

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

UNIVERSITÉ MOULOUD MAMMARI, TIZI-OUZOU

FACULTÉ DES SCIENCES

DÉPARTEMENT DE MATHÉMATIQUES



## **MEMOIRE DE MASTER**

SPÉCIALITÉ : MATHÉMATIQUES

OPTION : Probabilité et Statistique

**Présenté par GAHRAR Rachida**

Sujet

**Propriété asymptotique de l'estimation de la densité du noyau de Cauchy enveloppé  
circulaire**

Devant le jury d'examen composé de :

Mr MEHIRI Mohamed ;	MAA ;	UMMTO ;	Président
Mlle BEDOUHENE Kahina ;	MCB ;	UMMTO ;	Rapporteur
Mme MERABET Dalila ;	MCB ;	UMMTO ;	Examinatrice

Soutenue : le 01/07/2025

# Remerciements

Je tiens à exprimer ma profonde gratitude à ma directrice de mémoire, **Dr Bedouhene Kahina**, Maîtresse de Conférences à l'Université Mouloud Mammeri de Tizi-Ouzou, pour l'accompagnement rigoureux et bienveillant dont j'ai bénéficié tout au long de ce travail. Sa disponibilité, son expertise scientifique, ainsi que sa grande exigence intellectuelle ont été pour moi des sources d'enrichissement inestimables. Je suis sincèrement reconnaissante pour l'attention qu'elle a portée à mon travail et le temps qu'elle lui a consacré.

Mes remerciements les plus chaleureux s'adressent également à **Monsieur Mohamed Mehiri**, enseignant à l'Université Mouloud Mammeri, pour l'honneur qu'il me fait en acceptant de présider le jury de ce mémoire.

J'exprime également ma reconnaissance à **Madame Merabet Dalila**, Maîtresse de Conférences à l'UMMTO, pour sa disponibilité et pour l'intérêt qu'elle a manifesté à l'égard de mon travail en acceptant de siéger dans le jury.

Je souhaite adresser une mention toute particulière à l'ensemble des enseignants du département de Mathématiques de l'UMMTO, et plus spécialement à l'équipe de la spécialité ■ Probabilités et Statistiques ■ : **Pr H. Fellag, Dr L. Atil, Dr Ch. Belkacem, Dr K. Boullam, Dr D. Merabet, Dr M. Mehiri, Dr M. Semali, Pr A. Hamaz et Pr Dj. Hamadouche**, pour la qualité de leur enseignement et leur engagement pédagogique.

# Dédicace

Ce travail est bien plus qu'un simple mémoire. Il est le fruit d'un parcours long, parfois difficile, mais toujours porté par l'amour, la patience, la bienveillance et le soutien de celles et ceux qui me sont chers. À travers ces lignes, je souhaite leur exprimer ma plus profonde gratitude.

À mes chers parents, qui ont toujours été là, dans l'ombre ou en pleine lumière, pour m'encourager, me soutenir, me rassurer et me guider. Leur amour inconditionnel, leurs sacrifices quotidiens, leurs prières silencieuses et leur foi en moi ont été le socle de mon avancement. Ce mémoire est autant le leur que le mien. Que Dieu les protège et les comble de Ses bienfaits.

À chacun de mes frères, pour leur force tranquille, leur protection naturelle et leur fierté qu'ils n'ont jamais su cacher, même sans mots. Leur soutien discret mais sincère m'a souvent portée sans qu'ils ne le sachent.

À chacune de mes sœurs, pour leur affection profonde, leur écoute attentive, leurs gestes tendres et leurs paroles réconfortantes. Leur présence chaleureuse a été un appui précieux, surtout dans les moments de doute.

À mes amies, véritables compagnes de route, pour leur bienveillance, leur disponibilité, leurs conseils et leur amitié sincère. Merci pour vos encouragements constants, votre humour, vos messages rassurants et votre soutien sans faille. Vous avez rendu ce parcours plus doux et plus humain.

À mes camarades de spécialité en probabilités et statistiques, avec qui j'ai partagé tant d'efforts, de réflexions, de doutes, mais aussi de réussites et de fierté. Merci pour les échanges enrichissants, l'entraide dans les moments difficiles et la solidarité qui nous a liés.

Et enfin, avec une pensée chargée d'émotion, je dédie ce mémoire à la mémoire de mon grand-père, que la vie m'a arraché, mais que le cœur n'oubliera jamais. Sa sagesse, sa tendresse, sa patience et sa foi continuent de m'accompagner. Qu'Allah lui fasse miséricorde, lui accorde Sa lumière et lui ouvre les portes de Son vaste Paradis. Puisse-t-il être fier, là où il est, de ce petit pas que je fais aujourd'hui.

---

# Table des matières

---

<b>Introduction générale</b>	<b>8</b>
<b>1 Estimation non-paramétrique de la fonction densité par noyau circulaire</b>	<b>10</b>
1.1 Estimateur de la densité par noyau circulaire . . . . .	10
1.1.1 Propriétés de l'estimateur . . . . .	11
1.1.1.1 Biais . . . . .	12
1.1.1.2 Variance . . . . .	13
1.1.2 MSE et MISE . . . . .	14
1.1.2.1 Convergence au sens de MSE . . . . .	14
1.1.2.2 Convergence au sens de MISE . . . . .	14
1.1.3 Convergence en loi . . . . .	15
1.1.4 Estimateur à noyau de Cauchy enveloppé et ses propriétés . . . . .	15
1.1.4.1 Espérance de l'estimateur à noyau de Cauchy enveloppé . . . . .	17
1.1.4.2 Variance de l'estimateur à noyau de Cauchy enveloppé . . . . .	17
1.1.4.3 Biais de l'estimateur à noyau de Cauchy enveloppé . . . . .	18
1.1.4.4 Convergence de l'estimateur de Cauchy enveloppé au sens de MSE . . . . .	19
1.1.4.5 Convergence de l'estimateur de Cauchy enveloppé au sens de MISE . . . . .	19
1.1.5 Estimateur à noyau de von Mises et ses propriétés . . . . .	19
1.1.5.1 Biais de l'estimateur de Von Mises . . . . .	20
1.1.5.2 Variance de l'estimateur de Von Mises . . . . .	22

1.1.5.3	Convergence de l'estimateur de Von Mises au sens de MISE	22
1.2	Conclusion	23
<b>2</b>	<b>Choix du paramètre de concentration <math>\rho</math> dans l'estimation non-paramétrique de la densité par noyau circulaire</b>	<b>24</b>
2.1	Méthodes de ré-injection (plug-in)	24
2.1.1	Règle de référence (Rule of Thumb)	25
2.1.2	Méthode plug-in de Oliveira et al. (2012)	26
2.2	Méthodes de validation croisée	26
2.2.1	Validation croisée par le maximum de vraisemblance	27
2.2.2	Validation croisée non biaisée	27
2.2.3	Conclusion	29
<b>3</b>	<b>Étude de Simulation</b>	<b>30</b>
3.1	Résultats de la simulation	30
3.1.1	Expérience A : mélange de densités de Von Mises	31
3.1.2	Expérience B : densité de Cauchy enveloppée	32
3.2	Conclusion	37
	<b>Conclusion générale et perspectives</b>	<b>38</b>
	<b>Annexe</b>	<b>40</b>
	<b>Bibliographie</b>	<b>52</b>

---

# Table des figures

---

1.1	Noyau de Cauchy enveloppé pour $\theta = \pi/4$ (première rangée cas linéaire) et $\theta = \pi$ (deuxième rangée cas circulaire ) avec différentes valeurs de $\rho = \{0.3, 0.5, 0.7, 0.9\}$ . . . . .	16
1.2	Noyau de von Mises pour $\theta = \pi/4$ (première rangée cas linéaire) et $\theta = \pi$ (deuxième rangée cas circulaire) avec différentes valeurs de $\nu = \{3, 5, 10, 15\}$ . . . . .	21
3.1	Comparaison graphique des erreurs quadratiques moyennes intégrées ( <i>MISE</i> ) pour les expériences A et B. . . . .	35
3.2	Représentation des densités circulaires dans les expériences A et B . . . . .	36

---

# Liste des tableaux

---

3.1	Différence $\text{MISE}(\hat{f}^{VM}(\cdot; \kappa^*)) - \text{MISE}(\hat{f}^{WC}(\cdot; \rho^*))$ pour différentes valeurs de $\kappa$ et tailles $n$ . . . . .	32
3.2	Différence $\text{MISE}(\hat{f}^{VM}(\cdot; \kappa^*)) - \text{MISE}(\hat{f}^{WC}(\cdot; \rho^*))$ pour différentes valeurs de $\rho$ . . . . .	33

---

# Introduction générale

---

Les données directionnelles sont représentées par un angle  $\theta \in [0, 2\pi)$ , avec  $\theta = \theta + 2m\pi$ ,  $m \in \mathbb{Z}$ , ou par un vecteur unitaire  $\mathbf{x} = (\cos(\theta), \sin(\theta))^T$ . On les rencontre fréquemment dans des domaines tels que la météorologie (direction du vent), l'astronomie, l'écologie ou encore l'énergie (variations journalières de production, etc.).

