Statistics and Probability Letters 79 (2009) 2066–2075

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Local polynomial regression for circular predictors

Marco Di Marzio [a], Agnese Panzera [a], Charles C. Taylor [b,*]

[a] *DMQTE, Università di Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy*

[b] *Department of Statistics, University of Leeds, Leeds LS2 9JT, UK*

ARTICLE INFO

Article history:
Received 14 September 2008
Received in revised form 9 June 2009
Accepted 20 June 2009
Available online 1 July 2009

MSC:
62G07
62G08
62G20

ABSTRACT

We consider local smoothing of datasets where the design space is the d-dimensional ($d \geq 1$) torus and the response variable is real-valued. Our purpose is to extend least squares local polynomial fitting to this situation. We give both theoretical and empirical results.

1. Introduction

A *circular* observation can be regarded as a point on the unit circle, or a direction in the plane. Once an initial direction and an orientation of the unit circle have been chosen, any circular observation may be represented by an angle $\theta \in [0, 2\pi)$. Typical examples include flight direction of birds from a point of release, wind and ocean current direction, energy demand over a period of 24 h when the measurements are taken over a time interval much longer than the day and when the times of the day are recorded. A circular observation is periodic, i.e., $\theta = \theta + 2m\pi$ for $m \in \mathbb{Z}$. This periodicity sets apart circular statistical analysis from standard real-line methods. Recent accounts are given by Jammalamadaka and SenGupta (2001) and Mardia and Jupp (1999).

A much less studied subject is local regression in the case of circular predictors and real-valued responses. Its practical relevance is easily seen when considering the analysis of meteorological data, or more generally in earth and environmental sciences. Silverman (1986, sec. 2.10) suggests fitting data replicated along the interval $[-2\pi, 4\pi)$, with a smoothing degree depending on the original sample size. The only alternative approach appears to be periodic smoothing splines, introduced by Cogburn and Davis (1974). Nothing specific and reasonably simple appears to exist for the high-dimensional case, although this seems needed in many applications. For example, it could be of interest to predict ozone concentration given the wind directions at 6 am and at noon. In this example, the number of angles is $d = 2$, but this could easily be extended by considering more locations or time points for the explanatory wind directions; see Mardia and Jupp (1999, pp. 1–12) for further examples.

In this paper we extend least squares local polynomial fitting (Ruppert and Wand, 1994, for example) to the case when a design point $\boldsymbol{\theta}$ is a vector of angles $(\theta_1, \ldots, \theta_d)^{\top} \in [0, 2\pi)^d$, and the response is real-valued. Geometrically, $\boldsymbol{\theta}$ identifies a point of a d-dimensional torus made of the Cartesian product of d unit circles. Our strategy is twofold. We (i) introduce a class of circular weight functions (or *kernels*), and (ii) locally approximate the design density and the regression function by the pth degree polynomial

* Corresponding author.
E-mail addresses: mdimarzio@dmqte.unich.it (M. Di Marzio), agnesepanzera@yahoo.it (A. Panzera), charles@maths.leeds.ac.uk (C.C. Taylor).

0167-7152/$ – see front matter Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.06.014

$$\beta_0 + \sum_{j=1}^{d} \sum_{t=1}^{p} \beta_{jt} \sin^t(\cdot - \theta_j). \tag{1}$$

Point (ii) is motivated by the fact that the difference between two angular observations needs to be minimal at $2m\pi$, $m \in \mathbb{Z}$. Moreover, because $\sin(\theta) \simeq \theta$ as θ tends to 0, the polynomial (1) satisfies a Taylor series interpretation.

In Section 2 we define the kernels suitable for our polynomial fitting, and explore their efficiency properties. In Section 3 we consider the local linear ($p = 1$) regression estimator, along with conditional mean squared error and optimal smoothing. We also extend the analysis, for univariate predictors, to general p. Finally, Section 4 contains a small simulation study to illustrate the finite sample behaviour of the results.

2. Circular kernels

2.1. Definitions

We introduce our kernels in the one-dimensional setting. Such an approach seems adequate in that we will use weight functions which are products of univariate kernels, as the torus geometry allows for.

