

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ MOULOUD MAMMERRI, TIZI-OUZOU

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUES

Thèse de Doctorat LMD

SPÉCIALITÉ : MATHÉMATIQUES

OPTION : Analyse Mathématique et Applications

Présentée par BEDOUHENE Kahina

Sujet

**Sur l'estimation non-paramétrique et semi-paramétrique par noyau des
données circulaires : cas uni-varié et multi-varié**

Devant le jury d'examen composé de :

Mr. HAMADOUCHE Djamel ;	Professeur ;	UMMTO ;	Président
Mr. ZOUGAB Nabil ;	Maître de Conférences A ;	Univ. Bejaia ;	Rapporteur
Mme ACHEMINE Farida ;	Maître de Conférences A ;	UMMTO ;	Examineur
Mr. HAMAZ Abdelghani ;	Maître de Conférences A ;	UMMTO ;	Examineur
Mr. BOUALEM Mohamed ;	Professeur ;	Univ. Bejaia ;	Examineur

Soutenue : le 11/01/2021

Remerciements

Je tiens à exprimer ma gratitude à mon directeur de thèse Dr. Zougab Nabil, Maître de Conférences à l'Université de Bejaia, pour avoir suivi et dirigé mes recherches avec beaucoup d'intérêt et de disponibilité. Son soutien, son expérience, et sa minutie dans le travail ont été pour moi un apport assez précieux. Je suis consciente de l'énergie et du temps qu'il a donné pour mener à bien ma thèse et je lui ai très reconnaissante.

Mes remerciements les plus sincères vont également à Monsieur HAMADOUCHE Djamel, Professeur à l'Université Mouloud Mammeri de Tizi-Ouzou, pour l'honneur qu'il me fait en acceptant de présider le jury de cette thèse.

Je remercie profondément Madame ACHEMINE Farida, Monsieur HAMAZ Abdelghani, maîtres de conférences à l'UMMTO, et le professeur BOUALEM Mohamed de l'université de Bejaia, pour avoir accepté de participer dans le jury de ma thèse. J'apprécie l'intérêt que vous avez porté à mon travail.

Une mention très particulière à tous les enseignants de département de Mathématiques de l'UMMTO qui ont contribué à ma formation, en l'occurrence l'équipe de spécialité Analyse Mathématique et Applications : Pr. M. Morsli, Pr. L. Rahmani, Pr. F. Bedouhene, Dr. N. Zougab, Dr. A. Hamaz, Dr. F. Achemine, Dr. K. Merakeb, Dr. O. Mellah, Dr. M. Smaali, Dr. S. Smaili.

Je tiens à exprimer ma gratitude à tous les membres du Laboratoire de Mathématiques Pures et Appliquées (LMPA), en particulier l'équipe de "Modélisation Mathématique".

Enfin, je ne peux conclure sans exprimer mes plus chaleureux remerciements à tous les membres de ma famille pour leurs encouragements. Leur soutien m'a permis de garder une motivation toujours intacte.

Liste des travaux

1. **Kahina Bedouhene** and Nabil Zougab (2020). A Bayesian procedure for bandwidth selection in circular kernel density estimation. *Monte Carlo Methods Appl*, 26(1) : 69–82.
2. **Kahina Bedouhene** and Nabil Zougab (2019). Nonparametric multiplicative bias correction for von Mises kernel circular density estimator. *Communications in Statistics-Simulation and Computation*, DOI : 10.1080/03610918.2019.1659358.
3. Yasmina Ziane., Nabil Zougab., **Kahina Bedouhene** and Smail Adjabi (2020). Semiparametric von Mises kernel circular density estimator. *Soumis pour publication*.

Table des matières

Introduction générale	10
1 Estimation non-paramétrique de la fonction densité par noyau circulaire	17
1.1 Introduction	17
1.2 Estimateur de la densité par noyau circulaire	17
1.2.1 Propriétés de l'estimateur	18
1.2.1.1 Biais	19
1.2.1.2 Variance	20
1.2.2 MSE et MISE	21
1.2.2.1 Convergence au sens de MSE	21
1.2.2.2 Convergence au sens de MISE	21
1.2.3 Convergence en loi	22
1.2.4 Estimateur à noyau de von Mises et ses propriétés	22
1.2.4.1 Biais de l'estimateur de von Mises	24
1.2.4.2 Variance de l'estimateur de von Mises	24
1.2.4.3 Convergence de l'estimateur de von Mises au sens de MISE	25
1.2.5 Estimateur à noyau de Cauchy enveloppé et ses propriétés	25
1.2.5.1 Biais de l'estimateur à noyau de Cauchy enveloppé	27
1.2.5.2 Variance de l'estimateur de Cauchy enveloppé	27
1.2.5.3 Convergence de l'estimateur de Cauchy enveloppé au sens de MISE	27
1.3 Estimateur à noyau toroïdal (cas multi-varié)	28
1.4 Conclusion	29

2	Choix du paramètre de concentration ν dans l'estimation non-paramétrique de la densité par noyau circulaire	30
2.1	Méthodes de ré-injection (plug-in)	31
2.1.1	Règle de référence (Rule of Thumb)	31
2.1.2	Méthode plug-in de Oliveira et al. (2012)	32
2.2	Méthodes de validation croisée	32
2.2.1	Validation croisée par le maximum de vraisemblance	33
2.2.2	Validation croisée non biaisée	33
2.3	Approche bayésienne pour le choix du paramètre de concentration ν	35
2.3.1	Stratégie	35
2.3.1.1	Approximation de la vraisemblance	35
2.3.1.2	Choix de la loi a priori	36
2.3.1.3	Approximation de la loi a posteriori	36
2.3.2	MCMC pour estimer le paramètre de concentration	37
2.3.2.1	Principe de la méthode MCMC	37
2.3.2.2	Algorithme de Metropolis-Hasting (M-H)	38
2.3.3	Choix des paramètres de la loi a priori gamma	39
2.4	Résultats de la simulation	40
2.4.1	Implémentation bayésienne et convergence des méthodes MCMC	42
2.4.2	Approche bayésienne globale contre les approches classiques	44
2.5	Analyse de données réelles	48
2.6	Conclusion	52
3	Réduction du biais de l'estimateur non-paramétrique de la fonction densité à noyau circulaire	53
3.1	Introduction	53
3.2	Estimateurs MBC à noyau de von Mises de la fonction densité circulaire	53
3.2.1	Technique TS (Terrell et Scott)	54
3.2.2	Technique JLN (Jones, Linton et Neilson)	54
3.2.3	Propriétés asymptotiques des estimateurs MBC circulaires	55
3.2.3.1	Biais et variance	55

3.2.4	MISE des estimateurs MBC circulaires	61
3.2.5	Choix du paramètre de concentration pour les estimateurs MBC-VM	62
3.3	Etude de simulation	63
3.3.1	Modèles considérés dans l'étude de simulation	63
3.3.2	Résultats de simulation : comparaison entre ISE et ISB	63
3.4	Illustration avec des données réelles	67
3.5	Conclusion	70
4	Estimation semi-paramétrique par noyau circulaire de la fonction densité	71
4.1	Introduction	71
4.2	Estimateur semi-paramétrique par noyau circulaire de la fonction densité .	71
4.3	Estimateur semi-paramétrique HG à noyau circulaire de la fonction densité	73
4.3.1	Biais et variance de l'estimateur HG-VM	74
4.3.2	MISE de l'estimateur HG-VM	75
4.4	Estimateur semi-paramétrique JSH à noyau circulaire de la fonction densité	76
4.4.1	Biais et variance de l'estimateur JSH-VM	76
4.4.2	MISE de l'estimateur JSH-VM	77
4.5	Etude de simulation	78
4.6	Illustration avec des données réelles	84
4.7	Conclusion	86
	Conclusion générale et perspectives	87
	Annexe	89
	Bibliographie	96

Table des figures

1.1	Noyau de von Mises pour $\theta = \pi/4$ (première rangée) et $\theta = \pi$ (deuxième rangée) avec différentes valeurs de $\nu = \{3, 5, 10, 15\}$	23
1.2	Noyau de Cauchy enveloppé pour $\theta = \pi/4$ (première rangée) et $\theta = \pi$ (deuxième rangée) avec différentes valeurs de $\rho = \{0.3, 0.5, 0.7, 0.9\}$	26
2.1	Les modèles de simulation.	43
2.2	Densité a posteriori et la chaîne MCMC de ν obtenues à partir de données simulées avec le modèle M7 pour $n = 50$	44
2.3	Densités a posteriori et les chaînes MCMC générées pour ν obtenues à partir des données d'orientation des libellules (première rangée) et des données de lit croisé (deuxième rangée).	50
2.4	Diagrammes en rose (première rangée) et les estimateurs de densité à noyau circulaire de von Mises avec ν_{Byaes1} (tracés circulaires et linéaires donnés en secondes et troisième rangée, respectivement) pour les données d'orientation des libellules (panneau de gauche) et les données de lit croisé (panneau de droite).	51
3.1	Graphes des fonctions λ et ψ pour $c \in]0, 1[$	60
3.2	Estimation des densités des données de direction des tortues (première rangée) et des données d'orientation des libellules (deuxième rangée). Ligne pleine noire : estimateur TS-VM ; Ligne pleine grise : estimateur JLN-VM ; Ligne pointillé noire : estimateur standard de von Mises.	69

4.1	Les vraies densités et les estimateurs semi-paramétriques et non-paramétriques à noyau de von Mises (JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises) des modèles 3, 8, 11 et 17 avec $n = 200$. Trait noir continu : vraie densité; trait noir discontinu : estimateur standard de von Mises; ligne en pointillés noir : estimateur JLN-VM; trait gris continu : estimateur HG-VM et trait gris discontinu : estimateur JSH-VM.	83
4.2	Les estimateurs semi-paramétriques et non-paramétriques à noyau de von Mises (JHS-VM, HG-VM, JLN-VM et l'estimateur standard) des données réelles; Azimuts flare (première rangée), Orientations à grand axe des lattes de feldspath (deuxième rangée). Trait noir continu : estimateur standard de von Mises; trait noir discontinu : estimateur JLN-VM; trait gris continu : estimateur HG-VM et trait gris discontinu : estimateur JSH-VM	85

Liste des tableaux

2.1	Estimateurs bayésiens du paramètre ν et les indicateurs BMSE et SIF associés obtenus pour M7.	44
2.2	Valeurs empiriques de $ISE(\nu_{RT})$, $ISE(\nu_{PI})$, $ISE(\nu_{LCV})$, $ISE(\nu_{UCV})$, $ISE(\nu_{Bayes1})$, $ISE(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 25$	45
2.3	Valeurs empiriques de $ISE(\nu_{RT})$, $ISE(\nu_{PI})$, $ISE(\nu_{LCV})$, $ISE(\nu_{UCV})$, $ISE(\nu_{Bayes1})$, $ISE(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 50$	46
2.4	Valeurs empiriques $ISE(\nu_{RT})$, $ISE(\nu_{PI})$, $ISE(\nu_{LCV})$, $ISE(\nu_{UCV})$, $ISE(\nu_{Bayes1})$, $ISE(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 100$	46
2.5	Comparaison des temps d'exécution en secondes pour une replication à l'aide du modèle M7.	48
2.6	Estimateurs bayésiens de la largeur de bande et indicateurs BMSE et SIF associés obtenus à partir de jeux de données réelles.	49
3.1	Valeurs empiriques de $ISE (\times 100)$ et leurs écarts-types ($\times 100$) entre parenthèses basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 25$ et $n = 50$	64
3.2	Valeurs empiriques de $ISE (\times 100)$ et leurs écarts-types ($\times 100$) entre parenthèses basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 100$ et $n = 250$	65

3.3	Valeurs empiriques de ISB ($\times 100$) basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 25$ et $n = 50$.	65
3.4	Valeurs empiriques de ISB ($\times 100$) basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 100$ et $n = 250$.	66
3.5	Paramètre de concentration optimal avec LCV et log-vraisemblance pour les trois estimateurs utilisant les données de direction des tortues.	68
3.6	Paramètre de concentration optimal avec LCV et log-vraisemblance pour les trois estimateurs utilisant les données d'orientation des libellules.	68
4.1	Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 10$.	79
4.2	Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 50$.	79
4.3	Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 100$.	80
4.4	Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 200$.	80
4.5	Valeurs empiriques de ISB basées sur 500 replications pour les modèles considérés avec des tailles d'échantillon $n = 10$ et $n = 50$.	81
4.6	Valeurs empiriques de ISB basées sur 500 replications pour les modèles considérés avec des tailles d'échantillon $n = 100$ et $n = 200$.	81
4.7	Estimateurs de maximum de vraisemblance des paramètres du modèle de départ de von Mises pour les données réelles.	84

Introduction générale

Trois problèmes fondamentaux se posent en statistique : (i) concevoir des techniques statistiques scientifiques et valides pour la collecte de données, (ii) développer des méthodes statistiques pour l'inférence statistique à partir des données collectées, et (iii) améliorer les modèles statistiques pour une utilisation future basée sur ces expériences.

Les méthodes statistiques les plus répandues dans la littérature sont les méthodes classiques, dites aussi paramétriques, et les méthodes non paramétriques. Cependant, l'approche semi-paramétrique qui combine les deux approches paramétrique et non-paramétrique a été également développée dans la littérature.

Dans le cas classique, l'hypothèse est que l'échantillon d'observations provenait d'une population avec une famille paramétrique connue. Le problème est soit d'estimer les paramètres inconnus, soit de concevoir des tests et de dériver des régions de confiance, etc., pour les paramètres inconnus basés sur l'échantillon. Cette hypothèse est plutôt forte parce que le modèle paramétrique supposé n'a pas besoin d'être le modèle "réel" s'il y en a un, et les méthodes statistiques développées pour un modèle paramétrique particulier pourraient conduire à des conclusions erronées lorsqu'elles sont appliquées à un modèle légèrement perturbé. Ces problèmes ont entraîné une tendance récente à développer des méthodes statistiques non-paramétriques et semi-paramétriques pour l'analyse des données.

Le premier article dans le domaine de l'estimation fonctionnelle non paramétrique, en particulier l'estimation de la densité non paramétrique, est dû à [Rosenblatt \[1956\]](#) ; voir aussi [Parzen \[1962\]](#). Depuis la parution de ces deux articles, plusieurs méthodes ont été développées pour l'estimation non-paramétrique des fonctions de densité, des fonctions de distribution, des fonctions de régression, des taux de défaillance, etc.

L'estimation de la densité est une partie importante de l'estimation non-paramétrique

car elle est applicable que les données soient réelles, positives, discrètes, périodiques, etc. Plusieurs méthodes sont dédiées à l'estimation non-paramétrique de la fonction de densité et la fonction de régression. Nous citons à titre d'exemple :

- La méthode des histogrammes, qui sont un type d'estimateurs de densité, sont utilisés depuis de nombreuses années, à la fois par les statisticiens en général et les scientifiques des sciences sociales, physiques et biologiques, comme outils pour tirer des inférences à partir de données. Les propriétés des estimateurs par histogramme de la fonction de densité et de régression ont été étudiées par [Abou-Jouadé \[1976\]](#), [Geffroy \[1976\]](#), [Geoffroy \[1980\]](#), [Tukey \[1961\]](#) et [Lecoutre \[1982\]](#).
- La méthode d'estimation par les séries orthogonales proposée pour estimer des densités continues a été déployée à partir des travaux de [Cencov \[1962\]](#), et explorée par la suite par plusieurs auteurs ; notamment, [Schwartz \[1967\]](#), [Kronmal and Tarter \[1968\]](#), [Wahba \[1981\]](#), [Bosq \[2005\]](#) et [Saadi and Adjabi \[2009\]](#).
- La méthode de lissage par les fonctions splines développée et exploitée par [Reinsch \[1967\]](#), [Silverman \[1985\]](#), [Wahba \[1990\]](#) et [Eubank \[1999\]](#).
- La méthode d'estimation par noyau pour les fonctions de densité et de régression. Cette méthode dépend du noyau et du paramètre de lissage ; voir par exemple [Rosenblatt \[1956\]](#), [Parzen \[1962\]](#), [Silverman \[1986\]](#), [Nadaraya \[1964\]](#), [Watson \[1964\]](#), [Härdle and Marron \[1985\]](#) et [Zhang et al. \[2009\]](#) en utilisant les noyaux symétriques pour estimer des fonctions de densité et de régression à support réel ; [Chen \[2000\]](#), [Jin and Kawczak \[2003\]](#), [Scaillet \[2004\]](#), [Marchant et al. \[2013\]](#), [Saulo et al. \[2013\]](#), [Hirukawa and Sakudo \[2015\]](#), [Mousa et al. \[2016\]](#), [Igarashi \[2016\]](#), [Kakizawa and Igarashi \[2017\]](#), [Erçelik and Nadar \[2018\]](#) et [Shi and Son \[2016\]](#) en utilisant les noyaux asymétriques dans le cas des données à support positif (\mathbb{R}^+) pour éviter le problème des effets de bord ; [Kokonendji et al. \[2007\]](#), [Senga Kiessé \[2008\]](#), [Kokonendji and Senga Kiessé \[2011\]](#) et [Kokonendji et al. \[2009b\]](#) pour estimer des fonctions discrètes à support discret (par exemple \mathbb{N}), voir aussi [Aitchison and Aitken \[1976\]](#) pour le cas des données catégorielles ; [Taylor \[2008\]](#), [Di Marzio et al. \[2009\]](#), [Di Marzio et al. \[2011a\]](#) et [Oliveira et al. \[2012\]](#) en utilisant les noyaux circulaires pour estimer des densités à support $[0, 2\pi[$.

Dans plusieurs situations, l'approche semi-paramétrique par noyau peut être considérée comme une bonne alternative aux approches paramétrique et non-paramétrique par noyau, car elle représente un compromis entre ces deux approches. Notons que l'estimateur semi-paramétrique s'écrit comme le produit d'une partie paramétrique (modèle de départ) et d'un facteur de correction qu'il faut estimer par l'approche non-paramétrique du noyau pour laisser faire les données, voir par exemple, [Hjort and Glad \[1995\]](#), [Jones et al. \[1999\]](#), [Hagmann and Scaillet \[2007\]](#) et [Hirukawa and Sakudo \[2019\]](#) dans le cas des données linéaires ; voir aussi [Kokonendji et al. \[2009a\]](#) en utilisant les noyaux discrets.

Cette thèse est une contribution à l'estimation non paramétrique et semi-paramétrique par noyau pour les fonctions de densité dans le cas des données circulaires. Ce travail est motivé par plusieurs points. Les données circulaires peuvent apparaître dans divers domaines, comme par exemple, l'écologie, la météorologie, les sciences de l'environnement, etc (voir par exemple [Mardia and Jupp \[2000\]](#)). L'application des modèles circulaires paramétriques n'est pas satisfaisante, lorsque aucune information n'est disponible sur le phénomène étudié ou la taille de l'échantillon est petite ou modérée. Comme troisième motivation, les approches non-paramétrique et semi-paramétrique par noyau riment de succès au sein de la communauté des statisticiens, par rapport aux avantages qu'elles présentent, notamment :

1. les expressions théoriques des estimateurs non-paramétrique et semi-paramétrique sont simples puisque ils peuvent s'écrire comme la somme de n variables aléatoires indépendantes et identiquement distribuées (iid), en utilisant un noyau circulaire K et un paramètre de lissage (concentration) ν .
2. ils sont convergents en de nombreux sens.
3. les estimateurs à noyau ont plus de degré de liberté dans le sens où ils laissent à l'utilisateur une flexibilité dans le choix du noyau K et du paramètre de lissage ν .

Récemment, l'estimation de la fonction de densité de probabilité circulaire à l'aide de la méthode non-paramétrique du noyau circulaire a reçu une attention considérable dans la littérature statistique ; voir par exemple [Hall et al. \[1987\]](#) ; [Klemelä \[2000\]](#) ; [Taylor \[2008\]](#) ; [Di Marzio et al. \[2009\]](#) ; [Di Marzio et al. \[2011a\]](#) ; [Oliveira et al. \[2012\]](#) ; [Taylor et al. \[2012\]](#) ; [García-Portugués \[2013\]](#). Notons que les propriétés (asymptotiques, locales

et globales) de l'estimateur de densité circulaire telles que le biais, la variance, l'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne intégrée (MISE) ont également été données par de nombreux auteurs, voir par exemple Taylor [2008], Oliveira et al. [2012] et Bedouhene and Zougab [2019] pour l'estimateur spécifique du noyau de von Mises et Di Marzio et al. [2009, 2011a] pour l'estimateur à noyau circulaire; voir également Hall et al. [1987] pour l'estimation du noyau sphérique et Klemelä [2000] pour l'estimation de la densité et ses dérivées avec des données directionnelles. Plus récemment, García-Portugués et al. [2013] ont dérivé certaines propriétés d'un estimateur de densité à noyau conjoint pour les données directionnelles-linéaires, Amiri et al. [2017] ont proposé une version récursive pour l'estimateur directionnel de la densité du noyau et Tsuruta and Sagae [2017b] ont introduit une estimation de la densité par noyau d'ordre supérieur pour le cas circulaire. Naturellement, l'utilisation des noyaux circulaires est plus appropriée que les noyaux linéaires pour estimer n'importe quel densité circulaire. En tant que noyau circulaire, le noyau de von Mises peut être utilisé. Ce choix est motivé par le fait que le noyau de von Mises est couramment utilisé, et que les résultats de l'estimation de la densité circulaire utilisant un tel noyau sont alors disponibles.

Dans le cadre de cette thèse, nous nous focalisons sur le cas de l'estimation de la densité circulaire (cas uni-varié) et nous mettons l'accent sur les estimateurs à noyau de von Mises. Notre contribution est scindée en trois parties. Dans un premier temps, on s'est intéressé à une procédure bayésienne pour le choix du paramètre de concentration ν dans l'estimateur non-paramétrique à noyau de von Mises, lorsque l'algorithme d'échantillonnage de Monte Carlo par Chaînes de Markov (MCMC) est utilisé pour les estimations de Bayes. Cette approche est largement étudiée dans le cas linéaire, voir par exemple, Brewer [1998], Zhang et al. [2006], Li et al. [2015], Zougab et al. [2016]. Cependant, à notre connaissance, le cas circulaire n'a jamais été examiné. Par conséquent, la contribution de cette partie se résume comme suit :

1. Nous introduisons une sélection de paramètre de concentration bayésienne dans l'estimation de la densité du noyau pour les données circulaires. L'idée derrière cette méthode est de traiter le paramètre de lissage comme une quantité aléatoire avec une distribution préalable. La densité a posteriori et les estimateurs de Bayes

sous les fonctions de perte quadratique et entropique seront dérivés en utilisant les méthodes MCMC.

2. Nous examinons les performances de la procédure bayésienne et les approches classiques existantes telles que le plug-in et la validation croisée via des simulations de Monte Carlo en utilisant une variété de modèles circulaires, et en évaluant l'erreur quadratique intégrée (ISE). Notre étude est axée sur des échantillons de petite et moyenne taille.
3. Nous appliquons l'estimateur du noyau de von Mises pour deux ensembles de données réels, déjà discutés dans la littérature, lorsque le paramètre de concentration est choisi par une approche bayésienne. Les résultats de simulation obtenus montrent que l'approche bayésienne proposée est très compétitive par rapport aux méthodes globales classiques existantes, à savoir les techniques de plug-in et de validation croisée.

Dans un deuxième temps, notre contribution est étroitement liée au biais de l'estimateur non-paramétrique à noyau de von Mises dans le cas standard (sans correction du biais). Notons que le biais de l'estimateur standard de von Mises est $O(\nu^{-1})$ pour ν assez grand. Par conséquent, en utilisant les techniques dites de réduction du biais multiplicatif non paramétrique (MBC), nous proposons des améliorations dans l'estimation du noyau de von Mises qui réduisent l'ordre de grandeur du biais à $O(\nu^{-2})$, tandis que l'ordre de grandeur de la variance est maintenu.

Les approches MBC ont été initialement proposées et largement étudiées par plusieurs auteurs dans l'estimation de la densité par noyau linéaire symétrique et asymétrique, voir, par exemple [Terrell and Scott \[1980\]](#) ; [Jones et al. \[1995\]](#) ; [Hirukawa \[2010\]](#) ; [Hirukawa and Sakudo \[2014\]](#) ; [Hirukawa and Sakudo \[2015\]](#) ; [Funke and Kawka \[2015\]](#) ; [Zougab and Adjabi \[2016\]](#) ; [Harfouche et al. \[2018\]](#). De plus, les méthodes de réduction du biais multiplicatif et additif ont été récemment proposées par [Tsuruta and Sagae \[2017b\]](#) pour construire des estimateurs de densité de noyau d'ordre supérieur sur le cercle, suivant les idées de [Terrell and Scott \[1980\]](#) pour la méthode multiplicative et [Jones and Foster \[1993\]](#) pour l'approche additive. Ici, on se focalise principalement sur deux techniques MBC afin de réduire le reste du biais. Plus précisément,

1. La première technique revisite l'estimateur circulaire de [Tsuruta and Sagae \[2017b\]](#), en utilisant deux paramètres de concentration différents ν et $c\nu$, où $c \in]0, 1[$ est un paramètre de réglage. Notons qu'avec un choix optimal de c , le biais de notre estimateur circulaire TS proposé peut être réduit par rapport à l'estimateur [Tsuruta and Sagae \[2017b\]](#).
2. La deuxième technique MBC, attribuée à [Jones et al. \[1995\]](#), est basé sur l'idée d'exprimer

$$f(\theta) = \hat{f}(\theta) \left\{ f(\theta)/\hat{f}(\theta) \right\},$$

et l'estimation du terme de correction du biais

$$f(\theta)/\hat{f}(\theta),$$

de manière non-paramétrique. Ensuite, les propriétés asymptotiques et globales des estimateurs MBC à noyau de von Mises (MBC-VM) sont établies en utilisant les approximations de la série de Taylor. La sélection du paramètre de concentration est également étudiée en adoptant la méthode populaire de validation croisée de vraisemblance.

Dans un dernier temps, notre contribution consiste à développer l'approche semi-paramétrique par noyau circulaire pour estimer une densité circulaire inconnue, en suivant les travaux de [Hjort and Glad \[1995\]](#), [Hagmann and Scaillet \[2007\]](#), [Jones et al. \[1999\]](#) et [Hirukawa and Sakudo \[2019\]](#) proposés dans le cas des données linéaires. Plus précisément,

1. Nous proposons l'analogie circulaire de l'estimateur semi-paramétrique par noyau introduit par [Hjort and Glad \[1995\]](#) (estimateur HG) dans le cas linéaire, en utilisant les noyaux symétriques; voir aussi [Hagmann and Scaillet \[2007\]](#) en utilisant les noyaux asymétriques. Les propriétés asymptotiques et globales de l'estimateur semi-paramétrique HG de von Mises (HG-VM) sont établies.
2. Nous construisons une version améliorée (au sens du biais) de l'estimateur semi-paramétrique circulaire HG en utilisant la technique de [Jones et al. \[1999\]](#) (estimateur JSH) proposée pour le cas linéaire; voir aussi [Hirukawa and Sakudo \[2019\]](#). Nous examinons également les propriétés de l'estimateur semi-paramétrique JSH de von Mises (JSH-VM).

3. La sélection du paramètre de concentration pour les deux estimateurs proposés est étudiée en adoptant la méthode de validation croisée non biaisée.

Ce travail est structuré en quatre chapitres, une conclusion générale et une liste de références bibliographiques, que nous décrivons brièvement.

- **Le premier chapitre** concerne l'estimation non-paramétrique de la fonction de densité basée sur l'estimateur de type noyau circulaire, en particulier les estimateurs à noyau de von Mises et de Cauchy enveloppé. Nous nous focalisons surtout sur les propriétés asymptotiques, à savoir le biais, la variance, la MSE et la MISE. Les chapitres deux, trois et quatre comportent notre contribution.
- **Le chapitre deux** est dédié à l'approche bayésienne globale pour le choix du paramètre de concentration dans l'estimation non-paramétrique de la fonction de densité circulaire par noyau von Mises. Notons que les méthodes de choix du paramètre de concentration par les méthodes classiques (méthodes de plug-in, validation croisée) sont données aussi dans ce chapitre.
- **Le chapitre trois** concerne les techniques MBC de réduction du biais pour l'estimateur à noyau circulaire de von Mises de la fonction densité. On y présente entre autres quelques propriétés des estimateurs MBC-VM (biais, variance, MSE et MISE).
- **Le chapitre quatre** propose les estimateurs semi-paramétriques par noyau circulaire de von Mises, ainsi que leurs propriétés statistiques.

Cette thèse est achevée par une conclusion générale et quelques perspectives de recherche. Notons que dans chacun des chapitres deux, trois et quatre, les résultats seront illustrés à travers des simulations et appliqués sur des données réelles analysées dans la littérature. Nous donnons aussi dans l'annexe un bref rappel sur les modèles circulaires et les programmes informatiques mis en place sous logiciel R.

Chapitre 1

Estimation non-paramétrique de la fonction densité par noyau circulaire

1.1 Introduction

Dans ce chapitre, nous donnons la définition de l'estimateur à noyau des données circulaires dans le cas de l'estimation de la fonction densité, ainsi que ses propriétés statistiques (biais, variance, MSE et MISE). Par la suite, nous présentons les cas particuliers des estimateurs à noyau de von Mises et de Cauchy enveloppé. Enfin, nous donnons un aperçu sur l'estimateur à noyau toroïdal (cas multi-varié).

1.2 Estimateur de la densité par noyau circulaire

Nous donnons d'abord la définition d'un noyau circulaire. Ensuite, nous présentons l'estimateur circulaire.

Définition 1 (Noyau circulaire d'ordre r). *Un noyau circulaire d'ordre r et de paramètre de concentration $\nu > 0$ est la fonction $K_\nu : [0; 2\pi[\rightarrow \mathbb{R}$ telle que*

1. *Pour chaque $\theta \in [0, 2\pi[$, le noyau K_ν admet une représentation en série de Fourier*

convergente donnée par :

$$K_\nu(\theta) = 1/(2\pi) \left\{ 1 + 2 \sum_{j=1}^{\infty} \gamma_j(\nu) \cos(j\theta) \right\}, \quad (1.1)$$

où γ_j est le coefficient de Fourier associé donné par :

$$\gamma_j(\nu) = \int_0^{2\pi} K_\nu(\theta) \cos(j\theta) d\theta. \quad (1.2)$$

2. Soit $\eta_j(K_\nu) = \int_0^{2\pi} \sin^j(\theta) K_\nu(\theta) d\theta$, alors

$$\eta_0(K_\nu) = 1, \quad \eta_j(K_\nu) = 0 \quad \text{pour } 0 < j < r \quad \text{et} \quad \eta_r(K_\nu) \neq 0.$$

La condition 1 montre que le noyau circulaire est symétrique autour de 0 et la condition 2 joue le même rôle comme le $j^{\text{ème}}$ moment du noyau symétrique dans le cas linéaire (voir Silverman [1986]).

Dans la suite nous présentons quelques exemples de noyau circulaire. Nous mettons l'accent sur le noyau de von Mises, car c'est le noyau le plus utilisé dans la littérature en raison de sa simplicité; voir par exemple Taylor [2008] et Oliveira et al. [2012].

Définition 2 (Estimateur de la densité par noyau circulaire Di Marzio et al. [2009]). Soient $\Theta_1, \dots, \Theta_n$ un n -échantillon aléatoire d'une densité circulaire continue bornée f . Etant donné un noyau circulaire K_ν , l'estimateur par noyau de f en $\theta \in [0, 2\pi[$ est défini par :

$$\hat{f}_\nu(\theta) = \frac{1}{n} \sum_{i=1}^n K_\nu(\theta - \Theta_i). \quad (1.3)$$

où $\nu > 0$ est le paramètre de concentration qui joue un rôle très important pour la qualité de lissage et θ est la cible (l'angle où on veut estimer f).

1.2.1 Propriétés de l'estimateur

Dans cette section, nous examinons les différentes propriétés locales et globales de l'estimateur à noyau circulaire. Ensuite, nous traitons les cas particuliers des estimateurs à noyau de von Mises et de Cauchy enveloppé. Notons que ces propriétés sont obtenues par plusieurs auteurs; voir Hall et al. [1987], Taylor [2008] et Di Marzio et al. [2009], etc.

1.2.1.1 Biais

Le calcul des moments d'ordre 1 et 2 de l'estimateur circulaire (1.3) requière les hypothèses suivantes, à savoir :

1. $\lim_{n \rightarrow \infty} \gamma_j(\nu) = 1$, avec $j \in \mathbb{Z}^+$,
2. $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\nu) = 0$,
3. f'' continue et de carré intégrable.