L'estimation de densité pour ces données s'est d'abord développée par une approche paramétrique, notamment via la loi de von Mises (normale circulaire), caractérisée par un paramètre de moyenne  $\mu \in [0, 2\pi)$  et un paramètre de concentration  $\kappa > 0$ . Lorsque les données présentent une structure plus complexe (multimodalité, asymétrie), cette approche paramétrique peut être inadaptée.

D'où l'intérêt de recourir à une approche non-paramétrique, notamment à travers l'estimation par noyau, étendue au cas circulaire par divers auteurs [Hall et al. \[1987\]](#), [Taylor \[2008\]](#), [Di Marzio et al. \[2011a\]](#), [Oliveira et al. \[2012\]](#). Le noyau de Von Mises (VM) est alors souvent utilisé, bien qu'il ne possède pas certaines propriétés souhaitables comme la réductibilité.

[Di Marzio et al. \[2011a\]](#) ont introduit la notion de moments d'ordre sinusoidal pour classer les noyaux circulaires. Le noyau de Cauchy enveloppé (Wrapped Cauchy WC), qui appartient à cette classe, possède des queues plus lourdes que le VM, une propriété de réductibilité, et un taux de convergence asymptotique optimal de l'erreur quadratique moyenne intégrée (AMISE).

Bien que théoriquement le noyau VM présente de meilleures performances (MISE), les simulations montrent que le noyau WC peut surpasser le VM dans les situations complexes : densités multimodales, fortes concentrations ou queues lourdes.

C'est pourquoi ce travail s'intéresse aux performances comparées de ces deux estima-

teurs à noyau : VM et WC.

Deux grandes familles de méthodes ont été proposées dans la littérature pour choisir le paramètre de concentration :

- Les méthodes de type plug-in, qui visent à minimiser une approximation de la MISE, Taylor [2008], Oliveira et al. [2012] .
- Les méthodes de validation croisée, reposant uniquement sur les observations, Oliveira et al. [2012].

L'objectif de ce mémoire est double :

- Présenter et comparer différentes méthodes de sélection du paramètre de concentration sur des lois circulaires connues (VM, WC, mélange de VM) ;
- Étudier l'influence de la taille de l'échantillon sur les performances de ces méthodes.

Ce mémoire est structuré comme suit :

- Le premier chapitre décrit l'estimateur de densité circulaire par noyau, en mettant l'accent sur le noyau Wrapped Cauchy ;
- Le second chapitre présente les méthodes de sélection du paramètre de concentration ;
- Le troisième chapitre est consacré à l'étude par simulation et à l'interprétation des résultats obtenus.

Le mémoire se termine par une conclusion générale et quelques perspectives de recherche.

# Chapitre 1

---

## Estimation non-paramétrique de la fonction densité par noyau circulaire

---

Dans ce chapitre, nous introduisons d'abord la définition de l'estimateur à noyau appliqué aux données circulaires pour l'estimation de la densité, à titre exemple l'estimateur à noyau de Cauchy enveloppé et Von Mises. Ensuite, nous examinons ses propriétés asymptotiques, notamment le biais et la variance. Enfin, nous analysons les critères d'évaluation locale et globale de cet estimateur, à travers l'erreur quadratique moyenne (MSE) et l'erreur quadratique moyenne intégrée (MSIE).

### 1.1 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 tout  $\theta \in [0, 2\pi[$ , le noyau  $K_\nu$  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  $\gamma_j$  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 que 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  $K_\nu$ , 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  $\theta$  est la cible (l'angle où on veut estimer  $f$ ).

### 1.1.1 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.

### 1.1.1.1 Biais

Le calcul des moments d'ordre 1 et 2 de l'estimateur circulaire (1.3) requièrent 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  $K_\nu$  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 :

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

Donc la formule du biais est donnée par :

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

### 1.1.1.2 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  $K_\nu$  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 (voir annexe (3.1)), la formule de la variance de l'estimateur circulaire (1.3) est donnée par :

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

## 1.1.2 MSE et MISE

### 1.1.2.1 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  $\theta$  est donnée par :

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

**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  $\theta$  pour lequel la densité  $f$  est continue, c'est-à-dire :

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

### 1.1.2.2 Convergence au sens de MISE

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

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

**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 de carré intégrable, alors l'estimateur  $\widehat{f}_\nu(\theta)$  est consistant en moyenne quadratique intégrée, c'est-à-dire :

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

### 1.1.3 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.

### 1.1.4 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  $\theta$  et de paramètre de concentration  $\rho$  ; 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 1.1 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 \[2017a\]](#)) :

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

où  $\rho$  est le paramètre de concentration vérifiant  $\rho = \rho(n) \rightarrow 1$  quand  $n \rightarrow \infty$  et  $\theta$  est l'angle où on veut estimer la densité circulaire. Cet estimateur peut être interprété

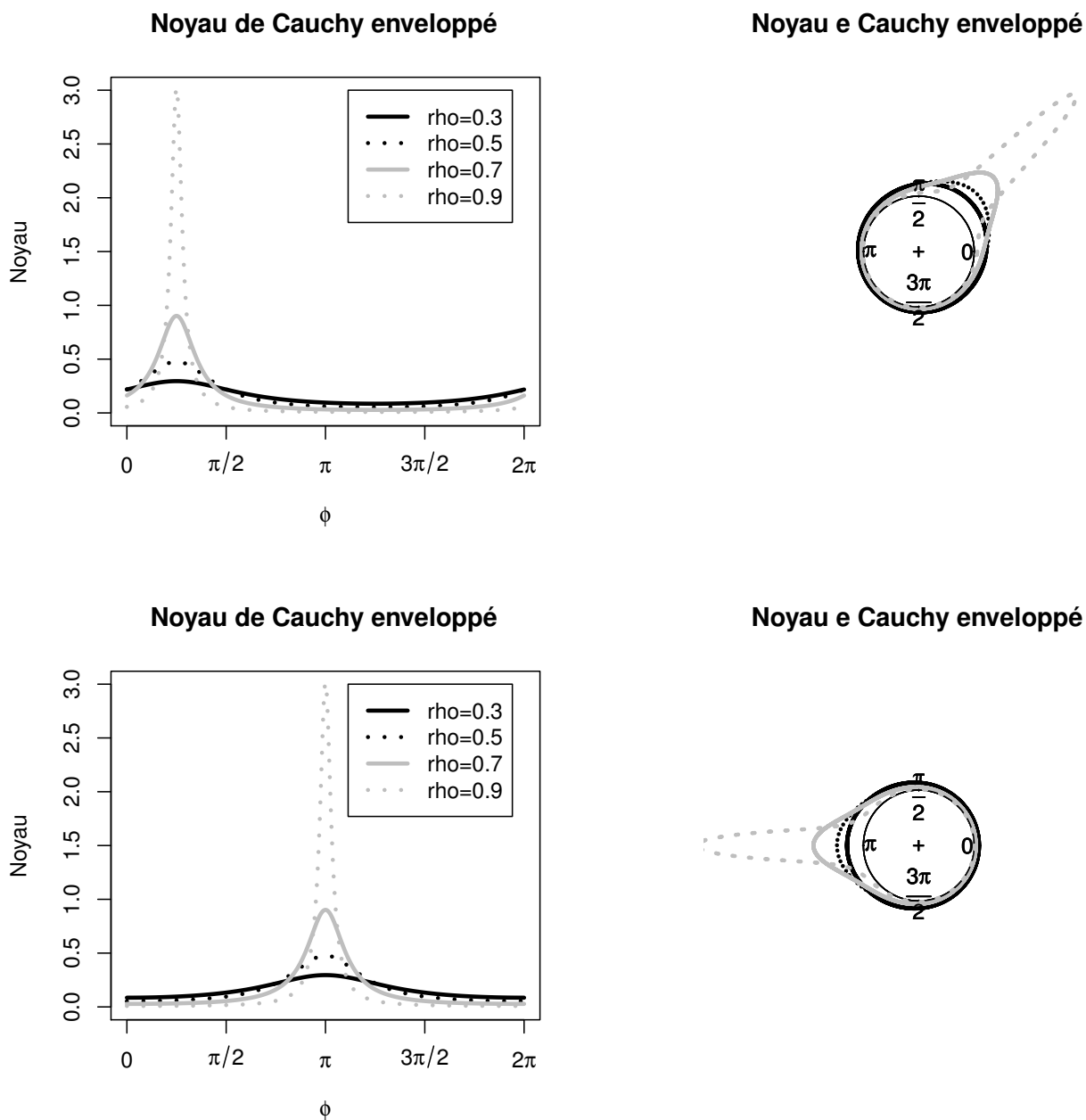


FIGURE 1.1: Noyau de Cauchy enveloppé pour  $\theta = \pi/4$  (première rangée cas linéaire) et  $\theta = \pi$  (deuxième rangée cas circulaire) avec différentes valeurs de  $\rho = \{0.3, 0.5, 0.7, 0.9\}$ .

comme un mélange de  $n$  distributions de Cauchy enveloppé, centré en  $\Theta_i$  et avec le même paramètre de concentration  $\rho$ .

