

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri de Tizi-Ouzou

Faculté des Sciences

Département de Mathématiques

# Mémoire de Master

Présenté par Aoudia Souhila

En vue de l'obtention du grade de Master 2

Option : Processus aléatoire et Statistique de la décision

Thème

---

APPROCHE NON PARAMÉTRIQUE PAR NOYAU DES DONNÉES  
CIRCULAIRES

---

Soutenu publiquement le 29/09/2015 devant le jury suivant :

<b>Président</b>	MAMOU MOHAMED	M.A.A	U. de Tizi-Ouzou
<b>Encadreur</b>	ZOUGAB NABIL	M.C.B	U. de Tizi-Ouzou
<b>Examineur</b>	BOUDIBA M.AREZKI	M.C.A	U. de Tizi-Ouzou
<b>Examineur</b>	GRAICHE FARID	M.C.B	U. de Tizi-Ouzou

Promotion 2014 - 2015

---

# Remerciements

---

Je remercie le bon Dieu tout puissant d'avoir guidé mes pas vers les portes du savoir tout en illuminant mon chemin, et de m'avoir donné suffisamment de courage et de persévérance pour pouvoir mener ce travail à terme.

Je tiens à exprimer ma profonde gratitude et mes plus sincères remerciements à Monsieur Zougab Nabil pour avoir accepté de m'encadrer, mais aussi pour sa disponibilité, sa gentillesse, et ses précieux conseils dès le début de ce travail.

Toute ma reconnaissance aux membres de jury pour avoir accepté de juger ce travail, je suis honoré de la présence de Monsieur Mamou Mohammed comme président de ce jury, Monsieur Boudiba M.Arezki et Monsieur Graïche Farid comme examinateur de ce travail. Qu'ils trouvent ici l'expression de mon profond respect.

Mes remerciements vont également à tous le personnel du département mathématique, particulièrement tous les enseignants qui nous ont donné des connaissances tout au long de notre cursus, et tous ceux et toutes celles qui ont contribué de près ou de loin à la réalisation de ce modeste travail surtout Soumia.

En fin un grand merci à ma chère (petite et grande) famille, particulièrement mes parents pour leur protection filiale et leur encouragement, et à tous mes amis(es) et camarades.

---

# Dédicaces

---

Je dédie ce modeste travail à tous ce qui sont chers :

A mes très chers parents Ali et Zohra à qui je dois tous ;

A mon mari Youcef qui ma toujours soutenue et aidé ;

A mes frères Tarik , Lakhdar , Yacine et Mourad ;

A ma soeur Kahina et son époux Mouloud sans oublier ses enfants (Assia, Sara, Zakaria) ;

A ma soeur Saliha et tout membre de ma famille ;

A ma soeur Fatima et son époux Idir sans oublier leur fille Iliane ;

A tous mes amis(es)de la fac en particulier (Nadia, Soumia, Malika, Samira, Naima, Nacera,

Cherifa, Haroun... etc) ;

A toute la promotion maths 2014 - 2015.

Souhila.A

---

# Table des matières

---

<b>Introduction générale</b>	<b>8</b>
<b>1 Estimateur à noyau des données circulaires de la fonction densité</b>	<b>11</b>
1.1 Estimateur à noyau circulaire de la fonction densité . . . . .	11
1.1.1 Distribution de Von Mises . . . . .	12
1.1.2 Estimateur à noyau de Von Mises . . . . .	13
1.2 Propriétés de l'estimateur . . . . .	14
1.2.1 Espérance . . . . .	15
1.2.2 Variance . . . . .	15
1.2.3 Biais . . . . .	16
1.2.4 Convergence au sens de l'erreur quadratique moyenne (MSE) . . . .	16
1.2.5 Convergence au sens de l'erreur quadratique moyenne intégrée(MISE)	16
1.3 Conclusion . . . . .	17
<b>2 Choix du paramètre de concentration <math>\nu</math></b>	<b>19</b>
2.1 Méthodes de ré-injection (plug-in) . . . . .	19
2.1.1 Règle de référence (Rule of Thumb) . . . . .	20
2.1.2 Méthode plug-in de Oliveira et al. (2012) . . . . .	20
2.2 Méthodes de validation croisée . . . . .	21
2.2.1 Validation croisée par le maximum de vraisemblance . . . . .	21
2.2.2 Validation croisée non biaisée . . . . .	22
2.3 Conclusion . . . . .	23
<b>3 Etude de simulation</b>	<b>24</b>

	4
<b>4 Application sur des données réelles</b>	<b>34</b>
Conclusion générale	38
Annexe	40
Bibliographie	44

---

# Table des figures

---

1.1	Représentation circulaire de la distribution Von Mises . . . . .	13
1.2	Représentation linéaire de la distribution Von Mises . . . . .	14
3.1	Densités circulaires (vraie) de <b>D1</b> et <b>D2</b> . . . . .	26
3.2	Densités circulaires (vraie) de <b>D3</b> et <b>D4</b> . . . . .	27
3.3	Densités circulaires (vraie) de <b>D5</b> et <b>D6</b> . . . . .	29
3.4	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D1</b> avec $n = 250$ . . . . .	30
3.5	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D2</b> avec $n = 250$ . . . . .	31
3.6	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D3</b> avec $n = 250$ . . . . .	31
3.7	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D4</b> avec $n = 250$ . . . . .	32
3.8	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D5</b> avec $n = 250$ . . . . .	32
3.9	Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour <b>D6</b> avec $n = 250$ . . . . .	33
4.1	Représentation linéaire de l'estimateur à noyau de Von Mises pour les données de l'orientation de libellules. Différentes méthodes pour le choix de paramètre de concentration sont utilisées et présentées. . . . .	36

4.2	Représentation circulaire de l'estimateur à noyau de Von Mises pour les données de l'orientation de libellules avec différentes méthodes pour le choix de paramètre $\nu$ . . . . .	37
-----	---	----

---

# Liste des tableaux

---

3.1	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D1</b> . . . . .	27
3.2	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D2</b> . . . . .	28
3.3	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D3</b> . . . . .	28
3.4	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D4</b> . . . . .	28
3.5	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D5</b> . . . . .	29
3.6	Résultats de simulation de $\widehat{ISE}$ pour le modèle <b>D6</b> . . . . .	30

---

# Introduction générale

---

Les données circulaires peuvent apparaître dans divers domaines, comme par exemple, l'écologie, la météorologie, les sciences de l'environnement (voir par exemple Mardia et Jupp (2000)). Comme dans le cas des données linéaires, l'analyse des données circulaires peut se faire en utilisant de l'estimation fonctionnelle telles que l'estimation de la fonction de densité ou la fonction de régression.

Afin d'estimer, par exemple, la densité inconnue  $f$ , une première approche dite paramétrique consiste à supposer que  $f$  appartient à une famille de densités continues linéaires ou circulaires qui peuvent être décrites par un certain nombre de paramètres réels. Pour le cas des données circulaires, la loi de Von Mises (Normale circulaire) de moyenne de direction  $0 \leq \mu < 2\pi$  et de paramètre de concentration  $\kappa > 0$  est considérée généralement comme la distribution standard pour l'analyse de ce type de données. Ce modèle paramétrique de Von Mises peut être modifié, en utilisant par exemple la mixturisation, lorsque les données présentent des phénomènes spécifiques (voir par exemple Mooney et al. (2003)). Le statisticien qui opte pour une telle approche possède une bonne connaissance a priori du phénomène aléatoire. Cependant, lorsque aucune information n'est disponible sur le phénomène étudié, l'application du modèle paramétrique de Von Mises n'est pas satisfaisant. Pour pallier les insuffisances et les défauts de cette approche paramétrique, une seconde approche dite méthode non-paramétrique par *noyau* consiste, à estimer, à partir des observations, une fonction inconnue sans spécifier de forme sur cette fonction à estimer. Cette seconde approche est proposée initialement pour le cas des données linéaires (voir Rosenblatt (1956), Parzen (1962), Silverman (1985), Chen (1999; 2000), Jin et Kawczack (2003) Scaillet (2004), Senga Kiessé (2009), Kokonendji et Senga Kiessé (2011) et Kokonendji et Libengué (2011)) et étendue pour le cas des données circulaires

par plusieurs auteurs (Hall et al. (1987), Klemelä (2000), Taylor (2008), Di Marzio et al. (2011) et Oliveira et al. (2012)). Comme dans le cas linéaire, l'approche non-paramétrique par noyau dans le cas des données circulaires rencontre beaucoup de succès auprès de la communauté des utilisateurs. Ceci peut s'expliquer essentiellement par : la simplicité de l'expression théorique de l'estimateur puisque il s'écrit comme la somme de  $n$  variables aléatoires indépendantes et identiquement distribuées, en utilisant la fonction noyau  $K$  et le paramètre de concentration  $\kappa$ , sa convergence en de nombreux sens, sa flexibilité dans la mesure où il laisse à l'utilisateur une grande latitude dans le choix du noyau  $K$  et du paramètre de concentration  $\kappa$ .

Le choix de noyau n'est pas très important. Il suffit de choisir un noyau adapté aux type de données analysées. Le noyau le plus utilisé dans le cas des données circulaires est le noyau de Von Mises. Cependant, il existe d'autres noyaux comme par exemple le noyau Cauchy enveloppé (wrapped Cauchy). Dans ce travail, nous utilisons le noyau de Von Mises à cause de sa simplicité.