Ces hypothèses permettent de calculer l'expression du biais $[\widehat{f}_\nu(\theta)]$. L'espérance mathématique de l'estimateur à noyau circulaire (1.3) est donnée par :

$$\begin{aligned} \mathbb{E}[\widehat{f}_\nu(\theta)] &= \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n K_\nu(\theta - \Theta_i) \right] \\ &= \frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n K_\nu(\theta - \Theta_i) \right]. \end{aligned}$$

Nous exprimons $K_\nu(\theta)$ via sa série de Fourier (1.1), et en remarquant que $\sin(u) \simeq u$ lorsque u tend vers 0, alors on a la formule de Taylor suivante :

$$f(u + \theta) = f(\theta) + \sin(u)f'(\theta) + \frac{1}{2} \sin^2(u)f''(\theta) + o\{\sin^2(u)\}.$$

Nous faisons le changement de variable $\psi = u + \theta$, le développement de K_ν en série de Fourier, et avec l'utilisation de l'hypothèse 1, à savoir

$$\lim_{n \rightarrow \infty} \gamma_j(\nu) = 1, \text{ avec } j \in \mathbb{Z}^+,$$

alors on obtient :

$$\begin{aligned}
\mathbb{E}[\widehat{f}_\nu(\theta)] &= \int_0^{2\pi} K_\nu(\psi - \theta) f(\psi) d\psi \\
&= \int_0^{2\pi} K_\nu(u) f(u + \theta) du \\
&= \int_0^{2\pi} K_\nu(u) \left[f(\theta) + f'(\theta) \sin(u) + \frac{f''(\theta)}{2} \sin^2(u) \right] + o(1) \\
&= f(\theta) + \frac{f''(\theta)}{2} \int_0^{2\pi} \sin^2(u) K_\nu(u) du + o(1) \\
&= f(\theta) + \frac{f''(\theta)}{2} \int_0^{2\pi} \left(\frac{1 - \cos(2u)}{2} \right) K_\nu(u) du + o(1) \\
&= f(\theta) + \frac{f''(\theta)}{2} \left[\frac{1}{2} \left(1 - \int_0^{2\pi} \cos(2u) K_\nu(u) du \right) \right] + o(1) \\
&= f(\theta) + \frac{f''(\theta)}{4} (1 - \gamma_2(\nu)) + o(1).
\end{aligned}$$

Donc la formule du biais est donnée par :

$$\text{biais}[\widehat{f}_\nu(\theta)] = \frac{f''(\theta)}{4} (1 - \gamma_2(\nu)) + o(1). \quad (1.4)$$

1.2.1.2 Variance

La variance de $\widehat{f}_\nu(\theta)$ est donnée par :

$$\begin{aligned}
\mathbb{V}[\widehat{f}_\nu(\theta)] &= \frac{1}{n} \int_0^{2\pi} \{K_\nu(\psi - \theta)\}^2 f(\psi) d\psi - \frac{1}{n} \left\{ \mathbb{E}[\widehat{f}(\theta; \nu)] \right\}^2 \\
&= \frac{1}{n} \int_0^{2\pi} K_\nu^2(u) f(u + \theta) du - \frac{1}{n} \left\{ \mathbb{E}[\widehat{f}(\theta; \nu)] \right\}^2 \\
&= \frac{1}{n} \int_0^{2\pi} K_\nu^2(u) \{f(\theta) + o(1)\} du - \frac{1}{n} \{f(\theta) + o(1)\}^2. \quad (1.5)
\end{aligned}$$

En remplaçant le noyau K_ν par son développement en série de Fourier dans (1.5), on obtient :

$$\mathbb{V}[\widehat{f}_\nu(\theta)] = \frac{1}{n} \int_0^{2\pi} \left(\frac{1 + 2 \sum_{j=1}^{\infty} \gamma_j(\nu) \cos(\theta j)}{2\pi} \right)^2 f(\theta) du + o(1). \quad (1.6)$$

Finalement, grâce à l'égalité de Parseval, la formule de la variance de l'estimateur circulaire (1.3) est donnée par :

$$\mathbb{V}[\widehat{f}_\nu(\theta)] = \frac{1}{2n\pi} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) \right\} f(\theta) + o(1). \quad (1.7)$$

1.2.2 MSE et MISE

1.2.2.1 Convergence au sens de MSE

En utilisant les expressions finales des deux termes, le biais et la variance, l'erreur quadratique moyenne en un point θ est donnée par :

$$\begin{aligned}
 \text{MSE}[\widehat{f}_\nu(\theta)] &= \text{biais}^2[\widehat{f}_\nu(\theta)] + \mathbb{V}[\widehat{f}_\nu(\theta)] \\
 &= \left\{ \frac{1}{4}(1 - \gamma_2(\nu))f''(\theta) + o(1) \right\}^2 + \frac{1}{2n\pi} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) \right\} f(\theta) + o(1) \\
 &= \left\{ \frac{1}{16}(1 - \gamma_2(\nu))^2 \{f''(\theta)\}^2 \right\} + \frac{1}{2n\pi} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) \right\} f(\theta) + o(1). \quad (1.8)
 \end{aligned}$$

Théorème 1.1. (*Di Marzio et al. [2009]*) Si $\lim_{n \rightarrow \infty} \gamma_j(\nu) = 1$ pour $j \in \mathbb{Z}^+$, $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\nu) = 0$ et f'' est continue, alors l'estimateur $\widehat{f}_\nu(\theta)$ est consistant en moyenne quadratique en tout point θ pour lequel la densité f est continue, c'est-à-dire :

$$\lim_{n \rightarrow \infty} \text{MSE}[\widehat{f}_\nu(\theta)] = 0.$$

1.2.2.2 Convergence au sens de MISE

L'erreur quadratique moyenne intégrée de \widehat{f}_ν est obtenue en intégrant la fonction MSE

$$\begin{aligned}
 \text{MISE}[\widehat{f}_\nu] &= \int_0^{2\pi} \text{MSE}[\widehat{f}_\nu(\theta)] d\theta \\
 &= \int_0^{2\pi} \left[\frac{1}{16}(1 - \gamma_2(\nu))^2 \{f''(\theta)\}^2 + \frac{1}{2n\pi} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) \right\} f(\theta) + o(1) \right] d\theta. \\
 &= \frac{1}{16}(1 - \gamma_2(\nu))^2 \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{1}{2n\pi} \left\{ 1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) \right\} + o(1). \quad (1.9)
 \end{aligned}$$

Théorème 1.2. (*Di Marzio et al. [2009]*) Si $\lim_{n \rightarrow \infty} \gamma_j(\nu) = 1$ pour $j \in \mathbb{Z}^+$, $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} \gamma_j^2(\nu) = 0$ et f'' est continue et carré intégrable, alors l'estimateur $\widehat{f}_\nu(\theta)$ est consistant en moyenne quadratique intégrée, c'est-à-dire :

$$\lim_{n \rightarrow \infty} \text{MISE} \left\{ \widehat{f}_\nu(\theta) \right\} = 0.$$

1.2.3 Convergence en loi

La convergence en loi de l'estimateur à noyau circulaire est établie par [Di Marzio et al. \[2011a\]](#); voir aussi [Tsuruta and Sagae \[2017a\]](#) en utilisant le noyau particulier de Cauchy enveloppé.

Théorème 1.3. (*Di Marzio et al. [2011a]*) Soit $f : [0, 2\pi[\rightarrow \mathbb{R}$ une densité circulaire, $\theta \in [0, 2\pi[$ et $\widehat{f}_\nu(\theta)$ est l'estimateur à noyau circulaire de f . Si $\lim_{n \rightarrow \infty} \gamma_j(\nu) = 1$ pour $j \in \mathbb{Z}^+$, alors on a

$$\frac{\widehat{f}_\nu(\theta) - \mathbf{E}\{\widehat{f}_\nu(\theta)\}}{\sqrt{\mathbf{V}\{\widehat{f}_\nu(\theta)\}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

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

1.2.4 Estimateur à noyau de von Mises et ses propriétés

Nous présentons le noyau de von Mises, l'estimateur à noyau de von Mises et ses propriétés (voir [Taylor \[2008\]](#); [Di Marzio et al. \[2009\]](#)). Le noyau de von Mises est obtenu en utilisant la distribution de von Mises de moyenne de direction, la cible θ et de paramètre de concentration ν ; l'expression de ce noyau est donnée par :

$$K_\nu^{VM}(\phi) = \frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \phi)\},$$

où $I_0(\nu)$ désigne la fonction modifiée de Bessel de première espèce de l'ordre zéro. La [Figure 1.1](#) montre la forme du noyau de von Mises pour $\theta = \{\pi/4; \pi\}$ et pour différentes valeurs de $\nu = \{3, 5, 10, 15\}$.

Soient $\Theta_1, \dots, \Theta_n$ un n -échantillon aléatoire d'une densité circulaire continue bornée f sur $[0, 2\pi[$. En utilisant le noyau de von Mises dans la formule (1.3), l'estimateur à noyau de von Mises est donné par (voir [Taylor \[2008\]](#) et [Oliveira et al. \[2012\]](#)) :

$$\widehat{f}_\nu^{VM}(\theta) = \frac{1}{n(2\pi)I_0(\nu)} \sum_{i=1}^n \exp\{\nu \cos(\theta - \Theta_i)\}, \quad 0 \leq \theta < 2\pi. \quad (1.10)$$

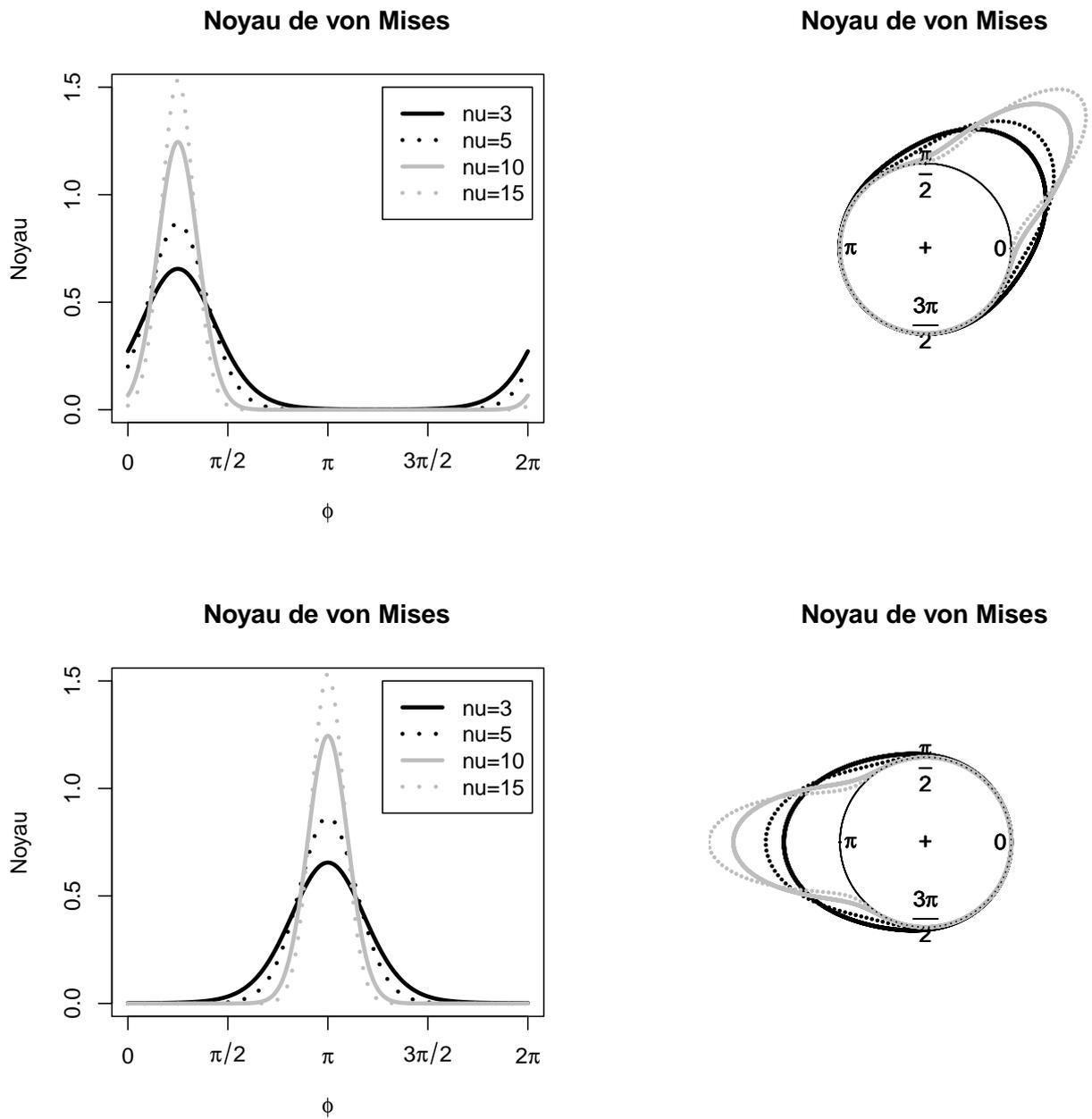


FIGURE 1.1: Noyau de von Mises pour $\theta = \pi/4$ (première rangée) et $\theta = \pi$ (deuxième rangée) avec différentes valeurs de $\nu = \{3, 5, 10, 15\}$.

où $I_j(\nu)$ est la fonction modifiée de Bessel d'ordre j , ν est le paramètre de concentration vérifiant $\nu = \nu(n) \rightarrow \infty$ quand $n \rightarrow \infty$ et θ est l'angle où on veut estimer le densité circulaire. Cet estimateur peut être interprété comme un mélange de n distributions de von Mises, centré en Θ_i et avec le même paramètre de concentration ν . Le coefficient γ_j donné dans (1.2) et la fonction de Bessel I_j sont reliés par la relation

$$\gamma_j(\nu) = \frac{I_j(\nu)}{I_0(\nu)}. \quad (1.11)$$

Cette relation peut être utilisée pour déduire les expressions du biais et de la variance de l'estimateur de von Mises.

1.2.4.1 Biais de l'estimateur de von Mises

Afin de déterminer le biais de l'estimateur de von Mises donné par (1.10), il suffit de remplacer $\gamma_2(\nu)$ par son expression donnée par (1.11), on obtient ainsi :

$$\text{biais}[\hat{f}_\nu^{VM}(\theta)] = \frac{f''(\theta)}{4} \left(1 - \frac{I_2(\nu)}{I_0(\nu)} \right) + o(1). \quad (1.12)$$

En utilisant $\frac{I_2(\nu)}{I_0(\nu)} = 1 - (1/\nu) + o(1/\nu)$ pour $\nu = \nu(n) \rightarrow \infty$ quand $n \rightarrow \infty$ (voir [Mardia and Jupp \[2000\]](#)), le biais asymptotique peut s'exprimer comme suit :

$$\text{biais}[\hat{f}_\nu^{VM}(\theta)] = \frac{f''(\theta)}{4\nu} + o(\nu^{-1}). \quad (1.13)$$

Notons que l'estimateur \hat{f}_ν^{VM} est asymptotiquement sans biais car $\lim_{\nu \rightarrow \infty} \text{biais}[\hat{f}_\nu^{VM}(\theta)] = 0$.

1.2.4.2 Variance de l'estimateur de von Mises

Pour obtenir la variance de l'estimateur de von Mises donné par (1.10), il suffit d'utiliser la relation $1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) = I_0(2\nu)/\{I_0(\nu)\}^2$ pour la distribution de von Mises (voir [Di Marzio et al. \[2009\]](#)), on obtient ainsi :

$$\mathbb{V}[\hat{f}_\nu^{VM}(\theta)] = \frac{I_0(2\nu)}{2n\pi\{I_0(\nu)\}^2} f(\theta) + o(1). \quad (1.14)$$

De plus, en utilisant $I_0(2\nu)/\{I_0(\nu)\}^2 \simeq \sqrt{\pi\nu}$ pour $\nu \rightarrow \infty$ (voir [Mardia and Jupp \[2000\]](#)), alors la variance peut être approximée par :

$$\mathbb{V}[\widehat{f}_\nu^{VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right). \quad (1.15)$$

La variance de \widehat{f}_ν^{VM} tend vers 0 quand $\nu^{1/2}/n \rightarrow 0$ lorsque $\nu \rightarrow \infty$ et $n \rightarrow \infty$.

1.2.4.3 Convergence de l'estimateur de von Mises au sens de MISE

En utilisant les expressions asymptotiques des deux termes, le biais et la variance, la (MISE) de l'estimateur de von Mises (1.10) est donnée par :

$$\text{MISE}[\widehat{f}_\nu^{VM}(\theta)] = \frac{1}{16\nu^2} \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{\nu^{1/2}}{2n\sqrt{\pi}} + o\left(\frac{1}{\nu^2} + \frac{\nu^{1/2}}{n}\right). \quad (1.16)$$

Théorème 1.4. *Si $\lim_{n \rightarrow \infty} \nu(n) = \infty$, $\lim_{n \rightarrow \infty} \nu^{1/2}(n)/n = 0$ et f'' est continue et carré intégrable, alors l'estimateur \widehat{f}_ν^{VM} est consistant en moyenne quadratique intégrée, c'est-à-dire :*

$$\lim_{n \rightarrow \infty} \text{MISE} \left\{ \widehat{f}_\nu^{VM}(\theta) \right\} = 0.$$

1.2.5 Estimateur à noyau de Cauchy enveloppé et ses propriétés

L'estimateur à noyau de Cauchy enveloppé et ses propriétés sont obtenues par [Tsuruta and Sagae \[2017a\]](#). Le noyau de Cauchy enveloppé est obtenu en utilisant la distribution de Cauchy enveloppé de moyenne de direction, la cible θ et de paramètre de concentration ρ ; l'expression de ce noyau est donnée par :

$$K_\rho^{WC}(\phi) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \phi)}.$$

La Figure 1.2 montre la forme du noyau de Cauchy enveloppé pour $\theta = \{\pi/4; \pi\}$ et pour différentes valeurs de $\rho = \{0.3, 0.5, 0.7, 0.9\}$.

Soient $\Theta_1, \dots, \Theta_n$ un n -échantillon aléatoire d'une densité circulaire continue bornée f sur $[0, 2\pi[$. En utilisant le noyau de Cauchy enveloppé dans la formule (1.3), l'estimateur à noyau de Cauchy enveloppé est défini comme suit (voir [Tsuruta and Sagae \[2017a\]](#)) :

$$\widehat{f}_\rho^{WC}(\theta) = \frac{1}{2n\pi} \sum_{i=1}^n \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \Theta_i)}, \quad 0 \leq \theta < 2\pi. \quad (1.17)$$

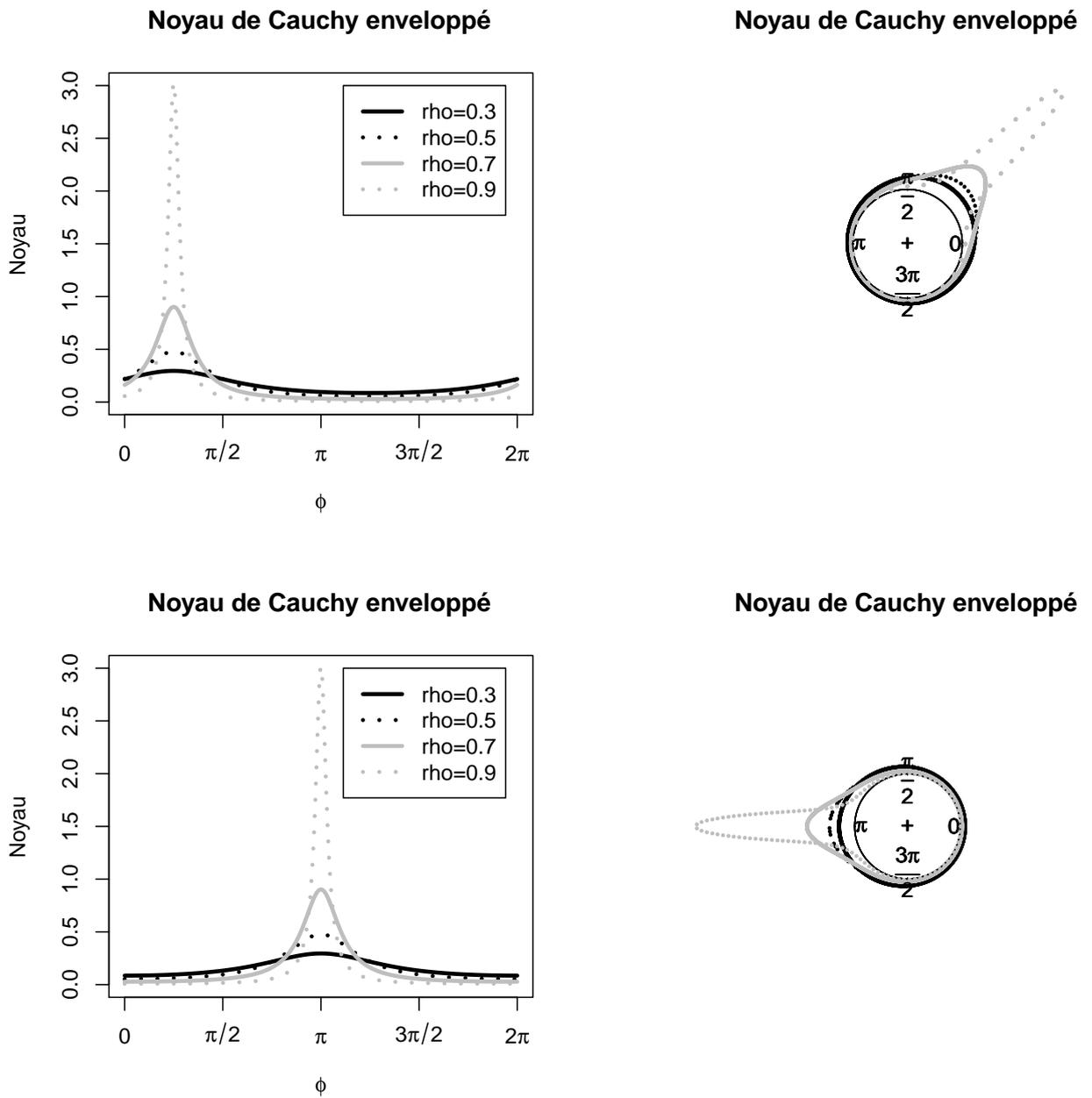


FIGURE 1.2: Noyau de Cauchy enveloppé pour $\theta = \pi/4$ (première rangée) et $\theta = \pi$ (deuxième rangée) avec différentes valeurs de $\rho = \{0.3, 0.5, 0.7, 0.9\}$.

où ρ est le paramètre de concentration vérifiant $\rho = \rho(n) \rightarrow 1$ quand $n \rightarrow \infty$ et θ est l'angle où on veut estimer la densité circulaire. Cet estimateur peut être interprété comme un mélange de n distributions de Cauchy enveloppé, centré en Θ_i et avec le même paramètre de concentration ρ .

1.2.5.1 Biais de l'estimateur à noyau de Cauchy enveloppé

Pour calculer le biais de l'estimateur de de Cauchy enveloppé donné par (1.17), il suffit de remplacer $1 - \gamma_2(\nu)$ par $1 - \rho^2$ (voir Tsuruta and Sagae [2017a] pour plus de détails), on obtient ainsi :

$$\text{biais}[\hat{f}_\rho^{WC}(\theta)] = \frac{f''(\theta)}{4} (1 - \rho^2) + o(1). \quad (1.18)$$

Notons que l'estimateur \hat{f}_ρ^{WC} est asymptotiquement sans biais puisque $\lim_{\rho \rightarrow 1} \text{biais}[\hat{f}_\rho^{WC}(\theta)] = 0$.

1.2.5.2 Variance de l'estimateur de Cauchy enveloppé

Pour approximer la variance de l'estimateur de Cauchy enveloppé donné par (1.17), il suffit d'utiliser la relation $1 + \sum_{j=1}^{\infty} \gamma_j^2(\nu) = (1 + \rho^2)/(1 - \rho^2) = \{2/(1 - \rho^2) - 1$ pour la distribution de Cauchy enveloppé (voir Tsuruta and Sagae [2017a]), on obtient :

$$\mathbb{V}[\hat{f}_\rho^{WC}(\theta)] = \frac{1}{n\pi(1 - \rho^2)} f(\theta) + o(1). \quad (1.19)$$

La variance de \hat{f}_ρ^{WC} tend vers 0 quand $n(1 - \rho^2) \rightarrow \infty$ lorsque $n \rightarrow \infty$.

1.2.5.3 Convergence de l'estimateur de Cauchy enveloppé au sens de MISE

En utilisant les expressions asymptotiques du biais et de la variance, la (MISE) de l'estimateur de Cauchy enveloppé (1.17) est donnée par :

$$\text{MISE}[\hat{f}_\rho^{WC}(\theta)] = \frac{\{1 - \rho^2\}^2}{16} \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{1}{n\pi\{1 - \rho^2\}} + o(\{1 - \rho^2\}^2 + n^{-1}\{1 - \rho^2\}). \quad (1.20)$$

Théorème 1.5. *Si $\lim_{n \rightarrow \infty} \rho(n) = 1$, $\lim_{n \rightarrow \infty} n(1 - \rho(n))^2 = \infty$ et f'' est continue et carré intégrable, alors l'estimateur \widehat{f}_ρ^{WC} est consistant en moyenne quadratique intégrée, c'est-à-dire :*

$$\lim_{n \rightarrow \infty} \text{MISE} \left\{ \widehat{f}_\rho^{WC}(\theta) \right\} = 0.$$

1.3 Estimateur à noyau toroïdal (cas multi-varié)

Dans cette section, nous présentons l'estimateur à noyau toroïdal (cas multi-varié) en utilisant la forme produit de la densité toroïdale f .

Définition 3 (Noyau toroïdal [Di Marzio et al. \[2011a\]](#)). *Un noyau toroïdal de dimension d et de vecteur de concentration $\mathcal{C} := \{\nu_s > 0, s = 1, \dots, d\}$ est le noyau produit $\mathbf{K}_{\mathcal{C}} := \prod_{s=1}^d K_{\nu_s}$, où K_{ν_s} est le noyau circulaire du paramètre de concentration $\nu_s > 0$ introduit par la Définition 1.*

Définition 4 (Estimateur de la densité par noyau toroïdal (cas multi-varié) [Di Marzio et al. \[2011a\]](#)). *Soient $\Theta_1, \dots, \Theta_n$ des vecteurs aléatoires iid d'une densité toroïdale f sur $([0, 2\pi]^d)$, où $\Theta_i = (\Theta_{i1}, \dots, \Theta_{id})^\top$. Etant donné un noyau toroïdal $\mathbf{K}_{\mathcal{C}}$, l'estimateur à noyau toroïdal (qu'on peut appeler aussi "estimateur à noyau circulaire multiple") de f en $\theta = (\theta_1, \dots, \theta_d)^\top \in ([0, 2\pi]^d)$ est défini par :*

$$\begin{aligned} \widehat{f}_{\mathcal{C}}(\theta) &= \frac{1}{n} \sum_{i=1}^n \mathbf{K}_{\mathcal{C}}(\theta - \Theta_i) \\ &= \frac{1}{n} \sum_{i=1}^n \prod_{s=1}^d K_{\nu_s}^{[s]}(\theta_s - \Theta_{is}). \end{aligned} \quad (1.21)$$

où $\mathcal{C} := \{\nu_s > 0, s = 1, \dots, d\}$ est le vecteur de concentration, $\theta = (\theta_1, \dots, \theta_d)^\top$ est le vecteur des angles, où on veut estimer f et $K_{\nu_s}^{[s]}$ est le s^{me} noyau circulaire.

En utilisant $K_{\nu_s}^{[s]}$ comme noyau de von Mises, nous obtenons l'estimateur à noyau de von Mises multiple donné par :

$$\widehat{f}_{\mathcal{C}}^{VM}(\theta) = \frac{1}{n} \sum_{i=1}^n \prod_{s=1}^d \frac{1}{(2\pi)I_0(\nu_s)} \exp\{\nu_s \cos(\theta_s - \Theta_{is})\}, \quad \theta \in ([0, 2\pi]^d).$$

Les propriétés asymptotiques de l'estimateur (1.21) sont établies par [Di Marzio et al. \[2011a\]](#).

1.4 Conclusion

Dans ce chapitre, nous avons exposé l'approche non-paramétrique par noyau circulaire dans l'estimation de la fonction de densité circulaire, particulièrement les estimateurs à noyaux de von Mises et de Cauchy enveloppé. Nous avons donné aussi les différents résultats théoriques concernant la convergence et la consistance de ces estimateurs (propriétés asymptotiques, convergence au sens de MSE et de MISE). Notons que l'estimateur à noyau circulaire dépend crucialement de paramètre de concentration. Dans le chapitre suivant, nous rappelons le principe des méthodes classiques pour le choix de paramètre de concentration, à savoir les méthodes de ré-injection (plug-in) et les méthodes de type validation croisée. En suite, nous développons l'approche bayésienne comme alternative de ces méthodes classiques pour la sélection du paramètre de concentration.

Chapitre 2

Choix du paramètre de concentration ν dans l'estimation non-paramétrique de la densité par noyau circulaire

Le choix du paramètre de concentration ν est d'une importance capitale dans l'estimation de la fonction densité par la méthode du noyau dans le cas des données circulaires. Plusieurs méthodes (dites classiques) pour choisir le paramètre de concentration ont été proposées dans la littérature, à savoir les méthodes de ré-injection (ou plug-in) et les techniques de validation croisée. Quelques études comparatives intéressantes ont été aussi effectuées sur plusieurs distributions circulaires. Nous présentons brièvement le principe de ces méthodes classiques pour le choix de ce paramètre dans le cas de l'estimateur à noyau de von Mises. En suite, nous proposons l'approche bayésienne pour estimer le paramètre de concentration dans l'estimation de la fonction densité par noyau circulaire. Au final, nous présentons une étude de simulation pour comparer les méthodes plug-in, les techniques de validation croisée et l'approche bayésienne proposée dans ce travail de thèse.

2.1 Méthodes de ré-injection (plug-in)

Dans le cas des méthodes plug-in, on cherche à approcher la valeur idéale du paramètre de concentration définie par :

$$\nu_{id} = \min_{\nu > 0} \text{MISE}(\nu). \quad (2.1)$$

Rappelons que l'expression asymptotique pour la MISE (AMISE) en utilisant le noyau de von Mises est donnée comme suit (voir [Oliveira et al. \[2012\]](#)) :

$$\text{AMISE}[\widehat{f}_\nu^{VM}(\theta)] = \frac{1}{16} \left[1 - \frac{I_2(\nu)}{I_0(\nu)} \right]^2 \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \left[\frac{I_0(2\nu)}{2n\pi(I_0(\nu))^2} \right]. \quad (2.2)$$

où f'' désigne la dérivée seconde de la densité f à estimer. Le paramètre de concentration (idéal) qui minimise l'AMISE n'est pas directement calculable car l'expression du AMISE dépend de la quantité inconnue f'' . Une façon classique de remédier à ce problème consiste à remplacer la quantité f'' par un estimateur approprié. Dans la suite, nous présentons deux techniques de plug-in.

2.1.1 Règle de référence (Rule of Thumb)

Cette technique a été proposée par [Taylor \[2008\]](#) en adoptant l'idée de [Silverman \[1986\]](#) proposée dans le cas des données linéaires. Elle consiste à choisir f comme étant la distribution de von Mises du paramètre de concentration k . Dans ce cas, l'expression AMISE donnée par (2.2) s'écrit comme suit :

$$\text{AMISE}(\nu) = \frac{3k^2 I_2(2k)}{32\pi\nu^2 I_0(k)^2} + \frac{\nu^{1/2}}{2n\pi^{1/2}}. \quad (2.3)$$

Par conséquent, le paramètre de concentration qui minimise l'AMISE donnée par (2.3) peut être estimé par

$$\hat{\nu}_{RT} = \left[\frac{3n\widehat{k}^2 I_2(2\widehat{k})}{4\pi^{1/2} I_0(\widehat{k})^2} \right]^{\frac{2}{5}}, \quad (2.4)$$

où \widehat{k} est estimé à l'aide des observations $\Theta_1, \Theta_2, \dots, \Theta_n$ en utilisant la méthode de maximum de vraisemblance. Cette approche donne de bons résultats si la population est réellement distribuée suivant une loi de von Mises. Mais elle peut donner une distribution trop lissée si la population est plutôt multimodale. Une autre méthode de plug-in plus robuste a été proposée par [Oliveira et al. \[2012\]](#) ; voir la section suivante.