#### 1.1.4.1 Espérance de l'estimateur à noyau de Cauchy enveloppé

Le noyau WC vérifie  $\gamma_j(\rho) = \rho^j$  et  $\eta_2(K_\rho) = \frac{1-\rho^2}{2}$ .

Puisque pour des petites valeurs de  $u$ , on a  $\sin(u) \simeq u$ , nous utilisons le développement suivant :

$$f(\theta + u) \simeq f(\theta) + f'(\theta) \sin(u) + \frac{1}{2} f''(\theta) \sin^2(u) + O(\sin^3(u)).$$

On obtient alors :

$$\begin{aligned} \mathbb{E}_f[K_\rho(\theta - Y)] &= \int K_\rho(\theta - y) f(y) dy \\ &= \int K_\rho(u) f(\theta + u) du \\ &= \int K_\rho(u) \left[ f(\theta) + f'(\theta) \sin(u) + \frac{1}{2} f''(\theta) \sin^2(u) + O(\sin^3(u)) \right] du \\ &= f(\theta) + \frac{1}{2} \eta_2(K_\rho) f''(\theta) + o(1) \\ \mathbb{E}[\widehat{f}_\rho^{WC}(\theta)] &= f(\theta) + \frac{1}{4} (1 - \rho^2) f''(\theta) + o(1). \end{aligned}$$

#### 1.1.4.2 Variance de l'estimateur à noyau de Cauchy enveloppé

La variance de l'estimateur à noyau de Cauchy enveloppé est donnée par :

$$\text{Var} \left( \frac{1}{n} \sum_{i=1}^n K_\rho(\theta - Y_i) \right) = \frac{1}{n} \text{Var} (K_\rho(\theta - Y))$$

$$\text{Var} (K_\rho(\theta - Y)) = \mathbb{E} [K_\rho^2(\theta - Y)] - (\mathbb{E} [K_\rho(\theta - Y)])^2$$

$$\mathbb{E} [K_\rho^2(\theta - Y)] = \int K_\rho^2(\theta - y) f(y) dy$$

$$\approx f(\theta) \int K_\rho^2(u) du = f(\theta) R(K_\rho).$$

Où

$$R(K_\rho) = \int_0^{2\pi} K_\rho^2(u) du = \frac{1 + \rho^2}{2\pi(1 - \rho^2)}$$

$$\mathbb{E}[K_\rho(\theta - Y)] \approx f(\theta)$$

$$\text{Var}(K_\rho(\theta - Y)) \approx f(\theta)R(K_\rho) - f(\theta)^2$$

$$\text{Var}\left(\frac{1}{n}\sum_{i=1}^n K_\rho(\theta - Y_i)\right) \approx \frac{1}{n} [f(\theta)R(K_\rho) - f(\theta)^2]$$

$$\mathbb{V}[\widehat{f}_\rho^{WC}(\theta)] \approx \frac{1}{n} [f(\theta)R(K_\rho) - f(\theta)^2]$$

En ce qui concerne la quantité  $R(K_\rho)$ , on a :

$$\begin{aligned} R(K_\rho) &= \frac{1 + 2\sum_{j=1}^{\infty}(\rho^j)^2}{2\pi} \\ &= \frac{1 + \frac{2\rho^2}{1-\rho^2}}{2\pi} \\ &= \frac{1 + \rho^2}{2\pi(1 - \rho^2)} \\ &= \frac{1}{\pi(1 - \rho^2)} - \frac{1}{2\pi}. \end{aligned}$$

### 1.1.4.3 Biais de l'estimateur à noyau de Cauchy enveloppé

Pour calculer le biais de l'estimateur de Cauchy enveloppé donné par (1.10), il suffit de remplacer  $1 - \gamma_2(\nu)$  par  $1 - \rho^2$  (voir [Tsuruta and Sagae \[2017a\]](#) pour plus de détails), on obtient ainsi :

$$\begin{aligned} \text{biais}[\widehat{f}_\rho^{WC}(\theta)] &= \mathbb{E}(\widehat{f}(\theta)) - f(\theta) \\ &= f(\theta) + \frac{f''(\theta)}{4}(1 - \rho^2) + o(1) - f(\theta) \\ &= \frac{f''(\theta)}{4}(1 - \rho^2) + o(1). \end{aligned}$$

Notons que l'estimateur  $\widehat{f}_\rho^{WC}$  est asymptotiquement sans biais puisque

$$\lim_{\rho \rightarrow 1} \text{biais}[\widehat{f}_\rho^{WC}(\theta)] = 0$$

#### 1.1.4.4 Convergence de l'estimateur de Cauchy enveloppé au sens de MSE

Par l'utilisation des expressions asymptotiques du biais et de la variance, la (MSE) de l'estimateur de Cauchy enveloppé (1.10) est donnée par :

$$\begin{aligned} \text{MSE}[\widehat{f}_\nu(\theta)] &= \text{biais}^2[\widehat{f}_\nu(\theta)] + \mathbb{V}[\widehat{f}_\nu(\theta)] \\ &= \frac{[f''(\theta)]^2}{16} [(1 - \rho^2)]^2 + \frac{1}{n\pi(1 - \rho^2)} f(\theta) + o(1) \end{aligned} \quad (1.11)$$

**Théorème 1.4.** Soit  $\widehat{f}_\nu(\theta)$  l'estimateur de densité circulaire par noyau de Cauchy enveloppé, défini en (1.10), avec un paramètre  $\rho = \rho_n \in (0, 1)$  dépendant de l'échantillon. Si les conditions suivantes sont satisfaites :

- $(1 - \rho_n^2) \rightarrow 0$  lorsque  $n \rightarrow \infty$ ,
- $n(1 - \rho_n^2) \rightarrow \infty$  lorsque  $n \rightarrow \infty$ ,

alors l'estimateur  $\widehat{f}_\nu(\theta)$  est convergent au sens de l'erreur quadratique moyenne :

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

#### 1.1.4.5 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 (1.10) est donnée par :

$$\text{MISE}[\widehat{f}_\rho^{WC}(\theta)] = \frac{\{1 - \rho^2\}^2}{16} \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{1}{n\pi\{1 - \rho^2\}} + o(\{1 - \rho^2\}^2 + n^{-1}\{1 - \rho^2\}). \quad (1.12)$$

**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  $\widehat{f}_\rho^{WC}$  est consistant en moyenne quadratique intégrée, c'est-à-dire :

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

#### 1.1.5 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, de cible  $\theta$  et de paramètre de concentration  $\nu$  ; 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 d'ordre zéro (voir annexe (3.8)). La Figure 1.2 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]) :

$$\hat{f}_\nu^{VM}(\theta) = \frac{1}{n(2\pi)I_0(\nu)} \sum_{i=1}^n \exp\{\nu \cos(\theta - \Theta_i)\}, \quad 0 \leq \theta < 2\pi. \quad (1.13)$$

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

$$\gamma_j(\nu) = \frac{I_j(\nu)}{I_0(\nu)}. \quad (1.14)$$

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

### 1.1.5.1 Biais de l'estimateur de Von Mises

Afin de déterminer le biais de l'estimateur de von Mises donné par (1.13), il suffit de remplacer  $\gamma_2(\nu)$  par son expression donnée par (1.14), on obtient ainsi :

$$\text{biais}[\hat{f}_\nu^{VM}(\theta)] = \frac{f''(\theta)}{4} \left(1 - \frac{I_2(\nu)}{I_0(\nu)}\right) + o(1). \quad (1.15)$$

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 :

