé, S. M. (2015). On multivariate associated kernels for smoothing general density function. arXiv : 1502.01173.
- [84] Kokonendji, C. C. and Varron, D. (2016). Performance of the discrete associated kernel estimator through the total variation distance. *Statistics and Probability Letters* 110, 225–235.
- [85] Kokonendji, C. C. and Zocchi, S. S. (2010). Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistics and Probability Letters* 80, 1655–1662.
- [86] Kulasekera, K. B. and Padgett, W. J. (2006). Bayes bandwidth selection in kernel density estimation with censored data. *Journal of Nonparametric Statistics*, 18 :129–143.
- [87] Kuruwita, C. N., Kulasekera, K. B., and Padgett, W. J. (2010) . Density estimation using asymmetric kernels and Bayes bandwidths with censored data. *Journal of Statistical Planning and Inference*, 140 :1765–1774.
- [88] Law, R., Illian, J., Burslem, D. F., Gratzner, G., Gunatilleke, C., and Gunatilleke, I. (2009). Ecological information from spatial patterns of plants : insights from point process theory. *Journal of Ecology*, 97(4), 616-628.
- [89] Lejeune, M., Sarda, P., (1992). Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis* 14, 457–471.
- [90] Li, Q. & Racine, J. (2007). Nonparametric Econometrics : *Theory and Practice*, Princeton University Press.

- [91] Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*. Ph.D. Thesis Manuscript (in French) to Université de Franche-Comté, Besançon, France & Université de Ouagadougou, Burkina Faso, June 2013, LMB no. 14334, Besançon, URL <http://hal.in2p3.fr/tel-01124288/document>.
- [92] Lo, P.H. (2018). *An iterative plug-in algorithm for optimal bandwidth selection in kernel intensity estimation for spatial data*. PhD thesis, Technical University of Kaiserslautern.
- [93] Loh, J. M. & Jang, W. (2010). Estimating a cosmological mass bias parameter with bootstrap bandwidth selection. *J. R. Stat. Soc. Ser. C. Appl. Stat.* 59, (5), 761–779.
- [94] Malec, P. and Schienle, M. (2014). Nonparametric kernel density estimation near the boundary. *Computational Statistics & Data Analysis*, 72 :57–76.
- [95] Maniya, G. H. (1961). Remarks on non parametric estimates of a bivariate probability density. *Soobshch. Akad. Nauk Gruz. SSR*, 27, 385–400.
- [96] Marron, J., and Harlde, W. (1986). “Random Approximation to Some Measures of Accuracy in Non-parametric Curve Estimation,” *Journal of Multivariate Analysis*, 20, 91–113.
- [97] Marron, J.S. & Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society* 4, 653–671.
- [98] Marron, J. S. & Wand, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.* 20, (2), 712–736.
- [99] Møller, J. and Toftaker, H. (2014). Geometric anisotropic spatial point pattern analysis and cox processes. *Scandinavian journal of statistics*, 41(2) :414–435.
- [100] Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. CRC Press.
- [101] Müller, H. G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* 78, 521–530.
- [102] Müller, H., and U. Stadtmuller (1999) : “Multivariate Boundary Kernels and a Continuous Least Squares Principle,” *Journal of the Royal Statistical Society, Series B*, 61, 439–458.
- [103] Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Applications* 9, 141–142.
- [104] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical association*, 83(401) :9–27.
- [105] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* 50, 379–402.
- [106] Ogata, Y. and Zhuang, J. (2006). Space–time etas models and an improved extension. *Tectonophysics*, 413(1), 13–23.
- [107] Ord, J. K. (1978). How many trees in a forest ? *Math. Sci.* 3, 23–33.
- [108] Parzen, E. (1962). On estimation of a probability density function and mode, *Annals of Mathematical Statistics* 33, 1065–1076.
- [109] Rao, P. B. (1983). *Nonparametric function estimation*, Academic Press, London.
- [110] Rao, P. B. (2014). *Nonparametric functional estimation*. Academic press.