En revanche, les performances de l'estimateur à noyau circulaire de la densité dépend crucialement du paramètre de concentration qui contrôle la qualité de lissage de l'estimateur. Lorsque le paramètre de concentration est grand l'estimateur est sous-lissé. Par contre, lorsque le paramètre de concentration est très petit, l'estimateur est sur-lissé. Deux catégories de méthodes ont été proposées dans la littérature pour choisir le paramètre de concentration. La première catégorie repose sur la minimisation de l'erreur quadratique moyenne intégrée (MISE) dites famille de ré-injection (plug-in en anglais). La deuxième catégorie est de type validation croisée, elle est intéressante en pratique car elle se laisse guider seulement par les observations.

Le premier objectif de ce travail est de faire le point sur les procédures de sélection du paramètre de concentration dans l'estimation de la fonction densité circulaire. Notons que ces procédures diffèrent au niveau du choix du critère à optimiser. Plusieurs méthodes ont été proposées dans la littérature :

- La classe des méthodes dites plug-in (ré-injection) : la méthode de règle de référence (Rule of Thumb) de Taylor (2008) et la méthode plug-in de Oliveira et al. (2012) qui peut être considérée comme la généralisation de la méthode de règle de référence.

- Les méthodes reposant sur la validation croisée. D’un point de vue pratique, un des principaux intérêts de ces méthodes est leur caractère direct. Contrairement aux méthodes plug-in, le paramètre de concentration estimé dépend des observations. Ces méthodes ont été proposées aussi dans Oliviera et al. (2012 ; 2014) : la technique de validation croisée par maximum de vraisemblance et la méthode de validation croisée non biaisée.

Le second objectif de ce travail est d’appliquer et de comparer les différentes méthodes de sélection du paramètre de concentration sur plusieurs distributions circulaires connues.

Nous avons structuré ce mémoire en trois parties principales. Le premier chapitre est consacré à la description de la méthode à noyau circulaire dans l’estimation de la fonction densité, en particulier l’estimateur à noyau de Von Mises.

Dans le second chapitre, on présente les différentes procédures pour le choix du paramètre de concentration en décrivant leurs principes.

Enfin, dans le chapitre trois, nous présentons les résultats des simulations conduites à partir de plusieurs densités circulaires connues : loi de Von Mises, mélange de Von Mises, afin de :

- Comparer les différents algorithmes de sélection du paramètre de concentration ;
- Etudier l’influence de la taille de l’échantillon sur ces différents algorithmes ;

Ce mémoire se termine par une conclusion générale et quelques perspectives.

# Chapitre 1

---

## Estimateur à noyau des données circulaires de la fonction densité

---

Dans ce chapitre, nous donnons d'abord la définition de l'estimateur à noyau des données circulaires dans le cas de l'estimation de la fonction densité, en particulier l'estimateur à noyau de Von Mises. En suite, nous présentons les propriétés asymptotiques (biais, variance) et les critères locale et globale (l'erreur quadratique moyenne et l'erreur quadratique moyenne intégrée) de cet estimateur.

### 1.1 Estimateur à noyau circulaire de la fonction densité

Nous donnons d'abord la définition d'un noyau circulaire.

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

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

$$\text{donnée par : } 1/(2\pi) \left\{ 1 + 2 \sum_{j=1}^{\infty} y_j(\nu) \cos(j\theta) \right\};$$

2. *Soit  $\eta_j(K_\nu) = \int_0^{2\pi} \sin^j(\theta) K_\nu(\theta) d\theta$ , alors  $\eta_0(K_\nu) = 1$ ,  $\eta_j(K_\nu) = 0$  pour  $0 < j < r$  et  $\eta_r(K_\nu) \neq 0$ .*

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

Soient  $\Theta_1, \Theta_2, \dots, \Theta_n$  un  $n$ -échantillon aléatoire d'angles indépendant et identiquement distribué issu de densité inconnue  $f$  sur  $T = [0, 2\pi[$  d'une variable aléatoire  $\Theta$ .

L'estimateur à noyau  $\hat{f}_\nu$  de  $f$  utilisant  $K_{\nu, \theta}$  est définie par :

$$\begin{aligned}\hat{f}_\nu(\theta) &= \frac{1}{n} \sum_{i=1}^n K_{\nu, \theta}(\Theta_i) \\ &= \frac{1}{n} \sum_{i=1}^n K_\nu(\theta - \Theta_i),\end{aligned}\tag{1.1}$$

où  $K_{\nu, \theta}$  est la fonction noyau circulaire (généralement choisie comme une densité de Von Mises),  $\nu > 0$  est le paramètre de concentration qui joue un rôle très important pour la qualité de lissage et  $\theta$  est la cible (le point où on veut estimer  $f$ ). Dans la suite nous présentons la distribution de Von Mises et l'estimateur à noyau de Von Mises car c'est le noyau le plus utilisé dans la littérature, voir Taylor (2008) et Oliveira et al. (2012).

### 1.1.1 Distribution de Von Mises

On dit qu'une variable aléatoire  $\Theta$  est de distribution Von Mises de paramètres  $\kappa > 0$  et  $\mu$  si sa fonction de densité de probabilité s'écrit comme suit (voir par exemple Jammalamadaka et SenGupta (2001)) :

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

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

$$\begin{aligned}I_0(\kappa) &= \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{\kappa \cos \theta\right\} d\theta \\ &= \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2.\end{aligned}$$

La distribution Von Mises notée  $\nu M(\mu, \kappa)$  est une distribution unimodal, symétrique et elle est caractérisée par sa moyenne de direction  $0 \leq \mu < 2\pi$  et son paramètre de concentration  $\kappa > 0$ . Cette distribution a été présentée comme un modèle statistique par Von

Mises (1918) discutée plus tôt par Langevin (1905). Quelques graphiques (représentation circulaire et linéaire) sont donnés dans les figures 1.1 et 1.2.

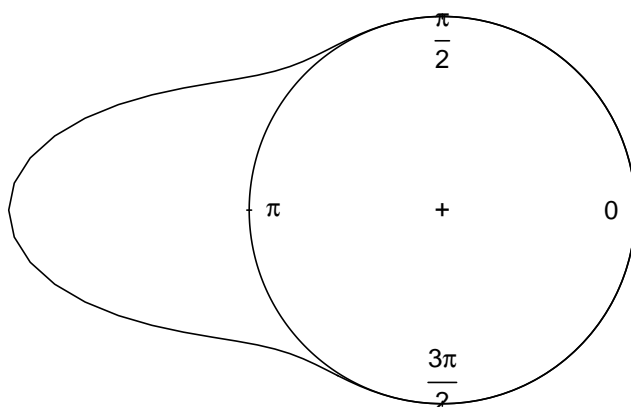


FIGURE 1.1: *Représentation circulaire de la distribution Von Mises*

### 1.1.2 Estimateur à noyau de Von Mises

En utilisant la distribution Von Mises de moyenne 0 et de paramètre de concentration  $\nu$  comme noyau dans la formule (1.1), l'estimateur à noyau de Von Mises est donné simplement par :

$$\hat{f}_\nu^{VM}(\theta) = \frac{1}{n(2\pi)I_0(\nu)} \sum_{i=1}^n \exp \left\{ \nu \cos(\theta - \Theta_i) \right\}, \quad 0 \leq \theta < 2\pi. \quad (1.2)$$

Cet estimateur peut être interprété comme un mélange de  $n$  variables aléatoires avec une distribution de Von Mises, centré en  $\Theta_i$  et avec le même paramètre de concentration  $\nu$ .

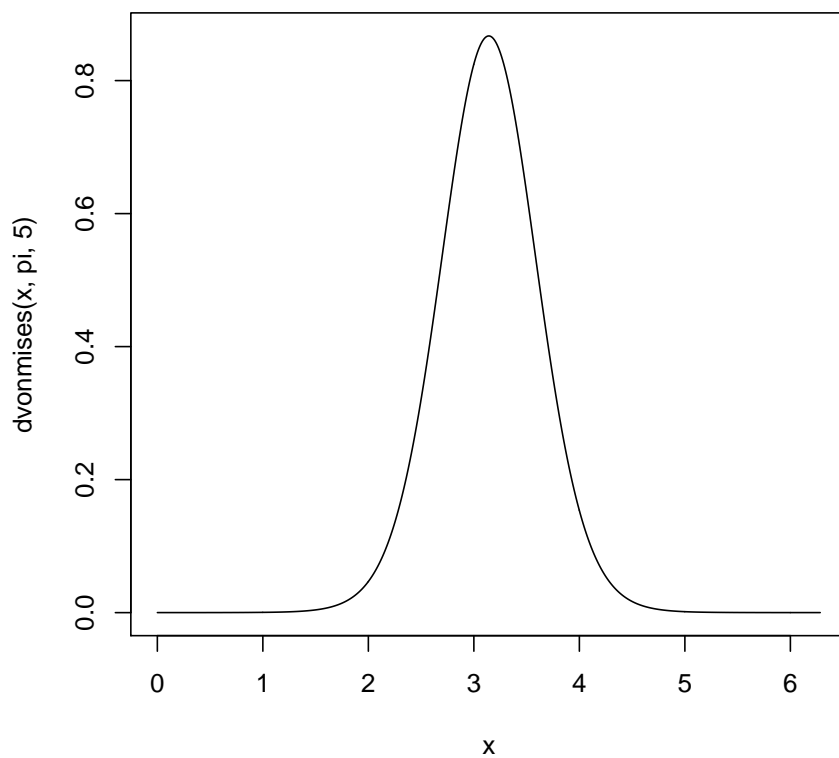


FIGURE 1.2: *Représentation linéaire de la distribution Von Mises*

## 1.2 Propriétés de l'estimateur

Dans cette section, nous présentons les propriétés asymptotiques, locale et globale de l'estimateur à noyau circulaire. En suite, nous traitons le cas particulier de l'estimateur à noyau de Von Mises. Notons que ces propriétés sont obtenues par plusieurs auteurs, voir Hall et al. (1987), Taylor (2008) et Di Marzio et al. (2009).