2.1.2 Méthode plug-in de Oliveira et al. (2012)

Cette alternative a été proposée par Oliveira et al. [2012] et peut être considérée comme une généralisation de l'approche de Taylor [2008]. Les auteurs choisissent ici d'estimer f par un mélange de densités de von Mises. La densité d'un mélange fini de M densités de von Mises $vM(\mu_r, k_r)$ est donnée par :

$$g(\theta) = \sum_{r=1}^M \alpha_r \frac{\exp\{k_r(\theta - \mu_r)\}}{2\pi I_0(k_r)}, \quad \text{avec} \quad \sum_{r=1}^M \alpha_r = 1. \quad (2.5)$$

La valeur approchée $\nu_{\hat{P}I}$ du paramètre de concentration ν est obtenue en utilisant les étapes suivantes :

- **Etape 1** Sélectionner le nombre M pour la distribution g donnée par (2.5).
- **Etape 2** Estimer l'AMISE comme suit :
 - **Etape 2.1** Estimer les paramètres (α_r, μ_r, k_r) de la distribution g donnée par (2.5).
 - **Etape 2.2** Calculer l'intégrale $\int (g''(\theta))^2 d\theta$ en remplaçant f par g .
 - **Etape 2.3** Substituer la quantité $\int (g''(\theta))^2 d\theta$ dans l'AMISE
- **Etape 3** Minimiser $AMISE(\nu)$ et obtenir $\nu_{\hat{P}I}$.

Notons que dans l'étape 1, le nombre M est calculé en utilisant le critère d'information d'Akaike (AIC). Dans l'étape 2.1, les paramètres α_r, μ_r et $k_r, r = 1, \dots, M$ sont estimés par la méthode de maximum de vraisemblance via l'algorithme EM. Finalement, l'intégrale $\int (g''(\theta))^2 d\theta$ est calculée en utilisant les méthodes de Monte Carlo. Notons que le paramètre de concentration ν_{RT} obtenu par la règle de référence correspond au nombre $M = 1$.

2.2 Méthodes de validation croisée

L'idée de base des méthodes de validation croisée consiste à trouver une fonction de score $CV(\nu)$ dont le calcul soit plus simple que la $MISE(\nu)$. Notons que les méthodes de validation croisée ne font pas usage des approximations des dérivées de f . Le paramètre de concentration ν obtenu par ce type de méthode n'est pas déterministe car il dépend des observations.

2.2.1 Validation croisée par le maximum de vraisemblance

Cette approche a été proposée par [Habbema et al. \[1974\]](#) dans le cas des données linéaires. Cette approche peut être adaptée facilement dans le cas des données circulaires (voir [Oliveira et al. \[2012\]](#) et [Oliveira et al. \[2014a\]](#)).

Soient $\Theta_1, \Theta_2, \dots, \Theta_n$ un n-échantillon aléatoire d'angles indépendant et identiquement distribué. Nous déterminons le paramètre de concentration optimal $\hat{\nu}_{LCV}$, tel que

$$\hat{\nu}_{lcv} = \max_{\nu} LCV(\nu),$$

qui maximise le critère

$$LCV(\nu) = \prod_{i=1}^n \hat{f}_{\nu,i}^{VM}(\Theta_i), \quad (2.6)$$

où

$$\hat{f}_{\nu,i}^{VM}(\Theta_i) = \frac{1}{(n-1)} \sum_{j \neq i} K_{\nu}(\Theta_i - \Theta_j), \quad (2.7)$$

est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation Θ_i .

En utilisant le noyau de von Mises, les étapes pour calculer le paramètre de concentration optimal $\hat{\nu}_{LCV}$ sont donnés comme suit :

- **Etape 1** Calculer $LCV(\nu) = \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j \neq i} \frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\Theta_i - \Theta_j)\}$.
- **Etape 2** Maximiser $LCV(\nu)$ et obtenir $\hat{\nu}_{LCV}$.

[Oliveira et al. \[2012\]](#) ont montré que la méthode de validation croisée par le maximum de vraisemblance $\hat{\nu}_{LCV}$ donne des résultats raisonnables sauf pour le cas des distributions qui présentent de fortes pic.

2.2.2 Validation croisée non biaisée

La méthode de validation croisée non biaisée a été proposée par [Rudemo \[1982\]](#) et [Bowman \[1984\]](#) pour le cas des données linéaires ; voir aussi [Scott and Terrel \[1987\]](#). Une adaptation pour le cas des données circulaires est donnée par [Oliveira et al. \[2012\]](#) et [Oliveira et al. \[2014a\]](#).

Considérons $\Theta_1, \Theta_2, \dots, \Theta_n$ un n-échantillon aléatoire d'angles indépendant et identiquement distribué. Cette approche consiste à choisir le paramètre de concentration qui

minimise un estimateur convenable de $ISE(\nu)$ qui est donné par :

$$\begin{aligned} ISE(\nu) &= \int \left(\widehat{f}_\nu^{VM}(\theta) - f(\theta) \right)^2 \\ &= \int (\widehat{f}_\nu^{VM}(\theta))^2 d\theta - 2 \int \widehat{f}_\nu^{VM}(\theta) f(\theta) d\theta + \int f^2(\theta) d\theta. \end{aligned}$$

Puisque $\int f^2(\theta) d\theta$ ne dépend pas du paramètre de concentration ν , alors on peut choisir le paramètre de concentration de façon à ce qu'il minimise un estimateur de

$$CV(\nu) = ISE(\nu) - \int f^2(\theta) d\theta.$$

Remarquons que $\frac{1}{n} \sum_{i=1}^n \widehat{f}_{\nu,i}^{VM}(\Theta_i)$ est un estimateur sans biais de $\int \widehat{f}_\nu^{VM}(\theta) f(\theta) d\theta$, où $\widehat{f}_{\nu,i}^{VM}(\Theta_i)$ est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation Θ_i donné par (2.7). Le critère à optimiser devient alors :

$$UCV(\nu) = \int \left\{ \frac{1}{n} \sum_{i=1}^n K_\nu(\theta - \Theta_i) \right\}^2 d\theta - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} K_\nu(\Theta_i - \Theta_j).$$

où K_ν est le noyau de von Mises. Les étapes pour calculer le paramètre de concentration optimal $\hat{\nu}_{UCV}$ sont donnés comme suit :

- **Etape 1** Calculer l'intégrale $\int \left\{ \frac{1}{n(2\pi)I_0(\nu)} \sum_{i=1}^n \exp\{\nu \cos(\theta - \Theta_j)\} \right\}^2 d\theta$.
- **Etape 2** Calculer $\frac{2}{n(n-1)(2\pi)I_0(\nu)} \sum_{i=1}^n \sum_{j \neq i} \exp\{\nu \cos(\Theta_i - \Theta_j)\}$.
- **Etape 3** Calculer $UCV(\nu)$.
- **Etape 4** Minimiser $UCV(\nu)$ et obtenir $\hat{\nu}_{UCV}$.

La méthode de validation croisée non biaisé se révèle variable d'un échantillon à l'autre et peut présenter plusieurs minimums locaux. Une étude de simulation conduite par [Oliveira et al. \[2012\]](#) a montré que la méthode de validation croisée par le maximum de vraisemblance ($\hat{\nu}_{LCV}$) est plus stable que la méthode de validation croisée non biaisé ($\hat{\nu}_{UCV}$).

Notons aussi que les techniques de validation croisée biaisé et lissée, proposées respectivement par [Scott and Terrel \[1987\]](#) et [Hall et al. \[1992\]](#) dans le cas des données linéaires peuvent être adaptées dans le cas de l'estimateur à noyau circulaire.

2.3 Approche bayésienne pour le choix du paramètre de concentration ν

2.3.1 Stratégie

La stratégie de l'estimation bayésienne pour le choix du paramètre de concentration dans l'estimation de la densité par noyau circulaire consiste à

1. Donner la forme de l'estimateur de la vraisemblance $\pi(\theta_1, \dots, \theta_n | \nu)$ des données $\theta_1, \dots, \theta_n$ sachant le paramètre de lissage ν ;
2. Choisir la loi a priori de ν de densité $\pi(\nu)$ sur le paramètre ν ;
3. Calculer la distribution a posteriori $\pi(\nu | \theta_1, \dots, \theta_n)$ du paramètre ν quand on dispose des données $\theta_1, \dots, \theta_n$ en utilisant le théorème de Bayes ;
4. Estimer le paramètre de lissage ν en utilisant la perte quadratique (ou la perte d'entropie) et les méthodes MCMC.

2.3.1.1 Approximation de la vraisemblance

L'estimateur de la vraisemblance est obtenu en utilisant l'estimateur à noyau circulaire de la fonction densité donné par (1.3) et la technique de validation croisée. Explicitement, nous considérons une suite de variable aléatoire $\Theta_1, \dots, \Theta_n$ indépendantes de même loi et de densité de probabilité circulaire inconnue f et des réalisations $\theta = (\theta_1, \dots, \theta_n)$. L'estimateur de la vraisemblance conditionnel des données θ sachant le paramètre ν , s'obtient en estimant $f(\theta_i)$ à partir de l'ensemble des points sauf le point θ_i en utilisant l'estimateur à noyau circulaire ; il s'écrit comme suit :

$$\begin{aligned} LCV(\theta_1, \dots, \theta_n; \nu) &= \widehat{\pi}(\theta_1, \dots, \theta_n | \nu) = \prod_{i=1}^n \widehat{f}_{\nu,i}(\theta_i) \\ &= \frac{1}{(n-1)^n} \prod_{i=1}^n \sum_{j=1, j \neq i}^n K_\nu(\theta_i - \theta_j). \end{aligned} \quad (2.8)$$

Cet estimateur de vraisemblance va servir dans l'estimation bayésienne au calcul de la loi a posteriori. Nous présentons dans la suite le choix de la loi a priori et le calcul de

la densité a posteriori, en utilisant le noyau circulaire de von Mises.

2.3.1.2 Choix de la loi a priori

L'inférence par l'approche bayésienne est caractérisée par le choix d'une loi a priori. Notons que le choix n'est pas unique. L'utilisation de la loi a priori conjuguée dans ce modèle n'est pas bénéfique, car la vraisemblance de validation croisée donnée par (2.8) s'écrit comme le produit de n termes, et le paramètre ν intervient dans chaque terme. Dans ce cas, le choix se fait pour que les méthodes MCMC fonctionnent correctement.

Maintenant, comme l'information est disponible a priori sur ν ($\nu > 0$ et $\nu \rightarrow \infty$, lorsque $n \rightarrow \infty$), alors la distribution gamma avec le paramètre de forme α large et le paramètre d'échelle β fixe, peut être un choix naturel comme une loi a priori sur le paramètre de concentration ν ; cette loi a priori est donné par

$$\pi(\nu|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \nu^{\alpha-1} \exp\{-\beta\nu\}, \quad \nu, \alpha, \beta > 0. \quad (2.9)$$

Le choix des paramètres α et β sera discuté dans la Section 2.3.3.

2.3.1.3 Approximation de la loi a posteriori

En utilisant le théorème de Bayes, la loi a posteriori de ν prend la forme suivante :

$$\hat{\pi}(\nu|\theta_1, \dots, \theta_n) = \frac{\pi(\nu) \prod_{i=1}^n \hat{f}_{\nu,i}(\theta_i)}{\int \pi(\nu) \prod_{i=1}^n \hat{f}_{\nu,i}(\theta_i) d\nu}. \quad (2.10)$$

En utilisant le noyau de von Mises, la densité a posteriori du paramètre ν sachant les données $(\theta_1, \dots, \theta_n)$ est donnée par :

$$\hat{\pi}(\nu|\theta_1, \dots, \theta_n) = \frac{\frac{\nu^{\alpha-1} \exp\{-\beta\nu\}}{(I_0(\nu))^n} \prod_{i=1}^n \sum_{j=1, i \neq j}^n \exp\{\nu \cos(\theta_i - \theta_j)\}}{\int \frac{\nu^{\alpha-1} \exp\{-\beta\nu\}}{(I_0(\nu))^n} \prod_{i=1}^n \sum_{j=1, i \neq j}^n \exp\{\nu \cos(\theta_i - \theta_j)\} d\nu}. \quad (2.11)$$

Les estimations bayésiennes de ν peuvent être calculés sous les fonctions de perte quadratique et d'entropie. Pour la fonction de perte quadratique, l'estimateur de Bayes est la moyenne posteriori donnée par

$$\nu_{Bayes1} = \mathbb{E}(\nu|\theta_1, \dots, \theta_n) = \int_0^\infty \nu \hat{\pi}(\nu|\theta_1, \dots, \theta_n) d\nu, \quad (2.12)$$

et pour la fonction de perte d'entropie, l'estimation de Bayes est donnée par

$$\nu_{Bayes2} = \left\{ \mathbb{E}(\nu^{-1} | \theta_1, \dots, \theta_n) \right\}^{-1} = \left\{ \int_0^\infty \nu^{-1} \hat{\pi}(\nu | \theta_1, \dots, \theta_n) d\nu \right\}^{-1}. \quad (2.13)$$

La densité a posteriori (2.11) n'est pas traitable analytiquement, et par conséquent, les estimations de Bayes données par (2.12) et (2.13) n'ont pas de solutions analytiques. Dans cette section, nous proposons d'utiliser les méthodes MCMC pour l'estimation bayésienne. Il convient de noter que les techniques d'échantillonnage MCMC ont été largement utilisées par plusieurs auteurs pour la sélection du paramètre de lissage dans l'estimation linéaire par noyau de la densité; voir, par exemple Brewer [1998], Brewer [2000], Zhang et al. [2006], Zougab et al. [2013], Zougab et al. [2016] et Li et al. [2015]. Notons que, les intégrales unidimensionnelles données par (2.12) et (2.13) peuvent être calculées en utilisant la quadrature numérique simple ou par la méthode de Monte Carlo avec des échantillons aléatoires générés à partir de la loi gamma. Cependant, nous préférons utiliser la technique MCMC car elle peut facilement être étendue pour le cas (circulaire) multi-varié; voir Zhang et al. [2006], Belaid [2016] et Belaid et al. [2016] pour l'estimation de matrice de lissage pour le cas (linéaire) multi-varié.

2.3.2 MCMC pour estimer le paramètre de concentration

La méthode MCMC constitue une des approches les plus utilisées par la communauté statistique. Cette méthode utilise une chaîne de Markov auxiliaire dont la distribution stationnaire est la distribution d'intérêt. Le principal intérêt de l'approche MCMC est sa facilité d'application à des distributions complexes et/ou de grandes dimensions. En plus, un autre grand avantage est le fait que la constante de normalisation de la densité cible ne doit généralement pas être spécifiée.

2.3.2.1 Principe de la méthode MCMC

La méthode MCMC se base sur la construction d'une chaîne de Markov $\{\nu^{(t)}\}$ ergodique à partir d'un noyau de transition qui garanti la convergence en loi vers la loi cible. Par conséquent, il suffit de choisir un état initial arbitraire $\{\nu^{(0)}\}$ et ensuite simuler une chaîne de Markov afin de générer un échantillon aléatoire provenant de la loi a posteriori

(la loi cible). Une première tranche d'observations appelée période de chauffe (*burn-in*), est laissée de côté parce que l'on considère que la chaîne n'a pas encore convergée. Au-delà de la période de chauffe, les valeurs sont considérées comme un échantillon provenant de la distribution cible.

2.3.2.2 Algorithme de Metropolis-Hasting (M-H)

L'algorithme de Metropolis-Hastings, introduit par [Metropolis et al. \[1953\]](#) et généralisé par [Hastings \[1970\]](#), est considéré comme l'une des premières méthodes MCMC. Il existe plusieurs algorithmes M-H : l'algorithme M-H indépendant, l'algorithme M-H à marche aléatoire et l'algorithme M-H dit à une variable à la fois.

Le principe de l'algorithme est de construire une chaîne de Markov à partir d'une loi de proposition (instrumentale) $q(\nu/\tilde{\nu})$ facilement simulable. Nous proposons dans ce travail d'utiliser l'algorithme M-H à marche aléatoire. Ces étapes sont données comme suit :

- **Étape 1.** initialiser ν^0 et $\chi(0)$.
- **Étape 2.** pour $k \in \{1, \dots, N\}$.
 - (a) Générer $\tilde{\nu} \sim N(\nu^{(k-1)}\tau_{k-1}^2)$,
 - (b) $\tilde{\nu} = |\tilde{\nu}|$,
 - (c) Calculer la probabilité d'acceptation

$$\rho = \min \left\{ 1, \frac{\hat{\pi}(\tilde{\nu}|\theta_1, \dots, \theta_n)}{\hat{\pi}(\nu^{(k-1)}|\theta_1, \dots, \theta_n)} \times \frac{q(\nu^{(k-1)}|\tilde{\nu}, \chi^{(k-1)})}{q(\tilde{\nu}|\nu^{(k-1)}, \chi^{(k-1)})} \right\} = \min \left\{ 1, \frac{\hat{\pi}(\tilde{\nu}|\theta_1, \dots, \theta_n)}{\hat{\pi}(\nu^{(k-1)}|\theta_1, \dots, \theta_n)} \right\},$$
 - d) Calculer l'état de la chaîne,

$$\nu^{(k)} = \begin{cases} \tilde{\nu}, & \text{si } u < \rho, \quad u \sim U_{[0,1]}, \\ \nu^{(k-1)}, & \text{sinon,} \end{cases}$$

- (e) Mettez à jour le paramètre de réglage χ comme suit :

$$\chi^{(k)} = \begin{cases} \chi^{(k-1)} + \frac{\chi^{(k-1)}}{k\tau}, & \text{si } \tilde{\nu} \text{ est accepté,} \\ \chi^{(k-1)} - \frac{\chi^{(k-1)}}{k(1-\tau)}, & \text{si } \tilde{\nu} \text{ est rejeté,} \end{cases}$$

- **Étape 3.** $k = k + 1$ et aller à 2.
- Calculer l'estimateur de Bayes

$$\nu_{Bayes1} = \frac{1}{N - N_0} \sum_{k=N_0+1}^N \nu^{(k)},$$

$$\nu_{Bayes2} = \left\{ \frac{1}{N - N_0} \sum_{k=N_0+1}^N \{\nu^{(k)}\}^{-1} \right\}^{-1},$$

où N_0 correspond à la période de chauffe et τ au taux d'acceptation égal à 0,44; voir [Roberts and Rosenthal \[2009\]](#) et [Garthwaite et al. \[2016\]](#). Afin d'éviter en pratique des dépassements numériques, nous avons également modifié la probabilité d'acceptation ρ comme suit (voir [Brewer \[1998\]](#)),

$$\hat{\rho} = \begin{cases} \min\{\varrho, 0\}, & \text{si } \hat{\pi}(\nu^{(k-1)}|\theta_1, \dots, \theta_n) > 0, \\ 0, & \text{si } \hat{\pi}(\nu^{(k-1)}|\theta_1, \dots, \theta_n) = 0, \end{cases}$$

où

$$\varrho = \log \{ \hat{\pi}(\tilde{\nu}|\theta_1, \dots, \theta_n) \} - \log \{ \hat{\pi}(\nu^{(i-1)}|\theta_1, \dots, \theta_n) \}.$$

La convergence des méthodes MCMC peut être examinée par l'erreur standard moyenne par lot (BMSE) et le facteur d'inefficacité de simulation (SIF) qui peuvent être interprétés comme le nombre de tirages nécessaires pour obtenir des tirages indépendants. Ces indicateurs ont été largement utilisés dans la littérature, voir, par exemple, [Zhang et al. \[2006\]](#), [Belaid et al. \[2016\]](#) et [Senga Kiessé et al. \[2016\]](#). Nous donnons un aperçu rapide du calcul de BMSE et SIF. Ces critères consistent à subdiviser la chaîne de Markov en m chaînes de longueur n_c :

$$\text{BMSE} = \sqrt{\hat{v}^2/(N - N_0)}; \quad \text{SIF} = \hat{v}^2/\tilde{v}^2,$$

où

$$\hat{v}^2 = n_c/(m - 1) \sum_{j=1}^m (\bar{\nu}_j - \bar{\nu})^2 \quad \text{et} \quad \tilde{v}^2 = 1/(N - N_0 - 1) \sum_{i=N_0+1}^N (\nu^{(i)} - \bar{\nu})^2,$$

avec

$$\bar{\nu} = 1/(N - N_0) \sum_{i=N_0+1}^N \nu^{(i)} \quad \text{et} \quad \bar{\nu}_j = 1/n_c \sum_{i=(j-1)n_c+1}^{jn_c} \nu^{(i)}, \quad j = 1, \dots, m.$$

2.3.3 Choix des paramètres de la loi a priori gamma

Notons que la mise en œuvre et les performances de l'approche bayésienne dépendent du choix des paramètres α et β de la loi a priori gamma. En suivant l'idée de [Zougab et al. \[2016\]](#), nous pouvons choisir $\alpha = n^{2/5}$ et fixer le paramètre d'échelle $\beta \ll \alpha$. Ce choix de

α qui dépend de la taille de l'échantillon n permet également la divergence des estimateurs bayésiens (2.12) et (2.13) à l'infini lorsque $n \rightarrow \infty$ avec un taux de convergence proche de celui du paramètre de concentration optimal au sens de (MISE) donné dans Taylor [2008] et Oliveira et al. [2012].

2.4 Résultats de la simulation

Dans cette section, nous établissons une étude de simulation pour comparer les performances de l'approche bayésienne globale à certains sélecteurs classiques pour le choix du paramètre de concentration ν , via plusieurs distributions circulaires de complexité variable (von Mises, cardioïde, diverses distributions enveloppées et leurs mélanges). En particulier, nous comparons notre approche avec les sélecteurs suivants :

- La règle de référence (RT) proposée par Taylor [2008] ;
- La règle de plug-in (PI) développée par Oliveira et al. [2012] ;
- Les techniques de validation croisée (LCV et UCV) (voir Hall et al. [1987], Taylor [2008] et Oliveira et al. [2012]).

Nous avons utilisé 20 modèles circulaires considérés par Oliveira et al. [2012] dans leur étude de simulation. Les formules de ces modèles sont données dans Jammalamadaka and SenGupta [2001] et Oliveira et al. [2012] classées en quatre groupes, en fonction de leur complexité (voir aussi la Figure 2.1 et l'annexe pour les expressions de ces distributions) :

1. Modèles simples : uniforme circulaire (M1) ; von Mises (M2) ; Normal enveloppé (M3) ; cardioïde (M4) ; Cauchy enveloppé (M5) et Normal enveloppé (M6).
2. Modèles à deux composants : mélanges de von Mises (M7, M8 et M9) ; mélange de von Mises et de Cauchy enveloppé (M10).
3. Modèles comportant plus de deux composants : mélanges de von Mises à trois composants (M11, M12 et M13) ; mélange de von Mises à quatre composants (M14) ; mélange de Cauchy enveloppé, Normal enveloppé, von Mises et Normal enveloppé (M15) ; mélange de von Mises à cinq composants (M16).
4. Autres modèles complexes : mélange de cardioïde et Cauchy enveloppé (M17) ; mélange de von Mises (M18 et M19) ; mélange de deux Normal enveloppés et deux

Cauchy enveloppés (M20).

Maintenant, les modèles (M1-M20) sont donnés comme suit (voir [Oliveira et al. \[2012\]](#) et [Oliveira et al. \[2014a\]](#)) :

1. M1 est un modèle uniforme circulaire.
2. M2 est un modèle de von Mises ; $vM(\pi, 1)$.
3. M3 est un modèle normal enveloppé ; $WN(\pi, 0.9)$.
4. M4 est un modèle cardioïde ; $C(\pi, 0.5)$.
5. M5 est un modèle de cauchy enveloppé ; $WC(\pi, 0.8)$.
6. M6 est un modèle skew-Normal enveloppé ; $WSN(\pi, 1, 20)$.
7. M7 est un mélange de deux distributions de von Mises ; $1/2vM(0, 4) + 1/2vM(\pi, 4)$.
8. M8 est un mélange de deux distributions de von Mises ; $1/2vM(2, 5) + 1/2vM(4, 5)$.
9. M9 est un mélange de deux distributions de von Mises ; $1/4vM(0, 2) + 3/4vM(\pi/\sqrt{3}, 2)$.
10. M10 est un mélange de von Mises et de distributions de Cauchy enveloppées ;

$$4/5vM(\pi, 5) + 1/5WC(4\pi/3, 0.9)$$
11. M11 est un mélange de trois distributions de von Mises ; $1/3vM(\pi/3, 6) + 1/3vM(\pi, 6) + 1/3vM(5\pi, 6)$.
12. M12 est un mélange de trois distributions de von Mises ; $2/5vM(\pi/2, 4) + 1/5vM(\pi, 5) + 2/5vM(3\pi, 4)$.
13. M13 est un mélange de trois distributions de von Mises ; $2/5vM(0.5, 6) + 2/5vM(3, 6) + 1/5vM(5, 24)$.
14. M14 est un mélange de quatre distributions de von Mises ; $1/4vM(0, 12) + 1/4vM(\pi/2, 12) + 1/4vM(\pi, 12) + 1/4vM(3\pi, 12)$.
15. M15 est un mélange de distributions Cauchy enveloppées, Normal enveloppé , von Mises et skew-Normal enveloppé ; $3/10WC(\pi - 1, 0.6) + 1/4WN(\pi + 0.5, 0.9) + 1/4vM(\pi + 2, 3) + 1/5WSN(6, 1, 3)$.
16. M16 est un mélange de cinq distributions de von Mises ; $1/5vM(\pi/5, 18) + 1/5vM(3\pi/5, 18) + 1/5vM(\pi, 18) + 1/5vM(7\pi/5, 18) + 1/5vM(9\pi/5, 18)$.

17. M17 est un mélange de distributions cardioïde et Cauchy enveloppé ; $2/3C(\pi, 0.5) + 1/3WC(\pi, 0.9)$.
18. M18 est un mélange de quatre distributions de von Mises ; $1/2vM(\pi, 1) + 1/6vM(\pi - 0.8, 30) + 1/6vM(\pi, 30) + 1/6vM(\pi + 0.8, 30)$.
19. M19 est un mélange de cinq distributions de von Mises ; $4/9vM(2, 3) + 5/36vM(4, 3) + 5/36vM(3.6, 50) + 5/36vM(4, 50) + 5/36vM(4.5, 50)$.
20. M20 est un mélange de deux distributions skew-Normal enveloppées et de deux distributions Cauchy enveloppées ; $1/3WSN(0, 0.7, 20) + 1/3WSN(\pi, 0.7, 20) + 1/6WC(3\pi/4, 0.9) + 1/6WC(7\pi/4, 0.9)$.

2.4.1 Implémentation bayésienne et convergence des méthodes MCMC

L'estimateur à noyau de von Mises combiné avec le paramètre de concentration optimal bayésien a été implémenté sous R. Rappelons que la performance de l'approche bayésienne dépend du choix des paramètres α et β de la loi a priori gamma. Pour assurer la convergence de l'estimateur à noyau de von Mises avec les sélecteurs bayésiens, nous avons choisi $\alpha = n^{2/5}$ et on a fixé $\beta \in \{1, 1.5, 2.5, 3, 3.5, 4\}$; voir Section 2.3.3. Cependant, pour économiser de l'espace, nous présentons que les résultats avec $\beta = 1.5$ (en général, les résultats pour les différentes valeurs de β sont assez similaires). Nous avons utilisé l'algorithme M-H à marche aléatoire présenté dans la Section 2.3.2. Le taux d'acceptation est fixé à 0.44, la période de chauffe N_0 contient 1000 itérations et les 4000 itérations suivantes ont été enregistrées pour l'estimation de ν_{Bayes1} et ν_{Bayes2} . Pour mesurer la convergence de la méthode MCMC, nous avons calculé les valeurs BMSE et SIF. Nous avons utilisé le modèle M7 pour illustrer les résultats. Le Tableau 2.1 indique la taille de l'échantillon, l'estimation de Bayes (ν_{Bayes1} et ν_{Bayes2}), le BMSE et le SIF. Les tracés de la chaîne $\nu^{(i)}$ pour $i = N_0 + 1, \dots, N$ et la densité a posteriori de ν sont également donnés dans la Figure 2.2 pour le modèle M7 avec une taille d'échantillon $n = 50$. Les valeurs des indicateurs et des graphiques montrent que les chaînes simulées se sont très bien mélangées.

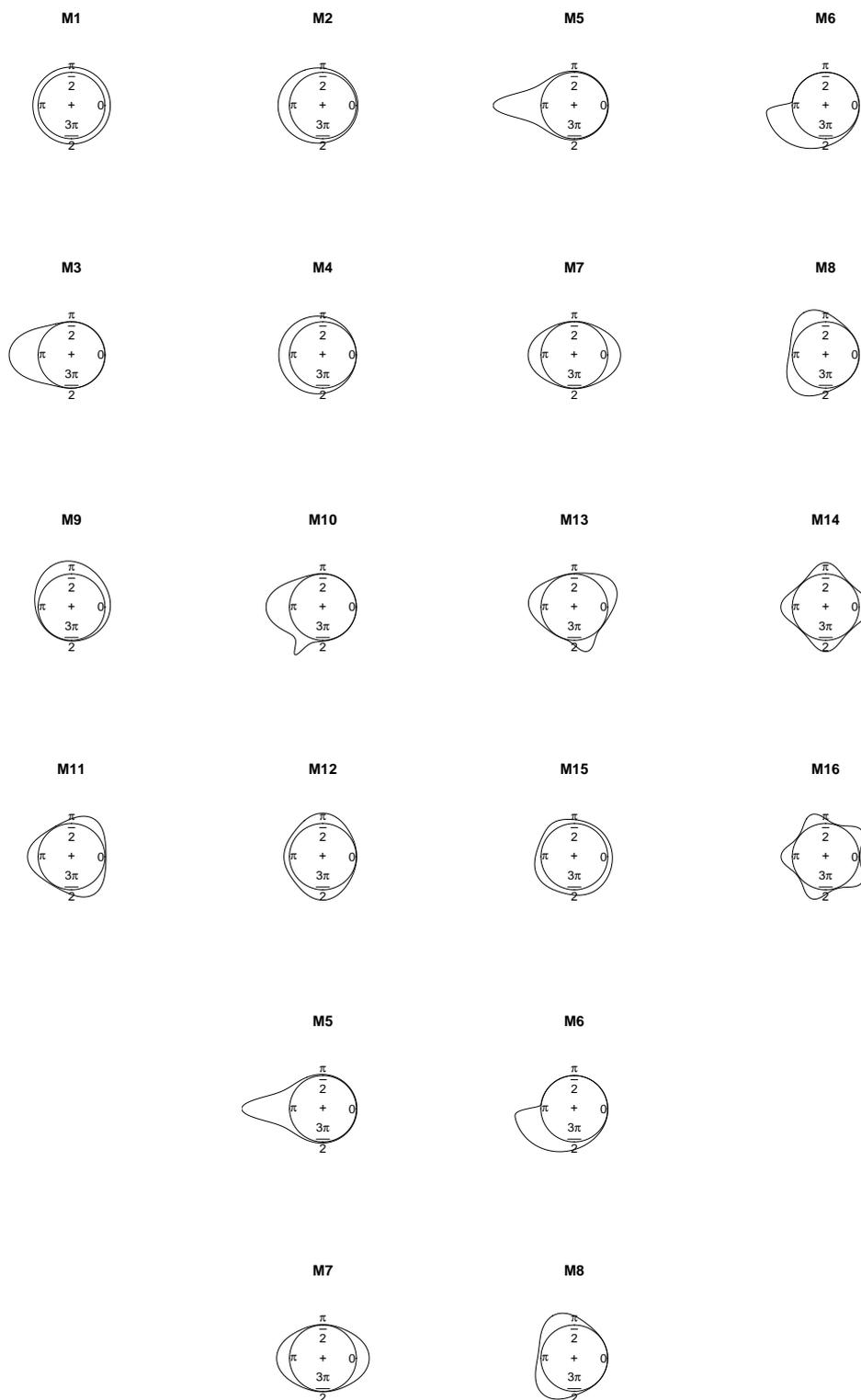


FIGURE 2.1: Les modèles de simulation.