Definition 1 (*Circular Kernels of Order r*). A circular kernel, of order r and concentration (smoothing) parameter $\kappa > 0$, is a function $K_\kappa : [0, 2\pi) \to \mathbb{R}$ such that

(i) it admits, at $\theta \in [0, 2\pi)$, a convergent Fourier series representation $1/(2\pi)\{1 + 2\sum_{j=1}^{\infty} \gamma_j(\kappa)\cos(j\theta)\}$;

(ii) denoting $\eta_j(K_\kappa) := \int_0^{2\pi} \sin^j(\theta) K_\kappa(\theta) \mathrm{d}\theta$, then

$$\eta_0(K_\kappa) = 1, \qquad \eta_j(K_\kappa) = 0 \quad \text{for } 0 < j < r, \quad \text{and} \quad \eta_r(K_\kappa) \neq 0;$$

(iii) as κ increases $\int_{-\epsilon}^{\epsilon} K_\kappa(\theta)\mathrm{d}\theta$ tends to 1 for $\epsilon \in (0, \pi)$.

Condition (i) specifies that the kernel is symmetric around the null mean direction. The quantity $\eta_j(K_\kappa)$ in (ii) plays a similar rôle as the jth moment of a symmetric kernel in the linear theory, being zero if j is odd.

Remark 1. Most of the usual circular densities, which are symmetric about the null mean direction, are included in Definition 1 as second-order kernels—this includes the kernel uniform on $[-\pi/\{\kappa + 1\}, \pi/\{\kappa + 1\})$. Dirichlet and Fejér kernels

$$D_\kappa(\theta) := \frac{\sin(\{\kappa + 1/2\}\theta)}{2\pi \sin(\theta/2)}, \qquad F_\kappa(\theta) := \frac{1}{2\pi(\kappa + 1)} \left[\frac{\sin(\{\kappa + 1\}\theta/2)}{\sin(\theta/2)}\right]^2, \quad \kappa \in \mathbb{N}$$

are both circular kernels. In particular, D_κ has order $\kappa + 1$ if κ is odd, and $\kappa + 2$ otherwise, while F_κ has order 2.

Remark 2. Our order definition is consistent with the techniques used for obtaining higher-order kernels starting from second-order ones. As an instance, we apply a technique of Lejeune and Sarda (1992), to get a result useful in Theorem 4. Given a second-order circular kernel K_κ, let $\boldsymbol{E}_\ell$ be a matrix of order $\ell + 1$ with (i, j) entry given by $\eta_{i+j-2}(K_\kappa)$, and $\boldsymbol{U}_\ell$ be the same as $\boldsymbol{E}_\ell$ with the first column replaced by $\{1, \sin(\theta), \ldots, \sin^\ell(\theta)\}^\mathsf{T}$. Then

$$\mathcal{K}_{(\ell)}(\theta) := \frac{|\boldsymbol{U}_\ell|}{|\boldsymbol{E}_\ell|} K_\kappa(\theta),$$

is a circular kernel of order $\ell + 1$ when ℓ is odd, and of order $\ell + 2$ otherwise.

Remark 3. The univariate setting allows for a comparison with previous work. Our kernels include kernels on the sphere which are functions of $\kappa\{1 - \cos(\theta)\}$ studied by Beran (1979), Hall et al. (1987), Bai et al. (1988) and Klemelä (2000). However, the kernels D_κ, F_κ and the wrapped Cauchy are not of this latter form, yet fulfill the conditions of Definition 1.

2.2. Kernel efficiency

We discuss the efficiency of our kernels in the density estimation setting to allow easy comparisons with the standard theory.

Definition 2 (*Kernel Circular Density Estimator*). Let $\Theta_1, \ldots, \Theta_n$ be a random sample from a bounded, continuous circular density f. Given a circular kernel K_κ, the kernel estimator of f at $\theta \in [0, 2\pi)$ is defined as

$$\hat{f}(\theta; \kappa) := \frac{1}{n} \sum_{i=1}^{n} K_\kappa(\theta - \Theta_i). \tag{2}$$

The efficiency theory of euclidean kernels (p. 42 Silverman, 1986, for example) is based on the fact that the bandwidth and the kernel have separable contributions to the mean integrated squared error $\mathsf{MISE}[\hat{g}] := \int \mathsf{E}[(\hat{g}-g)^2] \equiv \int (\mathsf{E}[\hat{g}]-g)^2 + \int \mathsf{Var}[\hat{g}]$, where $\hat{g}$ gives the kernel estimate of the curve g at a point of the domain. Unfortunately, this is not the case for the MISE of (2). In fact, we have

Theorem 1. *Given a random sample $\Theta_1, \ldots, \Theta_n$ drawn from a density f, let $\hat{f}(\cdot;\kappa)$ be the kernel circular density estimator equipped with the second-order kernel K_κ, if*