### 1.2.1 Espérance

L'espérance mathématique est donnée par :

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

En exprimons  $K_\nu(\theta)$  en terme de séries de Fourier et en remarquant que  $\sin(\mu) \sim \mu$  lorsque  $\mu$  tend vers 0, alors on a la formule de Taylor suivante :

$$f(\mu + \theta) = f(\theta) + \sin(\mu) f'(\theta) + \frac{1}{2} \sin^2(\mu) f''(\theta) + o\left(\sin^3(\mu)\right).$$

Maintenant, faisant le changement de variable  $\psi = \mu + \theta$  et avec l'utilisation de l'hypothèse

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

Alors on obtient

$$\begin{aligned} \mathbb{E} \left[ \widehat{f}_\nu(\theta) \right] &= \int_0^{2\pi} K_\nu(\psi - \theta) f(\psi) d\psi \\ &= \int_0^{2\pi} K_\nu(u) f(u + \theta) du \\ &= f(\theta) + \frac{1}{4} \left(1 - y_2(\nu)\right) f''(\theta) + o(1). \text{ (voir DiMarzio2009)} \end{aligned}$$

### 1.2.2 Variance

Avant de calculer la variance de  $\widehat{f}_\nu(\theta)$ , on suppose que

1.  $\lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^{\infty} y_j^2(\nu) = 0$ ,
2.  $f''$  continue et de carré intégrable.

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

$$\begin{aligned} \mathbb{V} \left[ \widehat{f}_\nu(\theta) \right] &= \frac{1}{n} \int_0^{2\pi} \left( K_\nu(\psi - \theta) \right)^2 f(\psi) d\psi - \frac{1}{n} \left( \mathbb{E} \left[ \widehat{f}_\nu(\theta) \right] \right)^2 \\ &= \frac{1}{n} \int_0^{2\pi} \{K_\nu(u)\}^2 \{f(\theta) + o(1)\} du - \frac{1}{n} \{f(\theta) + o(1)\}^2 \\ &= \frac{1}{2n\pi} \left\{ 1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu) \right\} f(\theta) + o(1). \text{ (voir DiMarzio2009)} \end{aligned}$$

### 1.2.3 Biais

Le biais de  $\widehat{f}_\nu(\theta)$  est donné comme suit :

$$\begin{aligned} \text{biais}\{\widehat{f}_\nu(\theta)\} &= \mathbb{E}\{\widehat{f}_\nu(\theta)\} - f(\theta) \\ &= f(\theta) + \frac{1}{4}\{1 - y_2(\nu)\}f''(\theta) + o(1) - f(\theta) \\ &= \frac{1}{4}\{1 - y_2(\nu)\}f''(\theta) + o(1). \end{aligned}$$

**Théorème 1.1.** Si  $\lim_{n \rightarrow +\infty} y_j(\nu) = 1$  pour  $j \in \mathbb{Z}^+$  et  $f''$  est continue, alors l'estimateur  $\widehat{f}_\nu(\theta)$  est asymptotiquement sans biais, c'est à dire  $\lim_{n \rightarrow \infty} \mathbb{E}[\widehat{f}_\nu(\theta)] = f(\theta)$ , en tout point  $\theta$  pour lequel la densité  $f$  est continue.

### 1.2.4 Convergence au sens de l'erreur quadratique moyenne (MSE)

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

$$\begin{aligned} \text{MSE}\{\widehat{f}_\nu(\theta)\} &= \text{biais}^2\{\widehat{f}_\nu(\theta)\} + V\{\widehat{f}_\nu(\theta)\} \\ &= \left[\frac{1}{4}\{1 - y_2(\nu)\}f''(\theta) + o(1)\right]^2 + \frac{1}{2n\pi} \left\{1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu)\right\} f(\theta) + o(1) \\ &= \left[\frac{1}{16}\{1 - y_2(\nu)\}^2\{f''(\theta)\}^2\right] + \frac{1}{2n\pi} \left\{1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu)\right\} f(\theta) + o(1) \end{aligned} \quad (1.3)$$

**Théorème 1.2.** Si  $\lim_{n \rightarrow +\infty} y_j(\nu) = 1$  pour  $j \in \mathbb{Z}^+$ ,  $\lim_{n \rightarrow +\infty} n^{-1} \sum_{j=1}^{\infty} y_j^2(\nu) = 0$  et  $f''$  est continue, alors l'estimateur  $\widehat{f}_\nu(\theta)$  est consistant en moyenne quadratique en tout point  $\theta$  pour lequel la densité  $f$  est continue, c'est à dire  $\lim_{n \rightarrow \infty} \text{MSE}\{\widehat{f}_\nu(\theta)\} = 0$ .

### 1.2.5 Convergence au sens de l'erreur quadratique moyenne intégrée (MISE)

L'erreur quadratique moyenne intégrée (MISE) de  $\widehat{f}_\nu(\theta)$  est obtenue en intégrant la fonction MSE.

$$\begin{aligned}
MISE \left\{ \widehat{f}_\nu(\theta) \right\} &= \int_0^{2\pi} MSE \left\{ \widehat{f}_\nu(\theta) \right\} d\theta \\
&= \int_0^{2\pi} \left[ \frac{1}{16} \{1 - y_2(\nu)\}^2 \{f''(\theta)\}^2 + \frac{1}{2n\pi} \left\{ 1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu) \right\} f(\theta) + o(1) \right] d\theta \\
&= \int_0^{2\pi} \left[ \frac{1}{16} \{1 - y_2(\nu)\}^2 \{f''(\theta)\}^2 \right] d\theta + \int_0^{2\pi} \left[ \frac{1}{2n\pi} \left\{ 1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu) \right\} f(\theta) + o(1) \right] d\theta \\
&= \frac{1}{16} \{1 - y_2(\nu)\}^2 \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{1}{2n\pi} \left\{ 1 + 2 \sum_{j=1}^{\infty} y_j^2(\nu) \right\} \int_0^{2\pi} f(\theta) d\theta + o(1). \tag{1.4}
\end{aligned}$$

**Théorème 1.3.** Si  $\lim_{n \rightarrow +\infty} y_j(\nu) = 1$  pour  $j \in \mathbb{Z}^+$ ,  $\lim_{n \rightarrow +\infty} n^{-1} \sum_{j=1}^{\infty} y_j^2(\nu) = 0$  et  $f''$  est continue, alors l'estimateur  $\widehat{f}_\nu(\theta)$  est consistant en moyenne quadratique en tout point  $\theta$  pour lequel la densité  $f$  est continue, c'est à dire  $\lim_{n \rightarrow \infty} MISE \left\{ \widehat{f}_\nu(\theta) \right\} = 0$ .

Nous donnons maintenant les expressions de MSE et MISE dans cas de l'estimateur à noyau de Von Mises. Notons qu'on peut montrer facilement que dans le cas de noyau de Von Mises,  $y_j(\nu) = I_j(\nu)/I_0(\nu)$  (voir Hall et al. (1987) et Di Marzio et al. (2009)), où  $I_j(\nu)$  désigne la fonction modifiée de Bessel de première espèce à l'ordre  $j$ . En remplaçant l'expression de  $y_j(\nu)$  dans (1.3) et (1.4), alors les expressions de MSE et MISE en utilisant le noyau de Von Mises sont données respectivement par :

$$MSE \left\{ \widehat{f}_\nu^{VM}(\theta) \right\} = \left\{ \frac{1}{16} \left[ 1 - \frac{I_2(\nu)}{I_0(\nu)} \right]^2 \left[ f''(\theta) \right]^2 + \frac{I_0(2\nu)}{2n\pi \left( I_0(\nu) \right)^2} \right\} f(\theta) + o(1). \tag{1.5}$$

$$MISE \left\{ \widehat{f}_\nu^{VM}(\theta) \right\} = \left\{ \frac{1}{16} \left[ 1 - \frac{I_2(\nu)}{I_0(\nu)} \right]^2 \int_0^{2\pi} \left[ f''(\theta) \right]^2 d\theta + \frac{I_0(2\nu)}{2n\pi \left( I_0(\nu) \right)^2} \right\} \int_0^{2\pi} f(\theta) d\theta + o(1) \tag{1.6}$$

### 1.3 Conclusion

Dans ce chapitre, nous avons exposé la méthode du noyau des données circulaires de la fonction densité, particulièrement l'estimateur à noyau de Von Mises. Nous avons donné aussi les différents résultats théoriques concernant la convergence et la consistance de cette

méthode (propriétés asymptotiques, convergence au sens de MSE et de MISE). Notons que cette méthode dépend cruciallement de paramètre de concentration. Dans le chapitre suivant, nous présenterons les différentes méthodes pour estimer ce paramètre.

