Electronic Journal of Statistics
Vol. 16 (2022) 3482–3522
ISSN: 1935-7524
https://doi.org/10.1214/22-EJS2024

Semiparametric inference for mixtures of circular data

Claire Lacour

LAMA, Univ Gustave Eiffel, Univ Paris Est Creteil, CNRS, F-77447 Marne-la-Vallée, France
e-mail: claire.lacour@univ-eiffel.fr

and

Thanh Mai Pham Ngoc

Université Paris-Saclay, CNRS, Laboratoire de mathématiques d'Orsay, 91405 Orsay, France
e-mail: thanh.pham_ngoc@math.u-psud.fr

Abstract: We consider $X_1, \ldots, X_n$ a sample of data on the circle $\mathbb{S}^1$, whose distribution is a two-component mixture. Denoting R and Q two rotations on $\mathbb{S}^1$, the density of the X_i's is assumed to be $g(x) = pf(R^{-1}x) + (1-p)f(Q^{-1}x)$, where $p \in (0,1)$ and f is an unknown density on the circle. In this paper we estimate both the parametric part $\theta = (p, R, Q)$ and the nonparametric part f. The specific problems of identifiability on the circle are studied. A consistent estimator of θ is introduced and its asymptotic normality is proved. We propose a Fourier-based estimator of f with a penalized criterion to choose the resolution level. We show that our adaptive estimator is optimal from the oracle and minimax points of view when the density belongs to a Sobolev ball. Our method is illustrated by numerical simulations.

MSC2020 subject classifications: 62H11, 62H30, 62G05.
Keywords and phrases: Mixture model, semiparametric estimation, circular data.

Received February 2021.

1. Introduction

Circular data are collected when the topic of interest is a direction or a time of day. These particular data appear in many applications: earth sciences (e.g. wind directions), medicine (e.g. circadian rhythm), ecology (e.g. animal movements), forensics (crime incidence). Different surveys on statistical methods for circular data can be found: Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), Ley and Verdebout (2017) or more recently Pewsey and García-Portugués (2021). In the present work, we consider a mixture model with two components equal up to a rotation. We observe $X_1, \ldots, X_n$ a sample of data on $\mathbb{S}^1$ with probability distribution function:

$$g(x) = p_0 f(R_0^{-1}x) + (1-p_0)f(Q_0^{-1}x) = p_0 f(x-\alpha_0) + (1-p_0)f(x-\beta_0). \quad (1)$$

In the right hand side we have identified $f : \mathbb{S}^1 \to \mathbb{R}$ and its periodized version on $\mathbb{R}$. Here R_0 and Q_0 are two unknown rotations of the circle. R_0 is a rotation with angle α_0 and Q_0 is a rotation with angle β_0. The aim is to estimate both $\theta_0 = (p_0, \alpha_0, \beta_0)$ and the nonparametric part f.

Bimodal circular data are commonly encountered in many scientific fields, for instance in climatology, animal orientations or in earth sciences. For the analysis of wind directions, see Hernández-Sánchez and Scarpa (2012) and for animal orientations, the dragonflies data set presented in Batschelet (1981). In geosciences, one can cite the cross-bed orientations data set obtained in the middle Mississipian Salem Limestone of central Indiana and which was presented by the Seminar Sedimentation (Sedimentation Seminar (1966)). Last but not least, the paper of Lark, Clifford and Waters (2014) analyzes some geological data sets and clearly favours for some of them a two component mixture of von Mises distributions.

Mixture models for describing multimodal circular data date back to Pearson (1894) and have been largely used since then. An important case in the literature is the mixture of two von Mises distributions which has been explored in numerous works. Let us cite among others papers by Bartels (1984), Spurr (1981) or Chen, Li and Fu (2008). From a practical point of view, algorithms have also been proposed to deal with mixture of two von Mises distributions, including maximum likelihood algorithms by Jones and James (1969) or a characteristic function based procedure by Spurr and Koutbeiy (1991). Note that on the unit hypersphere, Banerjee et al. (2005) investigated clustering methods for mixtures of von Mises Fisher distributions. In our framework, we shall not assume any parametric form of the density and hence the model is said to be semiparametric. To the best of our knowledge, this is the first work devoted to the study of the semiparametric mixture model for circular data. This semiparametric model is more complex and intricate than the usual parametric one encountered in the circular literature. In the spherical case, Kim and Koo (2000) studied the general mixture framework for a location parameter but assuming that the nonparametric part f is known. On the real line, this semiparametric model has been studied by Bordes, Mottelet and Vandekerkhove (2006), Hunter, Wang and Hettmansperger (2007), Butucea and Vandekerkhove (2014) or Gassiat and Rousseau (2016) for dependent latent variables. For the multivariate case, see for instance Hall and Zhou (2003), Hall et al. (2005), Gassiat, Rousseau and Vernet (2018), Hohmann and Holzmann (2013). When dealing with the specific case of one of the two components being parametric, one refers to work by Ma and Yao (2015) and references therein.

Note that we can rewrite model (1) as

$$X_i = Y_i + \varepsilon_i \pmod{2\pi}, \qquad i = 1, \dots, n, \tag{2}$$

where Y_i has density f and ε_i is a Bernoulli angle, which is equal to α_0 with probability p_0 and β_0 otherwise. Accordingly, model (1) can be viewed as a circular convolution model with unknown noise operator ε. The circular convolution model has been studied by Goldenshluger (2002) in the case of known noise operator whereas Johannes and Schwarz (2013) dealt with unknown error distribution but have at their disposal an independent sample of the noise to estimate this latter. It is worth pointing out that Goldenshluger (2002) and Johannes and Schwarz (2013) made the usual assumptions on the decay of the Fourier coefficients of the density of ε, whereas in model (1) the Fourier coefficients are not decreasing.

Identifiability questions are at the heart of the theory of mixture models and the circular context is no exception. Thus, our first task is to study the identifiability of the model. From a mathematical point of view, the topology of the circle makes the problem very different from the linear case. In the circular parametric case, Fraser, Hsu and Walker (1981) obtained identifiability results for the von Mises distributions, which were extended in Kent (1983) to generalized von Mises distributions while Holzmann, Munk and Stratmann (2004)) focused on wrapped distributions, basing their analysis on the Fourier coefficients. Here, the Fourier coefficients turn out to be very useful as well but the nonparametric paradigm makes the study quite different and intricate. Our identifiability results are obtained under mild assumptions on the Fourier coefficients. We require that the coefficients are real which can be related to the usual symmetry assumption in mixture models (see for instance Hunter, Wang and Hettmansperger (2007)) and we impose that only the first 4 coefficients do not vanish. Interestingly enough, some not intuitive phenomena appear. A striking case occurs when the angles α_0 and β_0 are distant from $2\pi/3$, model (1) is then nonidentifiable which is quite surprising at first sight.

Once the identifiability of the model is obtained, we resort to a contrast function in the line of Butucea and Vandekerkhove (2014) to estimate the Euclidian parameter θ_0. In that regard, we prove the consistency of our estimator and an asymptotic normality result. Thereafter, for the estimation of the nonparametric part, a penalized empirical risk estimation method is used. The estimator of the density turns out to be adaptive (meaning that it does not require the specification of the unknown smoothness parameter), a property which was not reached so far for this semiparametric model even in the linear case. The procedure devised is hence relevant for practical purposes. We prove an oracle inequality and minimax rates are achieved by our estimator for Sobolev regularity classes. Eventually, a numerical section shows the good performances of the whole estimation procedure.

The paper is organized as follows. Section 2 is devoted to the identifiability of the model. Section 3 tackles the estimation of the parameter θ_0 whereas Section 4 focuses on the estimation of the nonparametric part. Finally Section 5 presents numerical implementations of our procedure. Proofs are gathered in Section 6.

2. Identifiability

In this section, to keep the notation as light and clear as possible, we drop the subscript 0 in the parameters. For any function g and any angle α, denote $g_\alpha(x) := g(x - \alpha)$. For any complex number a, $\overline{a}$ is the complex conjugate of

a. For any integrable function $\phi : \mathbb{S}^1 \to \mathbb{R}$, for any $l \in \mathbb{Z}$, we denote by $\phi^{\star l}$ the Fourier coefficients $\phi^{\star l} = \int_{\mathbb{S}^1} \phi(x) e^{-ilx} \frac{dx}{2\pi}$. Note also that we use notation f and f' for two densities, where f' is not the derivative of f.

Let us now study the identifiability of our model (1) where the data have density $pf(x-\alpha) + (1-p)f(x-\beta)$. First, it is obvious that if $p = 0$, α is not identifiable, and if $p = 1$, β is not identifiable. In the same way, p is not identifiable if $\alpha = \beta$. Moreover, as explained in Hunter, Wang and Hettmansperger (2007) for a translation mixture on the real line, the case $p = 1/2$ has to be avoided. Indeed, denoting g a density and for instance $f = \frac{1}{2} g_1 + \frac{1}{2} g_{-1}$ and $f' = \frac{1}{2} g_2 + \frac{1}{2} g_{-2}$ we have

$$f_1 + f_5 = f'_2 + f'_4.$$

In addition, it is well known that, in such a mixture model, (p, α, β) cannot be distinguished from $(1-p, \beta, \alpha)$: it is the so-called *label switching* problem. So we will assume that $p \in (0, 1/2)$ (for mixtures on $\mathbb{R}$ it is assumed alternatively that $\alpha < \beta$ but ordering angles is less relevant).

Now let us study the specific problems of identifiability on the circle, that do not appear on $\mathbb{R}$. First, if f is the uniform probability, the model is not identifiable, so we have to exclude this case. Another case to exclude is the case of δ-periodic functions. Indeed in this case $f_\alpha = f_{\alpha+\delta}$. These functions have the property that $f^{\star l} = 0$ for all $l \notin (2\pi/\delta)\mathbb{Z}$. So we will require that the Fourier coefficients of f do not cancel out too much. Here we will assume

$$\text{for all } l \in \{1, 2, 3, 4\}, \qquad f^{\star l} \neq 0, \text{ and } \quad f^{\star l} = \overline{f^{\star l}}.$$

This last assumption can be related to the symmetry of f. Indeed if f is zero-symmetric then all its Fourier coefficients are real. Symmetry is a usual assumption in this mixture context, to distinguish between the translations of f: for any $\delta \in \mathbb{R}$,

$$pf(x-\alpha) + (1-p)f(x-\beta) = pf_\delta(x-\alpha+\delta) + (1-p)f_\delta(x-\beta+\delta).$$

More precisely, Hunter, Wang and Hettmansperger (2007) show that symmetry is a sufficient and necessary condition for identifiability of the model mixture on $\mathbb{R}$. In the circle framework, it is natural to work with Fourier coefficients rather than Fourier transform as on $\mathbb{R}$. A lot of circular densities have their Fourier coefficients real, provided that their location parameter is $\mu = 0$: for example the Jones-Pewsey density, which includes the cardioid, the wrapped Cauchy density, and the von Mises density. Here we require the assumption only for the first 4 Fourier coefficients of f (due to our proof), which is milder than symmetry.

Let us now state our identifiability result under these assumptions. Note that Holzmann, Munk and Stratmann (2004) have studied the identifiability of this model when f belongs to a parametric scale-family of densities, but here we face a nonparametric problem concerning f.

Theorem 1. *Assume that $\theta = (p, \alpha, \beta)$ and $\theta' = (p', \alpha', \beta')$ belong to*

$$\left\{(p, \alpha, \beta) \in (0, 1/2) \times \mathbb{S}^1 \times \mathbb{S}^1, \quad \alpha \neq \beta \pmod{2\pi}\right\}$$

and that f, f' belongs to

$$\left\{f : \mathbb{S}^1 \to \mathbb{R} \text{ density such that, for all } l \in \{1,2,3,4\},\ f^{\star l} \in \mathbb{R}\backslash\{0\}\right\}.$$

Suppose $pf_\alpha + (1-p)f_\beta = p'f'_{\alpha'} + (1-p')f'_{\beta'}$. *Then*

1. *either* $(p', \alpha', \beta') = (p, \alpha, \beta)$ *and* $f' = f$,
2. *or* $(p', \alpha', \beta') = (p, \alpha + \pi, \beta + \pi)$ *and* $f' = f_\pi$,
3. *or if* $\beta - \alpha = \pi \pmod{2\pi}$, *then* f' *is a linear combination of* f *and* f_π, *and either* $(\alpha', \beta') = (\alpha, \beta)$, *or* $(\alpha', \beta') = (\beta, \alpha)$,
4. *or if* $\beta - \alpha = \pm 2\pi/3 \pmod{2\pi}$, *then* f' *is a linear combination of* $f_{\pi/3}, f_{-\pi/3}$, f_π *and* $p' = (1-2p)/(2-3p)$ *and*
   (a) if $\beta - \alpha = 2\pi/3$, $(\alpha', \beta') = (\alpha + \pi, \beta - \pi/3)$ *or* $(\alpha', \beta') = (\alpha, \beta + 2\pi/3)$,
   (b) if $\beta - \alpha = -2\pi/3$, $(\alpha', \beta') = (\alpha + \pi, \beta + \pi/3)$ *or* $(\alpha', \beta') = (\alpha, \beta - 2\pi/3)$.

Case 2 arises from a specific feature of circular distributions: if f is symmetric with respect to 0 then it is symmetric with respect to π. Unlike the real case, a symmetry assumption does not exclude the case $f'(x) = f(x - \pi)$. To bypass this we could assume for instance $f^{\star 1} > 0$. Indeed for each $l \in \mathbb{Z}$, $(f_\pi)^{\star l} = f^{\star l}(-1)^l$, so the Fourier coefficients of f and f_π have opposite sign for any odd l. With our assumption, we recover among f and f_π the one with positive first Fourier coefficient, i.e. with positive mean resultant length. Nevertheless our estimation procedure begins with the parametric part so that this assumption concerning only the nonparametric part will not allow us to distinguish α from $\alpha + \pi$ in this first parametric estimation step. That is why we rather choose to assume that α and β belong to $[0, \pi) \pmod{\pi}$.

Case 3 concerns bipolar data since α and β are diametrically opposed (separated by π radians). In this case α and β are identifiable, but p and f not. Indeed, for any density f and any $0 < p' \leq p < 1/2$, we can find $q \in (0, 1]$ such that $f' = qf + (1-q)f_\pi$ verifies $pf_\alpha + (1-p)f_\beta = p'f'_{\alpha'} + (1-p')f'_{\beta'}$. Thus our result demonstrates that bimodal data sets with opposite modes lead to non-identifiability issues, and this highlights a fundamental issue in considering a too large class of possible densities.

Let us now discuss the case 4, which is the most curious (we shall only comment the first case (a), the other is similar). Let us set

$$f'(x) = (1-p)f\left(x - \frac{\pi}{3}\right) + (1-p)f\left(x + \frac{\pi}{3}\right) + (2p-1)f(x - \pi).$$

This function is symmetric if f is symmetric, verifies $\int_{\mathbb{S}^1} f' = 1$ and may be positive for some values of p (depending on f): see Figure 1. Then we can write $f'_{\pi/3}$:

$$f'\left(x - \frac{\pi}{3}\right) = (1-p)f\left(x - \frac{2\pi}{3}\right) + (1-p)f\left(x\right) + (2p-1)f\left(x - \frac{4\pi}{3}\right),$$

as well as f'_π:

$$f'(x - \pi) = (1-p)f\left(x - \frac{4\pi}{3}\right) + (1-p)f\left(x - \frac{2\pi}{3}\right) + (2p-1)f(x).$$

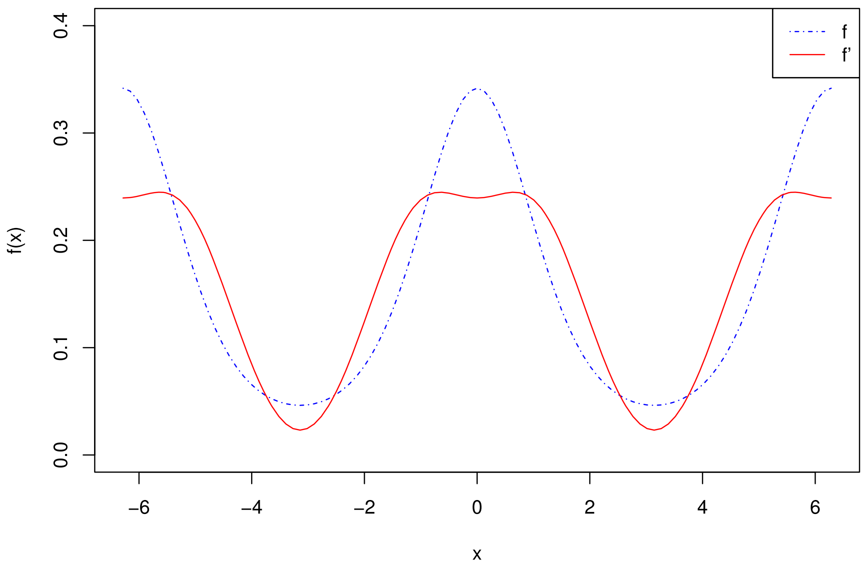

FIG 1. *Plot of a circular density f (dashed blue), and of $f' = (1-p)f_{\frac{\pi}{3}} + (1-p)f_{-\frac{\pi}{3}} + (2p-1)f_{\pi}$ (solid red). Here f is the von Mises density with mean 0 and concentration 1. In this case, f' is positive as soon as $p \geq 0.36$, here $p = 0.4$.*

Hence a mixture of f'_{π} and $f'_{\pi/3}$ gives a mixture of $f(x), f(x-\frac{2\pi}{3}), f(x-\frac{4\pi}{3})$:

$$\begin{aligned} p'f'(x-\pi) + (1-p')f'\left(x - \frac{\pi}{3}\right) &= [p'(2p-1) + (1-p')(1-p)]f(x) \\ &+[p'(1-p) + (1-p')(1-p)]f\left(x - \frac{2\pi}{3}\right) \\ &+[p'(1-p)+(1-p')(2p-1)]f\left(x - \frac{4\pi}{3}\right) \end{aligned}$$

If now $p' = (1-2p)/(2-3p)$, then $p'(1-p) + (1-p')(2p-1) = 0$ and the third component $f(x - \frac{4\pi}{3})$ vanishes. Thus

$$p'f'(x-\pi) + (1-p')f'\left(x - \frac{\pi}{3}\right) = pf(x) + (1-p)f\left(x - \frac{2\pi}{3}\right).$$

In such a particular case, we cannot identify θ nor f. However this happens only when $\beta - \alpha = \pm 2\pi/3$. So, to exclude these cases, we will now assume $\beta \neq \alpha$ (mod $2\pi/3$).