TABLE 2.1: *Estimateurs bayésiens du paramètre ν et les indicateurs BMSE et SIF associés obtenus pour M7.*

Modèle	n	ν_{Bayes1}	ν_{Bayes2}	$BMSE$	SIF
M7	10	4.9972	3.2890	0.1150	3.9997
	25	7.5104	6.4312	0.1148	6.0685
	50	7.0346	6.3706	0.0717	4.1781
	100	16.1461	13.0091	0.5391	10.6465
	200	21.7088	18.3260	0.2730	3.4843
	500	31.2948	27.6741	0.5992	10.2322

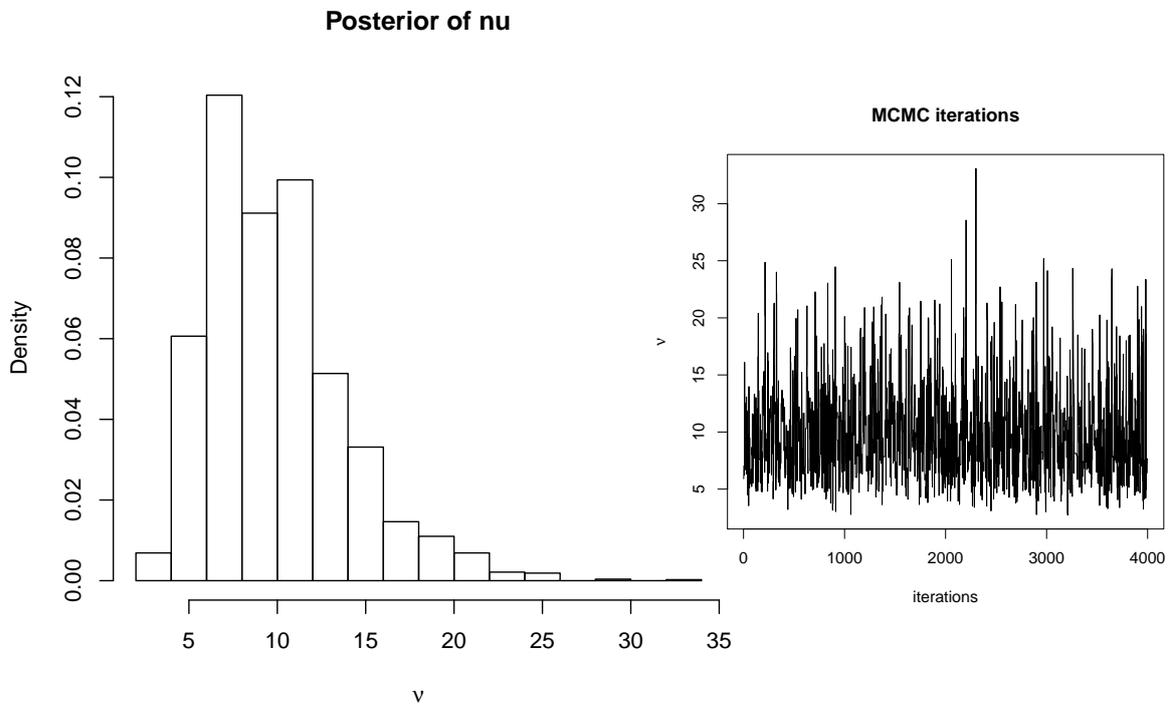


FIGURE 2.2: *Densité a posteriori et la chaîne MCMC de ν obtenues à partir de données simulées avec le modèle M7 pour $n = 50$.*

2.4.2 Approche bayésienne globale contre les approches classiques

Dans cette section, nous examinons la performance de notre approche bayésienne proposée par rapport à celle des approches classiques (techniques de plug-in et de validation

croisée) via l'erreur quadratique intégrée (ISE) donnée par :

$$\text{ISE}(\nu) = \int_0^{2\pi} \left(\widehat{f}_\nu(\theta) - f(\theta) \right)^2 d\theta.$$

Pour chaque modèle, 500 replications de taille $n = 25$, 50 et 100 sont générées.

Sur la base de ces données simulées, nous estimons les pdfs à l'aide de l'estimateur à noyau de von Mises avec différents sélecteurs (RT, PI, LCV, UCV, sélecteurs bayésiens avec fonctions de perte quadratique et d'entropie). Notons que les techniques RT, PI, LCV et UCV sont implémentés dans la bibliothèque NPCirc du logiciel R; voir [Oliveira et al. \[2014b\]](#).

Les Tables 2.2, 2.3 et 2.4 présentent la moyenne optimale de ISE sur 500 replications pour toutes les approches et tous les modèles avec la taille d'échantillon $n = 25$, 50 et 100.

TABLE 2.2: Valeurs empiriques de $\text{ISE}(\nu_{RT})$, $\text{ISE}(\nu_{PI})$, $\text{ISE}(\nu_{LCV})$, $\text{ISE}(\nu_{UCV})$, $\text{ISE}(\nu_{Bayes1})$, $\text{ISE}(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 25$.

Groupe	Modèle	$\text{ISE}(\nu_{RT})$	$\text{ISE}(\nu_{PI})$	$\text{ISE}(\nu_{LCV})$	$\text{ISE}(\nu_{UCV})$	$\text{ISE}(\nu_{Bayes1})$	$\text{ISE}(\nu_{Bayes2})$
G1	M1	0.0004 (0.0009)	0.0443(0.0540)	0.0056(0.0137)	0.0056(0.0152)	0.0018(0.0069)	0.0017(0.0027)
	M2	0.0186(0.0130)	0.0526(0.0424)	0.0208(0.0163)	0.0240(0.0195)	0.0168 (0.0124)	0.0174(0.0123)
	M3	0.0392(0.0257)	0.0655(0.0522)	0.0447(0.0268)	0.0530(0.0333)	0.0370 (0.0262)	0.0398(0.0274)
	M4	0.0135(0.0124)	0.0453(0.0349)	0.0173(0.0158)	0.0182(0.0139)	0.0130(0.0094)	0.0124 (0.0105)
	M5	0.1306(0.0391)	0.0711(0.0456)	0.1403(0.0628)	0.0706 (0.0448)	0.0812(0.0490)	0.0852(0.0515)
	M6	0.0480(0.0151)	0.0606(0.0243)	0.0459 (0.0144)	0.0540(0.0193)	0.0478(0.0145)	0.0520(0.0154)
G2	M7	0.1036(0.0065)	0.0489(0.0441)	0.0334(0.0264)	0.0379(0.0291)	0.0286 (0.0179)	0.0310(0.0197)
	M8	0.0646(0.0163)	0.0591(0.0297)	0.0440(0.0210)	0.0449(0.0233)	0.0394 (0.0186)	0.0407(0.0177)
	M9	0.0162(0.0105)	0.0559(0.0458)	0.0191(0.0135)	0.0314(0.0239)	0.0154(0.0108)	0.0149 (0.0093)
	M10	0.0570(0.0220)	0.0840(0.0544)	0.0649(0.0326)	0.0653(0.0291)	0.0595 (0.0219)	0.0618(0.0208)
G3	M11	0.0657(0.0028)	0.0743(0.0458)	0.0385(0.0172)	0.0438(0.0264)	0.0340 (0.0162)	0.0408(0.0189)
	M12	0.0454(0.0060)	0.0615(0.0495)	0.0378(0.0290)	0.0414(0.0278)	0.0310 (0.0169)	0.0316(0.0147)
	M13	0.1088(0.0013)	0.0667(0.0313)	0.0475 (0.0247)	0.0530(0.0256)	0.0477(0.0244)	0.0538(0.0281)
	M14	0.0827(0.0023)	0.0777(0.0419)	0.0488(0.0228)	0.0532(0.0267)	0.0471 (0.0216)	0.0566(0.0206)
	M15	0.0097 (0.0030)	0.0504(0.0470)	0.0183(0.0225)	0.0203(0.0259)	0.0174(0.0116)	0.0112(0.0054)
	M16	0.0788(0.0009)	0.0853(0.0343)	0.0695(0.0226)	0.0691(0.0211)	0.0650 (0.0187)	0.0680(0.0108)
G4	M17	0.1087(0.0276)	0.0984(0.0329)	0.1035(0.0264)	0.0898(0.0341)	0.0855 (0.0241)	0.0885(0.0246)
	M18	0.0530(0.0143)	0.0743(0.0446)	0.0547(0.0146)	0.0656(0.0310)	0.0518 (0.0127)	0.0532(0.0119)
	M19	0.0579(0.0146)	0.0743(0.0390)	0.0506(0.0177)	0.0541(0.0197)	0.0442 (0.0150)	0.0442 (0.0143)
	M20	0.1103(0.0028)	0.0981(0.0328)	0.0825 (0.0235)	0.0828(0.0270)	0.0826(0.0236)	0.0900(0.0214)

Note : Les valeurs en gras indiquent les meilleurs résultats.

En termes de ISE , les résultats obtenus basés sur une approche classique et bayésienne révèlent que :

1. En général, les valeurs de ISE basées sur 500 replications diminuent à mesure que la taille de l'échantillon n augmente pour toutes les méthodes.

TABLE 2.3: Valeurs empiriques de $ISE(\nu_{RT})$, $ISE(\nu_{PI})$, $ISE(\nu_{LCV})$, $ISE(\nu_{UCV})$, $ISE(\nu_{Bayes1})$, $ISE(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 50$.

Groupe	Modèle	$ISE(\nu_{RT})$	$ISE(\nu_{PI})$	$ISE(\nu_{LCV})$	$ISE(\nu_{UCV})$	$ISE(\nu_{Bayes1})$	$ISE(\nu_{Bayes2})$
G1	M1	0.0006 (0.0012)	0.0280(0.0242)	0.0079(0.0117)	0.0082(0.0125)	0.0098(0.0079)	0.0040(0.0057)
	M2	0.0082(0.0071)	0.0223(0.0204)	0.0088(0.0067)	0.0091(0.0069)	0.0081(0.0056)	0.0074 (0.0055)
	M3	0.0180 (0.0112)	0.0305(0.0202)	0.0189(0.0119)	0.0231(0.0117)	0.0186(0.0109)	0.0200(0.0110)
	M4	0.0098(0.0092)	0.0154(0.0182)	0.0097(0.0089)	0.0143(0.0139)	0.0097(0.0075)	0.0088 (0.0071)
	M5	0.1140(0.0474)	0.0569 (0.0458)	0.1144(0.0497)	0.0610(0.0534)	0.0702(0.0448)	0.0735(0.0452)
	M6	0.0423(0.0182)	0.0460(0.0244)	0.0414(0.0188)	0.0382 (0.0174)	0.0421(0.0176)	0.0444(0.0175)
G2	M7	0.1049(0.0035)	0.0284(0.0161)	0.0200(0.0095)	0.0232(0.0105)	0.0190 (0.0093)	0.0195(0.0095)
	M8	0.0504(0.0090)	0.0295(0.0183)	0.0241(0.0101)	0.0292(0.0145)	0.0234 (0.0098)	0.0242(0.0100)
	M9	0.0108(0.0087)	0.0221(0.0202)	0.0114(0.0084)	0.0148(0.0128)	0.0101(0.0079)	0.0100 (0.0078)
	M10	0.0401(0.0147)	0.0426(0.0226)	0.0419(0.0189)	0.0391 (0.0146)	0.0406(0.0156)	0.0421(0.0152)
G3	M11	0.0651(0.0004)	0.0370(0.0198)	0.0280(0.0152)	0.0300(0.0161)	0.0265 (0.0114)	0.0295(0.0151)
	M12	0.0435(0.0060)	0.0301(0.0188)	0.0202(0.0069)	0.0243(0.0132)	0.0182 (0.0068)	0.0187(0.0074)
	M13	0.1087(0.0011)	0.0358(0.0186)	0.0279(0.0148)	0.0274(0.0124)	0.0264 (0.0123)	0.0271(0.0122)
	M14	0.0822(0.0008)	0.0372(0.0131)	0.0338(0.0112)	0.0351(0.0108)	0.0325 (0.0098)	0.0369(0.0136)
	M15	0.0094 (0.0050)	0.0321(0.0239)	0.0138(0.0102)	0.0169(0.0145)	0.0123(0.0100)	0.0107(0.0086)
	M16	0.0785(0.0005)	0.0528(0.0210)	0.0436(0.0171)	0.0445(0.0173)	0.0414 (0.0157)	0.0427(0.0174)
G4	M17	0.0866(0.0132)	0.0618(0.0227)	0.0704(0.0220)	0.0631(0.0208)	0.0601 (0.0172)	0.0621(0.0172)
	M18	0.0403(0.0075)	0.0429(0.0126)	0.0394(0.0067)	0.0398(0.0099)	0.0380 (0.0060)	0.0390(0.0062)
	M19	0.0492(0.0096)	0.0374(0.0166)	0.0353(0.0129)	0.0356(0.0117)	0.0331 (0.0094)	0.0341(0.0093)
	M20	0.1095(0.0012)	0.0605(0.0228)	0.0582 (0.0180)	0.0596(0.0196)	0.0603(0.0158)	0.0678(0.0155)

Note : Les valeurs en gras indiquent les meilleurs résultats.

TABLE 2.4: Valeurs empiriques $ISE(\nu_{RT})$, $ISE(\nu_{PI})$, $ISE(\nu_{LCV})$, $ISE(\nu_{UCV})$, $ISE(\nu_{Bayes1})$, $ISE(\nu_{Bayes2})$ et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés avec $n = 100$.

Groupe	Modèle	$ISE(\nu_{RT})$	$ISE(\nu_{PI})$	$ISE(\nu_{LCV})$	$ISE(\nu_{UCV})$	$ISE(\nu_{Bayes1})$	$ISE(\nu_{Bayes2})$
G1	M1	0.0003 (0.0005)	0.0052(0.0099)	0.0031(0.0056)	0.0032(0.0057)	0.0049(0.0047)	0.0025(0.0035)
	M2	0.0058(0.0044)	0.0103(0.0106)	0.0055(0.0035)	0.0062(0.0055)	0.0055(0.0034)	0.0052 (0.0033)
	M3	0.0124(0.0082)	0.0158(0.0127)	0.0138 (0.0084)	0.0152(0.0083)	0.0139(0.0073)	0.0150(0.0072)
	M4	0.0044 (0.0021)	0.0092(0.0059)	0.0052(0.0025)	0.0048(0.0026)	0.0057(0.0029)	0.0051(0.0026)
	M5	0.0850(0.0280)	0.0338 (0.0197)	0.0671(0.0327)	0.0348(0.0190)	0.0374(0.0291)	0.0393(0.0301)
	M6	0.0337(0.0113)	0.0313(0.0151)	0.0293(0.0118)	0.0287(0.0120)	0.0303 (0.0108)	0.0313(0.0107)
G2	M7	0.1060(0.0033)	0.0135(0.0100)	0.0114(0.0070)	0.0133(0.0092)	0.0105(0.0060)	0.0104 (0.0057)
	M8	0.0349(0.0054)	0.0144(0.0107)	0.0122(0.0078)	0.0137(0.0084)	0.0115 (0.0067)	0.0119(0.0066)
	M9	0.0079 (0.0046)	0.0109(0.0080)	0.0094(0.0066)	0.0109(0.0088)	0.0083(0.0051)	0.0079 (0.0046)
	M10	0.0295(0.0064)	0.0250(0.0105)	0.0293(0.0095)	0.0247 (0.0097)	0.0301(0.0078)	0.0316(0.0078)
G3	M11	0.0648(0.0001)	0.0163(0.0089)	0.0146(0.0065)	0.0158(0.0075)	0.0140 (0.0058)	0.0142(0.0057)
	M12	0.0410(0.0051)	0.0132(0.0062)	0.0121(0.0068)	0.0136(0.0070)	0.0111 (0.0054)	0.0111 (0.0053)
	M13	0.1085(0.0100)	0.0223(0.0078)	0.0197(0.0072)	0.0193(0.0066)	0.0186 (0.0062)	0.0187(0.0061)
	M14	0.0820(0.0002)	0.0194(0.0082)	0.0180 (0.0066)	0.0182(0.0069)	0.0186(0.0066)	0.0192(0.0068)
	M15	0.0082 (0.0015)	0.0147(0.0122)	0.0115(0.0076)	0.0114(0.0078)	0.0094(0.0048)	0.0089(0.0037)
	M16	0.0783(0.0002)	0.0221(0.0063)	0.0212 (0.0063)	0.0215(0.0064)	0.0229(0.0069)	0.0237(0.0072)
G4	M17	0.0762(0.0138)	0.0325 (0.0088)	0.0497(0.0207)	0.0338(0.0094)	0.0341(0.0165)	0.0381(0.0170)
	M18	0.0347(0.0041)	0.0297(0.0118)	0.0312(0.0068)	0.0282 (0.0091)	0.0314(0.0058)	0.0328(0.0056)
	M19	0.0373(0.0057)	0.0241 (0.0053)	0.0256(0.0049)	0.0252(0.0064)	0.0254(0.0045)	0.0260(0.0044)
	M20	0.1091(0.0008)	0.0352 (0.0084)	0.0373(0.0087)	0.0369(0.0088)	0.0408(0.0081)	0.0423(0.0085)

Note : Les valeurs en gras indiquent les meilleurs résultats.

2. Pour les ensembles de données générés à partir de M2, M3, M4, M7, M8, M9, M11, M12, M16, M17, M18 et M19, l'approche bayésienne (Bayes1 ou Bayes2) est plus performante que les autres concurrents classiques (RT, PI, UCV et LCV) pour un petit échantillon $n = 25$ et 50 . Toutefois, pour un échantillon de taille modérée $n = 100$, les performances des approches bayésienne et classique sont assez similaires.
3. Pour les ensembles de données générés à partir de M1, M10 et M15, les premières et deuxièmes meilleures performances sont obtenues respectivement avec RT et l'approche bayésienne (Bayes1 ou Bayes2).
4. Pour les ensembles de données générés à partir de M5, les meilleures performances sont obtenues avec PI et UCV, et les secondes meilleures performances sont produites par Bayes1.
5. Pour les ensembles de données générés à partir de M6, la première meilleure performance est obtenue avec LCV et RT et la deuxième meilleure performance est produite par Bayes1.
6. Pour les ensembles de données générés à partir de M13 et M14, la première meilleure performance est obtenue avec LCV et Bayes1, et la deuxième meilleure performance est produite par UCV.
7. Pour les ensembles de données générés à partir de M20, les performances des UCV, LCV et Bayes1 sont assez similaires et présentent les meilleures performances.
8. Enfin, nous pouvons conclure que la performance de notre approche présentée ne dépend pas de la complexité du modèle et a présenté en général des résultats satisfaisants pour tous les modèles considérés dans cette étude.

Nous avons également comparé les temps d'exécution (en secondes) nécessaires pour évaluer le paramètre de concentration optimal ν selon différentes méthodes (ν_{PI} , ν_{LCV} , ν_{UCV} et ν_{Bayes1}). Les calculs ont été effectués par un ordinateur personnel standard à 2,5 GHz et par le logiciel R pour une replication. La comparaison est donnée dans la Tableau 2.5 et illustre les résultats pour le modèle $M7$, car la même conclusion est obtenue pour les autres modèles considérés.

Nous pouvons observer que le temps d'exécution de l'approche Bayesain est largement supérieur à celui des techniques de plug-in et de validation croisée. Cependant, dans la

pratique, le temps d'exécution bayésien peut être considéré comme raisonnable, en particulier pour les échantillons de petite taille et de taille moyenne.

TABLE 2.5: Comparaison des temps d'exécution en secondes pour une replication à l'aide du modèle M7.

Model	n	$t_{\nu_{PI}}$	$t_{\nu_{LCV}}$	$t_{\nu_{UCV}}$	$t_{\nu_{Bayes1}}$
M7	10	0.13	0.11	0.42	15.17
	25	0.86	0.19	0.61	35.24
	50	1.25	0.55	0.91	73.90
	100	1.13	1.40	2.94	163.74
	200	1.22	2.16	4.92	379.37

Note : Les valeurs en gras indiquent les meilleurs résultats.

2.5 Analyse de données réelles

Nous avons appliqué l'estimateur à noyau circulaire de von Mises et la méthode bayésienne pour la sélection de paramètre de concentration sur les ensembles de données circulaires examinés dans la littérature. Nous estimons d'abord la densité de l'orientation des libellules, puis nous estimons la densité des données de lits transversaux. Ces deux ensembles de données disponibles dans la bibliothèque NPCirc du logiciel R, ont déjà été étudiés par [Oliveira et al. \[2012\]](#) à l'aide de l'estimateur à noyau de von Mises avec les sélecteurs classiques (RT, PI, UCV et LCV).

- Le premier ensemble de données consiste en l'orientation de 214 libellules par rapport à l'azimut du soleil, voir [Batschelet \[1981\]](#).
- Le deuxième ensemble de données représente les mesures transversales de la mélasse de l'Himalaya au Pakistan présentées dans [Fisher \[1993\]](#). Cet ensemble de données recueille 104 mesures de grandes formes de lit de Chaudan Zam.

Nous appliquons l'estimateur non paramétrique du noyau de von Mises en considérant les sélecteurs bayésiens pour les deux ensembles de données réels. Pour l'estimation bayésienne, nous définissons les paramètres de la loi a priori gamma comme $(\alpha, \beta) = (n^{2/5}, 1.5)$ et nous utilisons la méthode MCMC développée basée sur 1000 itérations de période de chauffe et 5000 comme le nombre total d'itérations pour estimer le paramètre ν avec des

fonctions de perte quadratique et d'entropie. Pour vérifier la convergence de l'algorithme MCMC, nous avons utilisé les indicateurs BMSE et SIF. Les tracés de la chaîne $\nu^{(i)}$ pour $i = N_0 + 1, \dots, N$ et la loi a posteriori de ν sont également donnés dans la Figure 2.3.

Le Tableau 2.6 présente les résultats du MCMC bayésiens. Les valeurs de BMSE et SIF indiquent une performance de mélange raisonnable des chaînes simulées ; voir aussi la Figure 2.3.

TABLE 2.6: *Estimateurs bayésiens de la largeur de bande et indicateurs BMSE et SIF associés obtenus à partir de jeux de données réelles.*

Exemples de données réelles	n	ν_{Bayes1}	ν_{Bayes2}	BMSE	SIF
Données d'orientation des libellules	214	31.1566	29.2486	0.2466	3.9219
Lits transversaux (2)	104	5.3543	4.3452	0.1191	7.3976

La Figure 2.4 montre les diagrammes en rose et les estimateurs de densité à noyau circulaire de von Mises (graphiques circulaires et linéaires) avec le sélecteur ν_{Bayes1} (en utilisant la fonction de perte quadratique). Les graphiques pour l'orientation des libellules et l'ensemble des données des lits transversaux sont présentés dans les panneaux gauche et droit, respectivement.

À partir des diagrammes en rose, nous pouvons voir clairement que la distribution des données d'orientation des libellules est bi-modale et la distribution des données des lits transversaux est uni-modale.

Il ressort clairement de la Figure 2.4 que la méthode bayésienne est capable de reproduire la bi-modalité des données d'orientation des libellules et l'uni-modalité des données des lits transversaux, et la qualité de lissage de l'estimateur bayésien est également très satisfaisante.

Enfin, nous notons que les résultats obtenus avec l'approche bayésienne proposée sont proches de ceux obtenus par Oliveira et al. [2012] avec des règles de plug-in et des procédures de validation croisée par vraisemblance.

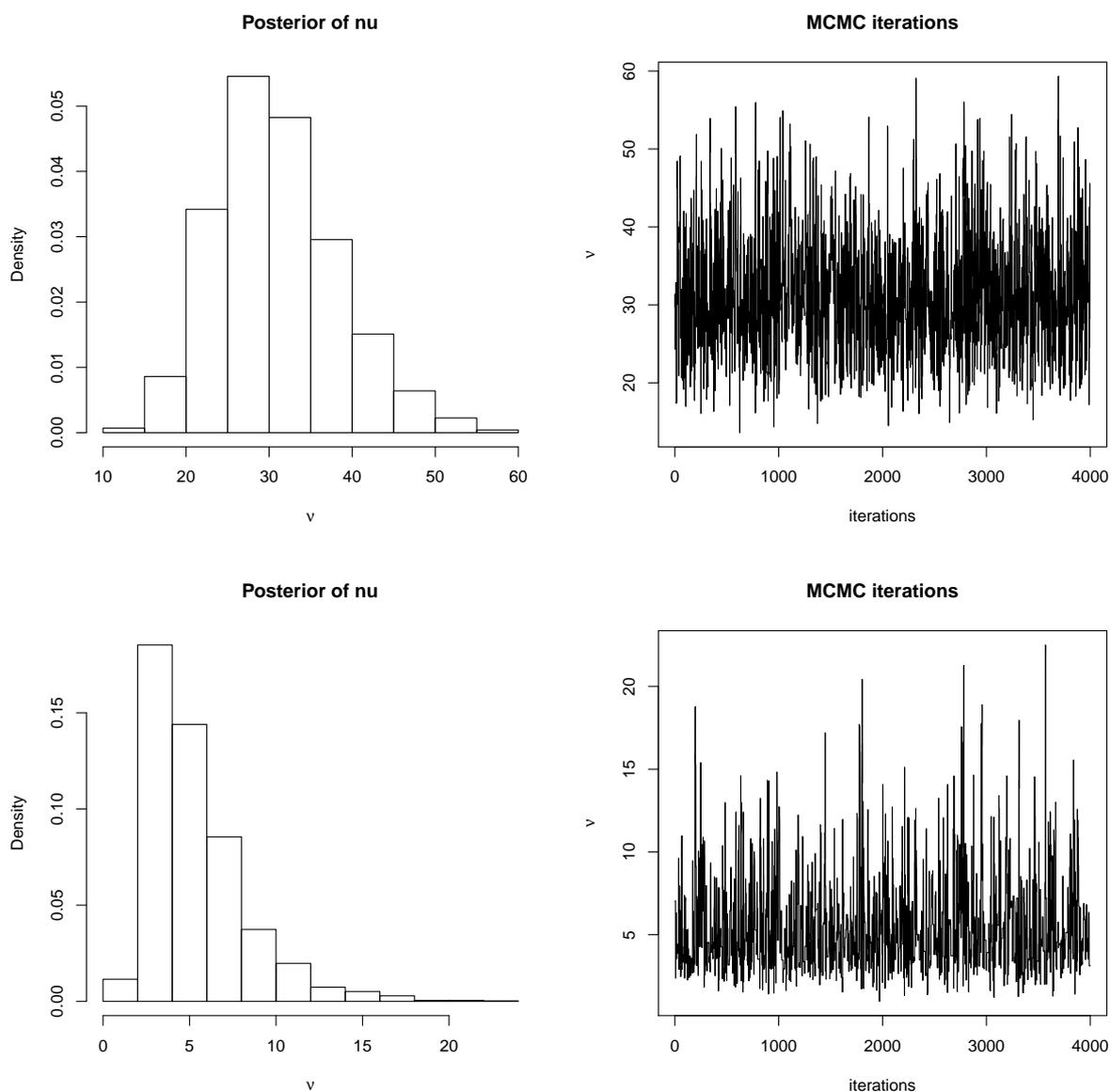


FIGURE 2.3: Densités a posteriori et les chaînes MCMC générées pour ν obtenues à partir des données d'orientation des libellules (première rangée) et des données de lit croisé (deuxième rangée).

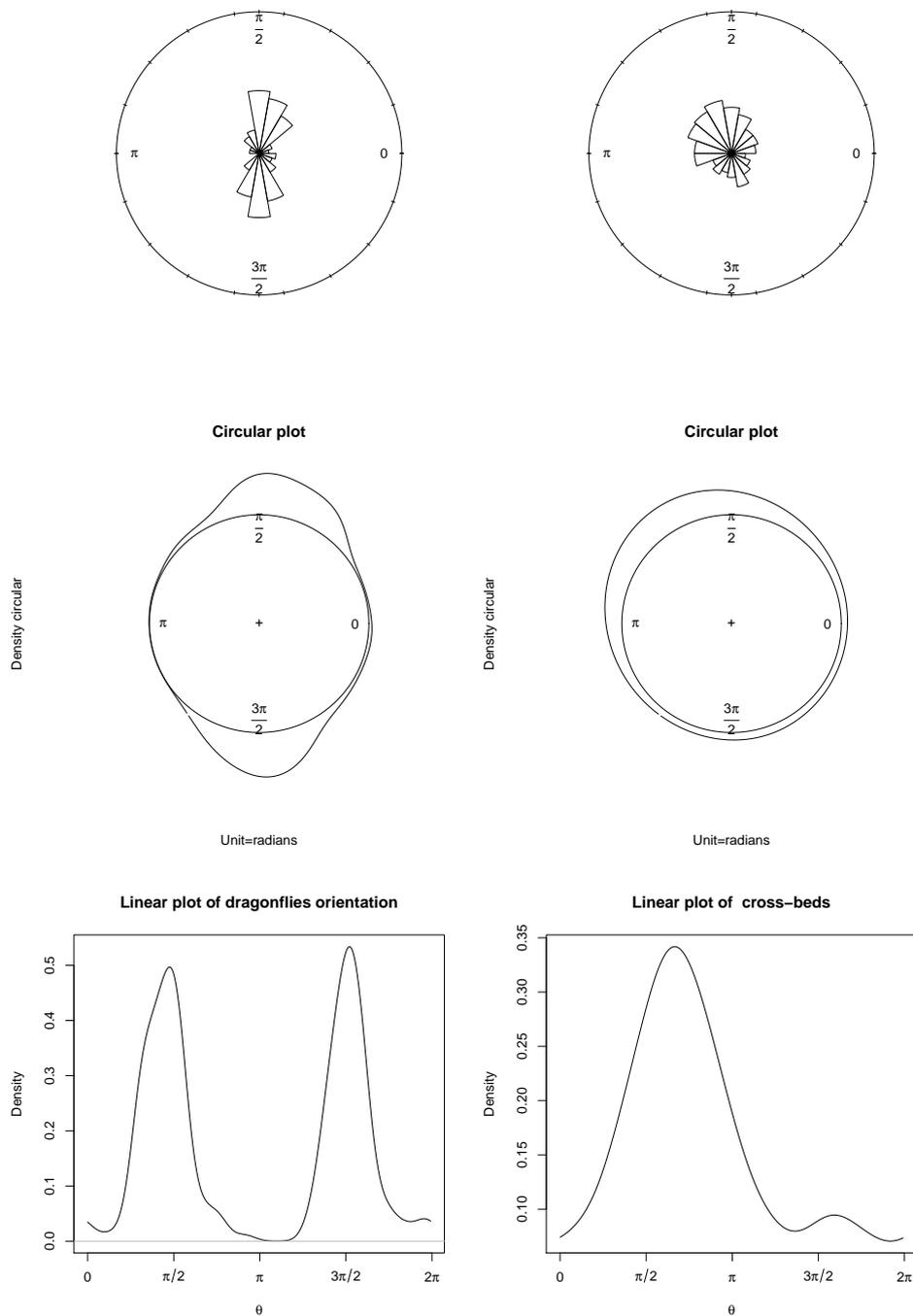


FIGURE 2.4: Diagrammes en rose (première rangée) et les estimateurs de densité à noyau circulaire de von Mises avec ν_{Byaes1} (tracés circulaires et linéaires donnés en secondes et troisième rangée, respectivement) pour les données d'orientation des libellules (panneau de gauche) et les données de lit croisé (panneau de droite).

2.6 Conclusion

Dans ce chapitre, nous avons introduit l'algorithme d'échantillonnage MCMC pour la sélection du paramètre de concentration dans l'estimation non-paramétrique de la densité par noyau circulaire. L'approche proposée peut être considérée comme une bonne alternative aux méthodes classiques existantes. Nous avons comparé la méthode présentée, les approches plug-in et les méthodes de validation croisée. Pour évaluer leurs performances, nous avons calculé le critère ISE, en utilisant des simulations pour des modèles circulaires de complexité variable. Nos études de simulation montrent que l'approche bayésienne présentée pour estimer le paramètre ν donne des résultats comparables à ceux obtenus avec la méthode plug-in et les techniques de validation croisée au sens de ISE. Nous avons également illustré notre procédure avec deux exemples de données circulaires réelles et les résultats obtenus sont satisfaisants.

Chapitre 3

Réduction du biais de l'estimateur non-paramétrique de la fonction densité à noyau circulaire

3.1 Introduction

Le but de ce chapitre est de construire deux nouveaux estimateurs à noyau circulaire (en utilisant le noyau de von Mises) basés sur les techniques MBC de réduction du biais de l'estimateur circulaire standard (sans correction du biais) donné par l'équation (1.10). Les propriétés statistiques des estimateurs MBC seront également étudiées. Une application sur des données simulées et des données réelles sera effectuée pour justifier la bonne performance des estimateurs MBC dans le cas circulaire.