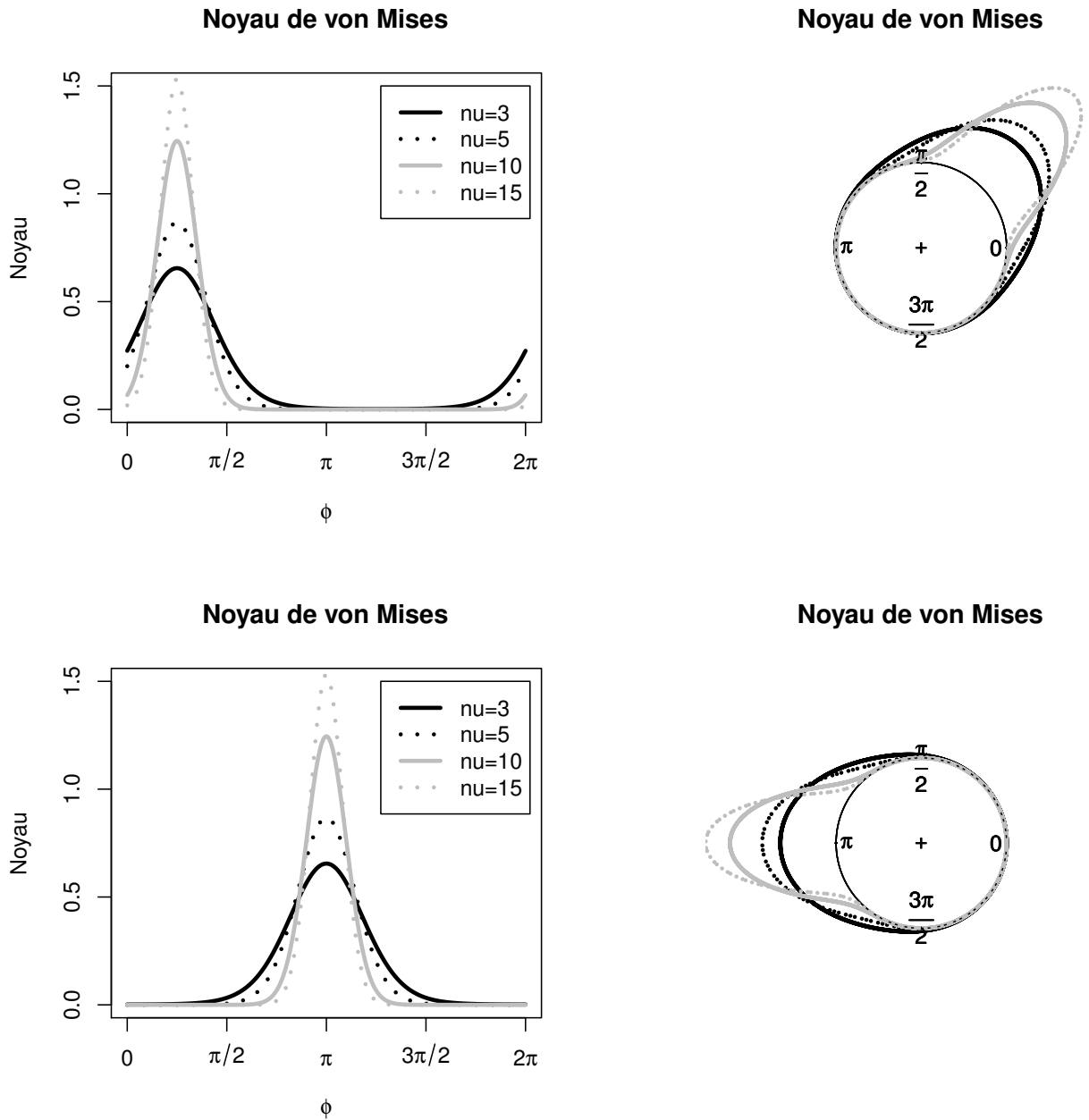


FIGURE 1.2: Noyau de von Mises pour  $\theta = \pi/4$  (première rangée cas linéaire) et  $\theta = \pi$  (deuxième rangée cas circulaire) avec différentes valeurs de  $\nu = \{3, 5, 10, 15\}$ .

$$\text{biais}[\widehat{f}_\nu^{VM}(\theta)] = \frac{f''(\theta)}{4\nu} + o(\nu^{-1}). \quad (1.16)$$

Notons que l'estimateur  $\widehat{f}_\nu^{VM}$  est asymptotiquement sans biais car

$$\lim_{\nu \rightarrow \infty} \text{biais}[\widehat{f}_\nu^{VM}(\theta)] = 0$$

.

### 1.1.5.2 Variance de l'estimateur de Von Mises

Pour obtenir la variance de l'estimateur de Von Mises donné par (1.13), 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 :

$$\mathbb{V}[\widehat{f}_\nu^{VM}(\theta)] = \frac{I_0(2\nu)}{2n\pi\{I_0(\nu)\}^2} f(\theta) + o(1). \quad (1.17)$$

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 :

$$\mathbb{V}[\widehat{f}_\nu^{VM}(\theta)] = \frac{\nu^{1/2}}{2n\sqrt{\pi}} f(\theta) + o\left(\frac{\nu^{1/2}}{n}\right). \quad (1.18)$$

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

### 1.1.5.3 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 (1.13) est donnée par :

$$\text{MISE}[\widehat{f}_\nu^{VM}(\theta)] = \frac{1}{16\nu^2} \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{\nu^{1/2}}{2n\sqrt{\pi}} + o\left(\frac{1}{\nu^2} + \frac{\nu^{1/2}}{n}\right). \quad (1.19)$$

**Théorème 1.6.** *Si  $\lim_{n \rightarrow \infty} \nu(n) = \infty$ ,  $\lim_{n \rightarrow \infty} \nu^{1/2}(n)/n = 0$  et  $f''$  est continue et de carré intégrable, alors l'estimateur  $\widehat{f}_\nu^{VM}$  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 Conclusion

Dans ce chapitre, nous avons présenté l'approche non paramétrique de l'estimateur à noyau circulaire de la densité circulaire, en particulier les estimateurs à noyaux de Cauchy enveloppé et de von Mises. Nous avons aussi exposé les principaux résultats théoriques concernant la convergence de ces estimateurs, notamment leurs propriétés asymptotiques ainsi que leur convergence au sens du MSE et du MISE.

Remarquons que l'estimateur à noyau circulaire dépend crucialement du paramètre de concentration. Dans le chapitre suivant, nous rappellerons les méthodes classiques de sélection de ce paramètre, à savoir les méthodes de réinjection (plug-in) et les méthodes de validation croisée.

## Chapitre 2

---

# Choix du paramètre de concentration $\rho$ dans l'estimation non-paramétrique de la densité par noyau circulaire

---

Le choix du paramètre de lissage joue un rôle crucial dans l'estimation de la densité par la méthode du noyau dans le cas de données circulaires. Dans ce chapitre, nous présentons différentes méthodes permettant de sélectionner ce paramètre dans le cas de l'estimateur à noyau de cauchy enveloppé.

### 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 :

$$\rho_{id} = \min_{\rho > 0} MISE(\rho). \quad (2.1)$$

Rappelons que l'expression asymptotique pour la MISE (AMISE) en utilisant le noyau de cauchy enveloppé est donnée comme suit (voir [Tsuruta and Sagae \[2017a\]](#)) :

$$AMISE[\hat{f}_\rho^{WC}(\theta)] = \frac{\{1 - \rho^2\}^2}{16} \int_0^{2\pi} \{f''(\theta)\}^2 d\theta + \frac{1}{n\pi\{1 - \rho^2\}} + o(\{1 - \rho^2\}^2 + n^{-1}\{1 - \rho^2\}). \quad (2.2)$$

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 cauchy enveloppé du paramètre de concentration  $\rho$ . Dans ce cas, l'expression AMISE donnée par (2.3) s'écrit comme suit :

$$AMISE(\rho) = \frac{h^2}{16} R(f'') + \frac{1}{n\pi h} \quad (2.3)$$

où  $h = 1 - \rho^2$  et  $R(f'') = \int_0^{2\pi} \{f''(\theta)\}^2 d\theta$ .