Finally, we shall assume that $f \in \mathcal{F}$ with some assumptions for $\mathcal{F}$:

Assumption 1.

$$\mathcal{F} \subset \left\{ f : \mathbb{S}^1 \to \mathbb{R} \textit{ density s.t. for all } l \in \{1,2,3,4\}, \quad f^{\star l} \in \mathbb{R}\backslash\{0\} \right\}$$

or

Assumption 2.

$$\mathcal{F} \subset \left\{f : \mathbb{S}^1 \to \mathbb{R} \text{ density s.t. for all } l \in \{1,2,3,4\}, \quad f^{\star l} \in \mathbb{R}\backslash\{0\}, \quad f^{\star 1} > 0\right\}$$

and we shall assume that $\theta \in \Theta$ with some assumptions for Θ:

Assumption 3.

$$\Theta \subset \left\{(p,\alpha,\beta) \in \left(0, \frac{1}{2}\right) \times \mathbb{S}^1 \times \mathbb{S}^1, \quad \alpha \neq \beta \pmod{\pi, 2\pi/3}\right\}$$

where $\alpha \neq \beta \pmod{2\pi/3, \pi}$ means $\beta - \alpha \notin \{-\frac{2\pi}{3}, 0, \frac{2\pi}{3}, \pi\} + 2\pi\mathbb{Z}$, or

Assumption 4.

$$\Theta \subset \left\{(p,\alpha,\beta) \in \left(0, \frac{1}{2}\right) \times [0,\pi) \times [0,\pi), \quad \alpha \neq \beta \pmod{2\pi/3}\right\}.$$

Note that Assumption 4 implies Assumption 3, and Assumption 2 implies Assumption 1. We can write the following result.

Corollary 2. *Under Assumptions 1 and 4, or under Assumptions 2 and 3, model (1) is identifiable. Under Assumptions 1 and 3, model (1) is identifiable modulo π, that is to say that if $pf_\alpha + (1-p)f_\beta = p'f'_{\alpha'} + (1-p')f'_{\beta'}$ then $p' = p$ and either $(\alpha', \beta') = (\alpha, \beta)$ and $f' = f$, or $(\alpha', \beta') = (\alpha + \pi, \beta + \pi)$ and $f' = f_\pi$.*

Moreover, the proof of Theorem 1 provides the following statement.

Lemma 3. *Under Assumption 3, denoting $M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}$, for all $\theta, \theta' \in \Theta$,*

$$\forall 1 \leq l \leq 4, \ \Im\left(M^l(\theta')\overline{M^l(\theta)}\right) = 0 \Leftrightarrow \theta' = \theta \text{ or } \theta' {=} \theta + \pi.$$

where $\theta' {=} \theta + \pi$ means $(p', \alpha', \beta') = (p, \alpha + \pi, \beta + \pi)$.

3. Estimation for the parametric part

Now, let us denote for all $l \in \mathbb{Z}$

$$M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}.$$

In model (1) the Fourier coefficients of g satisfy for any l:

$$g^{\star l} = (p_0 e^{-i\alpha_0 l} + (1-p_0)e^{-i\beta_0 l})f^{\star l}.$$

Thus $g^{\star l} = M^l(\theta_0)f^{\star l}$ and the previous lemma gives that $\theta = \theta_0$ (or $\theta_0 + \pi$) if and only if, for each $l \in \{1, \ldots, 4\}$,

$$\Im\left(M^l(\theta_0)\overline{M^l(\theta)}\right) = 0 \Leftrightarrow \Im\left(g^{\star l}\overline{M^l(\theta)}\right) = 0$$

using that $f^{\star l}$ are non-zero real numbers. This invites us to consider

$$S(\theta) := \sum_{l=-4}^{4} \left(\Im\left(g^{\star l}\overline{M^l(\theta)}\right)\right)^2 = \sum_{l=-4}^{4} \left(\Im\left(g^{\star l}\{pe^{i\alpha l} + (1-p)e^{i\beta l}\}\right)\right)^2.$$

Note that $g^{\star 0}\overline{M^0(\theta)} = 1/(2\pi)$ and that $\Im\left(g^{\star(-l)}\overline{M^{-l}(\theta)}\right) = \Im\left(\overline{g^{\star l}}M^l(\theta)\right) = -\Im\left(g^{\star l}\overline{M^l(\theta)}\right)$ so that we can also write

$$S(\theta) = 2\sum_{l=1}^{4} \left(\Im\left(g^{\star l}\overline{M^l(\theta)}\right)\right)^2.$$

The empirical counterpart of $S(\theta)$ is

$$\begin{aligned}
\tilde{S}_n(\theta) &= \sum_{l=-4}^{4} \left(\Im\left(\widehat{g^{\star l}}\overline{M^l(\theta)}\right)\right)^2 \\
&= \sum_{l=-4}^{4} \left(\Im\left(\frac{1}{2\pi n}\sum_{k=1}^{n} e^{-ilX_k}\overline{M^l(\theta)}\right)\right)^2 \\
&= \frac{1}{4\pi^2 n^2}\sum_{l=-4}^{4}\sum_{1\le k,j\le n} \Im\left(e^{ilX_k}M^l(\theta)\right)\Im\left(e^{ilX_j}M^l(\theta)\right).
\end{aligned}$$

Next, we consider a slightly modified version of $\tilde{S}_n(\theta)$ by removing the diagonal terms

$$S_n(\theta) = \frac{1}{4\pi^2 n(n-1)}\sum_{l=-4}^{4}\sum_{1\le k\neq j\le n} \Im\left(e^{ilX_k}M^l(\theta)\right)\Im\left(e^{ilX_j}M^l(\theta)\right). \quad (3)$$

Let us denote

$$Z_k^l(\theta) := \Im\left(\frac{e^{ilX_k}}{2\pi}M^l(\theta)\right) \quad \text{and} \quad J^l(\theta) := \Im\left(\overline{g^{\star l}}M^l(\theta)\right).$$

Hence

$$S_n(\theta) = \frac{1}{n(n-1)}\sum_{l=-4}^{4}\sum_{1\le k\neq j\le n} Z_k^l(\theta)Z_j^l(\theta).$$

Note that we have $\mathbb{E}(Z_k^l(\theta)) = J^l(\theta)$, and $S_n(\theta)$ is an unbiased estimator of $S(\theta)$.

Let the estimator of θ_0 be

$$\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta} S_n(\theta). \quad (4)$$

For this estimator we can prove the following consistency result.

Theorem 4. *Consider* Θ *a compact set included in*

$$\left\{(p,\alpha,\beta)\in(0,1/2)\times\mathbb{S}^1\times\mathbb{S}^1,\quad \alpha\neq\beta\pmod{2\pi/3,\pi}\right\}$$

and the estimator $\hat{\theta}_n=\operatorname{argmin}_{\theta\in\Theta}S_n(\theta)$. *We have* $\hat{\theta}_n\to\theta_0 \pmod{\pi}$ *in probability.*

The last convergence means that for all $\epsilon>0$, the probability $\mathbb{P}(\|\hat{\theta}_n-\theta_0\|\leq\epsilon$ or $\|\hat{\theta}_n-\theta_0-\pi\|\leq\epsilon)$ tends to 1 when n goes to $+\infty$, where $\|.\|$ denotes the Euclidean norm.

Proof. Θ is a compact set and S is continuous. Lemma 13 ensures that S_n is Lipschitz hence uniformly continuous, and Proposition 14 ensures that for all θ, $|S_n(\theta)-S(\theta)|$ tends to 0 in probability. Then it is sufficient to apply a classical Lemma to conclude. See the details in Section 6.2 □

From now on, we assume that Θ is a compact set included in $\left(0,\frac{1}{2}\right)\times[0,\pi)\times[0,\pi)$, as in Assumption 4. Then, $\theta_0+\pi$ is excluded and under Assumption 4, $\hat{\theta}_n\to\theta_0$ in probability. Moreover this estimator is asymptotically normal. We denote $\dot{\phi}(\theta)$ the gradient of any function ϕ with respect to $\theta=(p,\alpha,\beta)$, $\ddot{\phi}(\theta)$ the Hessian matrix and for any matrix A, we denote $A^\top$ its transpose.

Theorem 5. *Consider* Θ *a compact set included in*

$$\{(p,\alpha,\beta)\in(0,1/2)\times[0,\pi)\times[0,\pi),\quad \alpha\neq\beta\pmod{2\pi/3}\}$$

and the estimator $\hat{\theta}_n=\operatorname{argmin}_{\theta\in\Theta}S_n(\theta)$. *Assume that* $\theta_0\in\Theta$. *Let* $\mathcal{A}$ *be the Hessian matrix of* S *in* θ_0: $\mathcal{A}=\ddot{S}(\theta_0)=2\sum_{l=-4}^{4}\dot{J}^l(\theta_0)\dot{J}^l(\theta_0)^\top$. *Then, if* $\mathcal{A}$ *is invertible,*

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{d}\mathcal{N}(0,\Sigma),$$

where $\Sigma=\mathcal{A}^{-1}V\mathcal{A}^{-1}$, $V=4\mathbb{E}(UU^\top)$ *and* $U=\sum_{l=-4}^{4}\dot{J}^l(\theta_0)Z_1^l(\theta_0)$.

The proof can be found in Section 6.3. Note that $\mathcal{A}$ can be estimated by $\ddot{S}(\hat{\theta}_n)$ and V by

$$\frac{4}{n^3}\sum_{1\leq k,j,j'\leq n}\ \sum_{-4\leq l,l'\leq 4}Z_k^l(\hat{\theta}_n)Z_k^{l'}(\hat{\theta}_n)\dot{Z}_j^l(\hat{\theta}_n)(\dot{Z}_{j'}^{l'}(\hat{\theta}_n))^\top$$

(see details in Section 6.4). Thus we can estimate the covariance matrix Σ and deduce an asymptotic confidence region.

We also prove the following result on the quadratic risk of the estimator $\hat{\theta}_n$, which is useful for the sequel (see Section 6.5 for a proof).

Proposition 6. *Under the assumptions of Theorem 5, there exists a numerical constant* K *such that, for all* $\theta_0\in\Theta$ *and for all* $n\geq 1$

$$\mathbb{E}\|\hat{\theta}_n-\theta_0\|^2\leq Kn^{-1},$$

where the norm is the Euclidean norm in $\mathbb{R}^3$.

4. Nonparametric part

Let us now estimate the nonparametric part. We shall use the following norm: for any function ϕ, we denote $\|\phi\|_2 = \left(\frac{1}{2\pi}\int_{\mathbb{S}^1} \phi^2(x)dx\right)^{1/2}$. Recall that for all $l \in \mathbb{Z}$, $g^{\star l} = M^l(\theta_0)f^{\star l}$ where g is the density of the observations X_k and $g^{\star l}$ its Fourier coefficient. Then $f^{\star l} = g^{\star l}/M^l(\theta_0)$. We can verify that $M^l(\theta_0) \neq 0$. Indeed, for any $\theta \in \Theta$,

$$|M^l(\theta)|^2 = p^2 + (1-p)^2 + 2p(1-p)\cos[l(\beta-\alpha)] \geq (1-2p)^2 > 0.$$

Nevertheless this division by $M^l(\theta_0)$ requires us to impose a new assumption. We assume that there exists $P \in (0, 1/2)$ such that $0 < p < P$ for any p, i. e.

Assumption 5. Θ *is a compact set included in*

$$\{(p,\alpha,\beta) \in (0,P) \times [0,\pi) \times [0,\pi), \quad \alpha \neq \beta \pmod{2\pi/3}\}.$$

Under this assumption, $|M^l(\theta)|$ is always bounded from below by $1 - 2P$. Now, to estimate $g^{\star l} = \int_{\mathbb{S}^1} e^{-ilx}g(x)dx/(2\pi)$, it is natural to define

$$\widehat{g^{\star l}} = \frac{1}{2\pi n}\sum_{k=1}^{n} e^{-ilX_k}.$$

If $\hat{\theta} = \hat{\theta}_n$ is the previous estimator of the parametric part, we set the plugin estimator of the Fourier coefficient:

$$\widehat{f^{\star l}} = \frac{1}{2\pi n}\sum_{k=1}^{n} M^l(\hat{\theta})^{-1} e^{-ilX_k}.$$

Finally, for L an integer, set

$$\hat{f}_L(x) = \sum_{l=-L}^{L} \widehat{f^{\star l}} e^{ilx}.$$

To measure the performance of this estimator, we use Parseval equality to write

$$\|f - \hat{f}_L\|_2^2 = \sum_{|l|>L} |f^{\star l}|^2 + \sum_{l=-L}^{L} |f^{\star l} - \widehat{f^{\star l}}|^2$$

which is the classical bias variance decomposition. Moreover it is possible to prove that the variance term satisfies $\sum_{l=-L}^{L} \mathbb{E}|f^{\star l} - \widehat{f^{\star l}}|^2 = O(\frac{2L+1}{n})$ (see Lemma 18 below). To control the bias term we recall the definition of the Sobolev ellipsoid:

$$W(s,R) = \{f : \mathbb{S}^1 \to \mathbb{R}, \quad \sum_{l\in\mathbb{Z}} (1+l^2)^s |f^{\star l}|^2 \leq R^2\}.$$

For such a smooth f, the risk of estimator $\hat{f}_L$ is then bounded in the following way:

$$\mathbb{E}\|f - \hat{f}_L\|_2^2 \leq R^2 \left(1 + L^2\right)^{-s} + C\frac{2L+1}{n}.$$

It is clear that an optimal value for L is of order $n^{1/(2s+1)}$ but this value is unknown. We rather choose a data-driven method to select L. We introduce a classical minimization of a penalized empirical risk. Set

$$\widehat{L} = \operatorname*{argmin}_{L \in \mathcal{L}} \left\{ - \sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + \lambda \frac{2L+1}{n} \right\} \tag{5}$$

where $\mathcal{L}$ is a finite set of resolution level, and λ a constant to be specified later. The next theorem states an oracle inequality which highlights the bias variance decomposition of the quadratic risk and justifies our estimation procedure.

Theorem 7. *Assume Assumption 1 and Assumption 5. Assume that f belongs to the Sobolev ellipsoid $W(s,R)$ with $s \geq 1$. Let $\hat{L}$ defined in (5) with $\mathcal{L} = \{0, 1, \ldots, \lfloor cn^{\frac{1}{2s_0+1}} \rfloor\}$ for some $s_0 > 1$ and some positive constant c. Let $\epsilon > 0$. If the penalty constant verifies $\lambda > (3/\pi^2)(1+\epsilon^{-1})(1-2P)^{-2}$ then,*

$$\mathbb{E}\|\hat{f}_{\widehat{L}} - f\|_2^2 \leq (1+2\epsilon)\mathbb{E} \min_{L \in \mathcal{L}} \left\{ \|\hat{f}_L - f\|_2^2 + 2\lambda \frac{2L+1}{n} \right\} + \frac{C(1+R^2)}{n}$$

where C is a positive constant depending on $P, s_0, c, \epsilon, \lambda$. Moreover, if $s \geq s_0$,

$$\sup_{f \in W(s,R)} \sup_{\theta_0 \in \Theta} \mathbb{E}_{\theta_0, f} \|\hat{f}_{\widehat{L}} - f\|_2^2 = O\left(R^2 n^{-2s/(2s+1)}\right).$$

As a consequence our estimator has a quadratic risk in $n^{-2s/(2s+1)}$. Regarding the lower bound note that for any estimator $\tilde{f}_n$

$$\sup_{f \in W(s,R)} \sup_{\theta_0 \in \Theta} \mathbb{E}_{\theta_0, f} \|\tilde{f}_n - f\|_2^2 \geq \sup_{f \in W(s,R)} \mathbb{E}_{\theta, f} \|\tilde{f}_n - f\|_2^2$$

for some arbitrary $\theta \in \Theta$, so that the problem is reduced to a purely nonparametric lower bound. In the case of direct observations this quantity is lower bounded by $\underline{C} n^{-2s/(2s+1)}$, see Theorem 11 and its proof in Baldi et al. (2009) (case $d = 1$ for the circle $\mathbb{S}^1$). We can use this proof to prove the lower bound in our mixture case. Indeed, for any densities f_1 and f_2, if $g_i(x) = pf_i(x-\alpha) + (1-p)f_i(x-\beta)$ is the associated density of our observations, then the Kullback-Leibler divergence verifies

$$K(g_1 dx, g_2 dx) \leq \int \frac{(g_1 - g_2)^2}{g_2} \leq 2 \int \frac{(f_1 - f_2)^2}{f_2}$$

and the rest of the proof is identical. Thus

$$\sup_{f \in W(s,R)} \sup_{\theta_0 \in \Theta} \mathbb{E}_{\theta_0, f} \|\hat{f}_{\widehat{L}} - f\|_2^2 \geq \underline{C} n^{-2s/(2s+1)}$$

and our estimator is optimal minimax.