## Chapitre 2

---

# Choix du paramètre de concentration $\nu$

---

Le choix du paramètre de concentration est d'une importance capitale dans l'estimation de la fonction densité par la méthode du noyau dans le cas des données circulaires. Dans ce chapitre, nous présentons plusieurs méthodes pour le choix de ce paramètre dans le cas de l'estimateur à noyau de Von Mises.

### 2.1 Méthodes de ré-injection (plug-in)

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

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

Rappelons que l'expression asymptotique pour le MISE (AMISE) est donné comme suit :

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

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 pour le cas des données linéaires. Elle consiste à choisir  $f$  comme étant la distribution de Von Mises de paramètre de concentration  $k$ . Dans ce cas l'expression AMISE donnée par (2.2) s'écrit comme suit :

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

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

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

où  $\hat{k}$  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 dans la littérature.

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

Cette alternative a été proposée par Oliveira et al. (2012) et 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  $vM(\mu_r, k_r)$  de proportion  $\alpha_r$ ,  $r = 1, \dots, M$  est donnée par :

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

La valeur approchée  $\hat{\nu}_{PI}$  pour la paramètre de concentration  $\nu$  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  $(\alpha_i, \mu_i, k_i)$  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 (2.2).

**Etape 3** Minimiser  $\text{AMISE}(\nu)$  et obtenir  $\hat{\nu}_{PI}$ .

Noter 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  $\alpha_r, \mu_r$  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.

**Remarque 1.** Notez que le paramètre de concentration  $\nu_{RT}$  obtenu par la règle 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  $\text{CV}(\nu)$  dont le calcul soit plus simple que la  $\text{MISE}(\nu)$ . Le paramètre de concentration  $\nu$  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; 2014)).

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

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

est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation  $\Theta_i$ .

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 montrés que la méthode de validation croisée par le maximum de vraisemblance ( $\hat{\nu}_{LCV}$ ) donne de résultats raisonnable 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 Bowman (1984) pour le cas des données linéaires. Une adaptation pour le cas des données circulaires est donnée par Oliveira et al. (2012; 2014).

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 :

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

Puisque  $\int f^2(\theta) d\theta$  ne dépend pas de paramètre de concentration  $\nu$ , 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  $f_{\nu,i}^{VM}(\theta_i)$  est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation  $\Theta_i$  donné par (2.7). Le critère à optimiser devient alors :

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

Où  $K_\nu$  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  $\frac{2}{n(n-1)(2\pi)I_0(\nu)} \sum_{i=1}^n \sum_{j \neq i}^n \exp\{\nu \cos(\Theta_i - \Theta_j)\}$ .

**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 montrée 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 ( $\hat{\nu}_{UCV}$ ).

## 2.3 Conclusion

Dans ce chapitre, nous avons présenté deux classes de méthodes pour le choix de paramètre de concentration, à savoir les méthodes de ré-injection (plug-in) et les méthodes de type validation croisée. Le principe des méthodes plug-in consiste à approcher la valeur idéale de paramètre de concentration  $\nu_{id} = \min_{\nu} MISE(\nu)$  en remplaçant la quantité inconnue  $f$  par un estimateur approprié. Dans le cas des méthodes de validation croisée, l'idée de base consiste à trouver une fonction de score  $CV(\nu)$  dont le calcul soit plus simple que la  $MISE(\nu)$ .

## Chapitre 3

---

# Etude de simulation

---

Dans ce chapitre, nous présentons une étude de simulation pour illustrer les différentes méthodes pour le choix du paramètre de concentration  $\nu$  sur des données circulaires simulées à partir de distributions connues. Nous comparons les performances de l'estimateur à noyau de Von Mises en utilisant le paramètre de concentration  $\hat{\nu}_{RT}$ , le paramètre de concentration  $\hat{\nu}_{PI}$  avec  $M = 2$ , le paramètre de concentration  $\hat{\nu}_{LCV}$  et le paramètre de concentration  $\hat{\nu}_{UCV}$ . Nous utilisons comme mesure de comparaison l'erreur quadratique intégrée ISE donnée par :

$$\widehat{\text{ISE}} = \frac{1}{N_{sim}} \sum_{t=1}^{N_{sim}} \int_0^{2\pi} \left\{ \hat{f}_h^{[t]}(x) - f(x) \right\}^2, \quad (3.1)$$

où  $\hat{f}_h^{[t]}$  est l'estimateur à noyau de Von Mises obtenu en utilisant le  $t^{\text{ème}}$  échantillon et  $N_{sim}$  est le nombre de simulation (répétitions). Ici, nous avons choisis  $N_{sim} = 500$  répétitions.

Afin d'illustrer les performances des méthodes de choix du paramètre de concentration  $\nu$ , nous utilisons six densités tests. Nous avons choisis des densités présentant différents aspects. Les expressions des densités sont données comme suit :

1. **D1** est le modèle de Von Mises de paramètres  $\mu = \pi$  et  $\kappa = 1$  de densité de probabilité donnée par :

$$f_1(\theta) = \frac{1}{2\pi I_0(1)} \exp \{ \cos(\theta - \pi) \}, \quad 0 \leq \theta < 2\pi.$$

2. **D2** est le modèle de Cauchy enveloppé (wrapped Cauchy en anglais) de paramètres  $\mu = \pi$  et  $\rho = 0.9$  de densité de probabilité donnée par :

$$f_2(\theta) = \frac{1}{2\pi} \frac{1 - 0.8^2}{1 + 0.8^2 - 0.16 \cos(\theta - \pi)}, \quad 0 \leq \theta < 2\pi.$$

3. **D3** est le mélange de deux modèles de Von Mises de paramètres  $\mu_1 = 2$  et  $\kappa_1 = 5$  et  $\mu_2 = 4$  et  $\kappa_2 = 5$  de densité de probabilité donnée par :

$$f_3(\theta) = \frac{1}{2} \frac{1}{2\pi I_0(5)} \exp\{5 \cos(\theta - 4)\} + \frac{1}{2} \frac{1}{2\pi I_0(5)} \exp\{5 \cos(\theta - 2)\}, \quad 0 \leq \theta < 2\pi.$$

4. **D4** est le mélange d'un modèle de Von Mises de paramètres  $\mu_1 = \pi$  et  $\kappa_1 = 5$  et d'un modèle de Cauchy circulaire de paramètres  $\mu_2 = (4\pi)/3$  et  $\rho = 0.9$  de densité de probabilité donnée par :

$$f_4(\theta) = \frac{4}{5} \frac{1}{2\pi I_0(5)} \exp\{5 \cos(\theta - \pi)\} + \frac{1}{5} \frac{1}{2\pi} \frac{1 - 0.9^2}{1 + 0.9^2 - 0.18 \cos(\theta - \frac{4\pi}{3})}, \quad 0 \leq \theta < 2\pi.$$

5. **D5** est le mélange de trois modèles de Von Mises de paramètres  $\mu_1 = \pi/3$  et  $\kappa_1 = 6$ ,  $\mu_2 = \pi$  et  $\kappa_2 = 6$  et  $\mu_3 = (5\pi)/3$  et  $\kappa_3 = 6$  de densité de probabilité donnée par :

$$\begin{aligned} f_5(\theta) &= \frac{1}{3} \frac{1}{2\pi I_0(6)} \exp\{6 \cos(\theta - \frac{\pi}{3})\} + \frac{1}{3} \frac{1}{2\pi I_0(6)} \exp\{6 \cos(\theta - \pi)\} \\ &+ \frac{1}{3} \frac{1}{2\pi I_0(6)} \exp\{6 \cos(\theta - \frac{5\pi}{3})\}, \quad 0 \leq \theta < 2\pi. \end{aligned}$$

de densité de probabilité donnée par :

$$\begin{aligned} f_6(\theta) &= \frac{2}{5} \frac{1}{2\pi I_0(6)} \exp\{6 \cos(\theta - 0.5)\} + \frac{2}{5} \frac{1}{2\pi I_0(6)} \exp\{6 \cos(\theta - 3)\} \\ &+ \frac{1}{5} \frac{1}{2\pi I_0(24)} \exp\{24 \cos(\theta - 5)\}, \quad 0 \leq \theta < 2\pi. \end{aligned}$$

Nous présentons aussi les graphiques des vraies densités circulaires **D1**, **D2**, **D3**, **D4**, **D5** et **D6** dans les figures 3.9, 3.2 et 3.3.

Dans le but de comparer les performances des méthodes de sélection de paramètre de concentration, nous utilisons des données simulées pour chaque densité **D1**, **D2**, **D3**, **D4**, **D5** et **D6** de taille ( $n = 100, 250, 500$  et  $1000$ ). Nous calculons l'estimateur à noyau de

Von Mises donné par (1.2) en utilisant les paramètres de concentration obtenus par les méthodes plug-in et validation croisée. En suite, nous calculons le critère  $\widehat{\text{ISE}}$  donné par (3.1). Les résultats sont données dans les tableaux 3.1-3.6. Les valeurs à gauche donnent les  $\widehat{\text{ISE}}$ s et les valeurs à droite donnent les écarts type pour vérifier la stabilité des méthodes de choix du paramètre de concentration. Les résultats présentés, montrent que dans le cas de **D1**, la meilleure performance est obtenue par la méthode RT pour toutes les tailles d'échantillon. La deuxième meilleure performance est obtenue par PI sauf pour  $n = 100$ . Ces résultats sont expliqués par le fait que les données sont issues d'un modèle de Von Mises. Dans le cas de **D2**, les meilleurs résultats sont obtenus par RT pour  $n = 100$  et 1000 et par PI pour  $n = 250$  et 500. Dans le cas des mélanges de modèles (**D3**, **D4**, **D5** et **D6**), les meilleures performances sont obtenues par LCV, UCV et PI et la mauvaise performance est obtenue généralement par RT.