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

$$\hat{h} = \left( \frac{8}{\pi R(f'')n} \right)^{-\frac{1}{3}} \quad (2.4)$$

sous la condition que  $n > 8(\pi R(f''))^{-1}$ . Le taux de convergence optimale suivant d'AMISE de cauchy enveloppé est donné par (2.3) et (2.4) lorsque  $n$  est suffisamment grand :

$$AMISE = O(n^{-2/3}) \quad (2.5)$$

Nous devrions utiliser  $\hat{\rho}$  qui minimise (2.3) comme paramètre de concentration pratiquement optimale, puisque (2.4) est supérieur à un si  $n \leq 8(\pi R(f''))^{-1}$ .  $\hat{\rho}$  est donné (2.3)(2.4) comme suit

$$\hat{\rho} = \arg \min_{0 < \rho \leq 1} AMISE[\hat{f}(\cdot; \rho)] \quad (2.6)$$

où  $\hat{h}$  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 suit réellement une loi de cauchy enveloppée. 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 cauchy enveloppé. La densité d'un mélange fini de  $M$  densités de cauchy enveloppé  $vM(\mu_r, k_r)$  est donnée par :

$$g(\theta) = \sum_{r=1}^M \alpha_r \left( \frac{1}{2\pi} \frac{1 - k_r^2}{1 + k_r^2 - 2k_r \cos(\theta - \mu_r)} \right), \quad \text{avec} \quad \sum_{r=1}^M \alpha_r = 1. \quad (2.7)$$

La valeur approche  $\hat{\rho}_{PI}$  du paramètre de concentration  $\rho$  est obtenue en utilisant les étapes suivantes :

- **Étape 1** Sélectionner le nombre  $M$  pour la distribution  $g$  donnée par (2.7).
- **Étape 2** Estimer l'AMISE comme suit :
  - **Étape 2.1** Estimer les paramètres  $(\alpha_r, \mu_r, k_r)$  de la distribution  $g$  donnée par (2.7).
  - **Étape 2.2** Calculer l'intégrale  $\int (g''(\theta))^2 d\theta$  en remplaçant  $f$  par  $g$ .
  - **Étape 2.3** Substituer la quantité  $\int (g''(\theta))^2 d\theta$  dans l'AMISE
- **Étape 3** Minimiser  $AMISE(\rho)$  et obtenir  $\hat{h}_{PI}$ .

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  $\alpha_r, \mu_r$  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. Notons que le paramètre de concentration  $\rho_{RT}$  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  $CV(\rho)$  dont le calcul soit plus simple que la  $MISE(\rho)$ . Notons que les méthodes de

validation croisée ne font pas usage des approximations des dérivées de  $f$ . Le paramètre de concentration  $\rho$  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  $\hat{\rho}_{LCV}$ , tel que

$$\hat{\rho}_{lcV} = \max_{\rho} LCV(\rho),$$

qui maximise le critère

$$LCV(\rho) = \prod_{i=1}^n \hat{f}_{\rho,i}^{WC}(\Theta_i), \quad (2.8)$$

où

$$\hat{f}_{\rho,i}^{WC}(\Theta_i) = \frac{1}{(n-1)} \sum_{j \neq i} K_{\rho}(\Theta_i - \Theta_j), \quad (2.9)$$

est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation  $\Theta_i$ . En utilisant le noyau de Cauchy enveloppé, les étapes pour calculer le paramètre de concentration optimal  $\hat{\rho}_{LCV}$  sont données comme suit :

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

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

### 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; 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  $ISE(\rho)$  qui est donné par :

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

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

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

Remarquons que  $\frac{1}{n} \sum_{i=1}^n \widehat{f}_{\rho,i}^{WC}(\Theta_i)$  est un estimateur sans biais de  $\int \widehat{f}_\rho^{WC}(\theta) f(\theta) d\theta$ , où  $\widehat{f}_{\rho,i}^{WC}(\Theta_i)$  est l'estimateur de la densité calculé à partir de l'échantillon privé de l'observation  $\Theta_i$  donné par (2.9). Le critère à optimiser devient alors :

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

où  $K_\rho$  est le noyau de Cauchy enveloppé. Les étapes pour calculer le paramètre de concentration optimal  $\hat{\rho}_{UCV}$  sont données comme suit :

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

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 ( $\hat{\rho}_{LCV}$ ) est plus stable que la méthode de validation croisée non biaisé ( $\hat{\rho}_{UCV}$ ).

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.

### 2.2.3 Conclusion

Dans ce chapitre, nous avons présenté deux classes de méthodes pour le choix du 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  $\rho_{id} = \min_{\rho} \text{MISE}(\rho)$  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  $\text{CV}(\rho)$  dont le calcul soit plus simple que la  $\text{MISE}(\rho)$ .

## Chapitre 3

---

# Étude de Simulation

---

### 3.1 Résultats de la simulation

Le paramètre de concentration optimal dépend de  $R(f'')$ . La règle d'insertion (*plug-in rule*) est la procédure permettant d'estimer ce paramètre de concentration optimal à l'aide de  $\hat{R}(f'')$  comme estimateur de  $R(f'')$ . La règle la plus simple parmi les règles d'insertion consiste à supposer que la densité réelle  $f$  est la densité de von Mises  $f_{VM}$ . La règle d'insertion la plus simple utilise  $\hat{R}(f''_{VM}(\cdot; \hat{\kappa}))$  comme estimateur de  $R(f'')$ , où  $\kappa$  est le paramètre de concentration de  $f_{VM}$  et  $\hat{\kappa}$  est l'estimateur du maximum de vraisemblance de  $\kappa$ .  $\hat{R}(f''_{VM}(\cdot; \hat{\kappa}))$  correspond à la forme suivante :

$$\hat{R}(f''_{VM}(\cdot; \hat{\kappa})) = \frac{\hat{\kappa} [3\hat{\kappa}I_2(2\hat{\kappa}) + 2I_1(2\hat{\kappa})]}{8\pi I_0^2(\hat{\kappa})} \quad (3.1)$$

où  $\hat{\kappa}$  est l'estimateur du maximum de vraisemblance du paramètre de concentration  $\kappa$  de la loi de von Mises. Cette estimation permet ensuite de déterminer le paramètre de lissage optimal : soit  $\rho^*$  pour le noyau WC, soit  $\kappa^*$  pour le noyau VM.

La procédure standard de la règle plug-in est décrite comme suit :

1. Estimer  $\hat{\kappa}$  à partir de l'échantillon, puis calculer  $\hat{R}(f''_{VM}(\cdot; \hat{\kappa}))$  à partir de  $\hat{\kappa}$ .
2. Noyau WC : estimer  $\rho^*$  qui minimise (2.3) en remplaçant  $R(f'')$  par  $\hat{R}(f''_{VM}(\cdot; \hat{\kappa}))$ .
3. Noyau VM : estimer  $\kappa^*$  en remplaçant  $R(f'')$  dans (2.2) par  $\hat{R}(f''_{VM}(\cdot; \hat{\kappa}))$ .

Les deux expériences de simulation sont définies comme suit :

### 3.1.1 Expérience A : mélange de densités de Von Mises

#### 1. Pour le noyau WC :

(a) Supposons que la densité réelle est un mélange de deux lois de Von Mises :

$$f(\theta)_{MVVM} = \frac{1}{2}f_{VM1}(\theta; \mu_1, \kappa_1) + \frac{1}{2}f_{VM2}(\theta; \mu_2, \kappa_2), \quad (3.2)$$

où  $\mu_1 = \pi/2$ ,  $\mu_2 = 3\pi/2$ ,  $\kappa_1 = \kappa_2 = \kappa$ . Générer un échantillon de taille  $n$  selon cette densité.

(b) Estimer  $\rho^*$  à l'aide de la règle plug-in.

(c) Soit  $ISE$  l'erreur quadratique intégré donné par :

$$ISE(\hat{f}(\cdot; \rho)) = \int_0^{2\pi} \left( \hat{f}(\theta; \rho) - f(\theta) \right)^2 d\theta \quad (3.3)$$

Calculer  $ISE(\hat{f}(\cdot; \rho))$ .

(d) Répéter les étapes (a)–(c) 1000 fois et calculer :

$$MISE(\hat{f}(\cdot; \rho)) = \frac{1}{1000} \sum_{i=1}^{1000} ISE_i(\hat{f}(\cdot; \rho)) \quad (3.4)$$

#### 2. Pour le noyau VM :

Avec les mêmes procédures que (a) - (d) du noyau de cauchy enveloppé , calculer  $MISE(\hat{f}(\cdot; k^*))$ ,

#### 3. Comparer :

$$MISE(\hat{f}^{VM}(\cdot; \kappa^*)) - MISE(\hat{f}^{WC}(\cdot; \rho^*)), \quad (3.5)$$

en arrondissant au quatrième chiffre après la virgule.

### 3.1.2 Expérience B : densité de Cauchy enveloppée

1. Supposons que la densité réelle est une densité de Cauchy enveloppée :

$$f_{WC}(\theta) = \frac{1}{2\pi} \cdot \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}, \quad (3.6)$$

avec  $\mu = \pi$ .

2. Calculer la différence :

$$\text{MISE}(\hat{f}^{VM}(\cdot; \kappa^*)) - \text{MISE}(\hat{f}^{WC}(\cdot; \rho^*)), \quad (3.7)$$

en suivant les mêmes étapes que pour l'Expérience A.