Remark 1. *Note that the penalty only depends on P which is some safety margin around 1/2, that can be chosen by the statistician. For the practical choice of the penalty, see Section 5.*

Eventually, note that some densities may be *supersmooth*, in the following sense:

$$\sum_{l \in \mathbb{Z}} \exp(2b|l|^r)|f^{\star l}|^2 \leq R^2.$$

In this case, the quadratic bias is bounded by $R^2 \exp(-2bL^r)$ which gives the following fast rate of convergence:

$$\mathbb{E}\|\hat{f}_{\hat{L}} - f\|_2^2 = O\left(\frac{(\log n)^{1/r}}{n}\right).$$

5. Numerical results

All computations are performed with Matlab software and the Optimization Toolbox.

We shall implement our statistical procedure to both estimate the parameter θ_0 and the density f. We consider three popular circular densities, namely the von Mises density, the wrapped Cauchy and the wrapped normal densities. We remind their expression (see Ley and Verdebout (2017)). The von Mises density is given by:

$$f_{VM}(x) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(x-\mu)},$$

with $\kappa \geq 0$, $I_0(\kappa)$ the modified Bessel function of the first kind and of order 0. The wrapped Cauchy distribution has density:

$$f_{WC}(x) = \frac{1}{2\pi} \frac{1-\gamma^2}{1+\gamma^2 - 2\gamma\cos(x-\mu)},$$

with $0 \leq \gamma \leq 1$. The wrapped normal density expression is:

$$f_{WN}(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} e^{-\frac{(x-\mu+2k\pi)^2}{2\sigma^2}},$$

$\sigma > 0$. For more clarity, we set $\sigma^2 =: -2\log(\rho)$. Hence, we have $0 \leq \rho \leq 1$.

All these densities are characterized by a concentration parameter κ, γ or ρ and a location parameter μ. Remind that values $\kappa = 0$, $\gamma = 0$ and $\rho = 0$ correspond to the uniform density on the circle. To meet symmetry assumptions of Theorem 1, we consider in the sequel that the location parameter is set to $\mu = 0$.

First, let us focus on the parametric part. We set $\theta_0 = (p_0, \alpha_0, \beta_0) = (\frac{1}{4}, \frac{\pi}{8}, \frac{2\pi}{3})$. Obtaining the estimate $\hat{\theta}_n$ of θ_0 (see (4)) requires to solve a nonlinear minimization problem. To this end, we resort to the function *fmincon* of the Matlab Optimization toolbox. The function *fmincon* finds a constrained minimum of a function of several variables. Two parameters are to be specified: the domain over which the minimum is searched and an initial value. We consider the domain $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$. For more stability and to avoid possible local minimums, we perform the procedure over 10 initials values uniformly drawn on $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$. The final estimator $\hat{\theta}_n$ corresponds to the minimum value of the empirical contrast $S_n(\theta)$ given in (3) over the 10 trials.

Table 1 gathers mean squared errors for our estimation procedure. When analyzing Table 1, one clearly sees that increasing the number of observations improves noticeably the performances. As expected, von Mises densities with smaller concentration parameter are more difficult to estimate. Nonetheless, the overall performances are satisfying. Table 2 displays the performances of the method-of-moments estimation procedure developed by Spurr and Koutbeiy (1991) to handle the problem of estimating the parameters in mixtures of von Mises distributions. To fairly compare the two methods, Table 3 gives the Spurr and Koutbeiy (1991) performances but this time when estimating on the same domain than ours e.g $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$. At closer inspection, the Spurr and Koutbeiy (1991) method seems to behave better to estimate angles α_0 and β_0 while our method may appear more competitive for estimating p_0. It is worth noticing that the method by Spurr and Koutbeiy (1991) is completely parametric and takes advantage of the knowledge of the distributions. In this regard, our procedure which is semiparametric is competitive with a parametric method.

Figure 2 illustrates the asymptotic normality of our estimator $\hat{\theta}_n$ stated in Theorem 5.

TABLE 1
Mean squared errors for estimating parameter θ_0 over 50 Monte Carlo replications.

density	$n = 100$			$n = 1000$		
	p	α	β	p	α	β
f_{VM}, $\kappa = 2$	0.0121	0.6848	0.1131	0.0017	0.1919	0.0238
f_{VM}, $\kappa = 5$	0.0030	0.0285	0.0049	1.4632e-04	0.0017	4.4861e-04
f_{VM}, $\kappa = 7$	0.0033	0.0133	0.0031	1.6721e-04	0.0013	3.0102e-04
f_{WC}, $\rho = 0.8$	0.0029	0.0124	0.0024	2.0788e-04	8.5435e-04	1.8942e-04
f_{WN}, $\rho = 0.8$	0.0077	0.1679	0.0457	0.0020	0.0238	0.0037

TABLE 2
Spurr and Koutbeiy procedure: mean squared errors for estimating parameter θ_0 over 50 Monte Carlo replications on $\{(0, 1) \times [0, 2\pi) \times [0, 2\pi)\}$

density	$n = 100$			$n = 1000$		
	p	α	β	p	α	β
f_{VM}, $\kappa = 2$	0.0938	0.4212	0.1171	0.0116	0.0685	0.0062
f_{VM}, $\kappa = 5$	0.0031	0.0360	0.0049	2.9965e-04	0.0025	6.6273e-04
f_{VM}, $\kappa = 7$	0.0031	0.0084	0.0029	2.4553e-04	0.0014	3.5541e-04

Now, let us turn to the nonparametric estimation part namely the estimation of the density f. The estimator of f is given by $\hat{f}_{\hat{L}}$ (see Theorem 7). It requires

TABLE 3
Spurr and Koutbeiy procedure: mean squared errors for estimating parameter θ_0 *over* 50 *Monte Carlo replications on* $\{(0, \frac{1}{2}) \times [0, \pi) \times [0, \pi)\}$

density	$n = 100$			$n = 1000$		
	p	α	β	p	α	β
f_{VM}, $\kappa = 2$	0.0231	0.2117	0.0351	0.0112	0.0635	0.0081
f_{VM}, $\kappa = 5$	0.0032	0.0409	0.0042	4.1489e-04	0.0022	6.3122e-04
f_{VM}, $\kappa = 7$	0.0026	0.0094	0.0029	2.3197e-04	0.0010	2.8350e-04

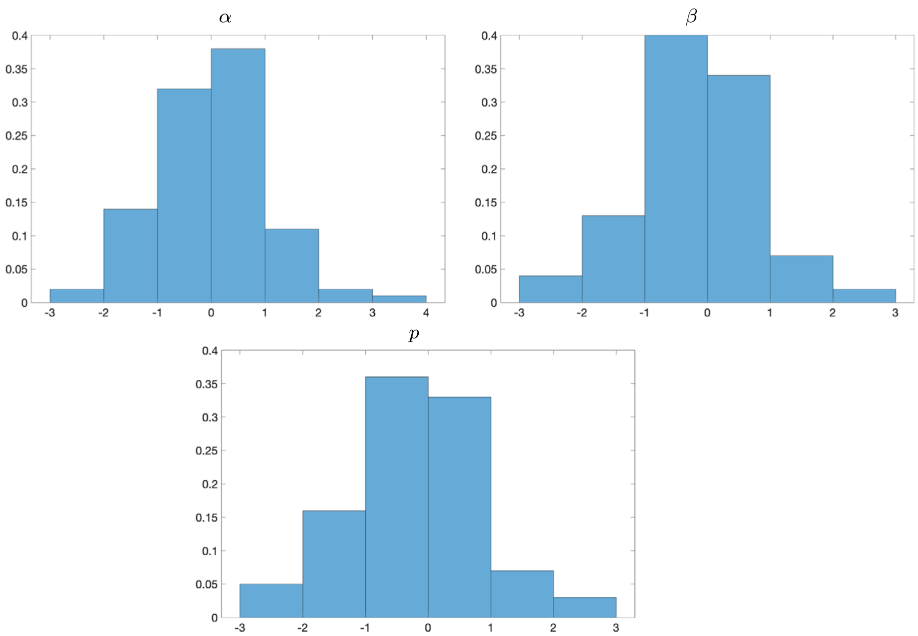

FIG 2. *Histograms of the centered and standardized statistics* $\hat{\theta}_n$ *for the von Mises density* f_{VM} *with* $\kappa = 5$, $n = 1000$ *observations and* 100 *Monte Carlo replications*

the computation of a data-driven resolution level choice $\widehat{L}$ (given in (5)) which implies a tuning parameter λ. To select the proper λ, we follow the data-driven slope estimation approach due to Birgé and Massart (see Birgé and Massart (2001) and Birgé and Massart (2007)). An overview in practice is presented in Baudry, Maugis and Michel (2012). To implement the slope heuristics method, one has to plot for $L = 0$ to $L_{\max}$ the couples of points $(\frac{2L+1}{n}, \sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2)$. For $L \geq L_0$, one should observe a linear behaviour (see Figure 3). Then, once the slope is estimated, say a, by a linear regression method, one eventually takes $\hat{\lambda} = 2a$ and the final resolution level is:

$$\widehat{L} = \underset{L \in \mathcal{L}}{\operatorname{argmin}} \left\{ - \sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + \hat{\lambda} \frac{2L+1}{n} \right\}.$$

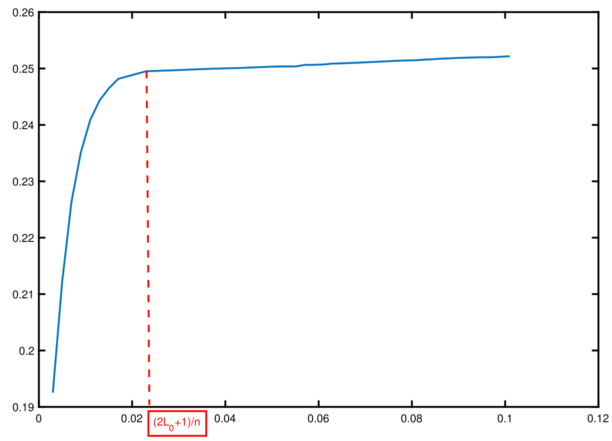

FIG 3. *For the wrapped Cauchy density f_{WC} with $\gamma = 0.8$ and $n = 1000$: plot of couples* $(\frac{2L+1}{n}, \sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2)$ *for* $L = \{1, \ldots, 50\}$.

Finally, Figure 4 shows reconstructions of the density f and the mixture density g as well. The estimates are good.

Remark 2. *Note that for the two exceptional cases, when $p_0 = 0$ or f is the uniform density, our procedure performs well. Indeed, if $p_0 = 0$, our method yields that $\alpha = \beta$ and retrieves that there is only one component in the mixture. When f is the uniform density, our algorithm selects $\hat{L} = 0$ which yields the uniform distribution.*

6. Proofs

6.1. Proof of Theorem 1 (identifiability)

Denote

$$M^l(\theta) := pe^{-i\alpha l} + (1-p)e^{-i\beta l}.$$

Suppose $pf(x-\alpha) + (1-p)f(x-\beta) = p'f'(x-\alpha') + (1-p')f'(x-\beta')$. The calculation of the Fourier coefficients gives, for all $l \in \mathbb{Z}$, $f^{\star l} M^l(\theta) = (f')^{\star l} M^l(\theta')$ which implies

$$f^{\star l} |M^l(\theta)|^2 = (f')^{\star l} M^l(\theta') \overline{M^l(\theta)}.$$

Then, our assumptions on f and f' entail

$$M^l(\theta')\overline{M^l(\theta)} \text{ is real} \quad \forall l \in \{1, 2, 3, 4\}.$$

Let us now study the consequence of this fact. Denote

$$\gamma_1 = \alpha' - \beta,\ \gamma_2 = \alpha' - \alpha,\ \gamma_3 = \beta' - \beta,\ \gamma_4 = \beta' - \alpha$$

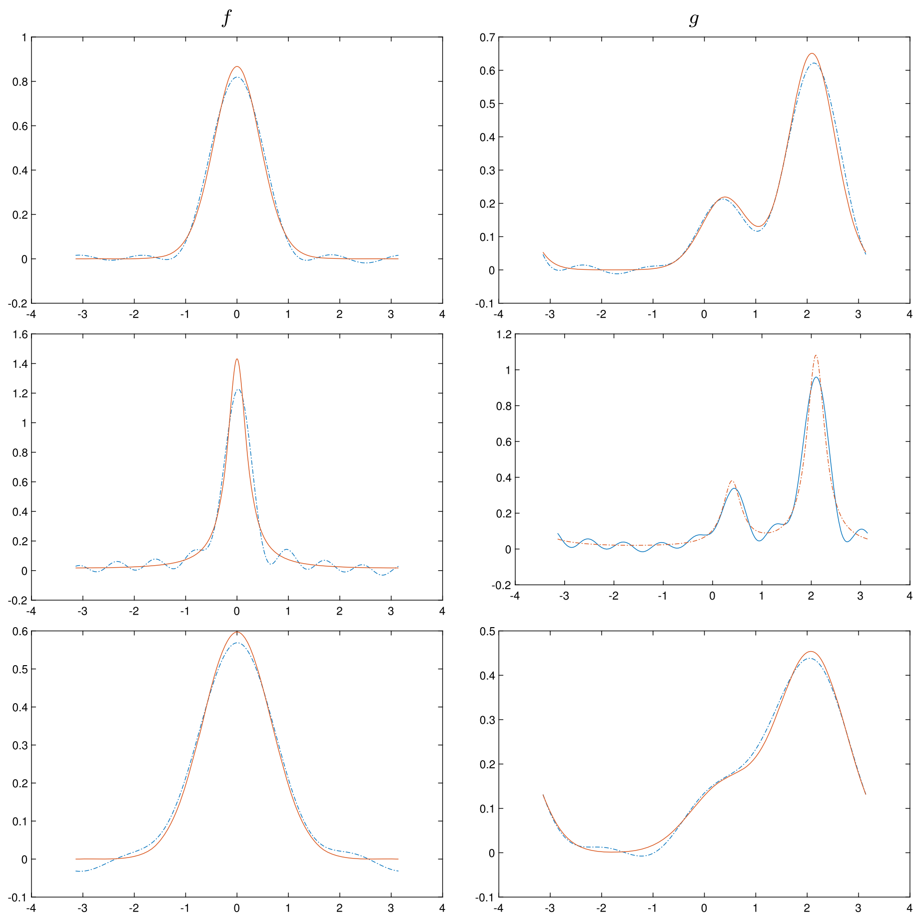

Fig 4. *Estimation of the density f and the mixture density g for $n = 1000$. In red, the density, in dotted lines its estimate. From top to bottom: the von Mises density with $\kappa = 5$, the wrapped Cauchy with $\gamma = 0.8$ and the wrapped normal density with $\rho = 0.8$.*

the 4 angles. Denote also the associated weights in $(0, 1)$:

$$\lambda_1 = p'(1-p),\ \lambda_2 = p'p,\ \lambda_3 = (1-p')(1-p),\ \lambda_4 = (1-p')p.$$

With this notation

$$M^l(\theta')\overline{M^l(\theta)} = \lambda_1 e^{-i\gamma_1 l} + \lambda_2 e^{-i\gamma_2 l} + \lambda_3 e^{-i\gamma_3 l} + \lambda_4 e^{-i\gamma_4 l}.$$

Then $M^l(\theta')\overline{M^l(\theta)}$ is real if and only if $\sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0$ and we have to solve the equations

$$\forall l = 1, 2, 3, 4, \qquad \sum_{k=1}^4 \lambda_k \sin(l\gamma_k) = 0. \tag{6}$$

This system of equations is studied in Lemmas 8 and 9 below.

Let us now reason with the representatives of the γ_k in $(-\pi, \pi]$. Lemma 9 says that the possible values for the γ_k's are $0, \pi, \gamma, -\gamma$, for some $\gamma \in (0, \pi)$. Note that here

$$\gamma_1 - \gamma_2 = \gamma_3 - \gamma_4 = \alpha - \beta \neq 0 \quad \text{and} \quad \gamma_1 - \gamma_3 = \gamma_2 - \gamma_4 = \alpha' - \beta' \neq 0 \tag{7}$$

and then the γ_k's take at least 2 different values: either 4 different values; or $\gamma_2 = \gamma_3$ and the other distinct; or $\gamma_1 = \gamma_4$ and the other distinct; or $\gamma_2 = \gamma_3$ and $\gamma_1 = \gamma_4$.