Afin de visualiser la qualité du lissage, nous avons présenté aussi sur le même graphe les courbes circulaires et linéaires des estimateurs à noyau de Von Mises, obtenus par différentes méthodes de sélection de paramètre de concentration  $\nu$ .



FIGURE 3.1: Densités circulaires (vraie) de **D1** et **D2**.



FIGURE 3.2: Densités circulaires (vraie) de **D3** et **D4**.

TABLE 3.1: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D1**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.006274 0.005168	0.009339 0.008178	0.006914 0.005477	0.007590 0.006269
<b>250</b>	0.002980 0.002454	0.003277 0.002447	0.003313 0.002530	0.003668 0.002749
<b>500</b>	0.001825 0.001247	0.001846 0.001165	0.002034 0.001369	0.002138 0.001454
<b>1000</b>	0.001075 0.000626	0.001008 0.000561	0.001098 0.000632	0.001121 0.000645

TABLE 3.2: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D2**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.029976 0.018147	0.028808 0.017784	0.064609 0.026467	0.030408 0.017610
<b>250</b>	0.016192 0.009418	0.014487 0.008350	0.029658 0.013275	0.021952 0.008794
<b>500</b>	0.010890 0.005741	0.010743 0.005041	0.020242 0.006323	0.020085 0.006110
<b>1000</b>	0.006640 0.003545	0.008300 0.003495	0.018476 0.004511	0.018476 0.004511

TABLE 3.3: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D3**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.035864 0.012097	0.016767 0.009188	0.015010 0.007662	0.015809 0.008113
<b>250</b>	0.033160 0.008470	0.006677 0.003236	0.006939 0.003516	0.006931 0.003407
<b>500</b>	0.027424 0.006645	0.003759 0.001678	0.003956 0.001822	0.0003897 0.001752
<b>1000</b>	0.022650 0.004455	0.002217 0.001009	0.002308 0.001076	0.002248 0.001052

TABLE 3.4: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D4**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.028116 0.009937	0.026903 0.010066	0.030506 0.011668	0.025992 0.010639
<b>250</b>	0.020056 0.005443	0.014643 0.005641	0.020666 0.006963	0.015334 0.005190
<b>500</b>	0.014861 0.003116	0.008880 0.003112	0.014154 0.004059	0.011102 0.002568
<b>1000</b>	0.011597 0.001915	0.006297 0.001630	0.009934 0.001785	0.009630 0.001488



FIGURE 3.3: Densités circulaires (vraie) de **D5** et **D6**.

TABLE 3.5: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D5**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.064003 0.000644	0.015002 0.006330	0.015358 0.006548	0.015728 0.006872
<b>250</b>	0.063397 0.000760	0.006832 0.002977	0.007074 0.003040	0.007223 0.003105
<b>500</b>	0.062653 0.000929	0.003981 0.001543	0.004043 0.001560	0.004106 0.001601
<b>1000</b>	0.061499 0.001057	0.002374 0.000972	0.002360 0.000967	0.002392 0.000977

TABLE 3.6: Résultats de simulation de  $\widehat{ISE}$  pour le modèle **D6**.

<b>n</b>	$\widehat{ISE}_{RT}$	$\widehat{ISE}_{PI}$	$\widehat{ISE}_{LCV}$	$\widehat{ISE}_{UCV}$
<b>100</b>	0.100153 0.008305	0.019511 0.007504	0.018560 0.006789	0.019003 0.006777
<b>250</b>	0.097705 0.008305	0.009538 0.003694	0.009255 0.003641	0.009455 0.003692
<b>500</b>	0.092877 0.005139	0.005386 0.002071	0.005363 0.002023	0.005425 0.002077
<b>1000</b>	0.084257 0.005475	0.003220 0.001140	0.003783 0.001204	0.003783 0.001204

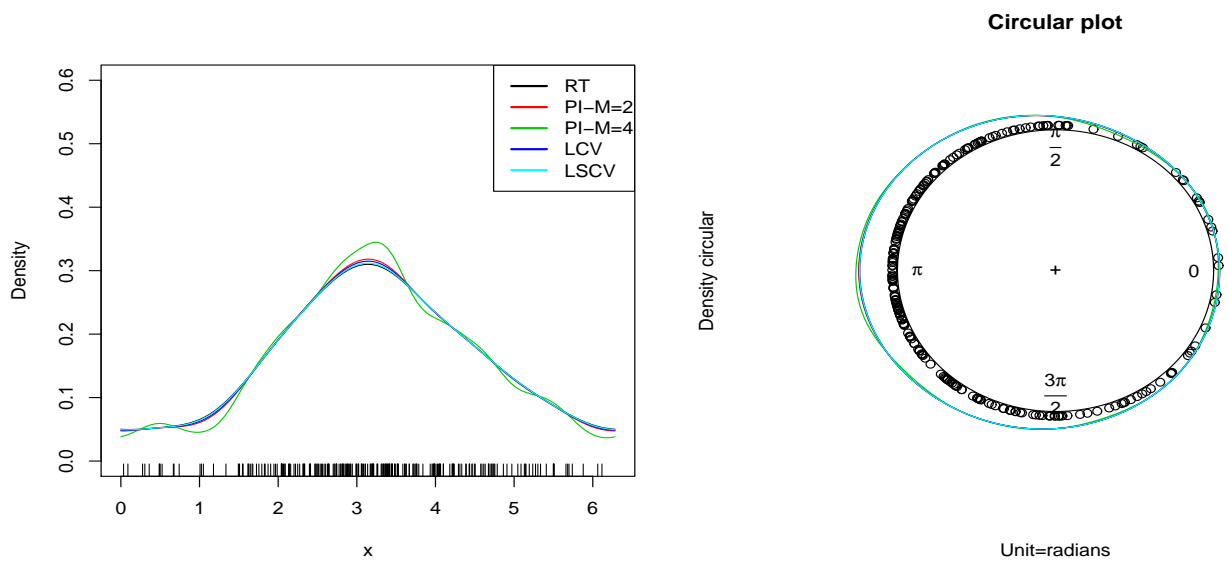


FIGURE 3.4: Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D1** avec  $n = 250$ .

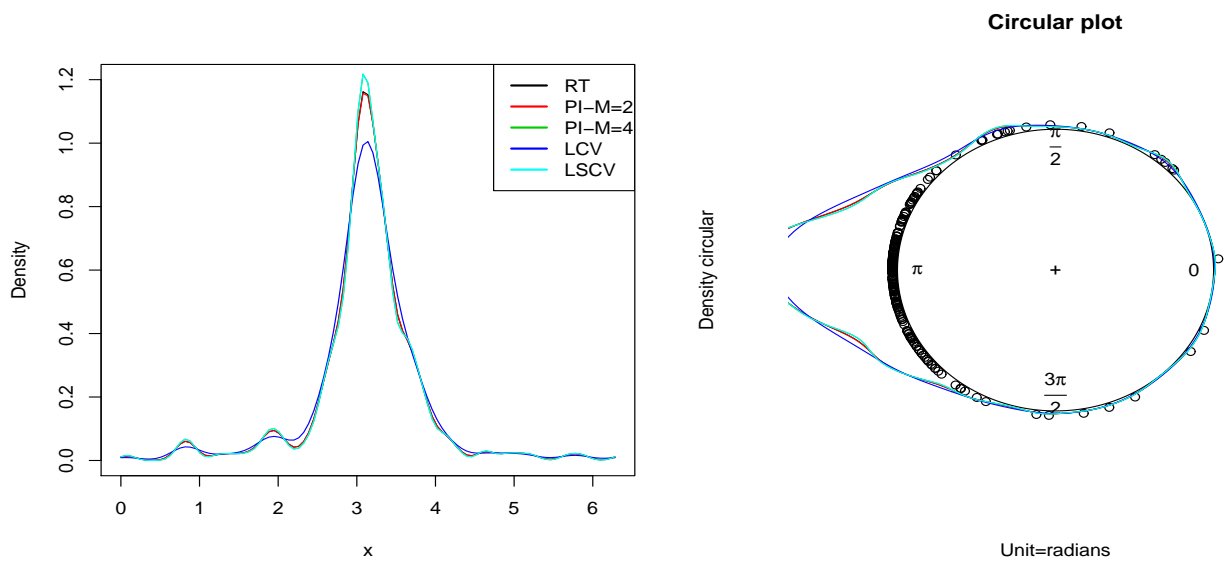


FIGURE 3.5: *Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D2** avec  $n = 250$ .*