TABLE 3.1: Différence  $\text{MISE}(\hat{f}^{VM}(\cdot; \kappa^*)) - \text{MISE}(\hat{f}^{WC}(\cdot; \rho^*))$  pour différentes valeurs de  $\kappa$  et tailles  $n$

$\kappa \backslash n$	10	50	100	200	500	1000
0.3	0.011	0.002	0.001	0	0	0
0.5	0.012	0.002	0.001	0	0	0
0.7	0.011	0.002	0.001	0	0	0
1	0.012	0.001	0	0	0	0
2	0.007	-0.004	-0.005	-0.005	-0.005	-0.005
5	-0.017	-0.023	-0.025	-0.024	-0.024	-0.023
10	-0.032	-0.036	-0.037	-0.037	-0.037	-0.037
15	-0.037	-0.044	-0.043	-0.042	-0.042	-0.042
20	-0.039	-0.045	-0.045	-0.045	-0.048	-0.046

TABLE 3.2: Différence  $\text{MISE}(\hat{f}^{VM}(\cdot; \kappa^*)) - \text{MISE}(\hat{f}^{WC}(\cdot; \rho^*))$  pour différentes valeurs de  $\rho$

$\rho \backslash n$	50	100	200	500	1000
0.5	-0.015	-0.009	-0.006	-0.003	-0.002
0.7	-0.011	-0.006	-0.003	-0.001	-0.001
0.8	0.007	0.012	0.013	0.011	0.009
0.9	0.133	0.141	0.134	0.112	0.093

Les résultats du Tableau 3.1 montrent que :

- Le noyau WC est supérieur au noyau VM lorsque  $n$  n'est pas suffisamment grand ( $n \leq 100$ ) et que la densité  $f$  est un mélange de lois de von Mises (cas multimodal) avec  $\kappa \leq 2$ .
- Cette différence entre le noyau WC et le noyau VM tend à diminuer à mesure que  $n$  augmente.

Les résultats du Tableau 3.2 montrent que :

- Le noyau WC est meilleur que le noyau VM lorsque  $n \leq 1000$ , que la densité  $f$  suit une loi de Cauchy enveloppée (avec queue lourde), et que  $\rho \geq 0,8$ .
- En particulier, le noyau WC présente un bien meilleur comportement autour du sommet de  $f$  que le noyau VM, même lorsque  $n$  est grand.

Les représentations graphiques des densités circulaires et linéaires des estimateurs von Mises et Wrapped Cauchy sont également données par les Figures 3.1, 3.2.

- Les Figures 3.1 et 3.2 montrent que pour les faibles valeurs de  $\kappa$ , le noyau Wrapped Cauchy est plus performant sur les petits échantillons. Lorsque  $\kappa$  augmente, le noyau de Von Mises devient préférable. L'écart entre les deux méthodes diminue avec l'augmentation de la taille  $n$ .
- La densité de Von Mises est aplatie pour  $\kappa = 0.3$  (faible concentration), plus resserrée pour  $\kappa = 2$ , et très pointue pour  $\kappa = 10$ . Cela illustre comment le paramètre

- $\kappa$  contrôle la concentration autour de la direction moyenne  $\mu$ .
- Les Figures 3.1 et 3.2 montrent également que pour  $\rho \leq 0.7$ , le noyau de Von Mises surpasse significativement le noyau de Cauchy enveloppé. Lorsque  $\rho \geq 0.8$ , les rôles s'inversent, en particulier pour  $\rho = 0.9$ , où la *MISE* du noyau de Cauchy enveloppé devient nettement meilleure.
  - Pour  $\rho = 0.5$ , la densité est très aplatie (faiblement concentrée); à  $\rho = 0.7$ , elle devient modérément concentrée; et à  $\rho = 0.9$ , elle est très concentrée autour de  $\mu$ . Cela montre que  $\rho$  joue un rôle analogue à  $\kappa$  dans le cas des lois de von Mises.
  - Lorsque la densité réelle est de type Wrapped Cauchy (fortement concentrée), le noyau WC devient supérieur, notamment pour les faibles valeurs de  $\rho$ .

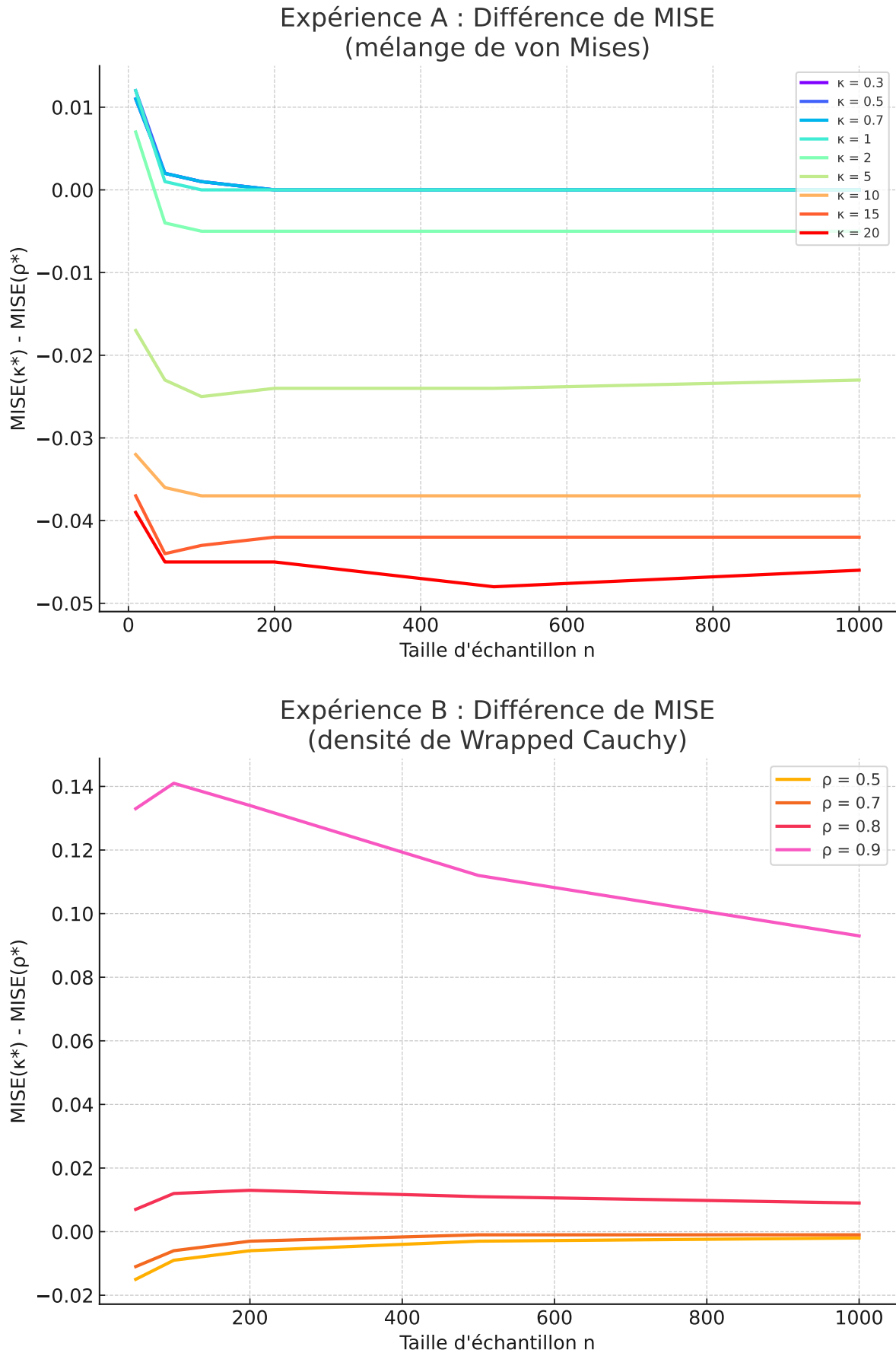


FIGURE 3.1: Comparaison graphique des erreurs quadratiques moyennes intégrées (MISE) pour les expériences A et B.