3.2 Estimateurs MBC à noyau de von Mises de la fonction densité circulaire

Dans cette section, nous adaptons deux classes de techniques MBC pour l'estimateur de densité à noyau de von Mises, initialement proposées par [Terrell and Scott \[1980\]](#) pour la technique "TS" et [Jones et al. \[1995\]](#) pour la technique "JLN" afin de réduire le biais

de l'estimateur standard de von Mises. Ces techniques MBC ont également été étendues et adaptées par [Hirukawa \[2010\]](#), [Hirukawa and Sakudo \[2014\]](#), [Funke and Kawka \[2015\]](#), [Zougab and Adjabi \[2016\]](#) pour l'estimation linéaire de densité par noyau asymétrique; voir aussi [Harfouche et al. \[2018\]](#) pour les cas des données discrètes. Plus récemment, [Tsuruta and Sagae \[2017b\]](#) ont utilisé la méthode additive de [Jones and Foster \[1993\]](#) et la méthode multiplicative de [Terrell and Scott \[1980\]](#) pour construire un estimateur de densité à noyau circulaire d'ordre supérieur. Dans cette étude nous révisons l'estimateur TS de [Tsuruta and Sagae \[2017b\]](#) et nous introduisons également l'estimateur JLN de [Jones et al. \[1995\]](#) pour l'estimation de la densité circulaire, en utilisant le noyau de von Mises.

3.2.1 Technique TS (Terrell et Scott)

En se basant sur la même idée de l'estimateur géométrique de [Terrell and Scott \[1980\]](#) et [Hirukawa \[2010\]](#), l'estimateur de densité à noyau MBC en utilisant le noyau von Mises, que nous appelons simplement l'estimateur de densité à noyau TS-VM, est défini comme suit :

$$\hat{f}_\nu^{TS-VM}(\theta) = \left\{ \hat{f}_\nu^{VM}(\theta) \right\}^{\frac{1}{1-c}} \left\{ \hat{f}_{c\nu}^{VM}(\theta) \right\}^{-\frac{c}{1-c}}. \quad (3.1)$$

où $\hat{f}_\nu^{VM}(\theta)$ et $\hat{f}_{c\nu}^{VM}(\theta)$ sont les estimateurs de densité à noyau de von Mises donnés par (1.10) avec les paramètres de concentration ν et $c\nu$, respectivement, où $c \in (0, 1)$ est un paramètre de réglage qui ne dépend pas de l'angle θ ; voir, par exemple [Hirukawa \[2010\]](#) pour les noyaux linéaires asymétriques. Notez que lorsque $c = \frac{1}{4}$, nous obtenons simplement l'estimateur de [Tsuruta and Sagae \[2017b\]](#) (noyau de von Mises). Le choix optimal de c peut réduire encore le biais de l'estimateur (3.1) par rapport à l'estimateur circulaire TS introduit dans [Tsuruta and Sagae \[2017b\]](#).

3.2.2 Technique JLN (Jones, Linton et Neilson)

La deuxième classe de techniques MBC pour les estimateurs symétriques de densité à noyau dans le cas linéaire est originalement proposée par [Jones et al. \[1995\]](#); voir aussi [Hirukawa \[2010\]](#) et [Hirukawa and Sakudo \[2014\]](#) pour les noyaux asymétriques. D'une manière analogue, en utilisant le noyau de von Mises, que nous appelons l'estimateur de

densité à noyau JLN-VM, est défini par :

$$\hat{f}_\nu^{JLN-VM}(\theta) = \hat{f}_\nu^{VM}(\theta) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_\nu(\theta - \Theta_i)}{\hat{f}_\nu^{VM}(\Theta_i)} \right\}, \quad (3.2)$$

où $K_\nu(\cdot)$ est le noyau de von Mises. Notons que l'estimateur (3.2) est obtenu en deux étapes.

- La première étape consiste à écrire $f(\theta) = \hat{f}(\theta)\{f(\theta)/\hat{f}(\theta)\}$.
- La deuxième étape consiste à estimer le terme de correction du biais $f(\theta)/\hat{f}(\theta)$ de manière non-paramétrique, en utilisant l'estimateur à noyau de von Mises.

3.2.3 Propriétés asymptotiques des estimateurs MBC circulaires

3.2.3.1 Biais et variance

Les théorèmes suivants présentent le biais et la variance des estimateurs TS-VM et JLN-VM. Pour approximer le biais et la variance de chaque estimateur MBC, nous supposons que

- A1.** f est quatre fois continuellement dérivables, de plus toutes ces dérivées sont bornées.
- A2.** La suite du paramètre de concentration $\nu = \nu(n)$ satisfait $\nu \rightarrow \infty$ et $\nu^{1/2}/n \rightarrow 0$, lorsque $n \rightarrow \infty$.

Théorème 3.1. Soit \hat{f}_ν^{TS-VM} l'estimateur TS-VM défini par (3.1). Pour $\theta \in [0, 2\pi)$ donné et sous les hypothèses **A1** et **A2**, alors :

(i) le biais de l'estimateur TS-VM est donné par :

$$\text{biais}[\hat{f}_\nu^{TS-VM}(\theta)] = \frac{1}{c} \left[\frac{1}{2} \left\{ \frac{f''(\theta)^2}{16f(\theta)} - \frac{f''(\theta)}{16} \right\} \right] \frac{1}{\nu^2} + o(\nu^{-2}),$$

(ii) la variance de l'estimateur TS-VM est donnée par :

$$\mathbb{V}[\hat{f}_\nu^{TS-VM}(\theta)] = \frac{1 + c + c^{5/2}(1 + c) - 2\sqrt{2}c^{3/2}(1 + c)^{1/2}}{2n\sqrt{\pi}(1 - c)^2(1 + c)} \nu^{1/2} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right).$$

Démonstration

- Biais de l'estimateur TS-VM

Notons que $H_\nu(\theta) = \mathbb{E}[\widehat{f}_\nu^{VM}(\theta)] = \mathbb{E}[f(\Psi)]$, où la variable aléatoire $\Psi \sim vM(\theta, \nu)$. En utilisant le développement de Taylor au point $\Psi = \theta$ pour $H_\nu(\theta)$ jusqu'à l'ordre 4, nous obtenons :

$$\begin{aligned} H_\nu(\theta) &= f(\theta) + \sum_{j=1}^4 \frac{f^{(j)}}{j!} \mathbb{E}(\Psi - \theta)^j + o(\mathbb{E}(\Psi - \theta)^4) \\ &= f(\theta) + f^{(1)} \mathbb{E}(\Psi - \theta) + \frac{f^{(2)}}{2} \mathbb{E}(\Psi - \theta)^2 + \frac{f^{(3)}}{6} \mathbb{E}(\Psi - \theta)^3 \\ &\quad + \frac{f^{(4)}}{24} \mathbb{E}(\Psi - \theta)^4 + o(\mathbb{E}(\Psi - \theta)^4). \end{aligned}$$

En utilisant maintenant les propriétés de la variable aléatoire de von Mises et le développement de Taylor lorsque $\nu \rightarrow \infty$ (voir [Mardia and Jupp \[2000\]](#)), nous avons

$$\begin{aligned} \mathbb{E}(\Psi - \theta) &= 0, \\ \mathbb{E}(\Psi - \theta)^2 &= 1 - \mathbb{E}(\cos(\Psi - \theta)) = 1 - \frac{I_1(\nu)}{I_0(\nu)} = \frac{1}{2\nu} + \frac{1}{8\nu^2} + o\left(\frac{1}{\nu^2}\right), \\ \mathbb{E}(\Psi - \theta)^3 &= \mathbb{E}(\sin 2(\Psi - \theta)) = 0, \\ \mathbb{E}(\Psi - \theta)^4 &= \mathbb{E}(\cos 2(\Psi - \theta)) - \{\mathbb{E}(\cos(\Psi - \theta))\}^4 = \frac{I_2(\nu)}{I_0(\nu)} - \left\{\frac{I_1(\nu)}{I_0(\nu)}\right\}^4 = o\left(\frac{1}{\nu^2}\right). \end{aligned}$$

En remplaçant les expressions de $\mathbb{E}(\Psi - \theta)^j$ pour $j = 1, \dots, 4$, $H_\nu(\theta)$ est approximé par

$$H_\nu(\theta) = f(\theta) \left\{ 1 + \frac{f''(\theta)}{4f(\theta)} \frac{1}{\nu} + \frac{f''(\theta)}{16f(\theta)} \frac{1}{\nu^2} + o\left(\frac{1}{\nu^2}\right) \right\}. \quad (3.3)$$

D'une manière similaire, $H_{c\nu}(\theta) = \mathbb{E}(\widehat{f}_{c\nu}^{VM}(\theta))$ peut être approximé par

$$H_{c\nu}(\theta) = f(\theta) \left\{ 1 + \frac{f''(\theta)}{4f(\theta)} \frac{1}{c\nu} + \frac{f''(\theta)}{16f(\theta)} \frac{1}{c^2\nu^2} + o\left(\frac{1}{\nu^2}\right) \right\}. \quad (3.4)$$

On définit

$$\widehat{f}_\nu^{VM}(\theta) = H_\nu(\theta) + Z,$$

et

$$\widehat{f}_{c\nu}^{VM}(\theta) = H_{c\nu}(\theta) + W.$$

L'estimateur \widehat{f}_ν^{VM} peut être écrit comme suit :

$$\widehat{f}_\nu^{VM}(\theta) = \{H_\nu(\theta)\}^{\frac{1}{1-c}} \left\{ 1 + \frac{Z}{H_\nu(\theta)} \right\}^{\frac{1}{1-c}} \{H_{c\nu}(\theta)\}^{-\frac{c}{1-c}} \left\{ 1 + \frac{W}{H_{c\nu}(\theta)} \right\}^{-\frac{c}{1-c}}.$$

En utilisant l'approximation $(1+t)^\alpha = 1 + \alpha t + O(t^2)$, on aura

$$\begin{aligned} \widehat{f}_\nu^{VM}(\theta)(\theta) &= \{H_\nu(\theta)\}^{\frac{1}{1-c}} \{H_{c\nu}(\theta)\}^{-\frac{c}{1-c}} + \frac{1}{1-c} Z \left\{ \frac{H_\nu(\theta)}{H_{c\nu}(\theta)} \right\}^{-\frac{c}{1-c}} \\ &\quad - \frac{c}{1-c} W \left\{ \frac{H_\nu(\theta)}{H_{c\nu}(\theta)} \right\}^{-\frac{c}{1-c}} + O\{(Z+W)^2\}. \end{aligned} \quad (3.5)$$

Sous la condition **A2**, nous pouvons montrer facilement que

$$\mathbb{E} \left(\widehat{f}_\nu^{VM}(\theta)(\theta) \right) = f(\theta) + \frac{1}{c} \left[\frac{1}{2} \left\{ \frac{f''(\theta)^2}{16f(\theta)} - \frac{f''(\theta)}{16} \right\} \right] \frac{1}{\nu^2} + o \left(\frac{1}{\nu^2} \right).$$

– Variance de l'estimateur TS-VM

Pour la variance, à partir de l'équation (3.5), on a :

$$\begin{aligned} \mathbb{V} \left(\widehat{f}_\nu^{TS-VM}(\theta) \right) &= \mathbb{E} \left(\frac{1}{1-c} Z - \frac{c}{1-c} W \right)^2 + O(n^{-1}) \\ &= \mathbb{V} \left(\frac{1}{1-c} \widehat{f}_\nu^{VM}(\theta) - \frac{c}{1-c} \widehat{f}_{c\nu}^{VM}(\theta) \right) + O(n^{-1}) \\ &= \frac{1}{(1-c)^2} \mathbb{V} \left(\widehat{f}_\nu^{VM}(\theta) \right) + \frac{c^2}{(1-c)^2} \mathbb{V} \left(\widehat{f}_{c\nu}^{VM}(\theta) \right) \\ &\quad - \frac{2c}{(1-c^2)} \text{cov} \left(\widehat{f}_\nu^{VM}(\theta), \widehat{f}_{c\nu}^{VM}(\theta) \right). \end{aligned}$$

Les termes $\mathbb{V} \left(\widehat{f}_\nu^{VM}(\theta) \right)$ et $\mathbb{V} \left(\widehat{f}_{c\nu}^{VM}(\theta) \right)$ sont définies par :

$$\mathbb{V} \left(\widehat{f}_\nu^{VM}(\theta) \right) = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o \left(\frac{\nu^{1/2}}{n} \right),$$

et

$$\mathbb{V} \left(\widehat{f}_{c\nu}^{VM}(\theta) \right) = \frac{(c\nu)^{1/2}}{2n\sqrt{\pi}} f(\theta) + o \left(\frac{\nu^{1/2}}{n} \right).$$

Maintenant,

$$\begin{aligned} \text{cov} \left(\widehat{f}_\nu^{VM}(\theta), \widehat{f}_{c\nu}^{VM}(\theta) \right) &= \frac{1}{n} \mathbb{E}(K_\nu(\theta - \Theta_i) + K_{c\nu}(\theta - \Theta_i)) + O(n^{-1}) \\ &= \frac{1}{n} \int_0^{2\pi} \frac{1}{4\pi^2 I_0(\nu) I_0(c\nu)} \exp\{(\nu + c\nu) \cos(\theta - \phi)\} f(\phi) d\phi + O(n^{-1}) \\ &= \frac{1}{n} \frac{I_0(\nu + c\nu)}{2\pi^2 I_0(\nu) I_0(c\nu)} \int_0^{2\pi} \frac{1}{2\pi^2 I_0(\nu + c\nu)} \\ &\quad \times \exp\{(\nu + c\nu) \cos(\theta - \phi)\} f(\phi) d\phi + O(n^{-1}) \\ &= \frac{1}{n} \frac{I_0(\nu + c\nu)}{2\pi^2 I_0(\nu) I_0(c\nu)} \mathbb{E}(f(\Phi)), \end{aligned}$$

où $\Phi \sim vM(\theta, \nu + c\nu)$. Notons que $\mathbb{E}(f(\Phi)) = f(\theta) + o(1)$. En utilisant le développement en séries de Taylor de $I_0(\nu)$, $I_0(c\nu)$ et $I_0(\nu + c\nu)$ quand $\nu \rightarrow \infty$ (voir [Mardia and Jupp \[2000\]](#))), on obtient alors

$$\text{cov} \left(\widehat{f}_\nu^{VM}(\theta), \widehat{f}_{c\nu}^{VM}(\theta) \right) = \frac{1}{n\sqrt{2\pi}} \left(\frac{c}{1+c} \right)^{1/2} \nu^{1/2} f(\theta).$$

Par conséquent, la variance de $\widehat{f}_\nu^{TS-VM}(\theta)$ est donnée par

$$\text{cov} \left(\widehat{f}_\nu^{TS-VM}(\theta) \right) = \frac{1+c+c^{5/2}(1+c) - 2\sqrt{2}c^{3/2}(1+c)^{1/2}}{2n\sqrt{\pi}(1-c)^2(1+c)} \nu^{1/2} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right),$$

ce qui correspond aux résultats du théorème [3.1](#).

Théorème 3.2. Soit \widehat{f}_ν^{JLN-VM} l'estimateur JLN-VM défini par [\(3.2\)](#). Pour $\theta \in [0, 2\pi)$ donnée et sous les hypothèses **A1** et **A2**, alors :

(i) le biais de JLN-VM est donné par :

$$\text{biais}[\widehat{f}_\nu^{JLN-VM}(\theta)] = -[f(\theta)q(\theta)]\frac{1}{\nu^2} + o(\nu^{-2}),$$

avec $q(\theta) = (f''(\theta)/\{4f(\theta)\})''$.

(ii) la variance de l'estimateur JLN-VM a la forme asymptotique suivante :

$$\mathbb{V}[\widehat{f}_\nu^{JLN-VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right).$$

Démonstration

– Biais de l'estimateur JLN-VM

Notons que l'estimateur \widehat{f}_ν^{JLN-VM} peut s'écrire sous la forme suivante (voir par exemple [Hirukawa \[2010\]](#)) :

$$\widehat{f}_\nu^{JLN-VM}(\theta) = f(\theta) \left\{ 1 + \frac{\widehat{f}_\nu^{VM}(\theta) - f(\theta)}{f(\theta)} \right\} \{1 + (\Omega(\theta) - 1)\}.$$

où $\Omega(\theta) = n^{-1} \sum_{i=1}^n K_\nu(\theta - \Theta_i) / \widehat{f}_\nu^{VM}(\Theta_i)$. De plus, on a

$$\begin{aligned} \mathbb{E} \left(\widehat{f}_\nu^{JLN-VM}(\theta) \right) &= f(\theta) + f(\theta) \mathbb{E} \left\{ \frac{\widehat{f}_\nu^{VM}(\theta) - f(\theta)}{f(\theta)} \right\} + f(\theta) \mathbb{E} \{ \Omega(\theta) - 1 \} \\ &+ f(\theta) \mathbb{E} \left\{ \left(\frac{\widehat{f}_\nu^{VM}(\theta) - f(\theta)}{f(\theta)} \right) (\Omega(\theta) - 1) \right\}. \end{aligned}$$

Sous l'hypothèse **A2** et en utilisant les propriétés de la variable aléatoire de von Mises, $\mathbb{E} \left(\widehat{f}_\nu^{JLN-VM}(\theta) \right)$ peut s'écrire sous la forme suivante :

$$\mathbb{E} \left(\widehat{f}_\nu^{JLN-VM}(\theta) \right) = f(\theta) - [f(\theta)q(\theta)] \frac{1}{\nu^2} + o(\nu^{-2}),$$

où $q(\theta) = (f''(\theta)/\{4f(\theta)\})''$.

– *Variance de l'estimateur JLN-VM*

Remarquons que $\widehat{f}_\nu^{JLN-VM}(\theta)$ est équivalent à (voir Hirukawa [2010] et Jones et al. [1995] pour le cas linéaire) :

$$\bar{f}_\nu^{JLN-VM}(\theta) = f(\theta) \frac{1}{n} \sum_{i=1}^n \frac{K_\nu(\theta - \Theta_i)}{f(\Theta_i)}.$$

Il suit que

$$\begin{aligned} \mathbb{V} \left(\bar{f}_\nu^{JLN-VM}(\theta) \right) &= f^2(\theta) \frac{1}{n} \mathbb{V} \left\{ \frac{K_\nu(\theta - \Theta_i)}{f(\Theta_i)} \right\} \\ &= f^2(\theta) \frac{1}{n} \mathbb{E} \left\{ \frac{K_\nu^2(\theta - \Theta_i)}{f^2(\Theta_i)} + O(n^{-1}) \right\}. \end{aligned}$$

Observons maintenant que

$$\begin{aligned} \mathbb{E} \left\{ \frac{K_\nu^2(\theta - \Theta_i)}{f^2(\Theta_i)} \right\} &= \int_0^{2\pi} \frac{1}{4\pi^2 I_0^2(\nu) f(\phi)} \exp\{2\nu \cos(\theta - \phi)\} d\phi \\ &= \frac{I_0(2\nu)}{2\pi I_0^2(\nu)} \int_0^{2\pi} \frac{1}{2\pi I_0(2\nu) f(\phi)} \exp\{2\nu \cos(\theta - \phi)\} d\phi \\ &= \frac{I_0(2\nu)}{2\pi I_0^2(\nu)} \mathbb{E} \left(\frac{1}{f(\Phi)} \right), \end{aligned}$$

où $\Phi \sim vM(\theta, 2\nu)$. En utilisant les séries de Taylor de $I_0(2\nu)$ et $I_0^2(\nu)$ lorsque $\nu \rightarrow \infty$ (voir Mardia and Jupp [2000]), on déduit que

$$\mathbb{V}[\widehat{f}_\nu^{JLN-VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right).$$

Remarque 1. *Le biais et la variance de l'estimateur TS-VM dépendent du paramètre de réglage c . Il est clair que le biais diminue avec $c \in]0, 1[$. De plus, on peut montrer facilement que la variance est strictement croissante sur $c \in]0, 1[$, car le facteur $\lambda(c) = \left\{ 1 + c + c^{5/2}(1 + c) - 2\sqrt{2}c^{3/2}(1 + c)^{1/2} \right\} / \{(1 - c)^2(1 + c)\}$ dans la variance, augmente de façon monotone dans $c \in]0, 1[$, avec $\lim_{c \rightarrow 0} \lambda(c) = 1$ et $\lim_{c \rightarrow 1} \lambda(c) = 1.6875$; voir aussi la Figure 3.1. Notons que l'estimateur TS-VM (3.1) peut être considéré comme une*

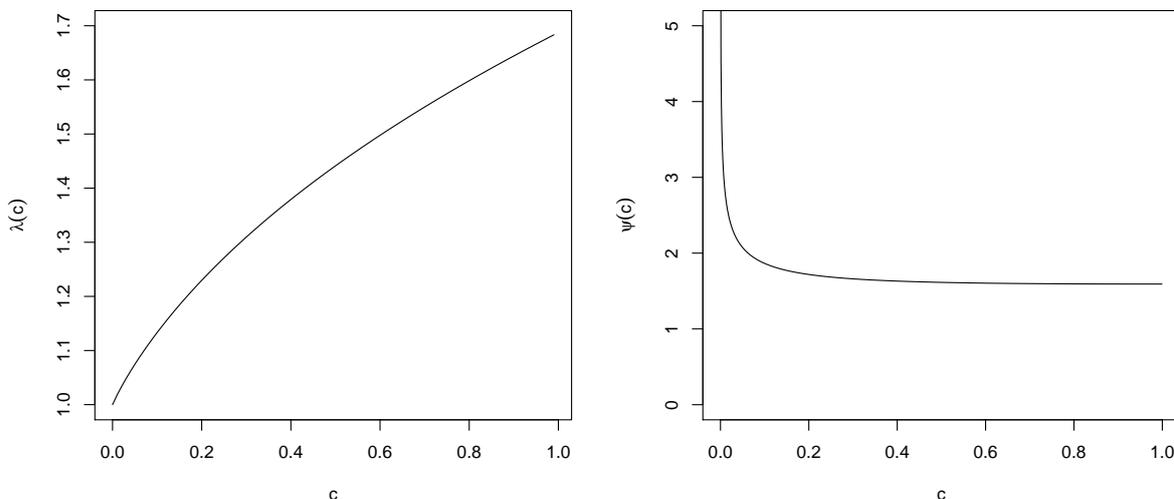


FIGURE 3.1: Graphes des fonctions λ et ψ pour $c \in]0, 1[$.

version révisée de l'estimateur proposé dans [Tsuruta and Sagae \[2017b\]](#); nous pouvons voir immédiatement que notre estimateur est plus performant au sens de biais lorsque $c > 1/4$, en comparaison avec l'estimateur [Tsuruta and Sagae \[2017b\]](#) introduit récemment. Pour notre étude, le choix approprié du paramètre de réglage c sera discuté ci-dessous.

Remarque 2. Les deux théorèmes montrent que les estimateurs TS-VM et JLN-VM réduisent l'ordre de grandeur du biais de $O(\nu^{-1})$ à $O(\nu^{-2})$, alors que leurs variances sont toujours d'ordre $O(\frac{\nu^{1/2}}{n})$. Notons que la variance de l'estimateur TS-VM a tendance à être plus grande que celle de l'estimateur standard de von Mises, mais la différence n'est pas significative car le facteur $\lambda(c)$ qui augmente dans $c \in]0, 1[$ prend ses valeurs dans $]1, 1.6875[$, voir la Remarque 1. En outre, la variance asymptotique de l'estimateur JLN-VM est exactement la même que l'estimateur standard de von Mises.

3.2.4 MISE des estimateurs MBC circulaires

Le critère à utiliser pour la propriété globale est la MISE définie comme

$$\text{MISE}(\widehat{f}_\nu^{MBC-VM}) = \int_0^{2\pi} \text{biais}^2[\widehat{f}_\nu^{MBC-VM}(\theta)]d\theta + \int_0^{2\pi} \mathbb{V}[\widehat{f}_\nu^{MBC-VM}(\theta)]d\theta, \quad (3.6)$$

où \widehat{f}_ν^{MBC-VM} est l'estimateur de densité de noyau TS-VM ou JLN-VM.

La MISE de l'estimateur TS-VM (3.1) est de la forme :

$$\begin{aligned} \text{MISE}(\widehat{f}_\nu^{TS-VM}) &= \frac{1}{4c^2\nu^4} \int_0^{2\pi} \left\{ \frac{f''(\theta)^2}{16f(\theta)} - \frac{f''(\theta)}{16} \right\}^2 d\theta \\ &+ \frac{1+c+c^{5/2}(1+c)-2\sqrt{2}c^{3/2}(1+c)^{1/2}}{2n\sqrt{\pi}(1-c)^2(1+c)} \nu^{1/2} \\ &+ o\left(\nu^{-4} + \frac{\nu^{1/4}}{n}\right). \end{aligned} \quad (3.7)$$

Le paramètre de concentration au sens de MISE de l'estimateur (3.1) est donné par :

$$\nu_{TS-VM}^{opt} = \left\{ \frac{4n\sqrt{\pi}(1-c)^2(1+c) \int_0^{2\pi} \left\{ \frac{f''(\theta)^2}{16f(\theta)} - \frac{f''(\theta)}{16} \right\}^2 d\theta}{c^2(1+c+c^{5/2}(1+c)-2\sqrt{2}c^{3/2}(1+c)^{1/2})} \right\}^{\frac{2}{9}}. \quad (3.8)$$

Par conséquent, la MISE optimale pour l'estimateur TS-VM devient :

$$\text{MISE}^{opt}(\widehat{f}_\nu^{TS-VM}) = \psi(c) \left(\frac{9\{4\sqrt{\pi}\}^{1/9}}{16\sqrt{\pi}} \right) \left(\int_0^{2\pi} \left\{ \frac{f''(\theta)^2}{16f(\theta)} - \frac{f''(\theta)}{16} \right\}^2 d\theta \right)^{\frac{1}{9}} n^{-\frac{8}{9}}. \quad (3.9)$$

où $\psi(c) = \left(\frac{1+c+c^{5/2}(1+c)-2\sqrt{2}c^{3/2}(1+c)^{1/2}}{(1-c)^2(1+c)} \right)^{8/9} c^{-2/9}$.

Notons que le paramètre de réglage c peut être choisi suivant [Funke and Kawka \[2015\]](#). Notons aussi qu'il n'y a pas de c optimal qui minimise la MISE optimale (3.9) car le facteur $\psi(c)$ qui apparaît dans la MISE optimale diminue strictement dans $c \in (0, 1)$; voir la Figure 3.1. Dans notre étude de simulation, nous proposons un choix approprié du paramètre de réglage $c = 1/2$. Nous avons considéré cette valeur pour les deux raisons suivantes :

1. Premièrement, le biais de l'estimateur TS-VM (3.1) peut être réduit par rapport à l'estimateur de [Tsuruta and Sagae \[2017b\]](#) (obtenu avec $c = (1/4)$).
2. Il n'y a pas beaucoup de différence entre $\psi(1/2) \approx 1.6142$ et $\lim_{c \rightarrow 1} \psi(c) \approx 1.5921$.

D'une manière similaire, la MISE de l'estimateur JLN-VM (3.2) est donnée par

$$\text{MISE}(\widehat{f}_\nu^{\text{JLN-VM}}) = \frac{1}{\nu^4} \int_0^{2\pi} f^2(\theta)q^2(\theta)d\theta + \frac{\nu^{1/2}}{2n\sqrt{\pi}} + o\left(\nu^{-4} + \frac{\nu^{1/4}}{n}\right) \quad (3.10)$$

En minimisant la MISE($\widehat{f}_\nu^{\text{TS-JLN}}$) donnée par (3.10) par rapport à ν , on obtient la valeur optimale

$$\nu_{\text{JLN-VM}}^{\text{opt}} = \left\{ 16n\sqrt{\pi} \int_0^{2\pi} f^2(\theta)q^2(\theta)d\theta \right\}^{2/9}. \quad (3.11)$$

où la quantité $q(\theta)$ est donnée dans le Théorème 3.2. En remplaçant le paramètre de concentration optimal (3.11) dans la MISE de l'estimateur JLN-VM donnée par (3.10), la MISE optimale pour l'estimateur JLN-VM est donnée par

$$\text{MISE}^{\text{opt}}(\widehat{f}_\nu^{\text{JLN-VM}}) = \left(\frac{1}{\{16\sqrt{\pi}\}^{\frac{8}{9}}} + \frac{16^{\frac{1}{9}}}{2\{\sqrt{\pi}\}^{\frac{8}{9}}} \right) \left(\int_0^{2\pi} f^2(\theta)q^2(\theta)d\theta \right)^{\frac{1}{9}} n^{-\frac{8}{9}}. \quad (3.12)$$

Notons que les paramètres de concentration (3.8) et (3.11) ne peuvent pas être appliqués dans la pratique. La section suivante présente une procédure pratique pour la sélection du paramètre de concentration dans les estimateurs TS-VM et JLN-VM.

3.2.5 Choix du paramètre de concentration pour les estimateurs MBC-VM

Comme les paramètres de concentration optimaux donnés par (3.8) et (3.11) dépendent de la densité inconnue f et de ses dérivées f' , f'' , f''' et f'''' , ils ne peuvent donc pas être exploités dans la pratique. Dans ce cas, nous adaptons la méthode de validation croisée par vraisemblance (LCV); voir par exemple Oliveira et al. [2012]). Dans le cas de la technique LCV , pour un estimateur donné $\widehat{f}_{\text{MBC-VM}}$, qui désigne les estimateurs TS-VM ou JLN-VM, le paramètre de concentration optimale ν_{lcv} de ν est obtenu par

$$\nu_{lcv} = \arg \max_{\nu} LCV(\nu),$$

où

$$LCV(\nu) = \prod_{i=1}^n \widehat{f}_{\nu,i}^{\text{MBC-VM}}(\Theta_i), \quad (3.13)$$

où $\widehat{f}_{\nu,i}^{\text{MBC-VM}}(\Theta_i)$ est l'estimateur calculé à partir de $\widehat{f}_{\nu,i}^{\text{MBC-VM}}(\cdot)$ en excluant l'observation Θ_i .

En maximisant la fonction LCV donnée par (3.13), nous obtenons le paramètre de concentration optimal LCV pouvant être utilisé dans des situations pratiques. Notons que l'approche de plug-in proposée Oliveira et al. [2012] dans le cas de l'estimateur VM standard peut également être adapté pour les estimateurs MBC-VM. Dans ce travail, nous avons choisi la technique LCV pour sa simplicité.

3.3 Etude de simulation

Dans cette section, nous examinons et comparons les performances des estimateurs MBC-VM avec l'estimateur standard de von Mises, à partir de différents modèles circulaires déjà considérés dans la littérature et dans le chapitre précédent de cette thèse.

3.3.1 Modèles considérés dans l'étude de simulation

Nous considérons les 20 modèles circulaires utilisés par Oliveira et al. [2012]; voir aussi la Section 2.4 pour la représentation graphique (Figure 2.1) et les expressions de ces modèles.

3.3.2 Résultats de simulation : comparaison entre ISE et ISB

Pour chaque modèle, 500 replications de la taille d'échantillon $n = 25, 50, 100$ et 250 sont générées. On utilise les estimateurs MBC-VM (TS-VM et JLN-VM) pour estimer la densité des données générées à partir des modèles considérés.

Notons que pour l'estimateur TS-VM, le paramètre de réglage c est choisi comme $c = 1/2$ suivant la Section 3.2.4. Pour comparaison, nous utilisons également l'estimateur standard de von Mises. Pour le choix du paramètre de concentration, nous utilisons la méthode LCV développée dans la section précédente.

Nous examinons les performances des estimateurs via l'erreur quadratique intégrée (ISE) et le biais quadratique intégré (ISB) donnés respectivement par

$$\text{ISE} := \int_0^{2\pi} \left[\widehat{f}_\nu(\theta) - f(\theta) \right]^2 d\theta \quad (3.14)$$

et

$$\text{ISB} := \int_0^{2\pi} \left[\mathbb{E}\{\widehat{f}_\nu(\theta)\} - f(\theta) \right]^2 d\theta, \quad (3.15)$$

où \widehat{f}_ν est l'estimateur MBC-VM ou l'estimateur standard de von Mises.

Les Tableaux 3.1 et 3.2 présentent les moyennes et les écarts types de ISE et les Tableaux 3.3 et 3.4 fournissent les valeurs ISB pour les estimateurs de von Mises (standard) et MBC-VM (TS-VM et JLN-VM).

TABLE 3.1: Valeurs empiriques de ISE ($\times 100$) et leurs écarts-types ($\times 100$) entre parenthèses basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 25$ et $n = 50$.