- [111] R Development Core Team (2015). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna : Austria, ISBN 3-900051-07-0 URL <http://www.R-project.org>.
- [112] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics* 27, 832–837.
- [113] Rudemo M, . (1982). "Empirical choice of Histograms and Kernel Density Estimators," *Scandinavian Journal of Statistics*, 9, 65-78
- [114] Sain, S. R., Baggerly, K. A. and Scott, D. W. (1994). Cross-validation of multivariate densities. *Journal of the American Statistical Association*, 89(427) :807–817, 1994.
- [115] Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics* 16, 217–226.
- [116] Schaap, W. E. & Van De Weygaert, R. (2000). Letter to the editor. Continuous fields and discrete samples : Reconstruction through Delaunay tessellations. *Astron. Astrophys.* 363, L29–L32.
- [117] Schlaifer, R. and Raiffa, H. (1961). *Applied statistical decision theory*.
- [118] Schuster, E.F. (1985). Incorporating support constraints into nonparametric estimators of densities, *Communications in Statistics - Theory and Methods* 40, 1123–1136.
- [119] Scott, D. (1992). Multivariate density estimation. *Theory, practice and visualization*, Wiley, New York.
- [120] Senga Kiese, T. (2008). *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. PhD thesis, Université de Pau et des Pays de l'Adour.
- [121] Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society : Series B (Methodological)*, 53(3) :683–690.
- [122] Silverman, B. (1986). *Density estimation for statistics and data analysis (monographs on statistics and applied probability)*, Chapman & Hall/CRC, New York.
- [123] Simonoff, J. S. (1996). *Smoothing Methods in Statistics*. Springer Science & Business Media.
- [124] Simonoff, J. S. and Tutz, G. (2000). Smoothing methods for discrete data. *Smoothing and regression : Approaches, computation, and application*, pages 193–228.
- [125] Singh, R. S. (1976). Non parametric estimation of mixed partial derivatives of a multivariate density. *Journal of Multivariate Analysis*, 6, 111–122.
- [126] Somé, S. M. (2015). *Estimation non-paramétrique par noyaux associés multivariés et applications*. Thèse de Doctorat, Université de Bourgogne Franche-Comté, France.
- [127] Stoyan, D. and Penttinen, A. (2000). Recent applications of point process methods in forestry statistics. *Statistical Science*, 15(1) :61-78.
- [128] Van Lieshout, M. N. M. (2011). A J-function for inhomogeneous point processes. *Statist. Neer.* 65, 183–201.
- [129] Van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. World Scientific.
- [130] Van Lieshout, M. N. M. (2012). On estimation of the intensity function of a point process. *Methodol. Comp. Appl. Prob.* 14, 567–78.

- [131] Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman-scott processes. *Biometrics* 63, (1), 252-258.
- [132] Wand, M. P. (1992). Error analysis for general multivariate kernel estimators. *J. Nonparametr. Stat.* 2, (1), 1-15.
- [133] Wand, M. P. and Jones, M. C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, 9, 97-116.
- [134] Wand, M. P., Jones, M. C., (1995). *Kernel Smoothing*. Chapman & Hall, New York.
- [135] Yue, Y. R. and Loh, J. M. (2011). Bayesian semiparametric intensity estimation for inhomogeneous spatial point processes. *Biometrics*, 67(3) :937-946.
- [136] Zhang, S. & Karunamuni, R.J. (2010). Boundary performance of the beta kernel estimators, *Journal of Nonparametric Statistics* 22, 81-104.
- [137] Zhang, S., Karunamuni, R. J. and Jones, M. C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, 94(448) :1231-1240.
- [138] Zhang, X., King, M. L. and Hyndman, R. J. (2006). A Bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 50, 3009-3031.
- [139] Zheng, Q., Kulasekera, K. B. and Gallagher, C. (2010). Local adaptive smoothing in kernel regression estimation. *Statistics & probability letters*, 80(7-8) :540-547, 2010.
- [140] Ziane Y, Adjabi S, Zougab N. (2015). Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data. *J Appl Stat*, 42, 1645-1658.
- [141] Zougab, N., Adjabi, S., Kokonendji, C. C. (2012). Binomial kernel and Bayes local bandwidth in discrete functions estimation. *J Nonparametric Stat*, 24, 783-795.
- [142] Zougab, N., Adjabi, S., Kokonendji, C. C. (2013). Adaptive smoothing in associated kernel discrete functions estimation using Bayesian approach. *J Stat Comput Simul*, 83, 2219-2231.
- [143] Zougab, N., Adjabi, S., Kokonendji, C. C. (2014). Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 75, 28-38.