(i) $\lim_{n\to\infty}\gamma_j(\kappa) = 1$, *for each* $j \in \mathbb{Z}^+$;
(ii) $\lim_{n\to\infty} n^{-1}\sum_{j=1}^{\infty}\gamma_j^2(\kappa) = 0$;
(iii) *f'' is continuous and square-integrable;*

then

$$\mathsf{MISE}\left[\hat{f}(\cdot;\kappa)\right] = \frac{1}{16}\{1-\gamma_2(\kappa)\}^2 \int_0^{2\pi}\left\{f''(\theta)\right\}^2 \mathrm{d}\theta + \frac{1+2\sum_{i=1}^{\infty}\gamma_j^2(\kappa)}{2n\pi} + o(1),$$

Proof. See Appendix. □

Remark 4. The MISE of Hall et al. (1987) is very similar to that above. For example, consider the von Mises kernel, for which $\gamma_j(\kappa) := \mathcal{I}_j(\kappa)/\mathcal{I}_0(\kappa)$, $\mathcal{I}_j(\cdot)$ being the modified Bessel function of the first kind and order j. Using the notation of (3.7) in Hall et al. (1987), we have: $c_0^2(\kappa)c_2(\kappa) = \mathcal{I}_0(2\kappa)/[2\pi\{\mathcal{I}_0(\kappa)\}^2] = \{1+2\sum_{i=1}^{\infty}\gamma_j^2(\kappa)\}/(2\pi)$ and $1-c_0(\kappa)c_1(\kappa) = 1-\mathcal{I}_1(\kappa)/\mathcal{I}_0(\kappa) = 1-\gamma_1(\kappa)$, consequently their asymptotic MISE differs from the leading terms in the above MISE of an order of $O(\kappa^{-4})$.

In our efficiency analysis we need

Result 1. *Let $\Theta_1, \ldots, \Theta_n$ be a random sample from a circular density f having Fourier series expansion $f(\theta) = 1/(2\pi)[1 + 2\sum_{j=1}^{\infty}\{\alpha_j\cos(j\theta)+\delta_j\sin(j\theta)\}]$ for $\theta \in [0, 2\pi)$. Then*

$$\mathsf{MISE}\left[\hat{f}(\cdot;\kappa)\right] = \frac{1}{\pi}\sum_{j=1}^{\infty}\{\gamma_j(\kappa)-1\}^2(\alpha_j^2+\delta_j^2) + \frac{1}{n\pi}\sum_{j=1}^{\infty}\gamma_j^2(\kappa)(1-\alpha_j^2-\delta_j^2).$$

Without loss of generality we can suppose that the mean direction is 0, and we consider only densities and kernels which are fully specified by their concentration parameters, respectively denoted as ρ and κ. For the above decomposition, when considering the (relative) efficiency of two circular kernels, the smoothing parameters do not "cancel" and so their equivalence needs first to be established as follows. For fixed ρ and n, we can obtain κ to minimize MISE for a given kernel function. The efficiency of one kernel relative to another may then be measured by taking the ratio of the minimized MISEs.

As the Dirichlet kernel ($\gamma_j(\kappa) = \mathbb{1}_{\{j\le\kappa\}}$) is of higher order for $\kappa > 1$ – and so expected to be asymptotically more efficient – we have measured the efficiency of other kernels relative to this one. In Fig. 1 we show the relative efficiency of the von Mises wrapped normal ($\gamma_j(\kappa) = \kappa^{j^2}$), and Fejér ($\gamma_j(\kappa) = \mathbb{1}_{\{j\le\kappa\}}(\kappa+1-j)/(\kappa+1)$) kernels for $n = 5, 25, 125, 625$ for the von Mises and wrapped Cauchy ($\alpha_j = \rho^j$; $\delta_j = 0$) distributions. Not surprisingly, the wrapped Normal and von Mises kernels are very similar, and both are better than the Fejér kernel. For small n, the von Mises kernel is more efficient that the Dirichlet kernel; markedly so for the Cauchy distribution, or for data with low concentration.

3. Local polynomial regression

3.1. Linear fitting with von Mises based kernels

Consider the dataset $\{(\boldsymbol{\Theta}_i, Y_i), i = 1, \ldots, n\}$, where $\boldsymbol{\Theta}_i := (\Theta_{i1}, \ldots, \Theta_{id})^\mathsf{T}$, and $Y_i \in \mathbb{R}$ are both observable, absolutely continuous, random variables taking values respectively in $[0, 2\pi)^d$ and $\mathbb{R}$. From now on we will assume that

$$Y_i = m(\boldsymbol{\Theta}_i) + \sigma(\boldsymbol{\Theta}_i)\varepsilon_i, \quad i = 1, \ldots, n$$

where $\sigma^2(\cdot)$ is the conditional variance of Y and ε_i's are real-valued random variables with zero mean and unit variance. Our objective is to construct an estimator of $m(\boldsymbol{\theta})$ as a function of the dataset when both $\boldsymbol{\Theta}_i$'s and ε_i's are i.i.d.