• Let us first study the case where all the γ_k's are distinct. There are 4!=24 ways of having $(\gamma_{i_1}, \gamma_{i_2}, \gamma_{i_3}, \gamma_{i_4}) = (-\gamma, 0, \gamma, \pi)$. But 16 combinations lead to $p = 1/2$ or $p' = 1/2$. For example, if $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (-\gamma, 0, \gamma, \pi)$ then (6) becomes

$$\lambda_1 \sin(-l\gamma) + \lambda_2 \sin(0) + \lambda_3 \sin(l\gamma) + \lambda_4 \sin(l\pi) = 0.$$

Thus $\lambda_1 = \lambda_3$, which gives $p' = 1/2$. In the same way, there are 4 possibilities giving $\lambda_1 = \lambda_3$, 4 possibilities giving $\lambda_1 = \lambda_2$, 4 possibilities giving $\lambda_2 = \lambda_4$, 4 possibilities giving $\lambda_3 = \lambda_4$. All of this is impossible, since $p, p' \in (0,1)\backslash\{1/2\}$. In addition, in the 4 cases where $\gamma_1 = -\gamma_4$, we obtain via (7) $\gamma_3 = -\gamma_2$ which is impossible if $\{\gamma_2, \gamma_3\} = \{0, \pi\}$. Idem if $\gamma_2 = -\gamma_3$ and $\{\gamma_1, \gamma_4\} = \{0, \pi\}$. Thus it is finally impossible that all the γ_k's are distinct.

• Let us now study the case where the γ_k's take 3 distinct values ($\gamma_2 = \gamma_3$ or $\gamma_1 = \gamma_4$) and belong to $\{0, \pi, \gamma\}$ or $\{0, \pi, -\gamma\}$. In the case where $\gamma_2 = \gamma_3$, coming back to equation (6), we understand that all the rearrangements lead to $\lambda_4 = 0$ or $\lambda_1 = 0$ or $\lambda_2 + \lambda_3 = 0$, which is impossible. In the same way, if $\gamma_1 = \gamma_4$, equation (6) leads to $\lambda_2 = 0$ or $\lambda_3 = 0$ or $\lambda_1 + \lambda_4 = 0$, which is impossible.

• The next case is when the γ_k's take 3 distinct values and belong to $\{0, \gamma, -\gamma\}$ or $\{\pi, \gamma, -\gamma\}$. If $\gamma_2 = \gamma_3$, we can then list the 6 cases:

γ_1	$\gamma_2 = \gamma_3$	γ_4	consequence
$-\gamma$	$0/\pi$	γ	$p = p', \alpha' - \alpha = \beta' - \beta = 0 \pmod{\pi}$
γ	$0/\pi$	$-\gamma$	$p = p', \alpha' - \alpha = \beta' - \beta = 0 \pmod{\pi}$
$-\gamma$	γ	$0/\pi$	$\lambda_1 = \lambda_2 + \lambda_3$
γ	$-\gamma$	$0/\pi$	$\lambda_1 = \lambda_2 + \lambda_3$
$0/\pi$	γ	$-\gamma$	$\lambda_4 = \lambda_2 + \lambda_3$
$0/\pi$	$-\gamma$	γ	$\lambda_4 = \lambda_2 + \lambda_3$

Note that $\lambda_1 = \lambda_2 + \lambda_3 \Leftrightarrow p'(2 - 3p) = 1 - p$, which is possible only if $p < 1/2$ and $p' > 1/2$ (recall that we suppose $p < 1/2$ and $p' < 1/2$). In the same way $\lambda_4 = \lambda_2 + \lambda_3 \Leftrightarrow p'(1 - 3p) = 1 - 2p$, which is possible only if $p > 1/2$ and $p' < 1/2$.

Finally, if $\gamma_1 = \gamma_4$, we have the 6 last cases:

γ_2	$\gamma_1 = \gamma_4$	γ_3	consequence
$-\gamma$	$0/\pi$	γ	$p' = 1 - p$
γ	$0/\pi$	$-\gamma$	$p' = 1 - p,$
$-\gamma$	γ	$0/\pi$	$p' = \frac{p}{3p-1}$
γ	$-\gamma$	$0/\pi$	$p' = \frac{p}{3p-1}$
$0/\pi$	γ	$-\gamma$	$p' = \frac{1-2p}{2-3p}, \beta - \alpha = \pm 2\pi/3$
$0/\pi$	$-\gamma$	γ	$p' = \frac{1-2p}{2-3p}, \beta - \alpha = \pm 2\pi/3$

Note that the 4 first lines of this table are impossible since $p, p' \in (0, 1/2)$ and $p' = p/(3p-1) \notin (0,1)$ if $0 < p < 1/2$. Let us detail the lines 5 and 6. In these cases, $\lambda_1 + \lambda_4 - \lambda_3 = 0$ which provides $p' = (1-2p)/(3-2p)$. Moreover (7) implies that $3\gamma_1 = \gamma_2 = 0 \pmod{\pi}$ and $2\gamma_1 = \beta - \alpha = \alpha' - \beta'$. According to the values of γ_1 and γ_2, there are 4 possibilities

◇ $\beta - \alpha = 2\pi/3$ and $(\alpha', \beta') = (\alpha, \beta + 2\pi/3)$,
◇ $\beta - \alpha = 2\pi/3$ and $(\alpha', \beta') = (\alpha + \pi, \beta - \pi/3)$
◇ $\beta - \alpha = -2\pi/3$ and $(\alpha', \beta') = (\alpha, \beta - 2\pi/3)$
◇ $\beta - \alpha = -2\pi/3$ and $(\alpha', \beta') = (\alpha + \pi, \beta + \pi/3)$

• The last case occurs when the γ_k's take 2 distinct values. If the γ_k's take exactly 2 different values, using (7), necessarily

$$\gamma_1 = \gamma_4 \text{ and } \gamma_2 = \gamma_3 \pmod{2\pi} \Rightarrow 0 = \gamma_1 - \gamma_4 + \gamma_3 - \gamma_2 = 2(\alpha - \beta) \pmod{2\pi}$$

which is possible only if $\alpha - \beta = \pi \pmod{2\pi}$ (recall that $\alpha - \beta$ is always assumed $\neq 0$). And in the same way $\alpha' - \beta' = \pi \pmod{2\pi}$. Then $\gamma_1 - \gamma_2 = \alpha - \beta = \pi \pmod{2\pi}$. Thus the two different values of the γ_k's are at a distant of π.

The first possibility is that these two values are 0 and π, which corresponds to the first case of Lemma 9. There are two subcases: 1a. $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (\pi, 0, 0, \pi)$ or 1b. $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0, \pi, \pi, 0)$. In the subcase 1a. $(\alpha', \beta') = (\alpha, \beta)$. Equations

$$\begin{cases} pf + (1-p)f_\pi = p'f' + (1-p')f'_\pi \\ pf_\pi + (1-p)f = p'f'_\pi + (1-p')f' \end{cases}$$

entails that f' is a linear combination of f and f_π. In the subcase 1*b*. $(\alpha', \beta') = (\alpha + \pi, \beta + \pi) = (\beta, \alpha)$.

The second possibility is that the two distinct values $\gamma_1 = \gamma_4$ and $\gamma_2 = \gamma_3$ are not multiples of π, which corresponds to the fourth case of Lemma 9. Then $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (\gamma_1, -\gamma_1, -\gamma_1, \gamma_1)$ and

$$\gamma_1 - (-\gamma_1) = \gamma_1 - \gamma_2 = \pi \pmod{2\pi}$$

which entails $\gamma_1 = \pi/2 \pmod{\pi}$. Equation (6) becomes

$$(\lambda_1 - \lambda_2 - \lambda_3 + \lambda_4)\sin(l\pi/2) = 0$$

so that $\lambda_1 + \lambda_4 = \lambda_2 + \lambda_3$, which gives

$$p'(1-p) + p(1-p') = p'p + (1-p')(1-p) \Rightarrow p' + p - 2pp' = 1/2 \Rightarrow p' = 1/2$$

which is impossible.

• Let us recap the only possible cases that we have obtained:

▹ $p = p', \alpha' - \alpha = \beta' - \beta = 0 \pmod{\pi}$,

▹ $p' = \frac{1-2p}{2-3p}, \beta - \alpha = \pm 2\pi/3$, with the four possibilities described above,

▹ $\beta - \alpha = \pi$, $(\alpha', \beta') = (\alpha, \beta)$ or $(\alpha', \beta') = (\beta, \alpha)$.

This completes the proof of the theorem.

Lemma 8. *Let $\gamma_1, \ldots, \gamma_4$ be four reals. Let A be the matrix $(\sin(i\gamma_j))_{1\leq i,j\leq 4}$. Then*

$$\det A = 64 \prod_{k=1}^{4} \sin(\gamma_k) \prod_{1\leq i<j\leq 4} (\cos(\gamma_i) - \cos(\gamma_j)).$$

Proof. From matrix A, doing line modification $L_3 \leftarrow L_3 - L_1$, and $L_4 \leftarrow L_4 - L_2$, we obtain (recall that $\sin(2p) = 2\sin(p)\cos(p)$ and $\sin(p) - \sin(q) = 2\sin(\frac{p-q}{2})\cos(\frac{p+q}{2})$)

$$\det A = \begin{vmatrix} \sin(\gamma_1) & \sin(\gamma_2) & \sin(\gamma_3) & \sin(\gamma_4) \\ 2\sin(\gamma_1)\cos(\gamma_1) & 2\sin(\gamma_2)\cos(\gamma_2) & 2\sin(\gamma_3)\cos(\gamma_3) & 2\sin(\gamma_4)\cos(\gamma_4) \\ 2\sin(\gamma_1)\cos(2\gamma_1) & 2\sin(\gamma_2)\cos(2\gamma_2) & 2\sin(\gamma_3)\cos(2\gamma_3) & 2\sin(\gamma_4)\cos(2\gamma_4) \\ 2\sin(\gamma_1)\cos(3\gamma_1) & 2\sin(\gamma_2)\cos(3\gamma_2) & 2\sin(\gamma_3)\cos(3\gamma_3) & 2\sin(\gamma_4)\cos(3\gamma_4) \end{vmatrix}.$$

Using 4-linearity of the determinant:

$$\det A = 8 \left(\prod_{j=1}^{4} \sin(\gamma_j)\right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ \cos(\gamma_1) & \cos(\gamma_2) & \cos(\gamma_3) & \cos(\gamma_4) \\ \cos(2\gamma_1) & \cos(2\gamma_2) & \cos(2\gamma_3) & \cos(2\gamma_4) \\ \cos(3\gamma_1) & \cos(3\gamma_2) & \cos(3\gamma_3) & \cos(3\gamma_4) \end{vmatrix}.$$

Now, denote $x_k = \cos(\gamma_k)$ and remark that $\cos(i\gamma_k) = T_i(\cos\gamma_k) = T_i(x_k)$ where T_i is the ith Chebyshev polynomial: $T_0 = 1, T_1 = X, T_2 = 2X^2 - 1, T_3 = 4X^3 - 3X$. We have $T_2 + T_0 = 2X^2$ and $T_3 + 3T_1 = 4X^3$. Then, doing $L_3 \leftarrow L_3 + L_1$, and $L_4 \leftarrow L_4 + 3L_2$:

$$\begin{aligned} \det A &= 8 \left(\prod_{j=1}^{4} \sin(\gamma_j)\right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ 2x_1^2 & 2x_2^2 & 2x_3^2 & 2x_4^2 \\ 4x_1^3 & 4x_2^3 & 4x_3^3 & 4x_4^3 \end{vmatrix} \\ &= 64 \left(\prod_{j=1}^{4} \sin(\gamma_j)\right) \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{vmatrix}. \end{aligned}$$

This is a Vandermonde matrix, hence

$$\det A = 64 \left(\prod_{j=1}^{4} \sin(\gamma_j)\right) \prod_{1\leq i<j\leq 4} (x_i - x_j)$$

$$= 64 \prod_{k=1}^{4} \sin(\gamma_k) \prod_{1 \le i < j \le 4} (\cos(\gamma_i) - \cos(\gamma_j)). \qquad \square$$

Lemma 9. *Let $\gamma_1, \dots, \gamma_4$ be four reals. Let $\lambda_1, \dots, \lambda_4 \in \mathbb{R} \backslash \{0\}$ such that*

$$\sum_{k=1}^{4} \lambda_k \sin(l\gamma_k) = 0, \qquad l = 1, \dots, 4. \tag{8}$$

Then, one of the following cases holds:

1. *All γ_k are multiples of π.*
2. *Exactly two γ_k are multiples of π: $\gamma_{i_1} = \gamma_{i_2} = 0 \pmod{\pi}$ and $\gamma_{i_3} = \pm\gamma_{i_4} \pmod{2\pi}$.*
3. *Only one γ_k is multiple of π: $\gamma_{i_1} = 0 \pmod{\pi}$ and $\gamma_{i_2} = \pm\gamma_{i_3} = \pm\gamma_{i_4} \pmod{2\pi}$.*
4. *No γ_k is multiple of π and $\gamma_1 = \pm\gamma_2 = \pm\gamma_3 = \pm\gamma_4 \pmod{2\pi}$.*

Proof. First observe that, since $\sum_{k=1}^{4} \lambda_k \sin(l\gamma_k) = 0$ with $\lambda \neq 0_{\mathbb{R}^4}$, necessarily det($A$)=0 where $A = (\sin(i\gamma_j))_{1 \le i,j \le 4}$. Using Lemma 8

$$\prod_{k=1}^{4} \sin(\gamma_k) \prod_{1 \le i < j \le 4} (\cos(\gamma_i) - \cos(\gamma_j)) = 0. \tag{9}$$

Now, let us study the various cases that make this quantity vanish.

For the first case, note that if three γ_k are multiples of π: $\gamma_{i_1} = \gamma_{i_2} = \gamma_{i_3} = 0 \pmod{\pi}$ then equation (8) becomes $\lambda_{i_4} \sin(l\gamma_{i_4}) = 0$ and the last angle is also null modulo π.

In case 2., equation (8) entails

$$\lambda_{i_3} \sin(l\gamma_{i_3}) + \lambda_{i_4} \sin(l\gamma_{i_4}) = 0, \qquad l = 1, 2$$

with $\gamma_{i_3} \neq 0 \pmod{\pi}, \gamma_{i_4} \neq 0 \pmod{\pi}$. Then, since $(\lambda_{i_3}, \lambda_{i_4}) \neq (0, 0)$,

$$0 = \begin{vmatrix} \sin(\gamma_{i_3}) & \sin(\gamma_{i_4}) \\ \sin(2\gamma_{i_3}) & \sin(2\gamma_{i_4}) \end{vmatrix} = 2\sin(\gamma_{i_3})\sin(\gamma_{i_4})(\cos(\gamma_{i_4}) - \cos(\gamma_{i_3})).$$

Then $\cos(\gamma_{i_3}) = \cos(\gamma_{i_4})$. Either $\gamma_{i_3} = \gamma_{i_4} \pmod{2\pi}$, or $\gamma_{i_3} = -\gamma_{i_4} \pmod{2\pi}$.

Let us now study case 3. For the sake of simplicity we assume that $\gamma_4 = 0 \pmod{\pi}$ and $\gamma_k \neq 0 \pmod{\pi}$ for $k = 1, 2, 3$. Equation (8) gives

$$\lambda_1 \sin(l\gamma_1) + \lambda_2 \sin(l\gamma_2) + \lambda_3 \sin(l\gamma_3) = 0, \qquad l = 1, 2, 3.$$

With the same proof as Lemma 8, we obtain

$$\prod_{k=1}^{3} \sin(\gamma_k) \prod_{1 \le i < j \le 3} (\cos(\gamma_i) - \cos(\gamma_j)) = 0.$$

Then $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$ or $\gamma_1 = \pm\gamma_3 \pmod{2\pi}$ or $\gamma_2 = \pm\gamma_3 \pmod{2\pi}$. Moreover, if, for example, $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$ then

$$(\lambda_1 \pm \lambda_2)\sin(l\gamma_1) + \lambda_3 \sin(l\gamma_3) = 0, \qquad l = 1, 2$$

We are reduced to the previous case, then $\gamma_1 = \pm\gamma_3 \pmod{2\pi}$.

In the case 4., equation (9) becomes $\prod_{1\leq i<j\leq 4}(\cos(\gamma_i) - \cos(\gamma_j)) = 0$, which provides 6 possible equalities. Assume, for example, $\cos(\gamma_1) - \cos(\gamma_2) = 0$ and consequently $\gamma_1 = \pm\gamma_2 \pmod{2\pi}$. Then

$$(\lambda_1 \pm \lambda_2)\sin(l\gamma_1) + \lambda_3 \sin(l\gamma_3) + \lambda_4 \sin(l\gamma_4) = 0, \qquad l = 1, 2, 3.$$

Reasoning as in previous case, $\gamma_1 = \pm\gamma_3 = \pm\gamma_4 \pmod{2\pi}$. □

6.2. Proof of Theorem 4 (consistency)

This proof and the following are inspired from Butucea and Vandekerkhove (2014). Let us denote $\tilde{\Theta} = (0, 1/2) \times \mathbb{S}^1 \times \mathbb{S}^1$. Denote by $\dot{\phi}(\theta)$ the gradient of any function ϕ with respect to $\theta = (p, \alpha, \beta)$, and by $\ddot{\phi}(\theta)$ the Hessian matrix.