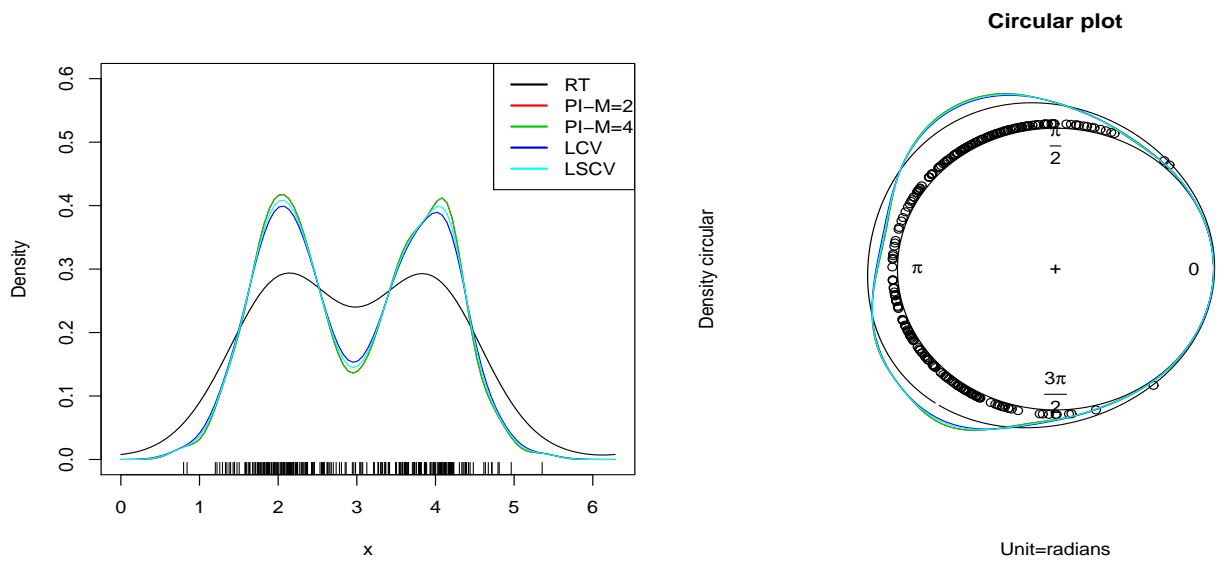


FIGURE 3.6: *Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D3** avec  $n = 250$ .*

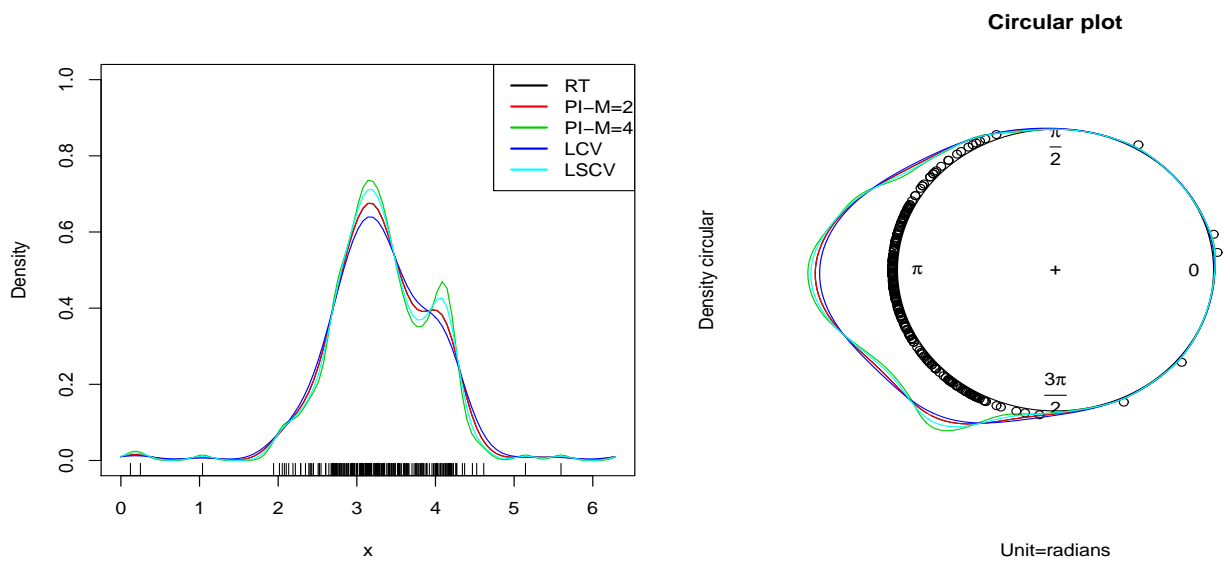


FIGURE 3.7: *Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D4** avec  $n = 250$ .*

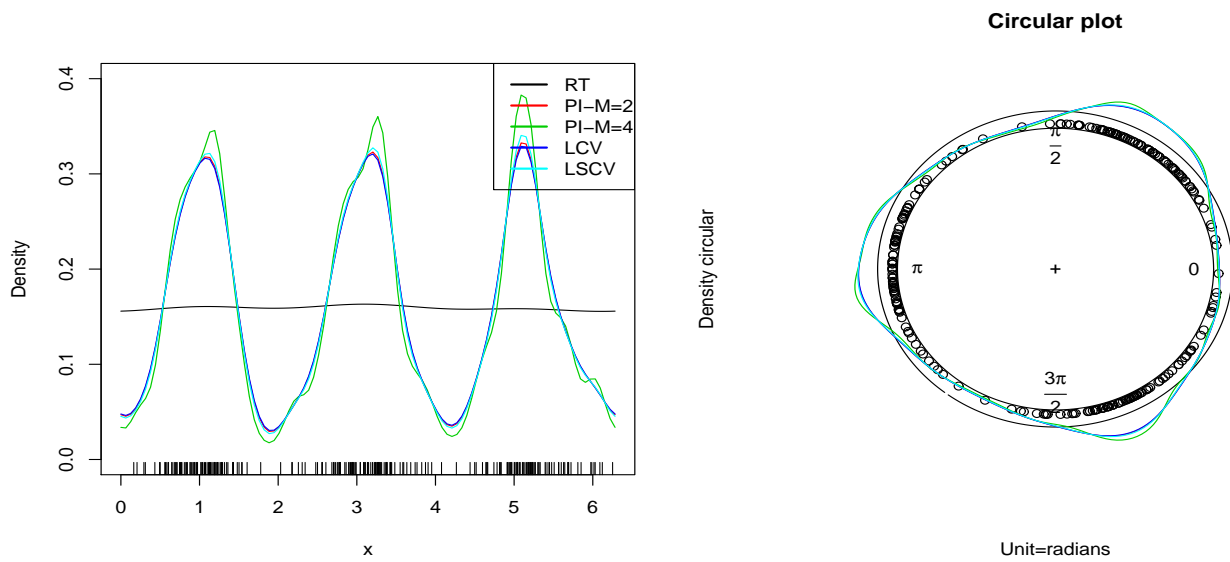


FIGURE 3.8: *Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D5** avec  $n = 250$ .*

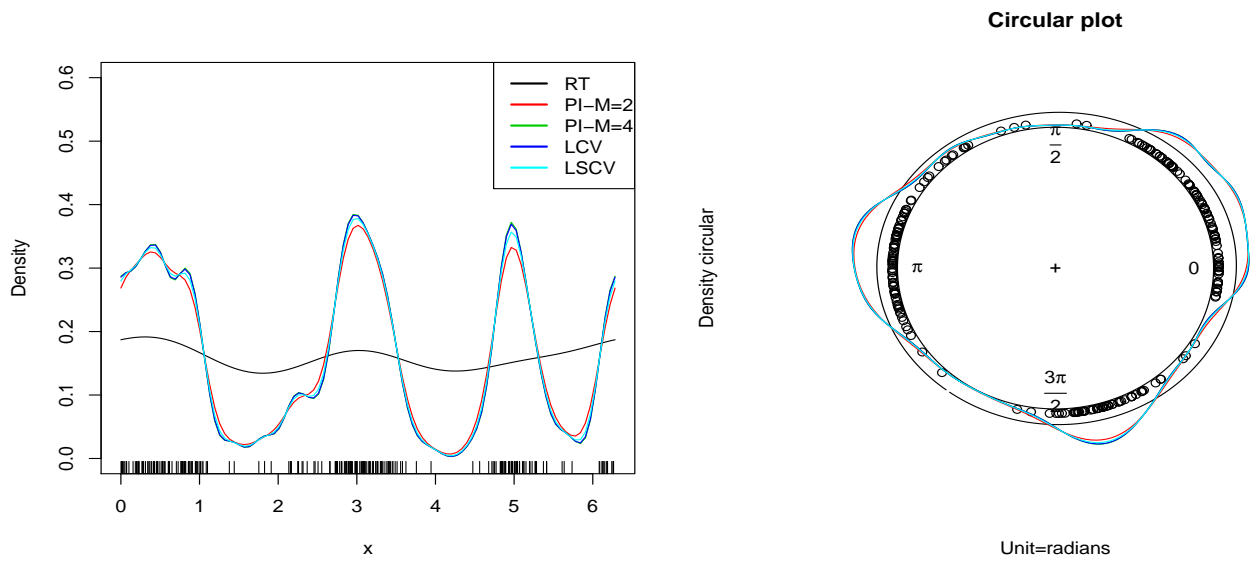


FIGURE 3.9: *Estimateur à noyau de Von Mises en utilisant différents paramètres de concentration pour **D6** avec  $n = 250$ .*