Groupe	modèle	$n = 25$			$n = 50$		
		\widehat{f}_ν^{VM}	\widehat{f}_ν^{TS-VM}	\widehat{f}_ν^{JLN-VM}	\widehat{f}_ν^{VM}	\widehat{f}_ν^{TS-VM}	\widehat{f}_ν^{JLN-VM}
G1	M1	1.321(2.390)	1.508(2.611)	0.573 (1.604)	0.657(1.401)	1.038(1.708)	0.125 (0.570)
	M2	1.843 (1.578)	1.957(1.878)	1.971(1.844)	1.167(1.483)	1.357(2.283)	0.996 (1.317)
	M3	4.580(4.456)	5.008(0.05635)	3.778 (4.342)	2.022(1.283)	2.516(1.287)	1.298 (0.808)
	M4	1.767(1.415)	2.181(2.137)	1.490 (1.335)	0.698(0.409)	0.651(0.437)	0.637 (0.430)
	M5	12.419 (4.919)	14.710(4.311)	13.098(5.476)	9.237 (4.230)	10.785(5.147)	9.849(4.302)
	M6	7.102 (3.443)	8.548(5.719)	7.412(3.266)	3.886(1.315)	4.954(1.455)	3.823 (1.393)
G2	M7	3.932 (2.676)	4.150(3.390)	4.040(2.701)	2.247(1.147)	2.216(1.409)	1.821 (0.999)
	M8	4.157(2.217)	4.617(3.008)	4.121 (2.288)	2.527(0.958)	2.683(1.205)	2.221 (0.960)
	M9	2.609(2.706)	2.604(2.150)	2.384 (3.013)	1.140(0.715)	1.079(0.737)	1.040 (0.644)
	M10	5.713(2.706)	5.774(2.167)	5.514 (3.313)	4.762(1.602)	5.194(1.785)	4.714 (1.507)
G3	M11	4.180 (2.261)	4.561(2.808)	5.514 (2.035)	2.925(1.285)	2.896(1.434)	3.496(1.989)
	M12	3.326 (1.909)	3.834(2.668)	4.605(1.157)	2.216 (1.086)	2.444(1.335)	2.229(1.293)
	M13	5.591 (2.193)	5.926(2.775)	7.812(3.294)	3.113(1.691)	3.188(1.872)	2.952 (1.466)
	M14	6.165 (2.468)	6.566(2.765)	6.414(2.494)	3.408(1.309)	3.376 (1.510)	3.405(1.504)
	M15	1.820(1.717)	2.497(2.461)	1.106 (0.382)	1.402(0.929)	1.844(1.467)	1.077 (0.420)
	M16	6.663 (2.723)	7.354(3.245)	7.181(2.047)	3.674 (1.363)	3.811(1.492)	3.834(1.920)
G4	M17	10.516 (2.041)	10.719(1.789)	10.803(1.801)	7.596 (2.226)	8.320(2.073)	8.391(1.439)
	M18	5.434(2.245)	4.965 (1.336)	5.300(2.730)	4.563(1.141)	4.391 (0.910)	4.583(0.908)
	M19	6.063(2.852)	6.912(3.596)	5.939 (2.923)	3.397(0.874)	3.732(1.245)	3.377 (0.848)
	M20	8.670 (2.145)	8.968(2.119)	10.356(1.223)	4.863 (1.062)	5.067(1.113)	6.831(2.665)

Note : Les valeurs en gras indiquent les meilleurs résultats.

- Les Tableaux 3.1 et 3.2 montrent que (performances ISE) :
 - Les moyennes de ISE basées sur 500 replications diminuent à mesure que la taille de l'échantillon n augmente pour tous les estimateurs considérés ;
 - L'estimateur JLN-VM fonctionne beaucoup mieux que les estimateurs TS-VM et von Mises standard pour les modèles M1, M3, M4 (sauf pour $n = 100$), M7 (sauf pour $n = 25$), M8, M9, M13 (sauf pour $n = 25$) et M14 (sauf pour $n = 25$) ;
 - Pour les modèles M5 (sauf pour $n = 250$), M6 (sauf pour $n = 50$) et M17, l'estimateur de von Mises standard est plus performant que les autres estimateurs (TS-VM et JLN-VM) ;

TABLE 3.2: Valeurs empiriques de ISE ($\times 100$) et leurs écarts-types ($\times 100$) entre parenthèses basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 100$ et $n = 250$.

Groupe	modèle	$n = 100$			$n = 250$		
		\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}
G1	M1	0.347(0.515)	0.701(0.745)	0.041 (0.144)	0.111(0.237)	0.259(0.414)	0.001 (0.006)
	M2	0.624(0.411)	0.526(0.420)	0.476 (0.336)	0.338(0.208)	0.263 (0.198)	0.275(0.213)
	M3	1.374(1.111)	2.052(0.662)	0.950 (1.034)	0.652(0.386)	1.620(0.487)	0.456 (0.316)
	M4	0.640(0.486)	0.542 (0.408)	0.621(0.434)	0.351(0.233)	0.317(0.357)	0.281 (0.144)
	M5	7.273 (3.668)	8.382(4.760)	7.293(3.494)	2.914(1.817)	3.998(2.935)	2.466 (1.655)
	M6	2.460 (0.839)	3.395(1.422)	2.817(1.008)	1.423 (0.476)	1.807(1.033)	1.680(0.597)
G2	M7	1.427(0.807)	1.283(0.991)	1.080 (0.647)	0.670(0.326)	0.576(0.265)	0.545 (0.290)
	M8	1.743(0.870)	1.778(0.858)	1.432 (0.713)	0.678(0.241)	0.719(0.290)	0.539 (0.168)
	M9	0.865(0.378)	0.791(0.371)	0.677 (0.306)	0.593(0.295)	0.647(0.492)	0.375 (0.221)
	M10	2.872 (0.950)	3.809(0.991)	3.039(0.891)	2.018 (0.670)	3.326(1.168)	2.093(0.643)
G3	M11	1.634(0.760)	1.458(0.731)	1.399 (0.634)	0.862(0.359)	0.763(0.477)	0.678 (0.274)
	M12	1.069(0.402)	1.087(0.459)	0.991 (0.456)	0.594(0.291)	0.629(0.341)	0.506 (0.253)
	M13	2.106(0.997)	2.060(1.079)	1.907 (0.948)	0.923(0.297)	0.887(0.303)	0.817 (0.328)
	M14	1.901(0.875)	1.739(0.910)	1.673 (0.725)	0.974(0.473)	0.843(0.437)	0.807 (0.393)
	M15	0.867 (0.407)	0.946(0.631)	0.872(0.175)	0.647 (0.326)	0.715(0.377)	0.764(0.255)
	M16	2.634(0.643)	2.589(0.740)	2.449 (0.558)	1.177(0.426)	1.059(0.429)	1.031 (0.413)
G4	M17	6.509 (2.030)	7.781(2.389)	7.486(1.444)	3.025 (0.654)	3.268(1.382)	3.810(0.693)
	M18	3.195 (0.694)	3.563(0.759)	3.555(0.623)	1.356 (0.528)	1.487(0.867)	2.040(0.815)
	M19	2.583 (0.779)	2.938(0.828)	2.671(0.826)	1.489 (0.333)	1.980(0.522)	1.772(0.275)
	M20	3.241 (0.968)	3.388(1.001)	3.483(0.961)	2.065 (0.632)	2.178(0.686)	2.209(0.722)

Note : Les valeurs en gras indiquent les meilleurs résultats.

TABLE 3.3: Valeurs empiriques de ISB ($\times 100$) basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 25$ et $n = 50$.

Groupe	modèle	$n = 25$			$n = 50$		
		\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}
G1	M1	0.070	0.081	0.041	0.035	0.057	0.005
	M2	0.635	0.282	0.794	0.247	0.101	0.200
	M3	0.642	0.996	0.477	0.755	1.489	0.379
	M4	0.159	0.102	0.308	0.103	0.098	0.147
	M5	8.184	10.987	8.369	7.250	8.248	8.015
	M6	3.021	3.092	4.147	1.628	2.284	2.085
G2	M7	0.434	0.211	0.748	0.384	0.299	0.373
	M8	1.395	1.353	1.779	0.758	0.752	0.576
	M9	0.327	0.254	0.451	0.370	0.217	0.312
	M10	3.023	3.243	3.060	2.074	2.377	2.449
G3	M11	0.988	0.495	2.903	0.572	0.341	1.072
	M12	0.733	0.513	2.811	0.431	0.314	0.607
	M13	1.746	1.234	3.912	0.590	0.494	0.784
	M14	2.190	1.396	3.386	0.712	0.390	0.926
	M15	0.410	0.316	0.811	0.304	0.256	0.727
	M16	2.912	1.926	5.298	0.758	0.413	1.256
G4	M17	8.021	8.129	8.684	5.651	5.794	7.291
	M18	3.554	3.165	3.601	3.387	3.032	3.665
	M19	2.241	2.198	2.810	1.429	1.379	1.852
	M20	5.978	4.973	9.284	2.303	1.867	4.351

Note : Les valeurs en gras indiquent les meilleurs résultats.

TABLE 3.4: Valeurs empiriques de ISB ($\times 100$) basées sur 500 replications pour les modèles considérés précédemment avec des tailles d'échantillon $n = 100$ et $n = 250$.

Groupe	modèle	$n = 100$			$n = 250$		
		\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{VM}	\hat{f}_ν^{TS-VM}	\hat{f}_ν^{JLN-VM}
G1	M1	0.010	0.022	0.002	0.004	0.009	3.084e-05
	M2	0.156	0.075	0.076	0.093	0.080	0.034
	M3	0.437	1.540	0.142	0.170	1.362	0.053
	M4	0.053	0.049	0.050	0.028	0.046	0.027
	M5	6.041	6.704	6.143	2.218	2.919	1.790
	M6	1.170	1.802	1.575	0.679	0.813	0.871
G2	M7	0.377	0.354	0.246	0.167	0.250	0.055
	M8	0.560	0.703	0.365	0.197	0.404	0.086
	M9	0.304	0.281	0.229	0.161	0.208	0.104
	M10	1.575	2.434	1.797	1.567	2.701	1.654
G3	M11	0.392	0.283	0.302	0.229	0.226	0.123
	M12	0.213	0.217	0.244	0.154	0.202	0.106
	M13	0.585	0.530	0.505	0.273	0.366	0.188
	M14	0.388	0.253	0.316	0.257	0.242	0.126
	M15	0.327	0.173	0.675	0.162	0.104	0.424
	M16	0.530	0.346	0.531	0.240	0.208	0.136
G4	M17	5.358	6.171	6.702	2.334	2.277	3.176
	M18	2.013	2.056	2.839	0.699	0.583	1.389
	M19	1.467	1.779	1.778	1.016	1.403	1.404
	M20	1.449	1.194	1.925	1.029	0.985	1.204

Note : Les valeurs en gras indiquent les meilleurs résultats.

- Pour le reste des modèles (M10, M11, M12, M15, M16, M18 et M19), les performances des trois estimateurs sont mixtes en fonction de la taille de l'échantillon n .
2. Les Tableaux 3.3 et 3.4 montrent clairement que (performances ISB) :
- Les valeurs ISB diminuent à mesure que la taille de l'échantillon n augmente pour les trois estimateurs ;
 - L'estimateur JLN-VM surpasse les estimateurs TS-VM et von Mises standard pour les modèles M1, M3, M8 (sauf pour $n = 25$) ;
 - L'estimateur TS-VM fonctionne mieux que les estimateurs JLN-VM et von Mises standard pour les modèles M2 (sauf pour $n = 250$), M3 (sauf pour $n = 250$), M7 (sauf pour $n = 250$), M11 (sauf pour $n = 250$), M12 (sauf pour $n = 250$), M13 (sauf pour $n = 250$), M14 (sauf pour $n = 250$), M15, M16 (sauf pour $n = 250$), M18 et M20 ;
 - L'estimateur von Mises standard donne de meilleurs résultats que les estimateurs MBC-VM (TS-VM et JLN-VM) pour les modèles M5 (sauf pour $n = 250$), M6, M10 et M17 (sauf pour $n = 250$) ;

- Pour le reste des modèles, les performances des trois estimateurs sont mixtes selon la taille de l'échantillon n .

Nous pouvons observer aussi que pour certains modèles et certaines valeurs de la taille d'échantillon n , en particulier pour $n = 25$, l'estimateur standard de von Mises (qui n'est pas meilleur en terme de ISB théorique) fonctionne mieux que les estimateurs MBC-VM au sens d'ISE. Ce résultat peut s'expliquer par la petite taille de l'échantillon ou par les valeurs de paramètre de concentration (calculées selon l'approche LCV).

3.4 Illustration avec des données réelles

Cette section présente l'application des estimateurs MBC-VM pour les deux jeux de données réels suivants :

- **Direction des tortues.** Ces données concernent les mesures des directions prises pour 76 tortues ; voir par exemple [Gould \[1957\]](#). L'ensemble de données des tortues a été discuté et analysé par de nombreux auteurs en considérant les modèles circulaires paramétriques ; voir par exemple [Fisher \[1993\]](#), [Mardia and Jupp \[2000\]](#) et [Rao and Sengupta \[2001\]](#)).
- **Orientation des libellules.** Ces données consistent en l'orientation de 214 libellules par rapport à l'azimut du soleil (voir par exemple [Batschelet \[1981\]](#) et le package `NPCirc` [Oliveira et al. \[2014b\]](#)) de R ([R Development Core Team \[2014\]](#)). Notons que ces données bimodales ont été analysées récemment par [Oliveira et al. \[2012\]](#) en utilisant l'estimateur standard de von Mises avec différents sélecteurs de paramètre de concentration (RT, PI, LCV et UCV), et très récemment par [Bedouhene and Zougab \[2020\]](#) en utilisant l'estimateur standard de von Mises avec le sélecteur bayésien pour le paramètre de concentration ; voir aussi la Section 2.5.

Nous appliquons maintenant les estimateurs MBC-VM (TS-VM et JLN-VM) développés dans ce travail pour estimer les densités des données d'orientation des tortues et des libellules. Pour l'estimateur TS-VM, nous avons utilisé trois valeurs de $c \in \{0.25, 0.5, 0.75\}$. À titre de comparaison, l'estimateur de von Mises (standard) est utilisé pour estimer les densités des données considérées. La technique LCV est utilisée pour le choix du paramètre

de concentration. Afin d'évaluer les performances des trois estimateurs, nous calculons le critère de log-vraisemblance. Les résultats sont reportés dans les Tableaux 3.5 et 3.6.

TABLE 3.5: *Paramètre de concentration optimal avec LCV et log-vraisemblance pour les trois estimateurs utilisant les données de direction des tortues.*

Direction des données des tortues avec $n = 76$					
Estimateur	\hat{f}^{VM}	$\hat{f}_{\nu;c=0.5}^{TS-VM}$	$\hat{f}_{\nu;c=0.25}^{TS-VM}$	$\hat{f}_{\nu;c=0.75}^{TS-VM}$	\hat{f}_{ν}^{JLN-VM}
Paramètre de concentration optimal avec LCV	8.49	3.89	5.21	3.21	4.97
log-vraisemblance	-106.31	-103.85	-104.23	-103.71	-109.84

TABLE 3.6: *Paramètre de concentration optimal avec LCV et log-vraisemblance pour les trois estimateurs utilisant les données d'orientation des libellules.*

Données d'orientation des libellules avec $n = 214$					
Estimateur	\hat{f}_{ν}^{VM}	$\hat{f}_{\nu;c=0.5}^{TS-VM}$	$\hat{f}_{\nu;c=0.25}^{TS-VM}$	$\hat{f}_{\nu;c=0.75}^{TS-VM}$	\hat{f}_{ν}^{JLN-VM}
Paramètre de concentration optimale avec LCV	35.36	11.11	14.73	9.37	21.15
log-vraisemblance	-240.76	-237.52	-237.60	-237.41	-248.99

Les Tableaux 3.5 et 3.6 montrent que pour les deux ensembles de données, l'estimateur TS-VM est plus performant que les autres estimateurs (JLN-VM et von Mises standard) au sens de log-vraisemblance. Pour l'estimateur TS-VM, les meilleurs résultats sont obtenus avec $c = 0.5$ et $c = 0.75$.

Les représentations graphiques des densités circulaires et linéaires des estimateurs MBC-VM (TS-VM et JLN-VM) et von Mises standard sont également données par la Figure 3.2. Nous pouvons voir que chacun des trois estimateurs est capable de saisir la forme bi-modale pour les deux jeux de données. La qualité de lissage des trois estimateurs peut être considérée comme similaire sauf sur certains points, en particulier pour les modes ; on observe que l'estimateur standard de von Mises peut présenter les phénomènes de sous et sur-lissage.

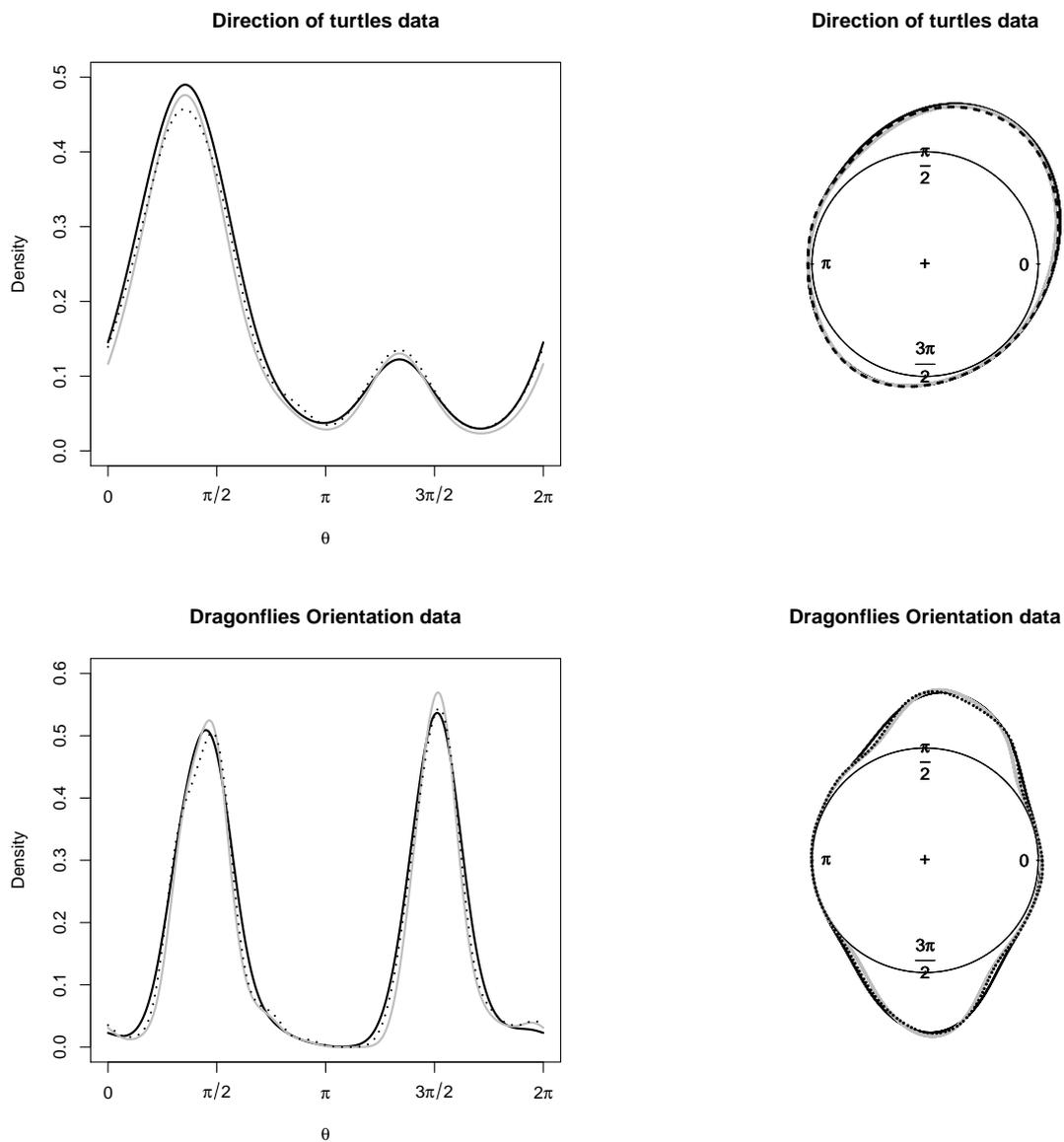


FIGURE 3.2: Estimation des densités des données de direction des tortues (première rangée) et des données d'orientation des libellules (deuxième rangée). Ligne pleine noire : estimateur TS-VM; Ligne pleine grise : estimateur JLN-VM; Ligne pointillé noire : estimateur standard de von Mises.

3.5 Conclusion

Dans ce chapitre, nous avons appliqué deux techniques MBC de réduction du biais. Ces techniques nous ont permis de construire deux nouveaux estimateurs à noyau circulaire de von Mises dont le biais est amélioré de $O(\nu^{-1})$ à $O(\nu^{-2})$. Nous avons également adapté la technique de validation croisée de vraisemblance pour le choix du paramètre de concentration des estimateurs développés. Les résultats sur des données simulées et sur des données réelles montrent la bonne performance des estimateurs MBC de von Mises au sens des critères ISE et ISB, pour plusieurs modèles circulaires et différentes tailles d'échantillon.

Chapitre 4

Estimation semi-paramétrique par noyau circulaire de la fonction densité

4.1 Introduction

L'objectif de ce chapitre est triple. Dans un premier temps, nous proposons l'estimateur semi-paramétrique par noyau circulaire, ainsi que ses propriétés statistiques. Dans un second temps, nous construisons un nouveau estimateur semi-paramétrique par noyau circulaire, en utilisant la technique JSH de [Jones et al. \[1999\]](#) afin d'accélérer la convergence du biais. Les propriétés statistiques de cet nouvel estimateur seront également étudiées. Comme troisième objectif, une application sur des données simulées et des données réelles sera effectuée pour comparer les performances des estimateurs développés.

4.2 Estimateur semi-paramétrique par noyau circulaire de la fonction densité

Nous développons les estimateurs à noyau circulaire de la fonction densité. L'approche semi-paramétrique par la méthode du noyau dans le cas des données linéaires a été pro-

4.2 Estimateur semi-paramétrique par noyau circulaire de la fonction densité

posée par [Hjort and Glad \[1995\]](#) (estimateur HG), en utilisant les noyaux symétriques (classiques); voir aussi [Hagmann and Scaillet \[2007\]](#) en utilisant les noyaux asymétriques pour estimer des densités à support \mathbb{R}^+ et [Kokonendji et al. \[2009a\]](#) en utilisant les noyaux discrets dans le cas des données de dénombrement. Afin d'améliorer l'estimateur semi-paramétrique HG au sens du biais, [Jones et al. \[1999\]](#) ont proposés un nouveau estimateur semi-paramétrique (estimateur JSH), en remplaçant le modèle paramétrique dans l'estimateur HG par le modèle semi-paramétrique HG. Plus récemment, [Hirukawa and Sakudo \[2019\]](#) ont développés l'estimateur semi-paramétrique JSH en utilisant les noyaux asymétriques dans le cas des données linéaires positives. Ces études ont montrés que l'estimation semi-paramétrique par noyau est compétitive aussi bien avec l'estimation paramétrique qu'avec l'estimation non-paramétrique. À notre connaissance, l'estimation semi-paramétrique par noyau circulaire n'a jamais été proposée dans la littérature. Donc, l'objectif principal de cette section est de proposer cette approche dans le cas des données circulaires.

Soit Θ une variable aléatoire circulaire de fonction densité circulaire inconnue f sur $[0, 2\pi[$. Considérons l'identité suivante

$$f(\theta) = g(\theta) \left\{ \frac{f(\theta)}{g(\theta)} \right\} := g(\theta)r(\theta), \quad (4.1)$$

où $g(\theta)$ est un modèle de départ (initial); voir ci-dessous la discussion sur le choix de g et $r(\theta)$ est un facteur de correction. L'approche semi-paramétrique par noyau circulaire consiste à estimer g par un modèle paramétrique ou semi-paramétrique circulaire et estimer f dans r par l'approche non-paramétrique en utilisant la méthode du noyau circulaire.

Soient $\Theta_1, \dots, \Theta_n$ un n -échantillon aléatoire d'une densité circulaire continue bornée f . Suivant l'identité (4.1), nous investiguons l'estimateur semi-paramétrique de la forme suivante :

$$\tilde{f}(\theta) := g(\theta)\hat{r}(\theta) := g(\theta) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_\nu(\theta - \Theta_i)}{g(\Theta_i)} \right\}, \quad (4.2)$$

où g est le modèle de départ qui peut être choisi comme une distribution circulaire uniforme, un modèle paramétrique circulaire connu (fixe), un modèle paramétrique circulaire inconnu ou encore un modèle semi-paramétrique circulaire, $K_\nu(\cdot)$ est un noyau circulaire

et ν est le paramètre de concentration. Selon le choix de g , plusieurs estimateurs peuvent en découler de l'estimateur (4.2) :

- Si g est une distribution circulaire uniforme; i.g., $g(\theta) = 1/2\pi$, alors l'estimateur (4.2) devient l'estimateur non-paramétrique circulaire standard donné par (1.3), étudié par Di Marzio et al. [2009], Di Marzio et al. [2011a] et Taylor [2008]; voir aussi la Définition 2 de la Section 1.2.
- Si g est un estimateur non-paramétrique circulaire standard, l'estimateur (4.2) devient l'estimateur non-paramétrique JLN donné par l'équation (3.2), étudié par Bedouhene and Zougab [2019]; voir aussi la Section 3.2.
- Si g est un modèle paramétrique circulaire inconnu, alors l'estimateur (4.2) est l'estimateur semi-paramétrique HG à noyau circulaire. Cet estimateur sera présenté et étudié dans la section 4.3.
- Si g est un modèle semi-paramétrique circulaire, alors l'estimateur donné par (4.2) est l'estimateur semi-paramétrique JSH à noyau circulaire. Cet estimateur sera présenté et étudié dans la Section 4.4.

4.3 Estimateur semi-paramétrique HG à noyau circulaire de la fonction densité

Suivant les travaux de Hjort and Glad [1995] et Haggmann and Scaillet [2007], l'estimateur semi-paramétriques circulaire HG de f est défini par :

$$\widehat{f}_\nu^{HG}(\theta) = g(\theta; \widehat{\vartheta}) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_\nu(\theta - \Theta_i)}{g(\Theta_i; \widehat{\vartheta})} \right\}. \quad (4.3)$$

où $\widehat{\vartheta}$ est l'estimateur de maximum de vraisemblance de $\vartheta \in \mathcal{V} \subseteq \mathbb{R}^p$ du modèle de départ g ; il peut être choisi comme un modèle de von Mises de paramètres inconnus $\vartheta = (\mu, k)$.

En utilisant le noyau circulaire de von Mises, l'estimateur (4.3) s'écrit comme suit :

$$\widehat{f}_\nu^{HG-VM}(\theta) = g(\theta; \widehat{\vartheta}) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \Theta_i)\}}{g(\Theta_i; \widehat{\vartheta})} \right\}. \quad (4.4)$$

4.3.1 Biais et variance de l'estimateur HG-VM

Pour examiner les propriétés de l'estimateur semi-paramétrique (4.4), nous introduisons des notations en plus. Soit ϑ_0 la pseudo vraie valeur qui minimise la distance de Kulleback-Leibler de $g(\theta, \vartheta)$ à partir de la vraie fonction $g(\theta)$. Notons aussi $g_0(\cdot) = g(\cdot, \vartheta_0)$ et $r(\cdot) = f(\cdot)/g_0(\cdot)$. Le théorème suivant présente le biais et la variance de l'estimateur HG-VM.

Théorème 4.1. *Soit \hat{f}_ν^{HG-VM} l'estimateur HG-VM défini par (4.4). Soit $g_0(\cdot) = g(\cdot, \vartheta_0)$ le meilleur départ approchant la fonction inconnue f au sens de la distance de Kullback-Leibler, et soit $r(\cdot) = f/g_0$ une fonction poids deux fois continuellement dérivables. Lorsque $\nu \rightarrow \infty$ et $\nu^{1/2}/n \rightarrow 0$ quand $n \rightarrow \infty$, alors :*

(i) le biais de l'estimateur HG-VM est donné par :

$$\text{biais}[\hat{f}_\nu^{HG-VM}(\theta)] = \frac{1}{4\nu}g_0(\theta)r''(\theta) + o(\nu^{-1}),$$

(ii) la variance de l'estimateur HG-VM est donné par :

$$\mathbb{V}[\hat{f}_\nu^{HG-VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}}f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right).$$

Démonstration

Un développement de Taylor permet d'avoir

$$\begin{aligned} \frac{g(\theta; \hat{\vartheta})}{g(\Theta_i; \hat{\vartheta})} &= \exp\{\log g(\theta; \hat{\vartheta}) - \log g(\Theta_i; \hat{\vartheta})\} \\ &\doteq \frac{g_0(\theta)}{g_0(\Theta_i)} + \frac{g_0(\theta)}{g_0(\Theta_i)} \{u_0(\theta) - u_0(\Theta_i)\}^\top (\hat{\vartheta} - \vartheta_0) \\ &= \frac{g_0(\theta)}{g_0(\Theta_i)} \left[1 - \{u_0(\Theta_i) - u_0(\theta)\}^\top (\hat{\vartheta} - \vartheta_0) \right], \end{aligned}$$

où $u_0(\theta) = \partial \log g(\theta; \vartheta_0)/\partial \vartheta$. Maintenant, l'estimateur semi-paramétrique (4.4) peut être approximer comme suit :

$$\hat{f}_\nu^{HG-VM}(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \Theta_i)\} \frac{g_0(\theta)}{g_0(\Theta_i)} \left[1 - \{u_0(\Theta_i) - u_0(\theta)\}^\top (\hat{\vartheta} - \vartheta_0) \right]. \quad (4.5)$$

En utilisant la représentation (4.5), les propriétés de la variable aléatoire de von Mises et le développement de Taylor lorsque $\nu \rightarrow \infty$ (voir [Mardia and Jupp \[2000\]](#)) et les mêmes étapes de calcul de [Hjort and Glad \[1995\]](#), nous obtenons les résultats donnés dans le théorème 4.1.

Remarque 3. *Notons que la variance asymptotique de l'estimateur HG-VM est similaire à celle de l'estimateur standard de von Mises.*

4.3.2 MISE de l'estimateur HG-VM

Le critère global à utiliser pour l'estimateur HG-VM (4.4) est la MISE. En utilisant les expressions du biais et de la variance données dans le théorème 4.1, la MISE de l'estimateur HG-VM s'exprime comme suit :

$$\text{MISE}[\widehat{f}_\nu^{\text{HG-VM}}] = \frac{1}{16\nu^2} \int_0^{2\pi} \{g_0(\theta)r''(\theta)\}^2 d\theta + \frac{\nu^{1/2}}{2n\sqrt{\pi}} + o\left(\nu^{-2} + \frac{\nu^{1/2}}{n}\right). \quad (4.6)$$

Le paramètre de concentration au sens de MISE de l'estimateur (4.4) est donné par :

$$\nu_{\text{HG-VM}}^{\text{opt}} = \left\{ \frac{n\sqrt{\pi}}{2} \int_0^{2\pi} \{g_0(\theta)r''(\theta)\}^2 d\theta \right\}^{2/5}. \quad (4.7)$$

En remplaçant le paramètre de concentration optimal (4.7) dans la MISE (4.6) de l'estimateur HG-VM, la MISE optimale pour l'estimateur HG-VM est donnée par

$$\text{MISE}^{\text{opt}}[\widehat{f}_\nu^{\text{HG-VM}}] = \frac{1}{2\sqrt{\pi}} \left(\frac{\sqrt{\pi}}{2} \int_0^{2\pi} \{g_0(\theta)r''(\theta)\}^2 d\theta \right)^{1/5} \left(1 + \frac{1}{4} \left\{ \int_0^{2\pi} \{g_0(\theta)r''(\theta)\}^2 d\theta \right\}^{-1} \right) n^{-4/5}.$$

L'ordre de convergence de l'estimateur HG-VM au sens de MISE est de l'ordre $O(-4/5)$. Notons que le paramètre de concentration (4.7) ne peut être utilisé dans la pratique car il dépend de la quantité inconnue r . Nous adaptons la méthode UCV pour approcher la valeur $\nu_{\text{HG-VM}}^{\text{opt}}$. Le paramètre de concentration ν_{ucv} dans le cas de l'estimateur HG-VM s'obtient en minimisant par rapport à ν la fonction $UCV(\nu)$ donnée par :

$$UCV(\nu) = \frac{1}{n^2} \int_0^{2\pi} \{\widehat{f}_\nu^{\text{HG-VM}}(\theta)\}^2 d\theta - \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} K_\nu(\Theta_i - \Theta_j) \frac{g(\Theta_i; \widehat{\vartheta})}{g(\Theta_j; \widehat{\vartheta})}.$$

où $K_\nu(\cdot)$ est le noyau de von Mises.