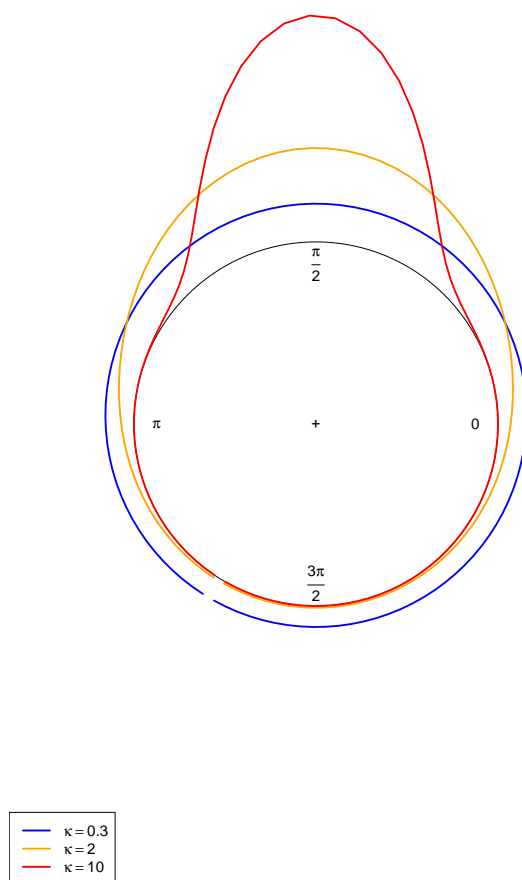
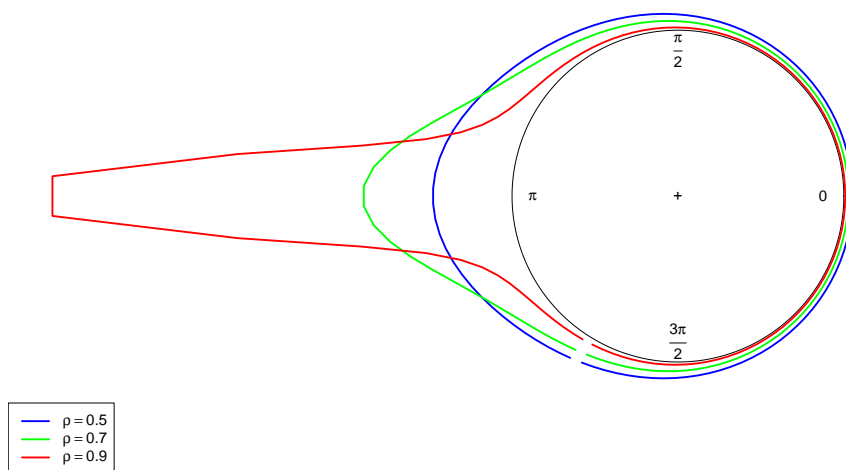
Expérience A : Densités de von Mises ( $\mu = \pi/2$ )Expérience B : Densités Wrapped Cauchy ( $\mu = \pi$ )

FIGURE 3.2: Représentation des densités circulaires dans les expériences A et B

## 3.2 Conclusion

Dans ce chapitre, nous avons étudié les performances asymptotiques de deux noyaux circulaires : le noyau de Cauchy enveloppé (Wrapped Cauchy, WC) et le noyau de von Mises (VM).

Le taux de convergence du noyau WC s'explique par (AMISE), qui correspond à celle du noyau de Cauchy classique. Cela justifie la comparaison directe entre les deux cadres linéaire et circulaire.

En définitive, bien que le noyau VM reste supérieur en termes théoriques de MISE asymptotique, nos résultats de simulation suggèrent que le noyau WC offre de meilleures performances dans les cas où la densité réelle est multimodale et présente des queues lourdes. Il apparaît donc comme un choix pertinent et robuste dans des contextes pratiques complexes.

---

# Conclusion générale et perspectives

---

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 simulées. 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  $f''$ .

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. Le noyau WC est plus adapté aux densités peu concentrées ou multimodales, en particulier pour de petits échantillons ;

2. Le noyau VM devient plus performant pour les densités fortement concentrées ou unimodales, notamment lorsque  $\rho$  est élevé ;
3. La méthode plug-in, bien qu'elle repose sur une approximation via une loi de von Mises, fournit des résultats fiables pour les deux noyaux, à condition que la forme de la densité réelle soit bien capturée ;
4. Le choix du noyau doit donc être guidé à la fois par la structure de la densité et la taille de l'échantillon disponible.

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 [Bedouhene and Zougab \[2020\]](#)).
2. Étudier l'estimateur à noyau circulaire de la fonction de régression.

---

# 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 von Mises (VM), Cauchy enveloppé (WC) ; 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 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 :

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa \cos \theta\} d\theta \tag{3.8}$$

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.

2. La fonction de densité de la distribution de Cauchy enveloppé 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 :

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

La distribution de Cauchy enveloppé 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  $\mu$ .

**Théorème 3.1** (Identité de Parseval). *Soit  $f$  une fonction  $2\pi$ -périodique et carrée-intégrable sur  $[-\pi, \pi]$ , admettant une série de Fourier complexe : Soit  $f \in L^2([-\pi, \pi])$  une fonction réelle et  $2\pi$ -périodique, admettant un développement en série de Fourier de la forme :*

$$f(\theta) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(k\theta) + b_k \sin(k\theta)),$$

alors l'identité de Parseval s'écrit :

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f^2(\theta) d\theta = \frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2).$$

## 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 la méthode plug-in .

### A.2.1 Étude de simulation : performance des estimateurs à noyaux circulaires (VM et WC).

Cette étude présente une **comparaison entre deux estimateurs de densité circulaire** basés sur des noyaux : le *noyau de von Mises (VM)* et le *noyau Wrapped Cauchy (WC)*.

Deux expériences de simulation sont conduites :

- **Expérience A** : les données sont générées à partir d'un *mélange de deux densités Von Mises*, ce qui permet d'évaluer les performances des estimateurs en situation bimodale.

- **Expérience B** : la densité réelle est une *densité de Cauchy enroulée*, permettant d'analyser les comportements en cas de forte concentration.

L'objectif est d'évaluer et de comparer les performances des deux estimateurs à travers le critère de l'**erreur quadratique moyenne intégrée (MISE)**, en fonction des paramètres de concentration et de la taille de l'échantillon.

```
#=====
# Comparaison des estimateurs de densité circulaire : noyaux VM vs WC
# Simulation pour les expériences A (mélange de von Mises) et B (densité WC)
#=====

library(CircStats)
library(circular)
library(pracma)

# ----- Fonctions communes -----
#=====
# Densité von Mises simple et mélange
#=====
von_mises <- function(theta, mu, kappa) {
  (1 / (2 * pi * bessell(kappa, 0))) * exp(kappa * cos(theta - mu))
}
dvm_mix <- function(theta, mu1 = pi/2, mu2 = 3*pi/2, kappa) {
  0.5 * von_mises(theta, mu1, kappa) + 0.5 * von_mises(theta, mu2, kappa)
}
#=====
# Densité Wrapped Cauchy
#=====
```

```
wrapped_cauchy <- function(theta, mu = pi, rho = 0.9) {
  (1 / (2 * pi)) * (1 - rho^2) / (1 + rho^2 - 2 * rho * cos(theta - mu))
}

#=====
# Estimateur de densité
#=====

wc_kernel <- function(theta, data, rho) {
  sapply(theta, function(t) mean(((1 - rho^2) /
    (2 * pi * (1 + rho^2 - 2 * rho * cos(t - data))))))
}

vm_kernel <- function(theta, data, kappa) {
  sapply(theta, function(t) mean(dvonmises(t, data, kappa)))
}

#=====
# ISE
#=====

compute_ISE <- function(f_est, f_true, theta_grid) {
  diff2 <- (f_est - f_true)^2
  trapz(theta_grid, diff2)
}

#=====
# Estimation de R(f'') pour von Mises
#=====

R_hat_f2_VM <- function(kappa_hat) {
  num <- kappa_hat * (3 * kappa_hat * besseli(2 * kappa_hat, 1)
    + 2 * besseli(2 * kappa_hat, 1))
  den <- 8 * pi * (besseli(kappa_hat, 0))^2
  num / den
}
```

```
#=====
# Estimation de kappa (approchée)
#=====
estimate_kappa <- function(sample) {
  Rbar <- rho.circular(sample)
  if (Rbar < 0.53) {
    2 * Rbar + Rbar^3 + (5 * Rbar^5) / 6
  } else if (Rbar < 0.85) {
    -0.4 + 1.39 * Rbar + 0.43 / (1 - Rbar)
  } else {
    1 / (Rbar^3 - 4 * Rbar^2 + 3 * Rbar)
  }
}

#=====
# ----- Expérience A -----
#=====
simulate_experience_A <- function(kappa_vals, n_vals, n_sim = 1000,
theta_grid = seq(0, 2*pi, length.out = 500)) {
  results <- matrix(0, nrow = length(kappa_vals), ncol = length(n_vals))
  rownames(results) <- paste0("kappa=", kappa_vals)
  colnames(results) <- paste0("n=", n_vals)

  for (i in seq_along(kappa_vals)) {
    kappa <- kappa_vals[i]
    for (j in seq_along(n_vals)) {
      n <- n_vals[j]
      ise_wc <- ise_vm <- numeric(n_sim)

      for (s in 1:n_sim) {
        z <- rbinom(n, 1, 0.5)
```

```

mu <- ifelse(z == 1, pi/2, 3*pi/2)
data <- rvonmises(n, mu, kappa)

f_true_vals <- dvm_mix(theta_grid, kappa = kappa)

best_rho <- which.min(sapply(seq(0.1, 0.99, by = 0.01), function(rho) {
  f_est <- wc_kernel(theta_grid, data, rho)
  compute_ISE(f_est, f_true_vals, theta_grid)
}))

best_kappa <- which.min(sapply(seq(0.1, 25, by = 0.5), function(k) {
  f_est <- vm_kernel(theta_grid, data, k)
  compute_ISE(f_est, f_true_vals, theta_grid)
}))

f_wc <- wc_kernel(theta_grid, data, seq(0.1, 0.99, by = 0.01)[best_rho])
f_vm <- vm_kernel(theta_grid, data, seq(0.1, 25, by = 0.5)[best_kappa])

ise_wc[s] <- compute_ISE(f_wc, f_true_vals, theta_grid)
ise_vm[s] <- compute_ISE(f_vm, f_true_vals, theta_grid)
}

results[i, j] <- round(mean(ise_vm) - mean(ise_wc), 4)
}
}

return(results)
}
#=====
# ----- Expérience B -----