## Chapitre 4

---

# Application sur des données réelles

---

Nous considérons les données circulaires suivantes : 0.05235988 0.38397244 0.68067841  
0.76794487 0.87266463 0.87266463 0.90757121 0.90757121 0.90757121 0.94247780 0.94247780  
0.94247780 0.94247780 0.94247780 0.97738438 0.97738438 0.97738438 0.97738438 1.01229097  
1.01229097 1.01229097 1.04719755 1.04719755 1.08210414 1.08210414 1.08210414 1.11701072  
1.11701072 1.15191731 1.15191731 1.15191731 1.15191731 1.15191731 1.15191731 1.22173048  
1.22173048 1.25663706 1.25663706 1.25663706 1.25663706 1.29154365 1.29154365 1.29154365  
1.32645023 1.32645023 1.32645023 1.32645023 1.32645023 1.36135682 1.36135682 1.36135682  
1.36135682 1.39626340 1.39626340 1.39626340 1.39626340 1.43116999 1.43116999 1.46607657  
1.46607657 1.46607657 1.50098316 1.50098316 1.50098316 1.50098316 1.53588974 1.53588974  
1.53588974 1.53588974 1.53588974 1.57079633 1.57079633 1.57079633 1.57079633 1.57079633  
1.57079633 1.60570291 1.60570291 1.60570291 1.60570291 1.64060950 1.64060950 1.64060950  
1.67551608 1.67551608 1.67551608 1.67551608 1.67551608 1.67551608 1.71042267 1.71042267  
1.71042267 1.71042267 1.71042267 1.71042267 1.74532925 1.74532925 1.74532925 1.83259571  
1.88495559 1.93731547 1.98967535 2.04203522 2.09439510 2.23402144 2.28638132 2.33874120  
2.40855437 2.47836754 2.87979327 4.06661716 4.13643033 4.13643033 4.18879020 4.25860337  
4.29350996 4.32841654 4.32841654 4.32841654 4.32841654 4.36332313 4.39822972 4.39822972  
4.39822972 4.43313630 4.43313630 4.43313630 4.43313630 4.43313630 4.46804289 4.46804289  
4.46804289 4.50294947 4.53785606 4.53785606 4.57276264 4.57276264 4.60766923 4.60766923  
4.60766923 4.60766923 4.60766923 4.64257581 4.64257581 4.64257581 4.64257581 4.64257581

4.67748240 4.67748240 4.67748240 4.67748240 4.67748240 4.67748240 4.71238898 4.71238898  
4.71238898 4.74729557 4.74729557 4.74729557 4.74729557 4.74729557 4.78220215 4.78220215  
4.78220215 4.81710874 4.81710874 4.81710874 4.81710874 4.81710874 4.81710874 4.85201532  
4.85201532 4.85201532 4.85201532 4.85201532 4.88692191 4.88692191 4.88692191 4.88692191  
4.92182849 4.92182849 4.92182849 4.92182849 4.92182849 4.95673508 4.95673508 4.95673508  
4.95673508 4.99164166 4.99164166 4.99164166 5.02654825 5.02654825 5.02654825 5.06145483  
5.06145483 5.06145483 5.06145483 5.11381471 5.11381471 5.11381471 5.11381471 5.18362788  
5.23598776 5.39306739 5.39306739 5.46288056 5.49778714 5.58505361 5.81194641 5.81194641  
6.12610567 6.12610567 6.17846555 qui décrivent l'orientation de  $n = 214$  libellules suivant  
l'azimut du soleil. Ces données ont été utilisés par Batschelet (1981), Pewsey (2004) et  
récemment par Oliveira et al (2012). Notons que ces données sont en radians. Nous avons  
calculé les valeurs de paramètre de concentration par les différentes méthodes présentées  
dans le chapitre 2 :  $\hat{\nu}_{RT} = 3.2262$ ,  $\hat{\nu}_{PI_{M=2}} = 41.2274$ ,  $\hat{\nu}_{PI_{M=4}} = 62.2724$ ,  $\hat{\nu}_{LCV} = 35.3681$   
et  $\hat{\nu}_{UCV} = 63.8650$ . Afin de comparer les différentes méthodes de choix du paramètre  
de concentration, nous avons tracé sur la même figure les estimateurs à noyau de Von  
Mises pour chaque méthode. Les figures 4.1 et 4.2 montrent la représentation linéaire et  
circulaire respectivement de l'estimateur à noyau de Von Mises pour ces données. Il est  
clair que la distribution pour cet exemple de données est bimodale. A travers les résultats,  
nous remarquons que les performances des méthodes de plug-in de Oliveira et al. (2012)  
(PI), la méthode de validation croisée par maximum de vraisemblance (LCV) et validation  
croisée sans biais (UCV) sont quasi-similaires. La mauvaise performance est obtenue par  
la méthode de règle de référence (RT) proposée par Taylor (2008), car elle a tendance  
à sous-lissée la distribution de ces données. Ceci, peut être expliqué par le fait que les  
données sont bimodale.

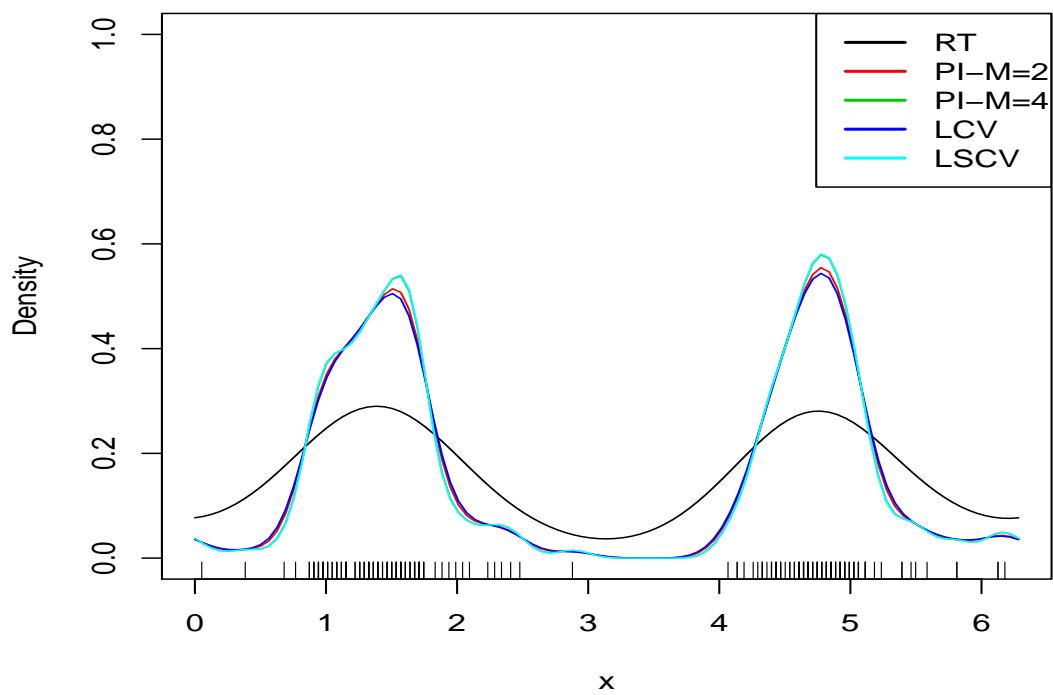


FIGURE 4.1: Représentation linéaire de l'estimateur à noyau de Von Mises pour les données de l'orientation de libellules. Différentes méthodes pour le choix de paramètre de concentration sont utilisées et présentées.

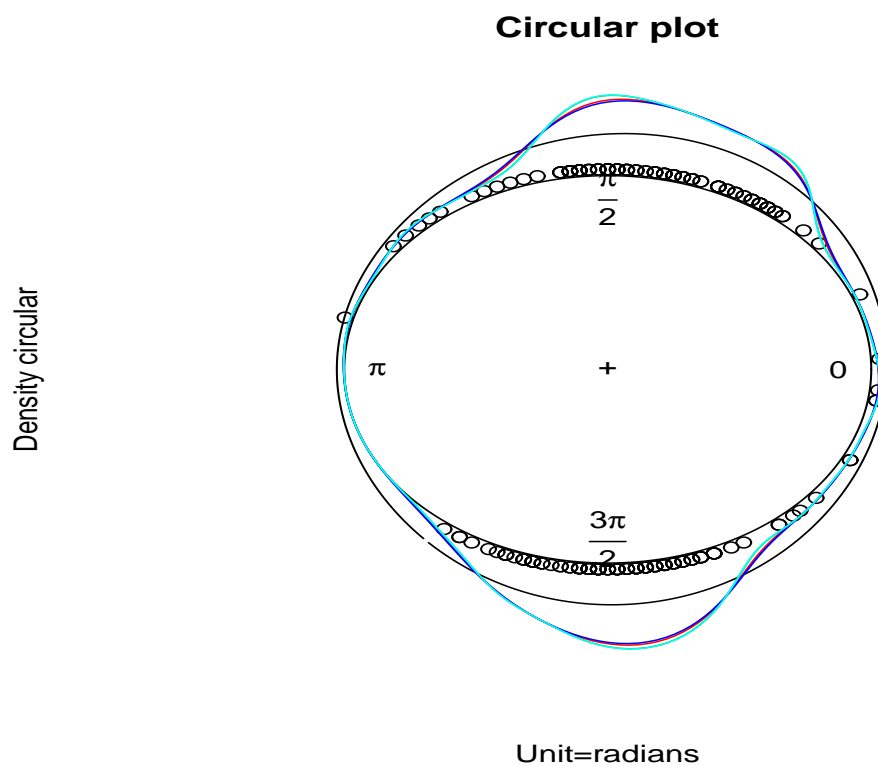


FIGURE 4.2: Représentation circulaire de l'estimateur à noyau de Von Mises pour les données de l'orientation de libellules avec différentes méthodes pour le choix de paramètre  $\nu$ .