The proof of Theorem 4 relies on some preliminary results, given in the sequel.

Proposition 10. *Under Assumption 3 the contrast function S verifies the following properties: $S(\theta) \geq 0$, and $S(\theta) = 0$ if and only if $\theta = \theta_0$ or $\theta = \theta_0 + \pi$.*

Proof. It is clear that $S(\theta) \geq 0$ and that

$$S(\theta_0) = \sum_{l=-4}^{4} \left(\Im\left(g^{\star l}\overline{M^l(\theta_0)}\right)\right)^2 = \sum_{l=-4}^{4} \left(\Im\left(f^{\star l}|M^l(\theta_0)|^2\right)\right)^2 = 0.$$

By Lemma 3, if $\theta \neq \theta_0 \pmod{\pi}$, there exists $l_1 \in \{1, \ldots, 4\}$ such that $\Im\left(M^{l_1}(\theta_0)\overline{M^{l_1}(\theta)}\right) \neq 0$ so that $S(\theta) \geq \left(\Im\left(g^{\star l_1}\overline{M^{l_1}(\theta)}\right)\right)^2 > 0$. □

Lemma 11. *1. For all θ in $\tilde{\Theta}$, $|M^l(\theta)| \leq 1$.*

2. For all $1 \leq k \leq n$, for all l in $\mathbb{Z}$,

$$\sup_{\theta\in\tilde{\Theta}} |Z_k^l(\theta)| \leq \frac{1}{2\pi}, \qquad \sup_{\theta\in\tilde{\Theta}} |J^l(\theta)| \leq \frac{1}{2\pi}.$$

3. For all $1 \leq k \leq n$, for all l in $\mathbb{Z}$,

$$\sup_{\theta\in\tilde{\Theta}} \|\dot{Z}_k^l(\theta)\| \leq \frac{2+|l|}{\sqrt{2}\pi}, \qquad \sup_{\theta\in\tilde{\Theta}} \|\dot{J}^l(\theta)\| \leq \frac{2+|l|}{\sqrt{2}\pi}.$$

where $\|.\|$ is the Euclidean norm.

4. For all $1 \leq k \leq n$, for all l in $\mathbb{Z}$,

$$\sup_{\theta\in\tilde{\Theta}} \|\ddot{Z}_k^l(\theta)\|_F \leq \frac{|l|+l^2}{\pi}, \qquad \sup_{\theta\in\tilde{\Theta}} \|\ddot{J}^l(\theta)\|_F \leq \frac{|l|+l^2}{\pi}.$$

where $\|.\|_F$ is the Frobenius norm.

Proof. Point 1 is straightforward.

2. Let us start with $Z_k^l(\theta)$. We recall that $Z_k^l(\theta) = \Im\left(\frac{e^{ilX_k}}{2\pi} M^l(\theta)\right)$. Then

$$|Z_k^l(\theta)| \le \frac{1}{2\pi}|M^l(\theta)| \le \frac{1}{2\pi}.$$

Furthermore

$$|J^l(\theta)| \le |g^{\star l}||M^l(\theta)| \le \frac{1}{2\pi}\int_{\mathbb{S}_1} g \le \frac{1}{2\pi}.$$

3. We have

$$\dot{Z}_k^l(\theta) = \frac{1}{2\pi}\Im\left(e^{ilX_k}\dot{M}^l(\theta)\right) = \frac{1}{2\pi}\Im\left(e^{ilX_k}\begin{pmatrix} e^{-il\alpha} - e^{-il\beta} \\ -ilpe^{-i\alpha l} \\ -il(1-p)e^{-i\beta l}\end{pmatrix}\right)$$

and

$$\dot{J}^l(\theta) = \Im\left(\overline{g^{\star l}}\dot{M}^l(\theta)\right) = \Im\left(\overline{g^{\star l}}\begin{pmatrix} e^{-il\alpha} - e^{-il\beta} \\ -ilpe^{-i\alpha l} \\ -il(1-p)e^{-i\beta l}\end{pmatrix}\right).$$

We get

$$\|\dot{Z}_k^l(\theta)\| \le \frac{1}{2\pi}\left(2^2 + p^2l^2 + (1-p)^2l^2\right)^{1/2} \le \frac{2+|l|}{\sqrt{2}\pi}$$

and we have the same bound for $\|\dot{J}^l(\theta)\|$.

4. We have

$$\begin{aligned}\ddot{Z}_k^l(\theta) &= \Im\left(\frac{e^{ilX_k}}{2\pi}\ddot{M}^l(\theta)\right) \\ &= \Im\left(\frac{e^{ilX_k}}{2\pi}\begin{pmatrix} 0 & -ile^{-il\alpha} & ile^{-il\beta} \\ -ile^{-il\alpha} & -l^2pe^{-il\alpha} & 0 \\ ile^{-il\beta} & 0 & -l^2(1-p)e^{-i\beta l}\end{pmatrix}\right).\end{aligned}$$

Thus

$$\|\ddot{Z}_k^l(\theta)\|_F \le \frac{1}{2\pi}\left(4l^2 + l^4p^2 + l^4(1-p)^2\right)^{1/2} \le \frac{|l| + l^2}{\pi}.$$

We bound $\|\ddot{J}^l(\theta)\|_F$ in the same way. This ends the proof of the lemma. □

Lemma 12. *There exists a numerical positive constant C such that the following inequalities hold.*

1. For all $1 \le k \le n$, for all l in $\mathbb{Z}$

$$\forall \theta, \theta' \in \tilde{\Theta} \qquad \|\dot{Z}_k^l(\theta) - \dot{Z}_k^l(\theta')\| \le C\|\theta - \theta'\|(1 + |l| + l^2).$$

2. We also have

$$\|\ddot{Z}_k^l(\theta) - \ddot{Z}_k^l(\theta')\|_F \le C\|\theta - \theta'\|(1 + |l| + l^2 + |l|^3).$$

Proof. We use Taylor expansions at first order and then apply same bounding techniques as in Lemma 11. □

Lemma 13. *1. The function S is Lipschitz continuous over $\tilde{\Theta}$.*

2. The function $S_n(\theta)$ is Lipschitz continuous over $\tilde{\Theta}$.

3. The function $\ddot{S}_n(\theta)$ is Lipschitz continuous over $\tilde{\Theta}$ with respect to Frobenius norm, with Lipschitz constant not depending on n.

Proof. We will write C for a numerical constant that may change from line to line but is numerical.

Let us start with point 1. We recall that $S(\theta) = \sum_l J^l(\theta)^2$. Let θ and θ' in $\tilde{\Theta}$. As $\tilde{\Theta}$ is a convex set, we get, thanks to the mean value theorem

$$\begin{aligned} |S(\theta) - S(\theta')| &= \left|\sum_{l=-4}^{4} J^l(\theta)^2 - J^l(\theta')^2\right| = \left|2(\theta-\theta')^\top \sum_{l=-4}^{4} J^l(\theta_u)\dot{J}^l(\theta_u)\right| \\ &\leq C\|\theta-\theta'\| \sum_{l=-4}^{4}(1+|l|) \leq C\|\theta-\theta'\| \end{aligned}$$

with θ_u lying on the line connecting θ to θ', and using Lemma 11.

Let us shift to point 2. Due to the mean value theorem, we have

$$\begin{aligned} |S_n(\theta) - S_n(\theta')| &= \left|\frac{1}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} \left(Z_k^l(\theta)Z_j^l(\theta) - Z_k^l(\theta')Z_j^l(\theta')\right)\right| \\ &= \left|\frac{1}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} \left((\theta-\theta')^\top \nabla[Z_k^l(\theta)Z_j^l(\theta)]|_{\theta=\theta_u}\right)\right| \\ &= \left|\frac{2(\theta-\theta')^\top}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} \dot{Z}_k^l(\theta_u)Z_j^l(\theta_u)\right|, \end{aligned}$$

with θ_u lying on the line connecting θ to θ'. Then using 1. and 2. of Lemma 11 we get

$$|S_n(\theta) - S_n(\theta')| \leq \frac{C\|\theta-\theta'\|}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} (1+|l|) \leq C\|\theta-\theta'\|$$

which ends the proof of the second point.

Concerning point 3. we have that

$$\ddot{S}_n(\theta) = \frac{2}{n(n-1)} \sum_{k\neq j} \sum_{l=4}^{4} (\ddot{Z}_k^l(\theta)Z_j^l(\theta) + \dot{Z}_k^l(\theta)\dot{Z}_j^l(\theta)^\top).$$

Hence

$$\|\ddot{S}_n(\theta) - \ddot{S}_n(\theta')\|_F \leq \frac{2}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} \Big(\|(\ddot{Z}_k^l(\theta) - \ddot{Z}_k^l(\theta'))Z_j^l(\theta)\|_F$$

$$+\|\ddot{Z}_k^l(\theta')(Z_j^l(\theta)-Z_j^l(\theta'))\|_F+\|\dot{Z}_k^l(\theta')(\dot{Z}_j^l(\theta)-\dot{Z}_j^l(\theta')^\top)\|_F$$
$$+\|(\dot{Z}_k^l(\theta')-\dot{Z}_k^l(\theta))\dot{Z}_j^l(\theta)^\top\|_F\Big)$$

Using Taylor expansions and Lemma 11 and 12, we get that

$$\|\ddot{S}_n(\theta)-\ddot{S}_n(\theta')\|_F \leq C\|\theta-\theta'\|\sum_{l=-4}^{4}(1+|l|+l^2+|l|^3). \qquad \square$$

Proposition 14. *There exist a positive constant C such that*

$$\sup_{\theta\in\tilde{\Theta}}\mathbb{E}[(S_n(\theta)-S(\theta))^2] \leq \frac{C}{n}.$$

Proof. The definitions of S_n and S provide

$$S_n(\theta)-S(\theta)=\frac{1}{n(n-1)}\sum_{l=-4}^{4}\sum_{k\neq j}\left(Z_k^l(\theta)Z_j^l(\theta)-J^l(\theta)^2\right)=T_n+V_n$$

where

$$T_n=\frac{2}{n(n-1)}\sum_{l=-4}^{4}\sum_{k<j}(Z_k^l(\theta)-J^l(\theta))(Z_j^l(\theta)-J^l(\theta))$$

and

$$V_n=\frac{2}{n}\sum_{l=-4}^{4}\sum_{k=1}^{n}(Z_k^l(\theta)-J^l(\theta))J^l(\theta).$$

Note that $\mathbb{E}(Z_k^l(\theta)-J^l(\theta))=0$ which entails $\mathbb{E}[T_nV_n]=0$. Then

$$\mathbb{E}\left[(S_n(\theta)-S(\theta))^2\right]=\mathbb{E}\left[(T_n+V_n)^2\right]=\mathbb{E}\left[T_n^2\right]+\mathbb{E}\left[V_n^2\right].$$

Now, since the variables $\left(\sum_{l=-4}^{4}(Z_k^l(\theta)-J^l(\theta))(Z_j^l(\theta)-J^l(\theta))\right)_{k<j}$ are uncorrelated,

$$\begin{aligned}\mathbb{E}[T_n^2] &= \frac{2}{n(n-1)}\mathbb{E}\left[\left(\sum_{l=-4}^{4}(Z_1^l(\theta)-J^l(\theta))(Z_2^l(\theta)-J^l(\theta))\right)^2\right]\\ &\leq \frac{2}{n(n-1)}\mathbb{E}\left[\left(\sum_{l=-4}^{4}\frac{2}{2\pi}\cdot\frac{2}{2\pi}\right)^2\right]\leq\frac{C}{2n}\end{aligned}$$

using Lemma 11. We focus now on V_n: in the same way

$$\mathbb{E}[V_n^2] = \frac{4}{n}\mathbb{E}\left[\left(\sum_{l=-4}^{4}(Z_1^l(\theta)-J^l(\theta))J^l(\theta)\right)^2\right]$$

$$\leq \quad \frac{4}{n}\mathbb{E}\left[\left(\sum_{l=-4}^{4}\frac{2}{2\pi}\cdot\frac{1}{2\pi}\right)^2\right] \leq \frac{C}{2n},$$

using Lemma 11 again. □

Theorem 4 is finally proved using the following lemma, its assumptions being ensured by Proposition 10, Lemma 13 and Proposition 14.

Lemma 15. *Assume that Θ is a compact set and let $S : \Theta \to \mathbb{R}$ be a continuous function. Assume that*

$$S(\theta) = \min_{\Theta} S \Leftrightarrow \theta = \theta_0 \text{ or } \theta = \theta_0'$$

where $\theta_0, \theta_0' \in \Theta$. Let $S_n : \Theta \to \mathbb{R}$ be a function which is uniformly continuous and such that for all θ $|S_n(\theta) - S(\theta)|$ tends to 0 in probability. Let $\hat{\theta}_n$ be a point such that $S_n(\hat{\theta}_n) = \inf_\Theta S_n$. Then $\hat{\theta}_n \to \theta_0$ or θ_0' in probability.

This is a classical result in the theory of minimum contrast estimators, when $\theta_0 = \theta_0'$ (see van der Vaart (1998) or Dacunha-Castelle and Duflo (1986)). We reproduce the proof since it is slightly adapted to the case of two argmins.

Proof. Let $\epsilon > 0$ and B be the union of the open ball with center θ_0 and radius ϵ and the open ball with center θ_0' and radius ϵ. Since S is continuous and $B^c \subset \Theta$ is a compact set, there exists $\theta_\epsilon \in B^c$ such that $S(\theta_\epsilon) = \inf_{B^c} S$. Using the assumption, since $\theta_\epsilon \neq \theta_0$, and $\theta_\epsilon \neq \theta_0'$

$$\delta := S(\theta_\epsilon) - S(\theta_0) > 0.$$

Since S_n is uniformly continuous, there exists $\alpha > 0$ such that

$$\forall \theta, \theta' \qquad \|\theta - \theta'\| < \alpha \Rightarrow |S_n(\theta) - S_n(\theta')| \leq \delta/2.$$

Moreover B^c is a compact set then there exists a finite set (θ_i) such that $B^c \subset \cup_{i=1}^I B(\theta_i, \alpha)$. Denote $\Delta_n := \max_{0\leq i\leq I} |S_n(\theta_i) - S(\theta_i)|$. The assumption ensures that Δ_n tends to 0 in probability. Let $\theta \in B^c$. There exists $1 \leq i \leq I$ such that $\|\theta - \theta_i\| < \alpha$, and then $|S_n(\theta) - S_n(\theta_i)| \leq \delta/2$. Thus

$$\begin{aligned} S_n(\theta) - S_n(\theta_0) &= (S_n(\theta) - S_n(\theta_i)) + (S_n(\theta_i) - S(\theta_i)) \\ &\quad + (S(\theta_i) - S(\theta_0)) + (S(\theta_0) - S_n(\theta_0)) \\ &\geq -\delta/2 - \Delta_n + \delta - \Delta_n \end{aligned}$$

using that $S(\theta_i) - S(\theta_0) \geq S(\theta_\epsilon) - S(\theta_0) = \delta$. Then

$$\inf_{\theta\in B^c} S_n(\theta) - S_n(\theta_0) \geq \delta/2 - 2\Delta_n.$$

Now, if $\|\hat{\theta}_n - \theta_0\| \geq \epsilon$ and $\|\hat{\theta}_n - \theta_0'\| \geq \epsilon$ then $\hat{\theta}_n \in B^c$ and

$$\inf_{\theta\in\Theta} S_n(\theta) = S_n(\hat{\theta}_n) = \inf_{\theta\in B^c} S_n(\theta).$$

In particular $\inf_{\theta \in B^c} S_n(\theta) \leq S_n(\theta_0)$ so that

$$\begin{aligned}\mathbb{P}(\|\hat{\theta}_n - \theta_0\| \geq \epsilon \text{ and } \|\hat{\theta}_n - \theta_0'\| \geq \epsilon) &\leq \mathbb{P}(0 \geq \inf_{\theta \in B^c} S_n(\theta) - S_n(\theta_0) \geq \delta/2 - 2\Delta_n) \\ &\leq \mathbb{P}(\Delta_n \geq \delta/4) \longrightarrow 0\end{aligned}$$

since Δ_n tends to 0 in probability. □

6.3. Proof of Theorem 5 (asymptotic normality)

The Taylor's theorem and the definition of $\hat{\theta}_n$ give

$$\dot{S}_n(\hat{\theta}_n) = \dot{S}_n(\theta_0) + \ddot{S}_n(\theta_n^*)(\hat{\theta}_n - \theta_0) = 0,$$

where θ_n^* lies in the line segment with extremities θ_0 and $\hat{\theta}_n$. Equivalently we have,

$$\ddot{S}_n(\theta_n^*)(\hat{\theta}_n - \theta_0) = -\dot{S}_n(\theta_0).$$

We recall that

$$S_n(\theta_0) = \frac{1}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^{4} Z_k^l(\theta_0) Z_j^l(\theta_0)$$

and

$$\dot{S}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^{4} \dot{Z}_k^l(\theta_0) Z_j^l(\theta_0)$$

and

$$\ddot{S}_n(\theta_0) = \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^{4} \ddot{Z}_k^l(\theta_0) Z_j^l(\theta_0) + \dot{Z}_k^l(\theta_0) \dot{Z}_j^l(\theta_0)^\top.$$