4.4 Estimateur semi-paramétrique JSH à noyau circulaire de la fonction densité

En suivant l'idée de Jones et al. [1999] et Hirukawa and Sakudo [2019] l'estimateur semi-paramétrique circulaire JSH de f est donné comme suit :

$$\widehat{f}_\nu^{JSH}(\theta) = \widehat{f}_\nu^{HG}(\theta) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{K_\nu(\Theta_i - \Theta_j)}{\widehat{f}_\nu^{HG}(\theta)} \right\}. \quad (4.8)$$

où $\widehat{f}_\nu^{HG}(\cdot)$ est l'estimateur circulaire HG donné par l'équation (4.3).

En utilisant le noyau circulaire de von Mises, l'estimateur (4.8) s'écrit comme suit :

$$\widehat{f}_\nu^{JSH-VM}(\theta) = \widehat{f}_\nu^{HG-VM}(\theta) \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \Theta_i)\}}{\widehat{f}_\nu^{HG-VM}(\theta)} \right\}. \quad (4.9)$$

Notons que l'estimateur semi-paramétrique JSH s'obtient en remplaçant g dans (4.2) par l'estimateur semi-paramétrique HG.

4.4.1 Biais et variance de l'estimateur JSH-VM

Le théorème suivant présente le biais et la variance de l'estimateur JSH-VM donné par l'équation (4.9). Nous considérons les mêmes notations de la section 4.3.1.

Théorème 4.2. *Soit \widehat{f}_ν^{JSH-VM} l'estimateur JSH-VM défini par (4.9). Soit $g_0(\cdot) = g(\cdot, \vartheta_0)$ le meilleur départ approchant la fonction inconnue f au sens de la distance de Kullback-Leibler, et soit $r(\cdot) = f/g_0$ une fonction quatre fois continuellement dérivables. Lorsque $\nu \rightarrow \infty$ et $\nu^{1/2}/n \rightarrow 0$ quand $n \rightarrow \infty$, alors :*

(i) *le biais de l'estimateur JSH-VM est donné par :*

$$\text{biais}[\widehat{f}_\nu^{JSH-VM}(\theta)] = -f(\theta)q(\theta, r) \frac{1}{\nu^2} + o\left(\frac{1}{\nu^2}\right),$$

$$\text{où } q(\theta, r) = \left(\frac{r''(\theta)}{4r(\theta)}\right)'',$$

(ii) *la variance de l'estimateur JSH-VM est donné par :*

$$\mathbb{V}[\widehat{f}_\nu^{JSH-VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right).$$

Démonstration

Il est facile de montrer que lorsque $\nu \rightarrow \infty$ et $\nu^{1/2}/n \rightarrow 0$ quand $n \rightarrow \infty$,

$$\mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \Theta_i)\}}{g(\Theta_i; \hat{\vartheta})} \right] \simeq r(\theta) + \frac{1}{4\nu} r''(\theta), \quad (4.10)$$

et

$$\mathbb{V} \left[\frac{1}{n} \sum_{i=1}^n \frac{\frac{1}{2\pi I_0(\nu)} \exp\{\nu \cos(\theta - \Theta_i)\}}{g(\Theta_i; \hat{\vartheta})} \right] \simeq \frac{\nu^{1/2}}{2n\sqrt{\pi}} \frac{r(\theta)}{g_0(\theta)}. \quad (4.11)$$

En combinant maintenant la preuve du théorème 3.2 et les approximations (4.10) et (4.11), nous établissons les résultats du théorème 4.2.

Remarque 4. *Le théorème 4.2 montre que l'estimateur JSH-VM réduit l'ordre de grandeur du biais de $O(\nu^{-1})$ à $O(\nu^{-2})$, alors que les variances des estimateurs JSH-VM et HG-VM sont similaires.*

4.4.2 MISE de l'estimateur JSH-VM

Nous utilisons la MISE comme critère global. La MISE de l'estimateur JSH-VM est donnée comme suit :

$$\text{MISE}[\hat{f}_\nu^{JSH-VM}] = \frac{1}{\nu^4} \mathbb{E} [f(\Theta) q^2(\Theta, r)] + \frac{\nu^{1/2}}{2n\sqrt{\pi}} + o\left(\nu^{-4} + \frac{\nu^{1/2}}{n}\right). \quad (4.12)$$

Le paramètre de concentration au sens de MISE de l'estimateur (4.9) est donné par :

$$\nu_{JSH-VM}^{opt} = \{16 n \sqrt{\pi} \mathbb{E} [f(\Theta) q^2(\Theta, r_0)]\}^{2/9}. \quad (4.13)$$

En remplaçant le paramètre de concentration optimal (4.13) dans la MISE (4.12) de l'estimateur JSH-VM, la MISE optimale pour l'estimateur JSH-VM est donnée par

$$\text{MISE}^{opt}[\hat{f}_\nu^{JSH-VM}] = \left\{ \frac{1}{(16\sqrt{\pi})^{8/9}} + \frac{1}{\sqrt{\pi}^{8/9} 16^{5/36}} \right\} \{ \mathbb{E} [f(\Theta) q^2(\Theta, r)] \}^{1/9} n^{-8/9}.$$

L'ordre de convergence de l'estimateur JSH-VM au sens de MISE est $O(n^{-8/9})$, qui est plus rapide que celui de l'estimateur HG-VM.

Le paramètre de concentration donné par (4.13) ne peut être exploité dans la pratique. Donc, nous adaptons également la technique UCV pour estimer ν pour l'estimateur JSH-VM. Le paramètre de concentration s'obtient aussi en minimisant par rapport à ν la fonction $UCV(\nu)$ donnée par :

$$UCV(\nu) = \frac{1}{n^2} \int_0^{2\pi} \{\widehat{f}_\nu^{JSH-VM}(\theta)\}^2 d\theta - \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} K_\nu(\Theta_i - \Theta_j) \frac{\widehat{f}_\nu^{HG-VM}(\Theta_i)}{\widehat{f}_\nu^{HG-VM}(\Theta_j)}.$$

où $K_\nu(\cdot)$ est le noyau de von Mises.

4.5 Etude de simulation

Dans cette section, nous examinons et comparons les performances des estimateurs JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises. Nous simulons des données suivant les 20 modèles circulaires déjà considérés dans les chapitres précédents (Chapitre 2 et 3); voir la Section 2.4 pour la représentation graphique (Figure 2.1) et les expressions de ces modèles.

Pour chaque modèle, 500 replications de la taille d'échantillon $n = 10, 50, 100$ et 200 sont générées. On utilise les estimateurs JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises pour estimer la densité des données générées à partir des modèles considérés.

Dans le cas des estimateurs semi-paramétriques HG-VM et JHS-VM, nous avons considéré le modèle de von Mises de paramètres μ et k comme modèle de départ. Les paramètres μ et k sont estimés par la méthode de maximum de vraisemblance.

Nous examinons les performances des estimateurs via l'erreur quadratique intégrée (ISE) et le biais quadratique intégré (ISB) donnés respectivement par

$$\text{ISE} := \int_0^{2\pi} \left[\widehat{f}_\nu(\theta) - f(\theta) \right]^2 d\theta \quad (4.14)$$

et

$$\text{ISB} := \int_0^{2\pi} \left[\mathbb{E}\{\widehat{f}_\nu(\theta)\} - f(\theta) \right]^2 d\theta, \quad (4.15)$$

où \widehat{f}_ν est l'estimateur JHS-VM, HG-VM, JLN-VM ou l'estimateur standard de von Mises.

TABLE 4.1: Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 10$

modèle	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	0.02850478 (0.04012697)	0.04462439 (0.04853553)	0.08559806 (0.04279636)	0.06505175 (0.04654755)
M2	0.04473020 (0.02772726)	0.05453249 (0.03829670)	0.04163577 (0.03285210)	0.05642209 (0.04345689)
M3	0.08091215 (0.06307888)	0.07751351 (0.04542814)	0.06671344 (0.04865496)	0.06143573 (0.02990166)
M4	0.03164352 (0.01871130)	0.03537833 (0.02020315)	0.03297867 (0.03069572)	0.03632635 (0.02175982)
M5	0.17595160 (0.03988868)	0.14875420 (0.05087804)	0.16527790 (0.04691107)	0.13641840 (0.05053080)
M6	0.08191310 (0.03095106)	0.07686330 (0.02400437)	0.07152971 (0.01955158)	0.07529404 (0.01923726)
M7	0.05637439 (0.02421021)	0.05487096 (0.01960684)	0.06706219 (0.04029468)	0.05718308 (0.02670783)
M8	0.07343544 (0.03444620)	0.06558899 (0.03115417)	0.06881742 (0.03042595)	0.06210159 (0.03364485)
M9	0.03808874 (0.02237555)	0.03738602 (0.02200587)	0.05825203 (0.02927019)	0.04340771 (0.02610141)
M10	0.10130150 (0.05048216)	0.10792480 (0.05797116)	0.10451900 (0.05881267)	0.10903300 (0.06235252)
M11	0.05150170 (0.01270330)	0.05767646 (0.01423613)	0.06909614 (0.03366367)	0.06769526 (0.03262917)
M12	0.06681441 (0.03873810)	0.07464401 (0.04263112)	0.10137160 (0.03964287)	0.10465530 (0.04097478)
M13	0.09751008 (0.03463271)	0.09630297 (0.03442454)	0.12658380 (0.06781179)	0.09534461 (0.04633361)
M14	0.09310941 (0.02139324)	0.08182814 (0.03025542)	0.11640170 (0.02926208)	0.09127352 (0.03469707)
M15	0.01899112 (0.01881756)	0.02567207 (0.02610819)	0.07152665 (0.06830387)	0.06649503 (0.05113047)
M16	0.08533305 (0.01426127)	0.09042381 (0.01499513)	0.13478820 (0.05539321)	0.10867360 (0.02809544)
M17	0.13517040 (0.05339199)	0.11943330 (0.04448334)	0.11110740 (0.01939895)	0.12119370 (0.04778508)
M18	0.07427035 (0.03749352)	0.07631341 (0.03790166)	0.10253500 (0.10578010)	0.09212988 (0.07697814)
M19	0.08008118 (0.02717786)	0.08591874 (0.03630872)	0.09025896 (0.02483154)	0.09431527 (0.03575715)
M20	0.12191930 (0.02969994)	0.12765390 (0.03915187)	0.15078870 (0.04786072)	0.14343000 (0.04014303)

TABLE 4.2: Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 50$

modèle	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	0.00186695 (0.00348982)	0.00623126 (0.00591743)	0.01391779 (0.00501099)	0.01122737 (0.00508161)
M2	0.01031364 (0.00856156)	0.01104468 (0.00994357)	0.00801816 (0.00604534)	0.01151185 (0.00903454)
M3	0.04183568 (0.01469642)	0.01786018 (0.00825009)	0.02830937 (0.01100455)	0.01419335 (0.00710870)
M4	0.00894551 (0.00804563)	0.01454852 (0.01196749)	0.01428344 (0.01501745)	0.01899995 (0.01363013)
M5	0.15422940 (0.01944151)	0.10999400 (0.02811483)	0.13981540 (0.01711945)	0.10353740 (0.02912551)
M6	0.05857521 (0.00995267)	0.04769840 (0.01159737)	0.04818761 (0.00638206)	0.05175939 (0.01690761)
M7	0.01958404 (0.00713345)	0.01435689 (0.00679112)	0.02257649 (0.00545522)	0.01470822 (0.00706927)
M8	0.02630641 (0.01247632)	0.01893797 (0.01413751)	0.03431719 (0.02150426)	0.02001173 (0.01708515)
M9	0.00966611 (0.00614507)	0.00994597 (0.00473333)	0.01356273 (0.00781459)	0.00947036 (0.00479179)
M10	0.05259985 (0.01119882)	0.04210774 (0.01328534)	0.04438010 (0.01035650)	0.04205748 (0.01487030)
M11	0.03150109 (0.00906005)	0.02460683 (0.01062329)	0.03634391 (0.01132911)	0.02505415 (0.01027391)
M12	0.01745451 (0.01179003)	0.01577841 (0.00786157)	0.02064748 (0.01096969)	0.01475578 (0.00571839)
M13	0.04200833 (0.00672970)	0.03118269 (0.00691712)	0.04853376 (0.00857784)	0.03252626 (0.00716394)
M14	0.05397754 (0.00715094)	0.04052399 (0.00781194)	0.06413296 (0.01562727)	0.04395908 (0.01056756)
M15	0.00917396 (0.00410681)	0.01230112 (0.00609103)	0.01912481 (0.00676723)	0.01779463 (0.00822804)
M16	0.07426608 (0.00769467)	0.06225453 (0.00958447)	0.07419915 (0.00892357)	0.06387056 (0.01024718)
M17	0.08134952 (0.00891426)	0.07171065 (0.01146586)	0.07832820 (0.01158603)	0.07180320 (0.01315921)
M18	0.04071024 (0.00590656)	0.04247675 (0.00628049)	0.04032444 (0.00515277)	0.04397529 (0.00753345)
M19	0.03576450 (0.00690464)	0.03509849 (0.00854059)	0.04189057 (0.01496249)	0.03581118 (0.00983269)
M20	0.08059858 (0.00596337)	0.07290565 (0.00740366)	0.08678004 (0.00506745)	0.07547261 (0.00707327)

TABLE 4.3: Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 100$

modèle	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	0.00145798 (0.00202581)	0.00310613 (0.00367138)	0.00713417 (0.00243086)	0.00821801 (0.00265626)
M2	0.00761880 (0.00769243)	0.00678811 (0.00590393)	0.00509324 (0.00500747)	0.00638528 (0.00554334)
M3	0.03766481 (0.00396919)	0.01369442 (0.00281471)	0.02601476 (0.00453876)	0.01050432 (0.00336118)
M4	0.00621909 (0.00424795)	0.00667069 (0.00533231)	0.00636381 (0.00390589)	0.00848213 (0.00516269)
M5	0.15407720 (0.01705888)	0.10638870 (0.01756875)	0.13848010 (0.01654077)	0.11923760 (0.04176512)
M6	0.05717849 (0.01060823)	0.04770299 (0.01698684)	0.04716118 (0.00770533)	0.05237416 (0.01985294)
M7	0.01753147 (0.00887950)	0.01060898 (0.00679619)	0.02180887 (0.01113304)	0.01107669 (0.00752539)
M8	0.02178118 (0.00535852)	0.01161516 (0.00504968)	0.02743689 (0.01046127)	0.01173104 (0.00578429)
M9	0.00618915 (0.00325880)	0.00519624 (0.00240630)	0.00764648 (0.00470137)	0.00487528 (0.00333116)
M10	0.04228818 (0.00819922)	0.03387987 (0.00550184)	0.03661454 (0.00663503)	0.03387627 (0.00580191)
M11	0.02312423 (0.00296936)	0.01297110 (0.00286685)	0.02857466 (0.00398228)	0.01394591 (0.00317897)
M12	0.01064811 (0.00506038)	0.00927723 (0.00455566)	0.01396192 (0.00486633)	0.00966330 (0.00421735)
M13	0.04023392 (0.00592697)	0.02488033 (0.00516262)	0.04787923 (0.00881022)	0.02647213 (0.00518729)
M14	0.04843290 (0.00630208)	0.03353572 (0.00796658)	0.05574819 (0.00589554)	0.03598268 (0.00765842)
M15	0.00763593 (0.00307544)	0.00817408 (0.00385788)	0.00874854 (0.00333546)	0.00827658 (0.00493149)
M16	0.06428624 (0.00348783)	0.05354444 (0.00468357)	0.06885617 (0.00315623)	0.05493533 (0.00306463)
M17	0.07976272 (0.00656549)	0.06691367 (0.00702965)	0.07616668 (0.00849390)	0.06791977 (0.00961850)
M18	0.03526437 (0.00577512)	0.03452853 (0.00553621)	0.03679798 (0.00509439)	0.03675539 (0.00570463)
M19	0.02945109 (0.00508890)	0.02667399 (0.00441191)	0.03084276 (0.00729825)	0.02666992 (0.00469730)
M20	0.07251194 (0.00413394)	0.06295764 (0.00473350)	0.07528865 (0.00360919)	0.06351900 (0.00388762)

TABLE 4.4: Valeurs empiriques de ISE et leurs écarts-types entre parenthèses sur la base de 500 replications pour les modèles considérés précédemment avec $n = 200$

modèle	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	0.00084057 (0.00176523)	0.00201282 (0.00265810)	0.00531320 (0.00259564)	0.00479899 (0.00270302)
M2	0.00535917 (0.00478603)	0.00492401 (0.00343834)	0.00208093 (0.00172013)	0.00492401 (0.00278308)
M3	0.03363875 (0.00613477)	0.01028555 (0.00224874)	0.02198236 (0.00442052)	0.00696548 (0.00129721)
M4	0.00391135 (0.00173614)	0.00339908 (0.00224373)	0.00317418 (0.00156170)	0.00376158 (0.00216936)
M5	0.14771130 (0.00342673)	0.09819085 (0.00522127)	0.13336910 (0.00388504)	0.10454800 (0.04227488)
M6	0.05567328 (0.00500733)	0.03931279 (0.00745352)	0.04548991 (0.00286847)	0.04248277 (0.01063370)
M7	0.01455506 (0.00352526)	0.00688927 (0.00273345)	0.01777946 (0.00510629)	0.00738230 (0.00326673)
M8	0.02062638 (0.00284873)	0.00947469 (0.00297511)	0.02374947 (0.00637659)	0.00893474 (0.00409067)
M9	0.00586936 (0.00375266)	0.00480553 (0.00323351)	0.00654871 (0.00373216)	0.00454161 (0.00257367)
M10	0.04066848 (0.00510494)	0.03189044 (0.00503707)	0.03539285 (0.00346167)	0.03199239 (0.00562031)
M11	0.02204371 (0.00339720)	0.01138591 (0.00341360)	0.02578125 (0.00593941)	0.01166707 (0.00404637)
M12	0.00782771 (0.00189468)	0.00454132 (0.00223176)	0.01217328 (0.00279871)	0.00488892 (0.00242274)
M13	0.03617597 (0.00311877)	0.02066516 (0.00270780)	0.04078696 (0.00416984)	0.02138143 (0.00271427)
M14	0.04639794 (0.00219153)	0.02977327 (0.00220602)	0.05079901 (0.00262050)	0.03103936 (0.00274746)
M15	0.00698435 (0.00296696)	0.00689334 (0.00377953)	0.00694808 (0.00453604)	0.00682923 (0.00441401)
M16	0.06107204 (0.00236975)	0.04923269 (0.00338616)	0.06460646 (0.00317860)	0.05081207 (0.00402884)
M17	0.07914440 (0.00708559)	0.06675516 (0.00706777)	0.07560362 (0.00660960)	0.06722664 (0.01131454)
M18	0.03493310 (0.00393830)	0.03218582 (0.00368674)	0.03336144 (0.00279237)	0.03294279 (0.00428510)
M19	0.02867237 (0.00218188)	0.02327596 (0.00164740)	0.03006528 (0.00358708)	0.02324531 (0.00192141)
M20	0.07106963 (0.00197397)	0.06044618 (0.00251849)	0.07298356 (0.00157604)	0.06078166 (0.00188464)

TABLE 4.5: Valeurs empiriques de ISB basées sur 500 replications pour les modèles considérés avec des tailles d'échantillon $n = 10$ et $n = 50$.

	$n = 10$				$n = 50$			
	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	0.00521960	0.00803873	0.02129694	0.01025768	0.00011973	0.00016872	0.00249991	0.00071457
M2	0.01434544	0.01025751	0.00185564	0.00229880	0.00331689	0.00187921	0.00036110	0.00087784
M3	0.04949739	0.03710505	0.03479424	0.02216711	0.03834652	0.01224761	0.02525677	0.00820153
M4	0.00867268	0.00743377	0.00559897	0.00651841	0.00092518	0.00240400	0.00375355	0.00384231
M5	0.15202280	0.11383630	0.14045980	0.10000760	0.14839500	0.10082000	0.13321750	0.09321688
M6	0.05601477	0.04347105	0.05059417	0.03841969	0.05362893	0.04001841	0.04336154	0.04166862
M7	0.02722916	0.02121128	0.02495260	0.01925643	0.01210378	0.00441852	0.01421415	0.00456675
M8	0.03894930	0.02776548	0.03099475	0.02135248	0.01836425	0.00777977	0.02582447	0.00848522
M9	0.01006418	0.00907109	0.02574393	0.01545152	0.00306616	0.00130062	0.00576021	0.00112297
M10	0.05218373	0.04545472	0.04291148	0.04362611	0.04757165	0.03372982	0.03885871	0.03255055
M11	0.02346317	0.02023976	0.02594665	0.01910157	0.02102211	0.01015460	0.02531545	0.01081299
M12	0.03408909	0.03581126	0.04828668	0.03761148	0.00891043	0.00501614	0.01126032	0.00475795
M13	0.04705021	0.04066109	0.05212273	0.03184710	0.03381570	0.02022416	0.03862079	0.02097765
M14	0.06512664	0.04378592	0.06698637	0.04575263	0.04492168	0.02909831	0.05301523	0.03172265
M15	0.00559726	0.00857453	0.03551415	0.02237170	0.00559726	0.00857453	0.03551415	0.02237170
M16	0.07371266	0.07097497	0.08068926	0.06507633	0.06549692	0.04918935	0.06405973	0.05070453
M17	0.10856960	0.08802480	0.09709504	0.07883091	0.07675503	0.06474014	0.07350764	0.06384469
M18	0.05021676	0.04951632	0.05633723	0.04995323	0.03285224	0.02967394	0.03268167	0.03118593
M19	0.03813590	0.03443349	0.03929432	0.03508721	0.02726156	0.02222512	0.03147367	0.02228077
M20	0.08391821	0.08091212	0.09096735	0.08359160	0.07265544	0.06117690	0.07821794	0.06364731

TABLE 4.6: Valeurs empiriques de ISB basées sur 500 replications pour les modèles considérés avec des tailles d'échantillon $n = 100$ et $n = 200$.

	$n = 100$				$n = 200$			
	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}	\hat{f}_ν^{VM}	\hat{f}_ν^{JLN-VM}	\hat{f}_ν^{HG-VM}	\hat{f}_ν^{JSH-VM}
M1	3.040821e-05	6.348471e-05	0.00150434	0.00048217	0.00013067	0.00042674	0.00293548	0.00118427
M2	0.00334685	0.00189913	0.00018506	0.00080073	0.00232021	0.00107231	0.00024007	0.00037914
M3	0.03558660	0.01031846	0.02382104	0.00682979	0.03285080	0.00904011	0.02115482	0.00562304
M4	0.00229039	0.00130052	0.00127339	0.00113470	0.00180822	0.00060162	0.00062766	0.00018327
M5	0.15110240	0.10201770	0.13546970	0.11104740	0.14673090	0.09651235	0.13217790	0.09876050
M6	0.05343394	0.04193338	0.04245233	0.04485323	0.05477977	0.03788451	0.04470470	0.04068343
M7	0.01205598	0.00341828	0.01516105	0.00362782	0.01189051	0.00304801	0.01472983	0.00340654
M8	0.01903571	0.00726165	0.02417563	0.00701984	0.01897185	0.00690209	0.02179473	0.00606151
M9	0.00334337	0.00209707	0.00383056	0.00186967	0.00325495	0.00130648	0.00349774	0.00116138
M10	0.03904712	0.02879505	0.03348378	0.02834418	0.03922783	0.02957774	0.03416939	0.02957662
M11	0.01938998	0.00787834	0.02442759	0.00880031	0.01927030	0.00757352	0.02260764	0.00784119
M12	0.00652242	0.00313679	0.00939304	0.00343940	0.00622938	0.00215727	0.01037140	0.00249353
M13	0.03524105	0.01921846	0.03196439	0.02059341	0.03378782	0.01778066	0.03810888	0.01846570
M14	0.04392108	0.02740814	0.05054139	0.02972560	0.04361718	0.02736435	0.04874774	0.02865401
M15	0.00445593	0.00290914	0.00482886	0.00185586	0.00433600	0.00290644	0.00367593	0.00136851
M16	0.05925803	0.04662084	0.06362063	0.04812077	0.05918882	0.04642142	0.06258305	0.04799654
M17	0.07653698	0.06344866	0.07317474	0.06374976	0.07425254	0.06393376	0.07370063	0.06183362
M18	0.03151345	0.02935020	0.03269273	0.03095942	0.03118848	0.02790362	0.03083753	0.02839715
M19	0.02669730	0.02210877	0.02811250	0.02201016	0.02654897	0.02003360	0.02776737	0.01992920
M20	0.06860961	0.05796681	0.07106940	0.05852636	0.06810613	0.05772155	0.07083166	0.05802073

- Les Tableaux 4.1, 4.2, 4.3 et 4.4 montrent que (performances ISE) :
 - Les moyennes de ISE basées sur 500 replications diminuent à mesure que la taille de l'échantillon n augmente pour tous les estimateurs considérés. ;

- L'estimateur JLN-VM fonctionne beaucoup mieux que les autres estimateurs pour les modèles M6, M7, M11, M13, M14, M16, M17 et M20 pour les tailles $n = 50$, $n = 100$ et $n = 200$;
 - Pour les modèles M3 (quelle que soit la taille), M5 (pour $n = 10$ et $n = 50$), M8 (pour $n = 10$ et $n = 200$), M9 (pour $n = 50$, $n = 100$ et $n = 200$), M10 (pour $n = 50$ et $n = 100$) et M19 (pour $n = 100$ et $n = 200$) l'estimateur JSH-VM est plus performant que les autres estimateurs ;
 - Pour $n = 10$, l'estimateur standard de von Mises est plus performant que les autres estimateurs pour les modèles M1, M4, M10, M11, M12, M16, M18, M19 et M20 ;
 - Pour le reste des modèles (par exemple M2, M15), les performances des quatre estimateurs sont mixtes en fonction de la taille de l'échantillon n .
2. Les Tableaux 4.5 et 4.6 montrent clairement que (performances ISB) :
- les valeurs ISB diminuent à mesure que la taille de l'échantillon n augmente pour les trois estimateurs ;
 - Les estimateurs JSH-VM et JLN-VM surpassent l'estimateur HG-VM et l'estimateur de von Mises standard sauf dans le cas des modèles M1, M2, M4, M10 et M12 pour $n = 10$, et M1, M2, M4 et M15 pour $n = 50$, et M1, M2 pour $n = 100$ et 200 ;
 - Les performances des estimateurs JSH-VM et JLN-VM sont mixtes selon la taille de l'échantillon n et le modèle étudié.

Afin de visualiser la qualité du lissage, nous avons tracé sur le même graphe (représentation circulaire et linéaire) la vraie densité et les estimateurs semi-paramétriques et non-paramétriques à noyau de von Mises (JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises) des modèles 3, 8, 11 et 17 avec $n = 200$. La Figure 4.1 montre que la qualité de lissage est satisfaisante.

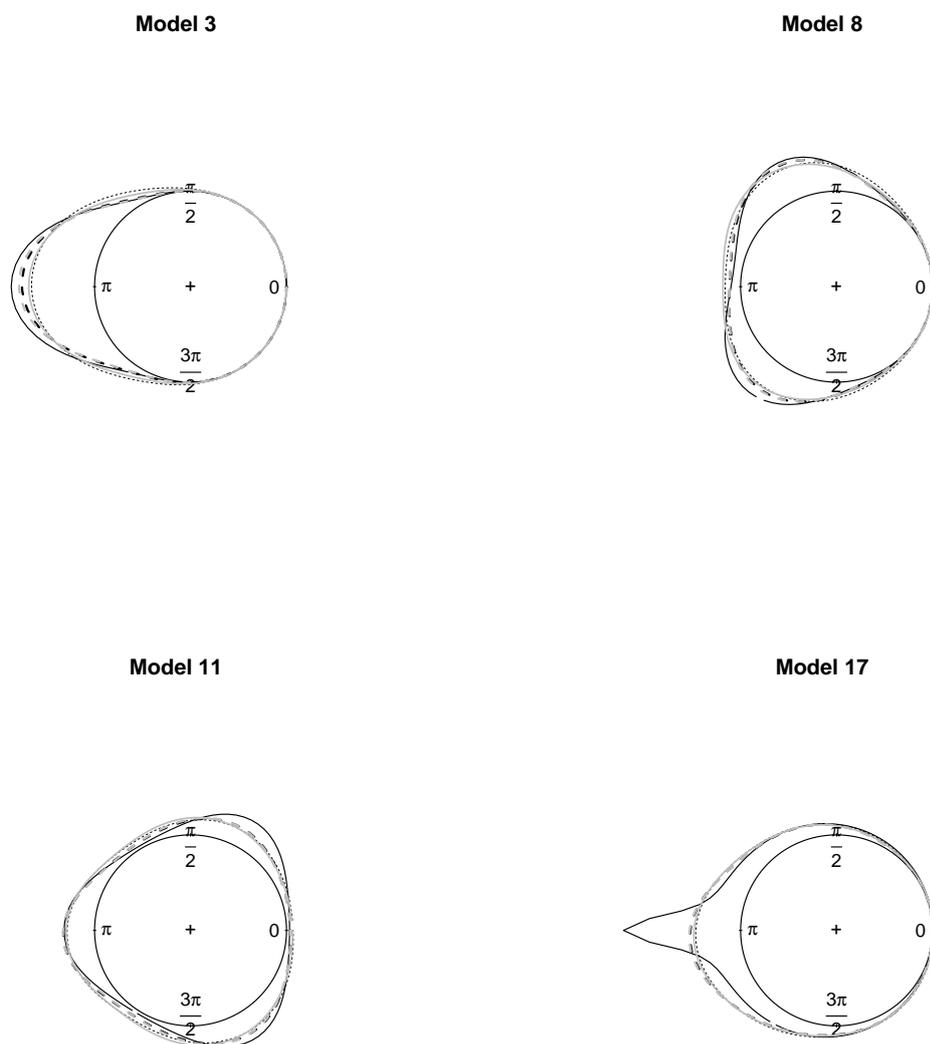


FIGURE 4.1: Les vraies densités et les estimateurs semi-paramétriques et non-paramétriques à noyau de von Mises (*JHS-VM*, *HG-VM*, *JLN-VM* et l'estimateur standard de von Mises) des modèles 3, 8, 11 et 17 avec $n = 200$. Trait noir continu : vraie densité; trait noir discontinu : estimateur standard de von Mises; ligne en pointillés noir : estimateur *JLN-VM*; trait gris continu : estimateur *HG-VM* et trait gris discontinu : estimateur *JSH-VM*.

4.6 Illustration avec des données réelles

Cette section présente l'application des estimateurs JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises pour les deux jeux de données réels suivants :

- **Azimuts flare.** Ces données présentent une série de mesures obtenues à partir d'une expérience pour évaluer la stabilité relative des assemblages de projectiles de fusée éclairante. Une fusée, attachée à un projectile, est lancée vers le haut à partir d'un point de lancement O dans une direction fixe. À un certain point P dans l'espace, la fusée commence à brûler. L'azimut de P par rapport à O donne une indication de la variabilité de l'assemblage car de plus en plus d'essais sont menés. Les données présentées sont basées sur 60 lancements successifs ; voir par exemple [Fisher \[1993\]](#).
- **Orientations à grand axe des lattes de feldspath.** Ces données présentent 133 mesures de lattes de feldspath dans le basalte, rapportées par [Smith \[1988\]](#) et présentées par [Fisher \[1993\]](#).

Nous appliquons maintenant les estimateurs JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises pour estimer les densités des données réelles. La technique UCV est utilisée pour le choix du paramètre de concentration ν . Pour les estimateurs semi-paramétriques, nous avons utilisé le modèle de départ de von Mises de paramètres inconnus μ et k . Ces paramètres sont estimés par la méthode de maximum de vraisemblance. Le tableau 4.7 présente les estimateurs de maximum de vraisemblance des paramètres μ et k pour les données réelles (Azimuts flare et Orientations à grand axe des lattes de feldspath).

TABLE 4.7: *Estimateurs de maximum de vraisemblance des paramètres du modèle de départ de von Mises pour les données réelles.*

	Azimuts flare	Orientations à grand axe des lattes de feldspath
$\hat{\mu}_0^{MV}$	3.050395	3.136083
\hat{k}_0^{MV}	1.343009	1.015225

Les représentations graphiques des densités circulaires et linéaires des estimateurs JHS-VM, HG-VM, JLN-VM et l'estimateur standard de von Mises sont données par la figures 4.2. Nous pouvons voir que chacun des estimateurs est capable de saisir la forme

uni-modale pour les deux jeux de données. Cependant, la meilleure qualité de lissage est obtenue par les estimateurs JHS-VM et JLN-VM.