---

# Conclusion générale

---

Dans ce travail, nous avons étudié l'applicabilité et l'efficacité des différentes techniques de sélection du paramètre de concentration dans l'estimation de la fonction densité de type circulaire par la méthode du noyau circulaire. Ces performances ont été mesurées numériquement à l'aide de jeux de données simulés. Une application sur des données réelles a été aussi présentée.

Dans une première partie, nous avons exposé la méthode du noyau circulaire dans le cas de l'estimation de la densité de type circulaire, en particulier, nous avons présenté l'estimateur à noyau de Von Mises à cause de sa popularité et sa simplicité. L'estimateur à noyau circulaire est fonction de deux paramètres : la fonction  $K$  appelée noyau et  $\nu$  appelé paramètre de concentration. Si le choix du noyau  $K$  n'est pas un problème, il n'en est pas de même pour le choix du paramètre de concentration  $\nu$  qui ne dépend que de la taille de l'échantillon.

Dans la deuxième partie, nous avons exposé les différentes méthodes de sélection du paramètre de concentration. Les méthodes reposant sur la validation croisée dont l'intérêt est le caractère direct, et l'autre classe de méthodes dite plug-in (ré-injection) qui reposent sur l'estimation de la quantité inconnue  $\int f''$ .

Enfin, afin de tester les différentes méthodes de sélection du paramètre de concentration, nous avons simulé des densités de probabilité circulaires tests présentant différents aspects (loi de Von Mises, loi de Cauchy enveloppée (Wrapped Cauchy), mélange de lois de Von Mises, ...) et nous avons calculé le critère de comparaison ISE. Les résultats numériques obtenus montrent que :

1. La méthode de règle de référence (RT) est supérieure aux autres méthodes lorsque les données sont issues d'un modèle de Von Mises.

2. Les méthodes plug-in de Oliveira et al. (2012) (PI), validation croisée par maximum de vraisemblance (LCV) et validation croisée sans biais (UCV) sont meilleures que la méthode de règle de référence (RT) dans le cas de mélange de modèles.
3. Il est essentiel de comprendre qu'il n'existe pas de méthode de sélection de paramètre de concentration qui soit meilleure que les autres dans tous les cas.

Parmi les perspectives de ce travail, nous pouvons dégager plusieurs axes intéressants :

1. Proposer l'approche bayésienne pour le choix de paramètre de concentration  $\nu$  (voir Zougab et al. (2012 ; 2013)).
2. Etudier l'estimateur à noyau circulaire de la fonction de régression.

---

# Annexe

---

## Code source avec R

Les programmes informatiques ont été implimentés sous R en utilisant les packages `circular` et `NPCirc`. Nous présentons les programmes mis en place pour le choix du paramètre de concentration par les méthodes plug-in et validation croisée dans le cas de l'estimation de la fonction densité.

```
#=====
##### choix du paramètre de concentration et calcul du critère ISE
#=====

library(circular)
library(NPCirc)

ISErt=0
ISEpi=0
ISElcv=0
ISEucv=0

i=1
Nsim=500 ### Changer le nombre de répétitions ici #####
while (i<=Nsim){
print(i)
n=100 ### Changer la taille de l'échantillon n ici #####
```

```
data<- rcircmix(n, model=2)

resrt <- density(data, bw=bw.rt(data,robust=TRUE),
control.circular=list(units="radians"))
respi <- density(data, bw=bw.pi(data, M=3),
control.circular=list(units="radians"))
reslcv <- density(data, bw=bw.CV(data, method="LCV", lower=0, upper=50),
control.circular=list(units="radians"))
resucv <- density(data, bw=bw.CV(data, method="LSCV", lower=0, upper=50,
tol=1e-2, np=500), control.circular=list(units="radians"))

xx=resrt$x
res1=dcircmix(xx, model=2)
res1rt=resrt$y
res1rt
res1pi=respi$y
res1pi
res1lcv=reslcv$y
res1lcv
res1ucv=resucv$y
res1ucv

ISErt[i]=2*pi*mean((res1-res1rt)^2)
ISEpi[i]=2*pi*mean((res1-res1pi)^2)
ISElcv[i]=2*pi*mean((res1-res1lcv)^2)
ISEucv[i]=2*pi*mean((res1-res1ucv)^2)

i=i+1}

RESrt=c(mean(ISErt),sd(ISErt))
```

```
RESrt
RESpi=c(mean(ISEpi),sd(ISEpi))
RESpi
RESlcv=c(mean(ISElcv),sd(ISElcv))
RESlcv
RESucv=c(mean(ISElcv),sd(ISElcv))
RESucv

#=====
##### Tracer les courbes de l'estimateur à noyau circulaire
##### (circulaire et linéaire)
#=====

library(circular)
library(NPCirc)

data(dragonfly)

theta1=circular(dragonfly$orientation)
bw1=bw.rt(theta1, robust=TRUE)
bw2=bw.pi(theta1, M=2)
bw3=bw.pi(theta1, M=4)
bw4=bw.CV(theta1, method="LCV", lower=0.5, upper=100)
bw5=bw.CV(theta1, method="LSCV", lower=0.5, upper=100)

est1 <- kern.den.circ(theta1, t=NULL, bw=bw1)
plot(est1, plot.type="circle", points.plot=TRUE,
shrink=1.3, main="Circular plot", xlab="Unit=radians")
est2 <- kern.den.circ(theta1, t=NULL, bw=bw2)
lines(est2, plot.type="circle",shrink=1.3,col=2)
```

---

```
est3<- kern.den.circ(theta1, t=NULL, bw=bw3)
lines(est3, plot.type="circle",shrink=1.3,col=3)
est4<- kern.den.circ(theta1, t=NULL, bw=bw4)
lines(est4, plot.type="circle",shrink=1.3,col=4)
est5<- kern.den.circ(theta1, t=NULL, bw=bw5)
lines(est5, plot.type="circle",shrink=1.3,col=5)
```

---

# Bibliographie

---

- E. Batschelet. *Circular Statistics in Biology*. Academic Press, New York, 1981.
- A W. Bowman. An Alternative Method of Cross-validation for the Smoothing of Density Estimates. *Biometrika*, 71 :353–360, 1984.
- S X. Chen. Beta kernels estimators for density functions. *Computational Statistics and Data Analysis*, 31 :131–145, 1999.
- S X. Chen. Gamma kernel estimators for density functions. *Annals of the Institute of Statistical Mathematics*, 52 :471–480, 2000a.
- M. Di Marzio, A. Panzera, and C C. Taylor. Local polynomial regression for circular predictors. *Statistics and Probability Letters*, 79 :2066–2075, 2009.
- E A. Epanechnikov. Nonparametric estimation of a multidimensional probability density. *Theory. Probab. Appl*, 14 :153–158, 1969.
- 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.
- P. Hall, G P. Watson, and J. Cabrera. Kernel density estimation for spherical data. *Biometrika*, 74 :751–762, 1987.
- 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.
- J. Klemelä. Estimation of densities and derivatives of densities with directional data. *Journal of Multivariate Analysis*, 73 :18–40, 2000.
- C C. Kokonendji and F G. Libengué. Méthode des noyaux associés continus et estimation de densité. *Journées de Statistique de la SFdS*, 6 pages, Tunis, 2011.
- C C. Kokonendji and T. Senga Kiessé. Discrete associated kernels method and extensions. *Statistical Methodology*, 8 :497–516, 2011.
- P. Langevin. Magnetisme et theorie des electrons. *Ann. Chim. Phys*, 5 :71—127, 1905.
- K V. Marida and P E. Jupp. *Directional Statistics*. Wiley, New York, 2000.
- J A. Mooney, P J. Helms, and L T. Jolliffe. Fitting mixtures of von mises distributions : a case study involving sudden infant death syndrome. *Computational Statistics and Data Analysis*, 41 :505–513, 2003.
- M. Oliveira, M.R. 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, M.R. Crujeiras, and A. Rodríguez-Casal. NPCirc : An R Package for Non-parametric Circular. *Journal of Statistical Software*, 61 :1–26, 2014.
- E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33 :1065–1076, 1962.
- A. Pewsey. Testing for circular reflective symmetry about a known median axis. *Journal of Applied Statistics*, 31 :575–585, 2004.
- M. Rosenblatt. Remarks in some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27 :832–837, 1956.

- O. Scaillet. Density estimation using inverse and reciprocal inverse gaussian kernels. *Journal of Nonparametric Statistics*, 16 :217–226, 2004.
- T. Senga Kiessé. *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Thèse de Doctorat, Université de Pau, France, 2008.
- B W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, 1986.
- C C. Taylor. Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*, 52 :3493–3500, 2008.
- R. Von Mises. Über die "ganzzahligkeit" der atomgewichte und verwandte fragen. *Physikal. Z*, 19 :490—500, 1918.
- N. Zougab, S. Adjabi, and C C. Kokonendji. Binomial kernel and Bayes local bandwidth in discrete functions estimation. *Journal of Nonparametric Statistics*, 24 :3 :783–795, 2012.
- 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.