Step 1- Let us prove that

$$\sqrt{n}\dot{S}_n(\theta_0) \xrightarrow{d} \mathcal{N}(0, V).$$

We remind by Lemma 3 that $J^l(\theta_0) = 0$. Hence

$$\mathbb{E}(\dot{S}_n(\theta_0)) = 2 \sum_{l=-4}^{4} \dot{J}^l(\theta_0) J^l(\theta_0) = 0.$$

We can break down $\dot{S}_n(\theta_0)$ in the following way:

$$\begin{aligned}\dot{S}_n(\theta_0) &= \frac{2}{n(n-1)} \sum_{k \neq j} \sum_{l=-4}^{4} (\dot{Z}_k^l(\theta_0) - \dot{J}^l(\theta_0) + \dot{J}^l(\theta_0)) Z_j^l(\theta_0) \\ &= \frac{4}{n(n-1)} \sum_{k<j} \sum_{l=-4}^{4} (\dot{Z}_k^l(\theta_0) - \dot{J}^l(\theta_0)) Z_j^l(\theta_0) + \frac{2}{n} \sum_{j=1}^{n} \sum_{l=-4}^{4} \dot{J}^l(\theta_0) Z_j^l(\theta_0)\end{aligned}$$

$$=: \quad A_n + B_n.$$

Note that A_n and B_n are centered variables. Let us show that $\sqrt{n}A_n = o_P(1)$. Note that the variables $W_{jk} := \left(\sum_{l=-4}^{4}(\dot{Z}_k^l(\theta_0) - \dot{J}^l(\theta_0))Z_j^l(\theta_0)\right)_{k<j}$ are centered and uncorrelated. Then

$$\mathbb{E}(\|A_n\|^2) = \mathbb{E}\left(\left\|\frac{4}{n(n-1)}\sum_{k<j} W_{jk}\right\|^2\right) = \frac{8}{n(n-1)}\mathbb{E}\|W_{12}\|^2.$$

Using Lemma 11, there exists $C > 0$ such that

$$\|W_{12}\| \leq \sum_{l=-4}^{4} \frac{2(1+|l|)}{\sqrt{2}\pi}\frac{1}{2\pi} \leq C$$

so that $\mathbb{E}(\|\sqrt{n}A_n\|^2) \leq 8C^2/(n-1)$. Finally, invoking Markov inequality we have that $\sqrt{n}A_n = o_P(1)$. We can write $\sqrt{n}B_n$ in the following way:

$$\sqrt{n}B_n = \frac{2}{\sqrt{n}}\sum_{k=1}^{n} U_k(\theta_0),$$

where we set $U_k(\theta_0) := \sum_{l=-4}^{4} \dot{J}^l(\theta_0)Z_k^l(\theta_0)$. Note that the $U_k(\theta_0)$'s are i.i.d and centered. Invoking the central limit theorem, we have that

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n} U_k(\theta_0) \xrightarrow{d} \mathcal{N}(0, V/4),$$

where $V/4$ is the covariance matrix of $U_1(\theta_0)$, equal to $\mathbb{E}(U_1(\theta_0)U_1(\theta_0)^\top)$.

Step 2- Let us prove that $\ddot{S}_n(\theta_n^*) \xrightarrow{P} \mathcal{A}(\theta_0)$ where $\mathcal{A}(\theta_0) = 2\sum_{l=-4}^{4} \dot{J}^l(\theta_0)$ $\dot{J}^l(\theta_0)^\top$. First, we have

$$\begin{aligned}\mathbb{E}(\ddot{S}_n(\theta_0)) &= \ddot{S}(\theta_0) = 2\sum_{l=-4}^{4}(\ddot{J}^l(\theta_0)\underbrace{J^l(\theta_0)}_{=0} + \dot{J}^l(\theta_0)\dot{J}^l(\theta_0)^\top) \\ &= 2\sum_{l=-4}^{4}\dot{J}^l(\theta_0)\dot{J}^l(\theta_0)^\top = \mathcal{A}(\theta_0).\end{aligned}$$

Next we write the decomposition

$$\|\ddot{S}_n(\theta_n^*) - \mathcal{A}(\theta_0)\|_F \leq \|\ddot{S}_n(\theta_n^*) - \ddot{S}_n(\theta_0)\|_F + \|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F.$$

We get due to the Lipschitz property of $\ddot{S}_n$ stated in Lemma 13 that

$$\mathbb{P}\left(\|\ddot{S}_n(\theta_0) - \ddot{S}_n(\theta_n^*)\|_F \geq \varepsilon\right) \leq \mathbb{P}\left(K\|\theta_n^* - \theta_0\| \geq \varepsilon\right) \to 0,$$

because $\hat{\theta}_n \xrightarrow{P} \theta_0$. Last, let us focus on the term $\|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F$. We remind that

$$\begin{aligned}
&\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0) \\
&\quad = \frac{2}{n(n-1)} \sum_{k\neq j} \sum_{l=-4}^{4} \left(\ddot{Z}_k^l(\theta_0) Z_j^l(\theta_0) + \dot{Z}_k^l(\theta_0) \dot{Z}_j^l(\theta_0)^\top - \dot{J}^l(\theta_0)\dot{J}^l(\theta_0)^\top \right).
\end{aligned}$$

From now on, we drop indices l and θ_0 to simplify the notation. We center the variables in order to find uncorrelatedness:

$$\begin{aligned}
\ddot{Z}_k Z_j + \dot{Z}_k \dot{Z}_j^\top - \dot{J}\dot{J}^\top = \underbrace{\left(\ddot{Z}_k - \ddot{J}\right) Z_j}_{A} + \underbrace{\ddot{J} Z_j}_{B} + \underbrace{(\dot{Z}_k - \dot{J})(\dot{Z}_j - \dot{J})^\top}_{C} \\
+ \underbrace{\dot{J}(\dot{Z}_j - \dot{J})^\top}_{D} + \underbrace{(\dot{Z}_k - \dot{J})(\dot{J})^\top}_{E}
\end{aligned}$$

(remind that $\mathbb{E}(Z_j) = J^l(\theta_0) = 0$). Then $\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0) = 2\sum_{l=-4}^{4}(A + B + C + D + E)$ where

$$\begin{aligned}
A &= \frac{2}{n(n-1)} \sum_{k<j} \left(\ddot{Z}_k - \ddot{J}\right) Z_j \\
B &= \frac{1}{n} \sum_{j=1}^{n} \ddot{J} Z_j \\
C &= \frac{2}{n(n-1)} \sum_{k<j} (\dot{Z}_k - \dot{J})(\dot{Z}_j - \dot{J})^\top \\
D &= \frac{1}{n} \sum_{j=1}^{n} \dot{J}(\dot{Z}_j - \dot{J})^\top \\
E &= \frac{1}{n} \sum_{k=1}^{n} (\dot{Z}_k - \dot{J})\dot{J}^\top = D^\top
\end{aligned}$$

Using the weak law of large numbers for uncorrelated centered variables, we obtain that $\|\ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0)\|_F \xrightarrow{P} 0$ which completes the step 2.

Finally it is sufficient to apply Slutsky's Lemma to obtain the theorem.

6.4. Estimation of the covariance

Proposition 16. *Consider notation and assumptions of Theorem 5. Let* $V = 4\mathbb{E}(U_1 U_1^\top)$ *where* $U_1 = \sum_{l=-4}^{4} Z_1^l(\theta_0)\dot{J}^l(\theta_0)$*. Then*

$$\frac{4}{n^3} \sum_{1\leq k,j,j'\leq n} \sum_{-4\leq l,l'\leq 4} Z_k^l(\hat{\theta}_n) Z_k^{l'}(\hat{\theta}_n) \dot{Z}_j^l(\hat{\theta}_n) (\dot{Z}_{j'}^{l'}(\hat{\theta}_n))^\top.$$

tends almost surely toward V *when* n *tends to* $+\infty$.

Thus we obtain a consistent estimator for V (that allows to estimate the covariance Σ). Nevertheless this estimator is biased. Notice that the quantity

$$\frac{4}{n(n-1)(n-2)}\sum_{k=1}^{n}\sum_{j\neq k}\sum_{j'\notin\{k,j\}}\sum_{-4\leq l,l'\leq 4} Z_k^l(\theta_0)Z_k^{l'}(\theta_0)\dot{Z}_j^l(\theta_0)(\dot{Z}_{j'}^{l'}(\theta_0))^\top$$

has expectation

$$4\sum_{-4\leq l,l'\leq 4}\mathbb{E}[Z_1^l(\theta_0)Z_1^{l'}(\theta_0)]\dot{J}^l(\theta_0)(\dot{J}^{l'}(\theta_0))^\top = V$$

and we could also prove (with some additional technicalities in the following proof about the uniform convergence in k) that it tends almost surely toward V. However, we lose the "unbiased" property when replacing θ_0 by $\hat{\theta}_n$.

Proof of Proposition 16

Let $U_k = \sum_{l=-4}^{4} Z_k^l(\theta_0)\dot{J}^l(\theta_0)$. The law of large numbers gives

$$V = \mathbb{E}(4U_1U_1^\top) = \lim_{n\to\infty}\frac{4}{n}\sum_{k=1}^{n}U_kU_k^\top$$

where the convergence is almost sure. Moreover

$$U_kU_k^\top = \sum_{-4\leq l,l'\leq 4} Z_k^lZ_k^{l'}\dot{J}^l(\dot{J}^{l'})^\top = \lim_{n\to\infty}\frac{1}{n^2}\sum_{-4\leq l,l'\leq 4}Z_k^lZ_k^{l'}\sum_{1\leq j,j'\leq n}\dot{Z}_j^l(\dot{Z}_{j'}^{l'})^\top$$

where the convergence is almost sure and we have dropped the θ_0 for the sake of simplicity. This convergence is uniform in k in the following sense: there exists a set with probability 1 for which for any $\varepsilon > 0$, there exists $N \geq 1$ such that for all $n \geq N$ and for all $1 \leq k \leq n$

$$\left\|\frac{1}{n^2}\sum_{-4\leq l,l'\leq 4}Z_k^lZ_k^{l'}\sum_{1\leq j,j'\leq n}\dot{Z}_j^l(\dot{Z}_{j'}^{l'})^\top - U_kU_k^\top\right\| \leq \varepsilon$$

Indeed

$$\begin{aligned}
&\left\|\frac{1}{n^2}\sum_{-4\leq l,l'\leq 4}Z_k^lZ_k^{l'}\sum_{1\leq j,j'\leq n}\dot{Z}_j^l(\dot{Z}_{j'}^{l'})^\top - U_kU_k^\top\right\| \\
&= \left\|\sum_{-4\leq l,l'\leq 4}Z_k^lZ_k^{l'}\left(\frac{1}{n^2}\sum_{1\leq j,j'\leq n}\dot{Z}_j^l(\dot{Z}_{j'}^{l'})^\top - \dot{J}^l(\dot{J}^{l'})^\top\right)\right\| \\
&\leq \frac{1}{4\pi^2}\sum_{-4\leq l,l'\leq 4}\left\|\left(\frac{1}{n^2}\sum_{1\leq j,j'\leq n}\dot{Z}_j^l(\dot{Z}_{j'}^{l'})^\top - \dot{J}^l(\dot{J}^{l'})^\top\right)\right\|.
\end{aligned}$$

Then we use the following lemma: "If $v_{nk} \to v_k$ uniformly, with (v_{nk}) and (v_k) bounded, and if $n^{-1}\sum_{k=1}^n v_k \to v$ then $n^{-1}\sum_{k=1}^n v_{nk} \to v$." To prove this lemma, notice that, for a given positive ε, for n large enough

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^n v_{nk} - v\right| &\leq \frac{1}{n}\sum_{k=1}^N |v_{nk}-v_k| + \frac{1}{n}\sum_{k=N+1}^n |v_{nk}-v_k| + \left|\frac{1}{n}\sum_{k=1}^n v_k - v\right| \\ &\leq \frac{N}{n}(\sup_{kn}|v_{nk}| + \sup_k |v_k|) + \frac{n-N}{n}\varepsilon + \varepsilon \leq 3\varepsilon.\end{aligned}$$

That provides

$$V = \lim_{n\to\infty} \frac{4}{n^3} \sum_{1\leq k,j,j'\leq n} \sum_{-4\leq l,l'\leq 4} Z_k^l Z_k^{l'} \dot{Z}_j^l (\dot{Z}_{j'}^{l'})^\top$$

where the convergence is almost sure. Here all the Z_k's are depending on θ_0, but we can use the consistency of $\hat{\theta}_n$ to finally assert

$$V = \lim_{n\to\infty} \frac{4}{n^3} \sum_{1\leq k,j,j'\leq n} \sum_{-4\leq l,l'\leq 4} Z_k^l(\hat{\theta}_n) Z_k^{l'}(\hat{\theta}_n) \dot{Z}_j^l(\hat{\theta}_n) (\dot{Z}_{j'}^{l'}(\hat{\theta}_n))^\top.$$

6.5. Proof of Proposition 6

We use the proof of Theorem 5. We have seen that

$$\ddot{S}_n(\theta_n^*)(\hat{\theta}_n - \theta_0) = -\dot{S}_n(\theta_0),$$

with θ_n^* in the line segment with extremities θ_0 and $\hat{\theta}_n$. Recall that $\dot{S}_n(\theta_0) = A_n + B_n$ with

$$\begin{aligned} A_n &= \frac{4}{n(n-1)} \sum_{k<j} \sum_{l=-4}^4 (\dot{Z}_k^l(\theta_0) - \dot{J}^l(\theta_0)) Z_j^l(\theta_0) \\ B_n &= \frac{2}{n}\sum_{k=1}^n U_k(\theta_0)\end{aligned}$$

where $U_k(\theta_0) := \sum_{l=-4}^4 \dot{J}^l(\theta_0) Z_k^l(\theta_0)$. Note that the $U_k(\theta_0)$'s are i.i.d and centered so that

$$\mathbb{E}\left\|\sum_k U_k(\theta_0)\right\|^2 = \sum_{j=1}^3 \mathrm{Var}\left(\sum_k U_{kj}(\theta_0)\right) = \sum_{j=1}^3 n\mathrm{Var}\left(U_{1j}(\theta_0)\right) \leq nc_1$$

using Lemma 11. Here c_1 is a numerical constant. Thus $\mathbb{E}\|B_n\|_1^2 \leq 4c_1/n$. In the same way the variables $W_{jk} := \left(\sum_{l=-4}^4 (\dot{Z}_k^l(\theta_0) - \dot{J}^l(\theta_0)) Z_j^l(\theta_0)\right)_{k<j}$ are centered and uncorrelated, and also bounded. Then

$$\mathbb{E}(\|n(n-1)A_n\|^2) = \mathbb{E}\left(\left\|\sum_{k<j} W_{jk}\right\|^2\right) = n(n-1)\mathbb{E}\left(\|W_{12}\|^2\right) \leq n(n-1)c_2.$$

Then $\mathbb{E}\left(\|A_n\|^2\right) \leq c_2/n$ and $\sup_n \mathbb{E}\left(n\|\dot{S}_n(\theta_0)\|^2\right) \leq 8c_1 + 2c_2 < \infty$.

In the proof of Theorem 5, we noted that $\ddot{S}_n(\theta_n^*)$ tends to $\ddot{S}(\theta_0)$ in probability. Actually we can prove that the convergence is almost sure. Indeed the strong law of large numbers is true for uncorrelated variables if their second moments have a common bound (see e.g. Chung and Zhong (2001)) so that

$$\ddot{S}_n(\theta_0) - \ddot{S}(\theta_0) = \ddot{S}_n(\theta_0) - \mathbb{E}\ddot{S}_n(\theta_0) \xrightarrow{a.s.} 0.$$

Since $\ddot{S}_n$ is continuous, it is sufficient to show that the convergence of $\hat{\theta}_n$ towards θ is almost sure and this will imply that $\ddot{S}_n(\theta_n^*)$ converges almost surely towards $\ddot{S}_n(\theta_0)$ (recall that θ_n^* in the line segment with extremities θ_0 and $\hat{\theta}_n$). To do this, remark first that $S_n(\theta) - S(\theta) \xrightarrow{a.s.} 0$ by the strong law of large numbers for uncorrelated variables again (see the decomposition of $S_n - S$ in the proof of Proposition 14). Now, we come back to the proof of Lemma 15 (in the case of a unique minimum θ_0), with this new assumption that $S_n(\theta)$ tends to $S(\theta)$ almost surely. The proof shows that for any $\epsilon > 0$ there exist $\delta(\epsilon) > 0$ and $\Delta_n(\epsilon)$ which tends to 0 almost surely such that

$$\|\hat{\theta}_n - \theta_0\| \geq \epsilon \Rightarrow \Delta_n(\epsilon) \geq \delta(\epsilon)/4.$$

Let $\Gamma = \cap_{p\geq 1}\{\Delta_n(1/p) \to 0\}$. This set has probability 1 and on this set, for any $\varepsilon > 0$, taking $p \geq 1/\varepsilon$, there exists $N \geq 1$ such that for any $n \geq N$

$$\Delta_n(1/p) < \delta(1/p)/4 \text{ and then } \|\hat{\theta}_n - \theta_0\| < (1/p) \leq \varepsilon.$$

This ensures that on the set Γ, $\hat{\theta}_n$ tends to θ_0, and finally $\ddot{S}_n(\theta_n^*)$ tends to $\ddot{S}(\theta_0)$ almost surely.