```

```
#=====
simulate_experience_B <- function(rho_vals, n_vals, n_sim = 1000, theta_grid =
seq(0, 2*pi, length.out = 500)) {
  results <- matrix(0, nrow = length(rho_vals), ncol = length(n_vals))
  rownames(results) <- paste0("rho=", rho_vals)
  colnames(results) <- paste0("n=", n_vals)

  for (i in seq_along(rho_vals)) {
    rho <- rho_vals[i]
    for (j in seq_along(n_vals)) {
      n <- n_vals[j]
      ise_wc <- ise_vm <- numeric(n_sim)

      for (s in 1:n_sim) {
        data <- rvonmises(n, pi, 1) # données WC simulées à partir
        d'une approximation von Mises si pas de fonction directe
        f_true_vals <- wrapped_cauchy(theta_grid, mu = pi, rho = rho)

        best_rho <- which.min(sapply(seq(0.1, 0.99, by = 0.01), function(r) {
          f_est <- wc_kernel(theta_grid, data, r)
          compute_ISE(f_est, f_true_vals, theta_grid)
        })))

        best_kappa <- which.min(sapply(seq(0.1, 25, by = 0.5), function(k) {
          f_est <- vm_kernel(theta_grid, data, k)
          compute_ISE(f_est, f_true_vals, theta_grid)
        })))

        f_wc <- wc_kernel(theta_grid, data, seq(0.1, 0.99, by = 0.01)[best_rho])
        f_vm <- vm_kernel(theta_grid, data, seq(0.1, 25, by = 0.5)[best_kappa])
      }
    }
  }
}
```

```
ise_wc[s] <- compute_ISE(f_wc, f_true_vals, theta_grid)
ise_vm[s] <- compute_ISE(f_vm, f_true_vals, theta_grid)
}

results[i, j] <- round(mean(ise_vm) - mean(ise_wc), 4)
}
}

return(results)
}
#=====
# Exemple d'utilisation
#=====
res_A <- simulate_experience_A(kappa_vals = c(0.3, 1, 5),
  n_vals = c(50, 100), n_sim = 100)
res_B <- simulate_experience_B(rho_vals = c(0.5, 0.9),
  n_vals = c(100, 500), n_sim = 100)

#=====
##### Représentation graphique linéaire et ccirculaire
#=====
library(circular)
library(qpdf)

# === Figure 1 : Différence de MISE - Expérience A ===
pdf("fig_experienceA.pdf", width = 8, height = 6)
```

```
kappa <- c(0.3, 0.5, 0.7, 1, 2, 5, 10, 15, 20)
n_values_A <- c(10, 50, 100, 200, 500, 1000)
mise_diff_A <- matrix(c(
  0.011, 0.002, 0.001, 0, 0, 0,
  0.012, 0.002, 0.001, 0, 0, 0,
  0.011, 0.002, 0.001, 0, 0, 0,
  0.012, 0.001, 0, 0, 0, 0,
  0.007, -0.004, -0.005, -0.005, -0.005, -0.005,
  -0.017, -0.023, -0.025, -0.024, -0.024, -0.023,
  -0.032, -0.036, -0.037, -0.037, -0.037, -0.037,
  -0.037, -0.044, -0.043, -0.042, -0.042, -0.042,
  -0.039, -0.045, -0.045, -0.045, -0.048, -0.046
), nrow = 9, byrow = TRUE)

matplot(n_values_A, t(mise_diff_A), type = "l", lty = 1, lwd = 2,
        col = rainbow(9), xlab = "Taille d'échantillon n", ylab = "MISE(?*) -
MISE(?*)",
        main = "Expérience A : Différence de MISE\n(mélange de von Mises)")
legend("topright", legend = paste("? =", kappa), col = rainbow(9),
lty = 1, lwd = 2)

dev.off()

# === Figure 2 : Différence de MISE - Expérience B ===
pdf("fig_experienceB.pdf", width = 8, height = 6)

rho <- c(0.5, 0.7, 0.8, 0.9)
n_values_B <- c(50, 100, 200, 500, 1000)
mise_diff_B <- matrix(c(
  -0.015, -0.009, -0.006, -0.003, -0.002,
```

```
-0.011, -0.006, -0.003, -0.001, -0.001,
  0.007,  0.012,  0.013,  0.011,  0.009,
  0.133,  0.141,  0.134,  0.112,  0.093
), nrow = 4, byrow = TRUE)

colors_B <- c("blue", "green", "orange", "red")
matplot(n_values_B, t(mise_diff_B), type = "l", lty = 1, lwd = 2,
        col = colors_B,
        xlab = "Taille d'échantillon n", ylab = "MISE(?*) - MISE(?*)",
        main = "Expérience B : Différence de MISE\n(densité de Wrapped Cauchy)")
legend("topright", legend = paste("? =", rho), col = colors_B,
       lty = 1, lwd = 2)

dev.off()

# === Figure 3 : Densités circulaires von Mises ===
pdf("fig_circular_expA.pdf", width = 8, height = 8)

angles <- seq(0, 2 * pi, length.out = 100)
kappa_vm <- c(0.3, 2, 10)
colors_vm <- c("blue", "orange", "red")

plot.circular(circular(0), xlim = c(-1.5, 1.5), ylim = c(-1.5, 1.5),
              main = "Expérience A - Densité von Mises", type = "n")

for (i in 1:length(kappa_vm)) {
  density <- dvonmises(circular(angles), mu = circular(pi / 2),
                      kappa = kappa_vm[i])
  lines.circular(circular(angles), density, col = colors_vm[i], lwd = 2)
}
```

```
legend("bottomleft", legend = paste("? =", kappa_vm),
col = colors_vm, lty = 1, lwd = 2)

dev.off()

# === Figure 4 : Densités circulaires Wrapped Cauchy ===
pdf("fig_circular_expB.pdf", width = 8, height = 8)

rho_wc <- c(0.5, 0.7, 0.9)
colors_wc <- c("blue", "green", "red")

plot.circular(circular(0), xlim = c(-1.5, 1.5), ylim = c(-1.5, 1.5),
              main = "Expérience B - Densité Wrapped Cauchy", type = "n")

density_wc <- function(theta, mu, rho) {
  (1 / (2 * pi)) * (1 - rho^2) / (1 + rho^2 - 2 * rho * cos(theta - mu))
}

for (i in 1:length(rho_wc)) {
  dens <- density_wc(angles, mu = pi, rho = rho_wc[i])
  lines.circular(circular(angles), dens, col = colors_wc[i], lwd = 2)
}

legend("bottomleft", legend = paste("? =", rho_wc),
col = colors_wc, lty = 1, lwd = 2)

dev.off()

# === Fusion et compression en ZIP ===
pdfs <- c("fig_experienceA.pdf", "fig_experienceB.pdf", "fig_circular_expA.pdf",
          "fig_circular_expB.pdf")
zip("figures_circulaires.zip", files = pdfs)
```



---

# Bibliographie

---

- 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.
- A W. Bowman. An Alternative Method of Cross-validation for the Smoothing of Density Estimates. *Biometrika*, 71 :353–360, 1984.
- 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.
- N I. Fisher. *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge, U.K. 1993.

- 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.
- P. Hall, J S. Marron, and B U. Park. Smoothed cross validation. *Probability Theory and Related Fields*, 92 :1–20, 1992.
- 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.*
- 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.
- J S. Rao and A. Sengupta. *Topics in circular statistics*. Singapore : World Scientific Publishing. 2001.
- 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.
- 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.
- 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.
- 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.