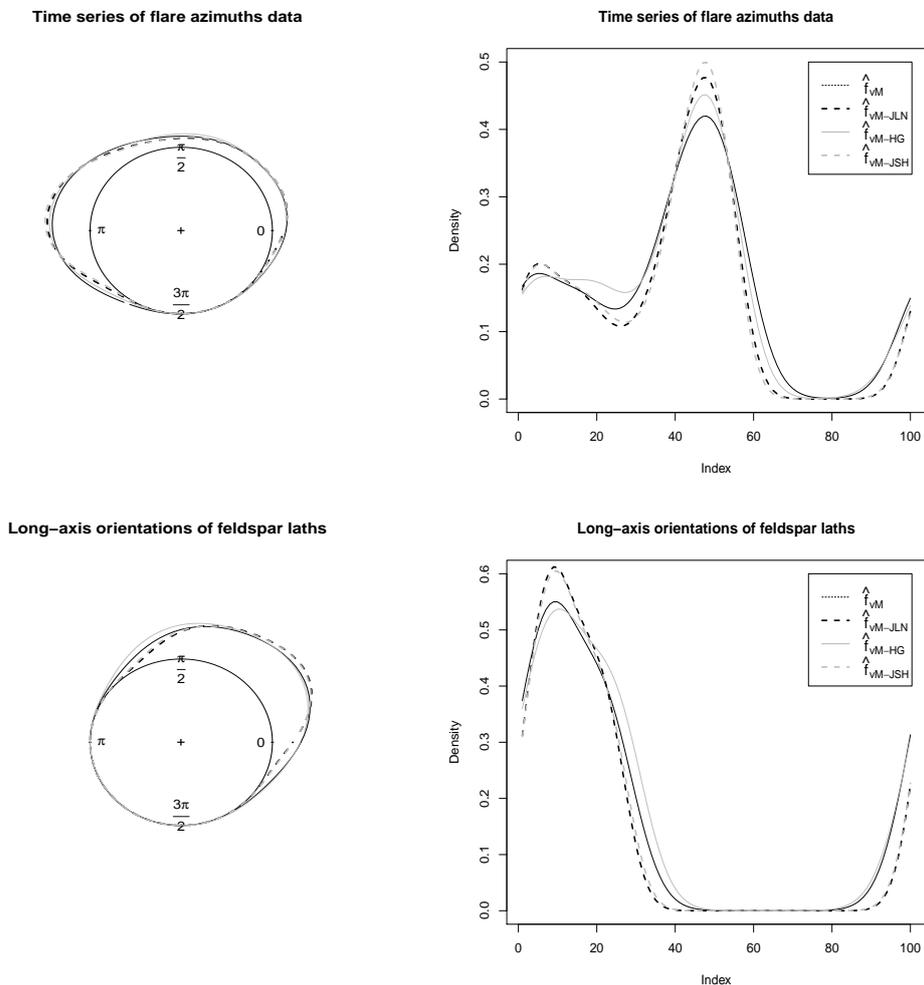


FIGURE 4.2: Les estimateurs semi-paramétriques et non-paramétriques à noyau de von Mises (JHS-VM, HG-VM, JLN-VM et l'estimateur standard) des données réelles; Azimuts flare (première rangée), Orientations à grand axe des lattes de feldspath (deuxième rangée). Trait noir continu : estimateur standard de von Mises; trait noir discontinu : estimateur JLN-VM; trait gris continu : estimateur HG-VM et trait gris discontinu : estimateur JSH-VM

4.7 Conclusion

Dans ce chapitre, nous avons proposé l'approche semi-paramétrique par noyau circulaire dans l'estimation de la fonction densité circulaire. Nous avons introduit les estimateurs semi-paramétriques HG et JSH, en utilisant le noyau circulaire de von Mises. Les propriétés statistiques (biais, variance et MISE) sont également établies. Nous avons montré que l'ordre du biais des estimateurs HG-VM et JSH-VM, sont respectivement $O(\nu^{-1})$ et $O(\nu^{-2})$, tandis que leurs variances sont asymptotiquement équivalentes. Pour évaluer leurs performances, nous avons calculé les critères ISE et ISB en utilisant des simulations pour des modèles circulaires de complexité variable. Nos études de simulation montrent que les estimateurs semi-paramétriques HG-VM et JSH-VM sont très compétitifs avec les estimateurs non-paramétriques (JLN-VM et standard de von Mises). Nous avons également illustré les estimateurs semi-paramétriques avec deux exemples de données circulaires réelles et les résultats obtenus sont satisfaisants.

Conclusion générale et perspectives

Par ce travail, nous estimons avoir contribué à l'étude de l'estimation non-paramétrique et semi-paramétrique de la densité par noyau circulaire de von Mises. Dans un premier temps, nous avons introduit et présenté l'algorithme de Metroplis-Hasting basé sur l'approche bayésienne globale pour le choix du paramètre de concentration.

Un autre aspect important de notre travail portait sur l'utilisation de deux techniques de correction du biais multiplicatif pour le noyau circulaire de von Mises. Nous avons montré que ces deux techniques améliorent (réduisent) l'ordre de grandeur du biais de $O(\nu^{-1})$ à $O(\nu^{-2})$.

Pour la dernière contribution de cette thèse, nous avons proposé l'approche semi-paramétrique en utilisant le noyau circulaire de von Mises pour estimer une densité circulaire inconnue. Dans plusieurs situations, l'estimateur semi-paramétrique est compétitif aussi bien avec l'estimateur paramétrique qu'avec l'estimateur non-paramétrique.

Les résultats de simulation obtenus dans cette thèse montrent que les approches proposées peuvent être considérée comme de bonnes alternatives aux méthodes classiques existantes.

De nombreuses questions demeurent encore posées dans ce domaine.

1. Le chapitre deux ne traite que du cas uni-varié. Comme suite à notre travail, une extension de l'approche bayésienne pour le choix de la matrice de lissage dans l'estimation des densités multidimensionnelle circulaire est envisagée.
2. On envisage aussi de considérer les techniques de corrections du biais dans le cas multi-varié.
3. Il est aussi question d'étudier et approfondir la correction optimale du biais et du

MSE.

4. Mettre en place des modèles de diagnostique permettant aux utilisateurs le choix entre une approche semi-paramétrique, paramétrique, ou non-paramétrique.
5. Il est aussi intéressant de proposer l'approche bayésienne pour le choix du paramètre de concentration dans l'estimation non-paramétrique par noyau circulaire de la fonction de régression.
6. Une autre question qui à notre connaissance demeure encore un problème ouvert est l'étude du cas de l'estimation semi-paramétrique de la fonction de régression dans le cas circulaire.

Enfin, nous souhaitons que notre modeste contribution puisse susciter un intérêt et donner lieu à d'autres études dans ce domaine où, nous le soulignons, de nombreuses questions fondamentales restent encore posées.

Annexe

A.1 Quelques modèles circulaires

Nous décrivons les modèles circulaires utilisés dans nos études de simulation. Nous donnons les expressions des densités de probabilités des modèles uniforme circulaire (CU), von Mises (VM), normal enveloppé (WN), cardioïde (C), Cauchy enveloppé (WC) et skew-normal enveloppé (WSN) ; voir [Mardia and Jupp \[2000\]](#), [Jammalamadaka and SenGupta \[2001\]](#) et [Rao and Sengupta \[2001\]](#)) pour plus de détails :

1. La fonction de densité de la distribution uniforme circulaire est donnée par :

$$f(\theta) = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi.$$

2. La fonction de densité de la distribution de von Mises (appelée aussi distribution normale circulaire) caractérisée par sa moyenne de direction $0 \leq \mu < 2\pi$ et de son paramètre de concentration $\kappa > 0$ notée $vM(\mu, \kappa)$ est exprimée comme suit :

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad 0 \leq \theta < 2\pi.$$

où $I_0(\kappa)$ désigne la fonction modifiée de Bessel de première espèce de l'ordre zéro, donnée par :

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa \cos \theta\} d\theta \tag{4.16}$$

La distribution de von Mises est une distribution uni-modale et symétrique. Elle a été largement étudiée et les techniques d'inférence sont bien développées. C'est

donc le modèle de choix pour les données circulaires dans la plupart des problèmes appliqués.

3. La fonction de densité de la distribution cardioïde de moyenne de direction $0 \leq \mu < 2\pi$ et de paramètre de concentration $|\rho| < 1/2$ notée $C(\mu, \rho)$ est donnée par :

$$f(\theta; \mu, \rho) = \frac{1}{\sqrt{2\pi}}(1 + 2\rho \cos\{\theta - \mu\}), \quad 0 \leq \theta < 2\pi; \quad |\rho| < \frac{1}{2}.$$

La distribution cardioïde est une distribution uni-modale et symétrique autour de μ .

4. La fonction de densité de la distribution normale enveloppée caractérisée par sa moyenne de direction $0 \leq \mu < 2\pi$ et de son paramètre de concentration $\rho \in]0, 1[$ notée $WN(\mu, \rho)$ est donnée par :

$$f(\theta; \mu, \rho) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{+\infty} \exp\left\{-\frac{(\theta - \mu + 2k\pi)^2}{2\sigma^2}\right\}, \quad 0 \leq \theta < 2\pi,$$

où $\sigma^2 = -2 \log(\rho)$. La distribution normale enveloppée est une distribution uni-modale et symétrique autour de μ .

5. La fonction de densité de la distribution de Cauchy enveloppée caractérisée par sa moyenne de direction $0 \leq \mu < 2\pi$ et de son paramètre de concentration $\rho \in]0, 1[$ notée $WC(\mu, \rho)$ s'écrit comme suit :

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta)}, \quad 0 \leq \theta < 2\pi.$$

La distribution de Cauchy enveloppée est une distribution uni-modale et symétrique. Lorsque $\rho \rightarrow 0$, la distribution se rapproche de la distribution uniforme; alors que lorsque $\rho \rightarrow 1$, la la densité se concentre autour de μ .

6. La fonction de densité de la distribution skewed-Normal enveloppée $WSN(\xi, \eta, \lambda)$ est

$$f(\theta; \xi, \eta, \lambda) = \frac{2}{\eta} \varpi\left(\frac{\theta + 2k\pi - \xi}{\eta}\right) \mathcal{W}\left(\frac{\theta + 2k\pi - \xi}{\eta}\right), \quad 0 \leq \theta < 2\pi.$$

où ϖ et \mathcal{W} désignent respectivement les fonctions normales de densité et de distribution normales.

A.2 Code source avec R

Les programmes informatiques ont été implémentés sous R. Nous présentons les programmes mis en place pour le choix du paramètre de concentration par l'approche bayésienne globale dans l'estimation non-paramétrique par noyau de von Mises de la densité, les estimateurs non-paramétriques MBC (TS et JLN). Enfin, nous présentons les estimateurs semi-paramétriques HG et JSH.

A.2.1 Approche bayésienne pour le choix du paramètre de concentration

Il est important de signaler que le programme MCMC pour le choix du paramètre de lissage est précédé par le programme de l'estimateur de la vraisemblance. Nous donnons les programmes de la vraisemblance estimée dans le cas continu et discret, en utilisant le noyau gaussien et binomial respectivement.

La fonction `lcv_vM(nu)` permet de calculer l'estimateur de la vraisemblance en utilisant le noyau circulaire de von Mises.

```
#=====
##### Estimateur de vraisemblance avec noyau de von Mises
#=====
lcv_vM=function(nu){
n=length(data1)
a=0
b=matrix(0,nrow=n, ncol=n)
for(i in 1:n){
b[i,]=dvonmises(data1,data1[i],nu)
b[i,i]=0
a[i]=sum(b[i,])
}
b
a
```

```

out=sum(log(a))-n*log(n-1)
return(out)
}

```

Maintenant, il suffit d'injecter `lcv_vM(nu)` dans le programme `MCMC_est(T,nu0)` pour pouvoir estimer le paramètre de concentration ν par MCMC de façon globale.

```

#=====
##### Estimation par MCMC
#=====

MCMC_est<-function(T,nu0){
chain_nu<-0
taux_accpt<-0
u<-0
t<-1
chain_nu[1]<-nu0
while (t<=T){
print(t)
h_cand<-abs(rnorm(1,chain_nu[t],16))
r<-lcv_vM(h_cand)+dgamma(h_cand,1.5,1/(n^(2/5)),log=TRUE)-lcv_vM(chain_nu[t])
-dgamma(chain_nu[t],1.5,1/(n^(2/5)),log=TRUE)
alpha<-r
u[t]<-log(runif(1))
if (u[t]<alpha){
chain_nu[t+1]<-h_cand
taux_accpt<-taux_accpt+1}
else{
chain_nu[t+1]<-chain_nu[t]}
t<-t+1}
chain_nu
ta=ta/T

```

```

result<-list(h=h, ta=ta)
result<-list(chain_nu=chain_nu, taux_accpt=taux_accpt)
}

```

A.2.1 Estimateurs non-paramétriques MBC (TS et JLN)

```

#=====
##### Estimateur non-paramétrique TS en utilisant le noyau de von Mises
#=====
TSvM=function(data,teta,nu){
c=0.5 # changer ici la valeur de c
est1=dvonmises(circular(data),circular(teta),nu)
est1=mean(est1)
est2=dvonmises(circular(data),circular(teta),c*nu)
est2=mean(est2)

est=(est1^(1/(1-c)))*(est2^(-c/(1-c)))

}

```

```

#=====
##### Estimateur non-paramétrique JLN en utilisant le noyau de von Mises
#=====
JLNvM=function(data,teta,nu){
n=length(data)
V=0
for(i in 1:n){
V[i]=mean(dvonmises(circular(data),circular(data[i]),nu))
}
V
est1=dvonmises(circular(data),circular(teta),nu)

```

```

est1=mean(est1)
est2=dvonmises(circular(data),circular(teta),nu)
est=est1*mean(est2/V)
est
}

```

A.2.1 Estimateurs semi-paramétriques HG et JSH

```

#=====
##### Modèle paramétrique (distribution de von Mises)
#=====
MP_vM=function(TT,kappa2,nu){
  for (i in 1:length(TT)){
    MP_g= dvonmises(TT,circular(nu), kappa2)
  }
  MP_g
}

#=====
##### Estimateur semi-paramétrique HG en utilisant le noyau de von Mises
#=====
densite_estime_semiP_vM_HG=function(TT,t,kappa,kappa2,nu){
  matrice=matrix(0,nrow=length(t),ncol=length(TT))
  division=matrix(0,nrow=length(t),ncol=length(TT))
  noy_vM=0
  for ( i in 1:length(t)){
    MP_vM1= dvonmises(circular(t),circular(nu), kappa2) #f(teta) Modèle paramétrique
    matrice[i,]=dvonmises(TT,circular(t[i]), kappa) #noyau von Mises
    division[i,]= (matrice[i,])/MP_vM(TT,kappa2,nu)
    noy_vM[i]=sum(division[i,])
  }
}

```

```

}
densite_vM_HG=MP_vM1*(1/length(TT))*noy_vM
  densite_vM_HG
}

#=====
##### Estimateur semi-paramétrique JSH en utilisant le noyau de von Mises
#=====

densite_estime_semiP_vM_JSH=function(TT,x,kappa,kappa2,nu){
matrice=matrix(0,nrow=length(x),ncol=length(TT))
division=matrix(0,nrow=length(x),ncol=length(TT))
noy_vM=0
densite_HG=densite_estime_semiP_vM_HG(TT,TT,kappa,kappa2,nu)
for ( i in 1:length(x)){
matrice[i,]=dvonmises(TT,circular(x[i]) , kappa) #noyau von Mises
division[i,]= (matrice[i,])/densite_HG
noy_vM[i]=sum(division[i,])
}
densite_vM_JSH=densite_estime_semiP_vM_HG(TT,x,kappa,kappa2,nu)*
(1/length(TT))*noy_vM
densite_vM_JSH
}

```

Bibliographie

- S. Abou-Jouadé. Sur une condition nécessaire et suffisante de l_1 convergence presque complète de l'estimateur de la partition fixe pour une densité. *C. R. Acad. Sci. Paris Sér A-B*, 283(16) :A1107–A1110, 1976.
- J. Aitchison and C G G. Aitken. Multivariate binary discrimination by the kernel method. *Biometrika*, 63 :413–420, 1976.
- A. Amiri, B. Thiam, and T. Verdebouti. On the estimation of the density of a directional data stream. *Scandinavian Journal of Statistics*, 44 (1) :249–67, 2017.
- E. Batschelet. *Circular Statistics in Biology*. Academic Press, New York. 1981.
- K. Bedouhene and N. Zougab. Nonparametric multiplicative bias correction for von mises kernel circular density estimator. *Communications in Statistics-Simulation and Computation*, DOI : 10.1080/03610918.2019.1659358 :1–19, 2019.
- K. Bedouhene and N. Zougab. A bayesian procedure for bandwidth selection in circular kernel density estimation. *Monte Carlo Methods and Applications*, 26 :69–82, 2020.

- N. Belaid. *Nonparametric estimation by Discrete Associated-Kernel method for multivariate probability mass function*. Thèse de Doctorat, Université de Bejaia, Algérie, 2016.
- N. Belaid, S. Adjabi, N. Zougab, and C C. Kokonendji. Bayesian bandwidth selection in discrete multivariate associated kernel estimators for probability mass functions. *Journal of the Korean Statistical Society*, 2016.
- D. Bosq. Estimation suroptimale de la densité par projection. *Can. J. Statist*, 33(1) : 21–37, 2005.
- A W. Bowman. An Alternative Method of Cross-validation for the Smoothing of Density Estimates. *Biometrika*, 71 :353–360, 1984.
- M J. Brewer. A modelling approach for bandwidth selection in kernel density estimation. *In : Proceedings of COMPSTAT Physica Verlag, Heidelberg*, pp :203–208, 1998.
- M J. Brewer. A bayesian model for local smoothing in kernel density estimation. *Statistics and Computing*, 10 :299–309, 2000.
- N. Cencov. Evaluation of an unknown distribution density from observations. *Sovet, Math*, 3 :1559–1562, 1962.
- S X. Chen. Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics*, 52 :471–480, 2000.
- M. Di Marzio, A. Panzera, and C C. Taylor. Local polynomial regression for circular predictors. *Statistics and Probability Letters*, 79 :2066–2075, 2009.
- M. Di Marzio, A. Panzera, and C C. Taylor. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141 :2156–2173, 2011a.
- M. Di Marzio, A. Panzera, and C C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40(2) :238–255, 2011b.
- E. Erçelik and M. Nadar. Nonparametric density estimation based on beta prime kernel. *Communications in Statistics-Theory and Methods*, 49 :325–342, 2018.

- R L. Eubank. *Nonparametric Regression and Spline Smoothing*. New York, 2nd ed, Marcel Dekker, Inc, 1999.
- N I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, U.K. 1993.
- B. Funke and R. Kawka. Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. *Computational Statistics and Data Analysis*, 92 :148–62, 2015.
- E. García-Portugués. Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7 :1655–85, 2013.
- E. García-Portugués, R M. Crujeiras, and González-Manteiga. Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, 121 :152–75, 2013.
- P H. Garthwaite, Y. Fan, and S A. Sisson. Adaptive optimal scaling of metropolis-hastings algorithms using the robbins-monro process. *Communications in Statistics-Theory and Methods*, 45 :5098–5111, 2016.
- J. Geffroy. Sur l'estimation d'une densité dans un espace métrique. *C. R. Acad. Sci. Paris Sér A-B*, 278 :1449–1452, 1976.
- J. Geoffroy. Etude de la convergence du régressogramme. *Publ. Inst. Statist. Univ. Paris*, 25 (1-2) :41–56, 1980.
- E. Gould. Orientation in box turtles, terrapenec. carolina (linneaus). *The Biological Bulletin*, 112(3) :336–48, 1957.
- J D F. Habbema, J. Hermans, and K. Vandenbroek. A stepwise discriminant analysis program using density estimation. *in Compstat, ed. G. Bruckmann, Physica-Verlag, Wien*, 1974.
- M. Hagmann and O. Scaillet. Local multiplicative bias correction for asymmetric kernel density estimators. *Journal of Econometrics*, 141 :213–249, 2007.

- P. Hall, G P. Watson, and J. Cabrera. Kernel density estimation for spherical data. *Biometrika*, 74 :751–762, 1987.
- P. Hall, J S. Marron, and B U. Park. Smoothed cross validation. *Probability Theory and Related Fields*, 92 :1–20, 1992.
- L. Harfouche, S. Adjabi, N. Zougab, and B. Funke. Multiplicative bias correction for discrete kernels. *Statistical Methods and Application*, 27 :253–76, 2018.
- W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57 :97–109, 1970.
- M. Hirukawa. Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval. *Computational Statistics and Data Analysis*, 54(2) :473–95, 2010.
- M. Hirukawa and M. Sakudo. Nonnegative bias reduction methods for density estimation using asymmetric kernels. *Computational Statistics and Data Analysis*, 75 :112–23, 2014.
- M. Hirukawa and M. Sakudo. Family of the generalised gamma kernels : A generator of asymmetric kernels for nonnegative data. *Journal of Nonparametric Statistics*, 27(1) : 41–63, 2015.
- M. Hirukawa and M. Sakudo. Another bias correction for asymmetric kernel density estimation with a parametric start. *Statistics and Probability Letters*, 154 :158–165, 2019.
- N L. Hjort and I K. Glad. Nonparametric density estimation with a parametric start. *Ann. Statist*, 23 :882–904, 1995.
- W. Härdle and J S. Marron. Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist*, 13 :1465–1481, 1985.
- G. Igarashi. Weighted log-normal kernel density estimation. *Communications in Statistics-Theory and Methods*, 45 :6670–6687, 2016.

- S R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific, Singapore. 2001.
- X. Jin and J. Kawczack. Birnbaum-saunders and lognormal kernel estimators for modeling durations in high frequency financial data. *Annals of Economics and Finance*, 4 : 103–124, 2003.
- X. Jin and J. Kawczak. Birnbaum-saunders and lognormal kernel estimators for modelling durations in high frequency financial data. *Ann. Econ. Fin*, 4 :103–124, 2003.
- M C. Jones and P J. Foster. Generalized jackknifing and higher-order kernels. *Journal of Nonparametric Statistics*, 3(1) :81–94, 1993.
- M C. Jones, O. Linton, and J P. Nielsen. A simple bias reduction method for density estimation. *Biometrika*, 82(2) :327–38, 1995.
- M. C. Jones, D. F. Signorini, and N.L. Hjort. On multiplicative bias correction in kernel density estimation. *Sankhyā : the indian journal of statistics*, 61 :422–430, 1999.
- Y. Kakizawa and G. Igarashi. Inverse gamma kernel density estimation for nonnegative data. *Journal of the Korean Statistical Society*, 46 :194–207, 2017.
- J. Klemelä. Estimation of densities and derivatives of densities with directional data. *Journal of Multivariate Analysis*, 73 :18–40, 2000.
- C C. Kokonendji and T. Senga Kiessé. Discrete associated kernels method and extensions. *Statistical Methodology*, 8 :497–516, 2011.
- C C. Kokonendji, T. Senga Kiessé, and S S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. *Journal of Nonparametric Statistics*, 19 :241–257, 2007.
- C C. Kokonendji, T. Senga Kiessé, and N. Balakrishnan. Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference*, 139 :3625–3638, 2009a.

- C C. Kokonendji, T. Senga Kiessé, and C G B. Demétrio. Appropriate kernel regression on a count explanatory variable and applications. *Advances and Applications in Statistics*, 12 :99–126, 2009b.
- R A. Kronmal and M E. Tarter. The estimation of probability densities and cumulatives by fourie series method. *J. Amerc. Stat. Assoc*, 63,p :925–952, 1968.
- J. Lecoutre. *Contribution à l'estimation non paramétrique de la régression*. Thèse de Doctorat de l'Université de Pierre et Marie Curie, 1982.
- S. Li, M J. Silvapulle, P. Silvapulle, and X. Zhang. Bayesian approaches to nonparametric estimation of densities on the unit interval. *Econometric Reviews*, 2015.
- C. Marchant, K. Bertin, V. Leiva, and H. Saulo. Generalized birnbaumī-saunders kernel density estimators and an analysis of financial data. *Computational Statistics and Data Analysis*, 63 :1–15, 2013.
- K V. Mardia and P E. Jupp. *Directional statistics*. New York : Wiley. 2000.
- N. Metropolis, A W. Rosenbluth, M N. Rosenbluth, A H. Teller, and E. Teller. Equations of state calculations by fast computing machine. *J. Chem. Phys*, 21 :1087–1093, 1953.
- A M. Mousa, MKh. Hassan, and A. Fathi. new non parametric estimator for pdf based on inverse gamma distribution. *Communications in Statistics-Theory and Methods*, 45 : 7002–7010, 2016.
- E A. Nadaraya. On estimating regression. *Theory. Probab. Appl*, 9 :141–142, 1964.
- E A. Nadaraya. On nonparametric estimation of density function and regression. *Theory. Probab. Appl*, 10 :186–190, 1965.
- M. Oliveira, R M. Crujeiras, and A. Rodríguez-Casal. A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data Analysis*, 56 :3898–3908, 2012.

- M. Oliveira, R M. Crujeiras, and A. Rodríguez-Casal. Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, 20 :1–17, 2013.
- M. Oliveira, R M. Crujeiras, and A. Rodríguez-Casal. Npcirc : An r package for nonparametric circular methods. *Journal of Statistical Software*, 61 :1–26, 2014a.
- M. Oliveira, R M. Crujeiras, and A. Rodríguez-Casal. Npcirc : Nonparametric circular methods. *R package version 2.0.1*, URL <http://www.CRAN.R-project.org/package=NPCirc>, 2014b.
- B U. Park and J S. Marron. Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, 85 :66–72, 1990.
- E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33 :1065–1076, 1962.
- R Development Core Team. A language and environment for statistical computin. *R Foundation for Statistical Computing*. Vienna : Austria. URL <http://www.R-project.org>, 2014.
- J S. Rao and A. Sengupta. *Topics in circular statistics*. Singapore : World Scientific Publishing. 2001.
- C. Reinsch. Smoothing by spline functions. *Numererisch Mathematik*, 10 :177–183, 1967.
- G O. Roberts and J S. Rosenthal. Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, 18 :349–367, 2009.
- M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 :832–837, 1956.
- M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, 9 :65–78, 1982.
- N. Saadi and S. Adjabi. On the estimation of the probability density by trigonometric series. *Communicatins in Statistics-Simulation and Computation*, 38 :3583–3595, 2009.

- H. Saulo, V. Leiva, F.A. Ziegelmann, and C. Marchant. A nonparametric method for estimating asymmetric densities based on skewed birnbaum–saunders distributions applied to environmental data. *Stoch Environ Res Risk Assess*, 7 :1479—1491, 2013.
- O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 16 :217–226, 2004.
- S C. Schwartz. Estimation of probability densities by an orthogonal series. *Ann. Math. Statist*, 38 :1261–1265, 1967.
- D W. Scott and G R. Terrel. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82 :1131–1146, 1987.
- T. Senga Kiessé. *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat, Université de Pau, France, 2008.
- T. Senga Kiessé, N. Zougab, and C C. Kokonendji. Bayesian estimation of bandwidth in semiparametric kernel estimation of unknown probability mass and regression functions of count data. *Computational Statistics*, 31 :189–206, 2016.
- J. Shi and W. Son. Asymptotic results in gamma kernel regression. *Communications in Statistics-Theory and Methods*, 45 :3489–3509, 2016.
- B W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. 13 :1–52, 1985.
- B W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.
- N M. Smith. Reconstruction of the tertiary drainage systems in the inverell region. *Unpublished B. Sc.(Hons.) thesis, Department of Geography, University of Sidney, Australia*, 1988.
- C C. Taylor. Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*, 52 :3493–3500, 2008.

- C C. Taylor, K V. Marida M. Di Marzio, and A. Panzera. Validating proteinstructure using kernel density estimates. *Journal of Applied Statistics*, 39 :2379–2388, 2012.
- G R. Terrell and D W. Scott. On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, 8(5) :1160–3, 1980.
- G R. Terrell and D W. Scott. Biased and unbiased cross-validation in density estimation. *Biometrika*, 69 :383–390, 1982.
- Y. Tsuruta and M. Sagae. Asymptotic property of wrapped cauchy kernel density estimation on the circle. *Bull. Inform. Cybernet. (accepted)*, 2017a.
- Y. Tsuruta and M. Sagae. Higher order kernel density estimation on the circle. *Statistics and Probability Letters*, 131 :46–50, 2017b.
- J. Tukey. Curves as parameters and touch estimation. *Proc ; of the 4th Berkeley Sump. on Math. Stat. Prob., Pages*, 1961.
- G. Wahba. Data based optimal smoothing of orthogonal serie density estimates. *Ann. Statist*, 9(1) :146–156, 1981.
- G. Wahba. *Spline models for observational data*. S.I.A.M., Philadelphia, 1990.
- G S. Watson. Smooth regression analysis. *Sankhyã Ser, A26* :359–372, 1964.
- X. Zhang, M L. King, and R J. Hyndman. A bayesian approach to bandwidth selection for multivariate kernel density estimation. *Computational Statistics and Data Analysis*, 50 :3009–3031, 2006.
- X. Zhang, R D. Brooks, and M L. King. A Bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation. *Journal of Econometrics*, 153(1) :21–32, 2009.
- N. Zougab and S. Adjabi. Multiplicative bias correction for generalized birnbaum–saunders kernel density estimators and application to nonnegative heavy tailed data. *Journal of the Korean Statistical Society*, 45(1) :51–63, 2016.

-
- N. Zougab, S. Adjabi, and C C. Kokonendji. A Bayesian approach to bandwidth selection in univariate associate kernel estimation. *Journal of Statistical Theory and Practice*, 2013.
- N. Zougab, S. Adjabi, and C C. Kokonendji. Comparison study to bandwidth selection in binomial kernel estimation using bayesian approaches. *Journal of Statistical Theory and Practice*, 10 :133–153, 2016.

Résumé : Cette thèse est dédiée à l'étude de l'estimation non-paramétrique et semi-paramétrique, par noyau circulaire de type von Mises pour la fonction de densité de probabilité circulaire (pdf). Dans un premier temps, nous avons introduit et présenté l'algorithme de Metropolis-Hasting basé sur l'approche bayésienne globale pour le choix du paramètre de concentration dans l'estimateur non-paramétrique à noyau circulaire de von Mises. Dans un second temps, nous avons proposé deux techniques (TS et JLN) de réduction du biais appelées techniques MBC pour l'estimateur à noyau circulaire de von Mises. Les propriétés statistiques (biais, variance, et l'erreur quadratique moyenne intégrée) des deux estimateurs proposés sont présentées. Le paramètre de concentration est estimé par la technique de validation croisée par le maximum de vraisemblance. Enfin, nous avons développé l'approche semi-paramétrique en utilisant le noyau circulaire de von Mises pour estimer une densité circulaire inconnue. On a présenté également les propriétés statistiques des estimateurs semi-paramétriques HG et JSH de von Mises. Le paramètre de concentration est estimé par la méthode de validation croisée non biaisée. Tout au long de ce travail, les performances des estimateurs de von-Mises proposés sont illustrées par des études de simulation et des applications réelles pour des données circulaires.

Mots-clés : Approche bayésienne ; Densité circulaire ; Monté Carlo par chaînes de Markov (MCMC) ; Noyau de von Mises ; Paramètre de concentration ; Techniques MBC ; Validation croisée.

Abstract : This thesis is dedicated to the study of non-parametric and semi-parametric estimation, by circular von Mises-type kernel for the circular probability density function (pdf). First, the Bayesian procedure for bandwidth selection in kernel circular density estimation is investigated, when the Markov chain Monte Carlo (MCMC) sampling algorithm is utilized for Bayes estimate. Second, we apply the multiplicative bias correction (MBC) techniques for von Mises kernel density estimator in the context of circular data. Some properties of the MBC-von Mises kernel circular density estimators (bias, variance and mean integrated squared error) are shown. The choice of bandwidth is investigated by adapting the popular cross-validation technique. Third, the semi-parametric approach based on von Mises kernel is developed. Some properties of the semi-parametric HG and JSH von Mises circular kernel density estimators are provided. The popular cross validation technique is adapted for bandwidth selection. The performances of the proposed von Mises estimators are illustrated by a simulation studies and real applications for circular data.

Key words : Bandwidth ; Bayesian approach ; Circular density ; Cross validation ; MBC techniques ; Markov chain Monte Carlo (MCMC) ; von Mises kernel.