Now, since $\ddot{S}(\theta_0)$ is assumed invertible, there exists n_1 such that for all $n \geq n_1$, $\ddot{S}_n(\theta_n^*)$ is invertible and $\|\ddot{S}_n(\theta_n^*)^{-1}\|_{op} \leq 2\|\ddot{S}_n(\theta_0)^{-1}\|_{op} := C(\theta_0)$ a.s. Then

$$n\|\hat{\theta}_n - \theta_0\|^2 \leq C(\theta_0)^2 n\|\dot{S}_n(\theta_0)\|^2 \quad \text{a.s.}$$

and

$$\mathbb{E}(n\|\hat{\theta}_n - \theta_0\|^2) \leq C(\theta_0)^2\mathbb{E}(n\|\dot{S}_n(\theta_0)\|^2) \leq C(\theta_0)^2(8c_1 + 2c_2).$$

Moreover $\sup_{\theta\in\Theta} C(\theta) < \infty$ because Θ is a compact set and $\theta \mapsto \|\ddot{S}_n(\theta)^{-1}\|_{op}$ is continuous.

6.6. Proof of Theorem 7 (nonparametric estimation)

The proof of the oracle inequality is based on Lemma 17 and Lemma 18 below. The conclusion follows, choosing $2\gamma = \epsilon/(1+\epsilon)$ and $\lambda = \gamma^{-1}\kappa(1-2P)^{-2} = 2\kappa(1+\epsilon^{-1})(1-2P)^{-2}$, and $q = (2s_0+1)/3$.

Let us derive the rate of convergence, which is the second result of Theorem 7. We use the notation of Lemma 17 and the notation $\|.\|_\ell$ for the natural norm of $\ell^2(\mathbb{C}^{\mathbb{Z}})$. Let $L \in \mathcal{L}$. Since $\nu_n(t) = \sum_{l\in\mathbb{Z}} \overline{t_l}(\widehat{f^{\star l}} - f^{\star l})$,

$$\sum_{l=-L}^{L} |\widehat{f^{\star l}} - f^{\star l}|^2 = \nu_n(\widehat{f_L^\star} - f_L^\star) \leq \sup_{t\in B_L} \nu_n(t)\|\widehat{f_L^\star} - f_L^\star\|_\ell$$

where we denote by $f_L^\star$ the sequence in $\mathbb{C}^\mathbb{Z}$ such that $(f_L^\star)_l = f^{\star l}$ if $-L \leq l \leq L$ and 0 otherwise. Hence $\|\widehat{f_L^\star} - f_L^\star\|_\ell^2 \leq \sup_{t\in B_L} \nu_n(t)\|\widehat{f_L^\star} - f_L^\star\|_\ell$ so that $\|\widehat{f_L^\star} - f_L^\star\|_\ell \leq \sup_{t\in B_L} \nu_n(t)$. Then, using Lemma 18

$$\begin{aligned}\mathbb{E}\sum_{l=-L}^{L} |\widehat{f^{\star l}} - f^{\star l}|^2 &= \mathbb{E}\|\widehat{f_L^\star} - f_L^\star\|_\ell^2 \leq \frac{\kappa}{(1-2P)^2}\frac{2L+1}{n} + \frac{C(1+R^2)}{n} \\ &\leq C'(1+R^2)\frac{2L+1}{n}.\end{aligned}$$

Using Parseval's identity,

$$\begin{aligned}&\mathbb{E}\|f - \hat{f}_L\|_2^2 \\ &\quad = \sum_{|l|>L} |f^{\star l}|^2 + C'(1+R^2)\frac{2L+1}{n} \leq R^2(1+L^2)^{-s} + C'(1+R^2)\frac{2L+1}{n}.\end{aligned}$$

Thus, the oracle inequality gives

$$\begin{aligned}\mathbb{E}\|\hat{f}_{\hat{L}} - f\|_2^2 &\leq (1+2\epsilon)\min_{L\in\mathcal{L}}\left\{R^2(1+L^2)^{-s} + (C'(1+R^2)+2\lambda)\frac{2L+1}{n}\right\} \\ &\quad + \frac{C(1+R^2)}{n} \\ &\leq C''R^2 n^{-2s/(2s+1)}\end{aligned}$$

choosing $L = L_0 = \lfloor Cn^{1/(2s+1)}\rfloor$. This choice is possible since $s \geq s_0$ and then L_0 belongs to $\mathcal{L}$.

Lemma 17. *Let $\lambda > 0$ and $\mathcal{L}$ be a finite set of resolution level and define*

$$\widehat{L} = \operatorname*{argmin}_{L\in\mathcal{L}}\left\{-\sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + \lambda\frac{2L+1}{n}\right\}.$$

Then, for all $0 < \gamma < 1/2$,

$$\begin{aligned}(1-2\gamma)\|\hat{f}_{\hat{L}} - f\|_2^2 \leq \min_{L\in\mathcal{L}}&\left\{(1+2\gamma)\|\hat{f}_L - f\|_2^2 + 2\lambda\frac{2L+1}{n}\right\} \\ &+\frac{1}{\gamma}\max_{L\in\mathcal{L}}\left(\sup_{t\in B_L}\nu_n^2(t) - \lambda\gamma\frac{2L+1}{n}\right)\end{aligned}$$

where $B_L = \{t \in \mathbb{C}^\mathbb{Z},\ \sum_{l\in\mathbb{Z}} |t_l|^2 = 1,\ t_l = 0$ if $|l| > L\}$ and $\nu_n(t) = \sum_{l\in\mathbb{Z}} \overline{t_l}(\widehat{f^{\star l}} - f^{\star l})$.

Proof. We recall that the dot product $\langle f, g\rangle$ means $\frac{1}{2\pi}\int \overline{f(x)}g(x)dx$ and that $\|.\|_2$ is the associated norm. Usual Fourier analysis gives for any L:

$$\|\hat{f}_L - f\|_2^2 = -\|\hat{f}_L\|_2^2 + 2(\|\hat{f}_L\|_2^2 - \langle\hat{f}_L, f\rangle) + \|f\|_2^2$$

$$= -\sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + 2\sum_{l=-L}^{L} \overline{\widehat{f^{\star l}}}(\widehat{f^{\star l}} - f^{\star l}) + \|f\|_2^2$$
$$= -\sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + 2\nu_n(\widehat{f}_L^\star) + \|f\|_2^2$$

where we denote by $\widehat{f}_L^\star$ the sequence in $\mathbb{C}^{\mathbb{Z}}$ such that $(\widehat{f}_L^\star)_l = \widehat{f^{\star l}}$ if $-L \le l \le L$ and 0 otherwise.

Now let L be an arbitrary resolution level in $\mathcal{L}$. Using the definition of $\widehat{L}$,

$$-\sum_{l=-\widehat{L}}^{\widehat{L}} |\widehat{f^{\star l}}|^2 + \lambda\frac{2\widehat{L}+1}{n} \le -\sum_{l=-L}^{L} |\widehat{f^{\star l}}|^2 + \lambda\frac{2L+1}{n}.$$

Thus

$$\|\hat{f}_{\widehat{L}} - f\|_2^2 - 2\nu_n(\hat{f}_{\widehat{L}}^\star) + \lambda\frac{2\widehat{L}+1}{n} \le \|\hat{f}_L - f\|_2^2 - 2\nu_n(\hat{f}_L^\star) + \lambda\frac{2L+1}{n}$$

which leads to

$$\|\hat{f}_{\widehat{L}} - f\|_2^2 \le \|\hat{f}_L - f\|_2^2 + 2\nu_n(\hat{f}_{\widehat{L}}^\star - \hat{f}_L^\star) - \lambda\frac{2\widehat{L}+1}{n} + \lambda\frac{2L+1}{n}.$$

But, denoting by $\|.\|_\ell$ the natural norm of $\ell^2(\mathbb{C}^{\mathbb{Z}})$

$$2\nu_n(\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star) = 2\nu_n\left(\frac{\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star}{\|\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star\|_\ell}\right)\|\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star\|_\ell$$
$$2\left|\nu_n(\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star)\right| \le \gamma\|\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star\|_\ell^2 + \frac{1}{\gamma}\left|\nu_n\left(\frac{\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star}{\|\widehat{f}_{\widehat{L}}^\star - \widehat{f}_L^\star\|_\ell}\right)\right|^2$$
$$\le 2\gamma(\|\hat{f}_{\widehat{L}} - f\|_2^2 + \|f - \hat{f}_L\|_2^2) + \frac{1}{\gamma}\sup_{t\in B_{L\vee\widehat{L}}} |\nu_n(t)|^2$$

where $L \vee \widehat{L} = \max(L, \widehat{L})$. Thus

$$\begin{aligned}
\|\hat{f}_{\widehat{L}} - f\|_2^2(1-2\gamma) \quad &\le \quad \|\hat{f}_L - f\|_2^2(1+2\gamma) \\
&+ \quad \frac{1}{\gamma}\sup_{t\in B_{L\vee\widehat{L}}} |\nu_n(t)|^2 - \lambda\frac{2\widehat{L}+1}{n} + \lambda\frac{2L+1}{n} \\
&\le \quad \|\hat{f}_L - f\|_2^2(1+2\gamma) \\
&+ \quad \frac{1}{\gamma}\left(\sup_{t\in B_{L\vee\widehat{L}}} |\nu_n(t)|^2 - \lambda\gamma\frac{2\widehat{L}+2L+2}{n}\right) + 2\lambda\frac{2L+1}{n} \\
&\le \quad \|\hat{f}_L - f\|_2^2(1+2\gamma) + 2\lambda\frac{2L+1}{n} \\
&+ \quad \frac{1}{\gamma}\max_{L'\in\mathcal{L}}\left(\sup_{t\in B_{L'}} |\nu_n(t)|^2 - \lambda\gamma\frac{2L'+1}{n}\right). \quad \square
\end{aligned}$$

Lemma 18. *Assume Assumption 1 and Assumption 5. Assume that f belongs to the Sobolev ellipsoid $W(s,R)$ with $s \geq 1$. Assume that $\mathcal{L} = \{0, \dots, L_n\}$ with L_n such that $L_n^3 \leq C_{\mathcal{L}} n^{1/q}$ for some $q > 1$. Then, with the notation of Lemma 17, for all $\kappa > 3/(2\pi^2)$,*

$$\mathbb{E} \max_{L \in \mathcal{L}} \left(\sup_{t \in B_L} |\nu_n(t)|^2 - \frac{\kappa}{(1-2P)^2} \frac{2L+1}{n} \right) \leq \frac{C(1+R^2)}{n},$$

where C is a positive constant depending on $P, q, C_{\mathcal{L}}, \kappa$.

Proof. Denote $R^l = \frac{1}{M^l(\hat{\theta})} - \frac{1}{M^l(\theta_0)}$. First note that

$$\nu_n(t) = \frac{1}{2\pi n} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \overline{t_l} \left(\frac{e^{-ilX_k}}{M^l(\hat{\theta})} - \frac{2\pi g^{\star l}}{M^l(\theta_0)} \right) = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t)$$

where

$$\nu_{n,1}(t) = \frac{1}{2\pi n} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \overline{t_l} \left(\frac{e^{-ilX_k} - 2\pi g^{\star l}}{M^l(\theta_0)} \right)$$

$$\nu_{n,2}(t) = \frac{1}{2\pi n} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \overline{t_l} \left(e^{-ilX_k} - 2\pi g^{\star l} \right) R^l$$

$$\nu_{n,3}(t) = \frac{1}{n} \sum_{k=1}^{n} \sum_{l \in \mathbb{Z}} \overline{t_l} g^{\star l} R^l = \sum_{l \in \mathbb{Z}} \overline{t_l} g^{\star l} R^l.$$

Thus $|\nu_n|^2 \leq 3|\nu_{n,1}|^2 + 3|\nu_{n,2}|^2 + 3|\nu_{n,3}|^2$, and, if $\kappa_1 = \kappa/3$,

$$\begin{aligned} &\mathbb{E} \max_{L} \left(\sup_{B_L} |\nu_n|^2 - \frac{\kappa}{(1-2P)^2} \frac{2L+1}{n} \right) \\ &\leq 3\mathbb{E} \sum_{L} \left(\sup_{B_L} |\nu_{n,1}|^2 - \frac{\kappa_1}{(1-2P)^2} \frac{2L+1}{n} \right)_+ \\ &\quad + 3\mathbb{E} \max_{L} \left(\sup_{B_L} |\nu_{n,2}|^2 \right) + 3\mathbb{E} \max_{L} \left(\sup_{B_L} |\nu_{n,3}|^2 \right) \end{aligned}$$

where $a_+ = \max(a, 0)$ denotes the positive part of a.

Control of $\nu_{n,3}$ First note that

$$\left| g^{\star l} R^l \right| = \left| f^{\star l} \frac{M^l(\theta_0) - M^l(\hat{\theta})}{M^l(\hat{\theta})} \right| \leq \frac{|f^{\star l}|}{1-2P} \left| M^l(\theta_0) - M^l(\hat{\theta}) \right|.$$

Thus, using Schwarz inequality

$$\sup_{t \in B_L} |\nu_{n,3}(t)|^2 \leq \sum_{l=-L}^{L} \frac{|f^{\star l}|^2}{(1-2P)^2} \left| M^l(\theta_0) - M^l(\hat{\theta}) \right|^2.$$

Moreover

$$\begin{aligned}|M^l(\theta_0)-M^l(\hat\theta)| &\leq \Big|(p_0-\hat p)e^{-i\alpha_0 l}+\hat p(e^{-i\alpha_0 l}-e^{-i\hat\alpha l})+(1-p_0-1+\hat p)e^{-i\beta_0 l}\\ &\quad +(1-\hat p)(e^{-i\beta_0 l}-e^{-i\hat\beta l})\Big|\\ &\leq |p_0-\hat p|+|e^{-i\alpha_0 l}-e^{-i\hat\alpha l}|+|p_0-\hat p|+|e^{-i\beta_0 l}-e^{-i\hat\beta l}|\\ &\leq 2|p_0-\hat p|+|l||\alpha_0-\hat\alpha|+|l||\beta_0-\hat\beta|\leq 2|l|\|\theta_0-\hat\theta\|_1\end{aligned}$$

(note that it is also true for $l=0$ since $M^0(\theta_0)=M^0(\hat\theta)=1$).

Thus, for any $L\in\mathcal{L}$

$$\sup_{t\in B_L}|\nu_{n,3}(t)|^2\leq\sum_{l=-L_n}^{L_n}\frac{|f^{\star l}|^2}{(1-2P)^2}4|l|^2\|\theta_0-\hat\theta\|_1^2.$$

Since $s\geq 1$

$$\max_{L\in\mathcal{L}}\left(\sup_{B_L}|\nu_{n,3}|^2\right)\leq\frac{4}{(1-2P)^2}\sum_{l=-L_n}^{L_n}|f^{\star l}|^2|l|^{2s}\|\theta_0-\hat\theta\|_1^2\leq\frac{4R^2}{(1-2P)^2}\|\theta_0-\hat\theta\|_1^2.$$

According to Proposition 6 and inequality $\|x\|_1^2\leq 3\|x\|^2$, there exists a constant $K>0$ such that $\mathbb{E}(n\|\hat\theta-\theta_0\|_1^2)\leq K$. Then

$$\mathbb{E}\max_{L\in\mathcal{L}}\left(\sup_{B_L}|\nu_{n,3}|^2\right)\leq\frac{C_3R^2}{n}$$

with $C_3=4K/(1-2P)^2$.

Control of $\nu_{n,2}$ Note that

$$|R^l|\leq\frac{1}{(1-2P)^2}\left|M^l(\theta_0)-M^l(\hat\theta)\right|\leq\frac{2}{(1-2P)^2}|l|\|\hat\theta-\theta_0\|_1,$$

so for $t\in B_L$,

$$|\nu_{n,2}(t)|^2\leq\left(\sum_{l=-L}^{L}|\overline{t_l}(\widehat{g^{\star l}}-g^{\star l})R^l|\right)^2\leq\frac{4}{(1-2P)^4}\sum_{l=-L}^{L}|\widehat{g^{\star l}}-g^{\star l}|^2l^2\|\hat\theta-\theta_0\|_1^2.$$

Then, for any $L\in\mathcal{L}$

$$\sup_{t\in B_L}|\nu_{n,2}(t)|^2\leq\frac{4}{(1-2P)^2}\sum_{l=-L}^{L}|\widehat{g^{\star l}}-g^{\star l}|^2l^2\|\hat\theta-\theta_0\|_1^2$$

and

$$\max_{L\in\mathcal{L}}\sup_{t\in B_L}|\nu_{n,2}(t)|^2\leq\frac{4}{(1-2P)^2}\sum_{l=-L_n}^{L_n}|\widehat{g^{\star l}}-g^{\star l}|^2l^2\|\hat\theta-\theta_0\|_1^2.$$

Using Hölder's inequality, for any $p, q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E} \max_{L\in\mathcal{L}} \sup_{t\in B_L} |\nu_{n,2}(t)|^2 \leq \frac{4}{(1-2P)^2} \sum_{l=-L_n}^{L_n} l^2 \mathbb{E}^{1/p}(|\widehat{g^{\star l}} - g^{\star l}|^{2p}) \mathbb{E}^{1/q} \|\hat{\theta} - \theta_0\|_1^{2q}.$$

But Proposition 6 gives us

$$\mathbb{E}\|\hat{\theta} - \theta_0\|_1^{2q} \leq (1 + 2\pi + 2\pi)^{2q-2} \mathbb{E}\left(3\|\hat{\theta} - \theta_0\|^2\right) \leq K'(q) n^{-1}.$$

Moreover, we can apply the Rosenthal inequality to the variables $Y_k = e^{ilX_k} - \mathbb{E}(e^{ilX_k})$: there exists $C(2p) > 0$ such that

$$\begin{aligned} \mathbb{E}\left|\sum_{k=1}^{n} Y_k\right|^{2p} &\leq C(2p)\left(\sum_{k=1}^{n} \mathbb{E}|Y_k|^{2p} + \left(\sum_{k=1}^{n} \mathbb{E}|Y_k|^2\right)^p\right) \\ &\leq C(2p)\left(n 2^{2p} + (4n)^p\right) \leq C'(p) n^p \end{aligned}$$

that provides

$$\mathbb{E}(|\widehat{g^{\star l}} - g^{\star l}|^{2p}) = \mathbb{E}\left((2\pi)^{-2p}\left|\frac{1}{n}\sum_{k=1}^{n} Y_k\right|^{2p}\right) \leq (2\pi)^{-2p} C'(p) n^{-p}.$$

Thus

$$\begin{aligned} \mathbb{E} \max_{L\in\mathcal{L}} \sup_{t\in B_L} |\nu_{n,2}(t)|^2 &\leq \frac{4}{(1-2P)^2} \sum_{l=-L_n}^{L_n} l^2 (2\pi)^{-2} C'(p)^{1/p} n^{-1} K'(q)^{1/q} n^{-1/q} \\ &\leq \frac{C''(q)}{(1-2P)^2} n^{-1-1/q} L_n^3. \end{aligned}$$

Since $L_n^3 \leq C_{\mathcal{L}} n^{1/q}$, we obtain

$$\mathbb{E} \max_{L\in\mathcal{L}} \sup_{t\in B_L} |\nu_{n,2}(t)|^2 \leq \frac{C_2}{n}$$

with $C_2 = C''(q) C_{\mathcal{L}}/(1-2P)^2$.

Control of $\nu_{n,1}$

To control $\nu_{n,1}$, we need Talagrand's inequality.

Lemma 19. *Let $X_1, \ldots, X_n$ be i.i.d. random variables, and define $\nu_n(t) = \frac{1}{n}\sum_{k=1}^{n} \psi_t(X_k) - \mathbb{E}[\psi_t(X_k)]$, for t belonging to a countable class $\mathcal{B}$ of real-valued measurable functions. Then, for $\delta > 0$, there exist three constants c_l, $l = 1, 2, 3$, such that*

$$\mathbb{E}\left[\left(\sup_{t\in\mathcal{B}} |\nu_n(t)|^2 - c(\delta) H^2\right)_+\right] \leq c_1 \left\{\frac{v}{n} \exp\left(-c_2 \delta \frac{nH^2}{v}\right)\right. \qquad (10)$$

$$+\frac{M_1^2}{C^2(\delta)n^2}\exp\left(-c_3 C(\delta)\sqrt{\delta}\frac{nH}{M_1}\right)\Bigg\},$$

with $C(\delta)=(\sqrt{1+\delta}-1)\wedge 1$, $c(\delta)=2(1+2\delta)$ *and*

$$\sup_{t\in\mathcal{B}}\|\psi_t\|_\infty \le M_1,\ \mathbb{E}\left[\sup_{t\in\mathcal{B}}|\nu_n(\psi_t)|\right]\le H,\ \textit{and}\ \sup_{t\in\mathcal{B}} Var(\psi_t(X_1))\le v.$$

Inequality (10) is a classical consequence of Talagrand's inequality given in Klein and Rio (2005): see for example Lemma 5 (page 812) in Lacour (2008). Using density arguments, we can apply it to the unit sphere of a finite dimensional linear space.

Here $\nu_{n,1}(t)=\frac{1}{n}\sum_{k=1}^n \psi_t(X_k)-\mathbb{E}[\psi_t(X_k)]$ with

$$\psi_t(X)=\frac{1}{2\pi}\sum_{l\in\mathbb{Z}}\overline{t_l}\frac{e^{-ilX}}{M^l(\theta_0)},\qquad \mathbb{E}(\psi_t(X))=\sum_{l\in\mathbb{Z}}\overline{t_l}\frac{g^{\star l}}{M^l(\theta_0)}.$$

Let us compute M_1, H and v.

- Using Cauchy Schwarz inequality, for $t\in B_L$,

$$|\psi_t(u)|^2=\left|\frac{1}{2\pi}\sum_{l=-L}^{L}\overline{t_l}\frac{e^{-ilu}}{M^l(\theta_0)}\right|^2\le\frac{1}{4\pi^2}\sum_{l=-L}^{L}|t_l|^2\sum_{l=-L}^{L}\left|\frac{e^{-ilu}}{M^l(\theta_0)}\right|^2$$
$$\le\frac{1}{4\pi^2(1-2p_0)^2}(2L+1),$$

thus $M_1=\frac{1}{2\pi(1-2p_0)}\sqrt{2L+1}$.

- Using Cauchy Schwarz inequality, for $t\in B_L$,

$$\sup_{t\in B_L}\left|\frac{1}{2\pi n}\sum_{k=1}^{n}\sum_{l\in\mathbb{Z}}\overline{t_l}\left(\frac{e^{-ilX_k}}{M^l(\theta_0)}-\mathbb{E}\left(\frac{e^{-ilX_k}}{M^l(\theta_0)}\right)\right)\right|^2$$
$$\le\sum_{l=-L}^{L}\left|\frac{1}{2\pi n}\sum_{k=1}^{n}\left(\frac{e^{-ilX_k}}{M^l(\theta_0)}-\mathbb{E}\left(\frac{e^{-ilX_k}}{M^l(\theta_0)}\right)\right)\right|^2,$$

then

$$\mathbb{E}\left(\sup_{t\in B_L}|\nu_{n,1}(\psi_t)|^2\right)\le\sum_{l=-L}^{L}\mathrm{Var}\left(\frac{1}{2\pi n}\sum_{k=1}^{n}\frac{e^{-ilX_k}}{M^l(\theta_0)}\right)$$
$$\le\sum_{l=-L}^{L}\frac{1}{4\pi^2 n}\mathrm{Var}\left(\frac{e^{-ilX_1}}{M^l(\theta_0)}\right)$$
$$\le\frac{1}{4\pi^2 n}\sum_{l=-L}^{L}\mathbb{E}\left|\frac{e^{-ilX_1}}{M^l(\theta_0)}\right|^2\le\frac{1}{4\pi^2(1-2p_0)^2}\frac{2L+1}{n},$$

thus by Jensen's inequality $H^2=\frac{1}{4\pi^2(1-2p_0)^2}\frac{2L+1}{n}$.

- It remains to control the variance. If $t \in B_L$

$$\mathrm{Var}(\psi_t(X)) \le \mathbb{E}\left|\frac{1}{2\pi}\sum_{l=-L}^{L}\overline{t_l}\frac{e^{-ilX}}{M^l(\theta_0)}\right|^2 = \frac{1}{4\pi^2}\sum_{l,l'} t_l\overline{t_{l'}}\frac{\mathbb{E}(e^{-ilX}\overline{e^{-il'X}})}{M^l(\theta_0)\overline{M^{l'}(\theta_0)}}$$
$$= \frac{1}{2\pi}\sum_{l,l'} t_l\overline{t_{l'}}\frac{g^{\star(l-l')}}{M^l(\theta_0)\overline{M^{l'}(\theta_0)}}.$$

Using twice Schwarz inequality

$$\begin{aligned}\mathrm{Var}(\psi_t(X)) &\le \frac{1}{2\pi}\sqrt{\sum_l\left|\frac{t_l}{M^l(\theta_0)}\right|^2\sum_l\left|\sum_{l'}\frac{\overline{t_{l'}}}{\overline{M^{l'}(\theta_0)}}g^{\star(l-l')}\right|^2} \\ &\le \frac{1}{2\pi(1-2p_0)}\sqrt{\sum_l\left|\sum_{l'}\frac{\overline{t_{l'}}}{\overline{M^{l'}(\theta_0)}}g^{\star(l-l')}\right|^2} \\ &\le \frac{1}{2\pi(1-2p_0)}\sqrt{\sum_l\sum_{l'}\left|\frac{\overline{t_{l'}}}{\overline{M^{l'}(\theta_0)}}\right|^2\sum_{l'}\left|g^{\star(l-l')}\right|^2} \\ &\le \frac{1}{2\pi(1-2p_0)}\sqrt{\sum_l\frac{1}{|1-2p_0|^2}\sum_{j\in\mathbb{Z}}|g^{\star j}|^2} \\ &\le \frac{R}{2\pi(1-2p_0)^2}\sqrt{2L+1},\end{aligned}$$

since $\sum_{j\in\mathbb{Z}}\left|g^{\star j}\right|^2 \le \sum_{j\in\mathbb{Z}}\left|f^{\star j}\right|^2 \le R^2$. Thus $v = \frac{R}{2\pi(1-2p_0)^2}\sqrt{2L+1}$.

Inequality (10) becomes

$$\begin{aligned}&\mathbb{E}\left[\left(\sup_{t\in B_L}|\nu_{n,1}(t)|^2 - \frac{c(\delta)}{4\pi^2(1-2p_0)^2}\frac{2L+1}{n}\right)_+\right] \\ &\le c_1\left\{\frac{R\sqrt{2L+1}}{2\pi(1-2p_0)^2 n}\exp\left(-c_2\delta\frac{\sqrt{2L+1}}{2\pi R}\right)\right. \\ &\qquad\left.+\frac{2L+1}{4\pi^2(1-2p_0)^2C^2(\delta)n^2}\exp\left(-c_3C(\delta)\sqrt{\delta n}\right)\right\} \\ &\le \frac{K\max(R,1)}{n}\left\{\sqrt{2L+1}\exp\left(-c\sqrt{2L+1}\right)+\frac{2L+1}{n}\exp\left(-c\sqrt{n}\right)\right\}\end{aligned}$$

with K and c positive constants depending on P, c_1, c_2, c_3, δ. This ends the control of $\nu_{n,1}$ with $\kappa_1 = \frac{c(\delta)}{4\pi^2}$ since

$$\sum_{L\in\mathcal{L}}\left\{\sqrt{2L+1}e^{-c\sqrt{2L+1}}+\frac{2L+1}{n}e^{-c\sqrt{n}}\right\}$$

$$\leq \sum_{L=0}^{\infty} \sqrt{2L+1} e^{-c\sqrt{2L+1}} + \sharp\mathcal{L} e^{-c\sqrt{n}} = O(1).$$

Finally it is sufficient to take

$$\kappa \geq 3\kappa_1 = \frac{3(2+4\delta)}{4\pi^2} = \frac{3}{2\pi^2} + \frac{3\delta}{\pi^2}$$

to conclude the proof. Since δ can be chosen arbitrary small, and we have assumed $\kappa > 3/(2\pi^2)$, this condition is satisfied. □

Acknowledgment

The authors would like to thank the Editors and one anonymous referee for valuable comments and suggestions leading to corrections and improvements of the article.

References

Baldi, P., Kerkyacharian, G., Marinucci, D. and Picard, D. (2009). Adaptive density estimation for directional data using needlets. *Ann. Statist.* **37** 3362–3395. MR2549563

Banerjee, A., Dhillon, I. S., Ghosh, J. and Sra, S. (2005). Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.* **6** 1345–1382. MR2249858

Bartels, R. (1984). Estimation in a bidirectional mixture of von Mises distributions. *Biometrics* **40** 1345–1382.

Batschelet, E. (1981). *Circular Statistics in Biology.* Academic Press, London. MR0659065

Baudry, J.-P., Maugis, C. and Michel, B. (2012). Slope heuristics: overview and implementation. *Stat. Comput.* **22** 455–470. MR2865029

Birgé, L. and Massart, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)* **3** 203–268. MR1848946

Birgé, L. and Massart, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138** 33–73. MR2288064

Bordes, L., Mottelet, S. and Vandekerkhove, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.* **34** 1204–1232. MR2278356

Butucea, C. and Vandekerkhove, P. (2014). Semiparametric mixtures of symmetric distributions. *Scand. J. Stat.* **41** 227–239. MR3181141

Chen, J., Li, P. and Fu, Y. (2008). Testing homogeneity in a mixture of von Mises distributions with a structural parameter. *Canad. J. Statist.* **36** 129–142. MR2406935

Chung, K. L. and Zhong, K. (2001). *A course in probability theory.* Academic press. MR1796326

Dacunha-Castelle, D. and Duflo, M. (1986). *Probability and statistics. Vol. II.* Springer-Verlag, New York Translated from the French by David McHale. MR837655

Fraser, M. D., Hsu, Y. S. and Walker, J. J. (1981). Identifiability of finite mixtures of von Mises distributions. *Ann. Statist.* **9** 1130–1131. MR628771

Gassiat, E. and Rousseau, J. (2016). Nonparametric finite translation hidden Markov models and extensions. *Bernoulli* **22** 193–212. MR3449780

Gassiat, E., Rousseau, J. and Vernet, E. (2018). Efficient semiparametric estimation and model selection for multidimensional mixtures. *Electron. J. Stat.* **12** 703–740. MR3769193

Goldenshluger, A. (2002). Density deconvolution in the circular structural model. *J. Multivariate Anal.* **81** 360–375. MR1906385

Hall, P. and Zhou, X.-H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.* **31** 201–224. MR1962504

Hall, P., Neeman, A., Pakyari, R. and Elmore, R. (2005). Nonparametric inference in multivariate mixtures. *Biometrika* **92** 667–678. MR2202653

Hernández-Sánchez, E. and Scarpa, B. (2012). A wrapped flexible generalized skew-normal model for a bimodal circular distribution of wind directions. *Chilean Journal of Statistics* **3** 129–141. MR2982066

Hohmann, D. and Holzmann, H. (2013). Two-component mixtures with independent coordinates as conditional mixtures: nonparametric identification and estimation. *Electron. J. Stat.* **7** 859–880. MR3044502

Holzmann, H., Munk, A. and Stratmann, B. (2004). Identifiability of finite mixtures—with applications to circular distributions. *Sankhyā* **66** 440–449. MR2108200

Hunter, D. R., Wang, S. and Hettmansperger, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.* **35** 224–251. MR2332275

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in circular statistics. Series on Multivariate Analysis* **5**. World Scientific Publishing Co., Inc., River Edge, NJ. MR1836122

Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli* **19** 1576–1611. MR3129026

Jones, T. A. and James, W. (1969). Analysis of bimodal orientation data. *Math. Geol.* **1** 129–135.

Kent, J. T. (1983). Identifiability of finite mixtures for directional data. *Ann. Statist.* **11** 984–988. MR707948

Kim, P. T. and Koo, J.-Y. (2000). Directional mixture models and optimal estimation of the mixing density. *Canad. J. Statist.* **28** 383–398. MR1792056

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312

Lacour, C. (2008). Adaptive estimation of the transition density of a particular hidden Markov chain. *J. Multivariate Anal.* **99** 787–814. MR2405092

Lark, R. M., Clifford, D. and Waters, C. N. (2014). Modelling complex geological circular data with the projected normal distribution and mixtures of von Mises distributions. *Solid Earth* **5** 631–639.

Ley, C. and Verdebout, T. (2017). *Modern directional statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3752655

Ma, Y. and Yao, W. (2015). Flexible estimation of a semiparametric two-component mixture model with one parametric component. *Electron. J. Stat.* **9** 444–474. MR3326131

Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics. Wiley Series in Probability and Statistics*. John Wiley & Sons, Ltd., Chichester. MR1828667

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Phil. Trans. Royal Soc., A* **185** 71–110.

Pewsey, A. and García-Portugués, E. (2021). Recent advances in directional statistics. *TEST* **30** 1–58. MR4242171

Sedimentation Seminar (1966). Cross-bedding in the Salem Limestone of Central Indiana. *Sedimentology* **6** 95–114.

Spurr, B. D. (1981). On estimating the parameters in mixtures of circular normal distributions. *J. Internat. Assoc. Math. Geol.* **13** 163–173. MR613759

Spurr, B. D. and Koutbeiy, M. A. (1991). A comparison of various methods for estimating the parameters in mixtures of von Mises distributions. *Comm. Statist. Simulation Comput.* **20** 725–741. MR1129093

van der Vaart, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. MR1652247