Let $\mathcal{P}_{\boldsymbol{\theta}}(\cdot;\boldsymbol{\beta}) := \beta_0 + \sum_{j=1}^{d}\beta_j\sin(\cdot-\theta_j)$, and suppose that $m(\boldsymbol{\psi}) \simeq \mathcal{P}_{\boldsymbol{\theta}}(\boldsymbol{\psi};\boldsymbol{\beta})$ for $\boldsymbol{\psi}$ in a neighborhood of $\boldsymbol{\theta}$. Here $\mathcal{P}_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{\beta}) = \beta_0$, which motivates estimating $m(\boldsymbol{\theta})$ by $\hat{\beta}_0$. Recalling that for very small values of θ we have $\sin(\theta) \simeq \theta$, then a Taylor series expansion justifies both $\hat{\beta}_0$ and the values $\hat{\beta}_j$, $j = 1, \ldots, d$, as estimates of the partial derivatives $\beta_j = \partial m(\boldsymbol{\theta})/\partial\theta_j$. Viewed as local least squares estimators, $\hat{\beta}_0, \ldots, \hat{\beta}_d$ minimize $\sum_{i=1}^{n}\{Y_i - \mathcal{P}_{\boldsymbol{\theta}}(\boldsymbol{\Theta}_i;\boldsymbol{\beta})\}^2 w(\boldsymbol{\Theta}_i, \boldsymbol{\theta})$ where

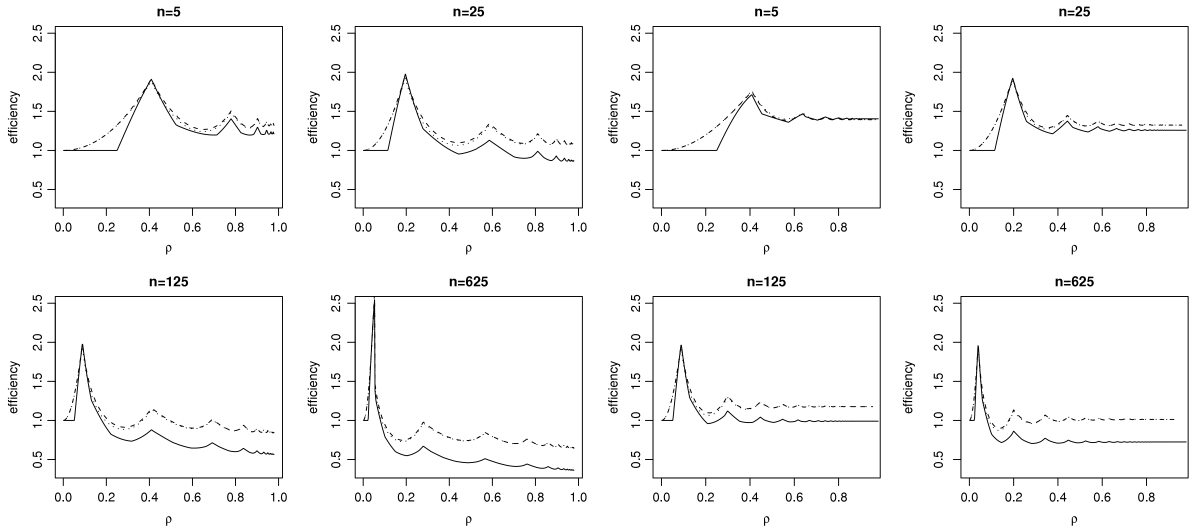

Fig. 1. Relative efficiency of Fejér (—), wrapped normal (- - - - -), and von Mises (· · ·) kernels to the Dirichlet kernel, for various values of n. With respect to the underlying true density, the left group corresponds to the von Mises distribution with $\rho = \mathcal{I}_1(\nu)/\mathcal{I}_0(\nu)$, while the right group corresponds to the wrapped Cauchy distribution.

$w(\boldsymbol{\Theta}_i, \boldsymbol{\theta})$ is the weight function, (a symmetric, continuous function integrating to 1) which, if strictly positive, decreases with some distance between $\boldsymbol{\Theta}_i$ and $\boldsymbol{\theta}$. Now we provide an explicit expression for $\hat{\beta}_0$ together with its L_2 properties.

Let $\boldsymbol{y} := (Y_1, \ldots, Y_n)^\mathsf{T}$ be the response vector,

$$\boldsymbol{\Theta} := \begin{bmatrix} 1 & \sin(\Theta_{11} - \theta_1) & \cdots & \sin(\Theta_{1d} - \theta_d) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \sin(\Theta_{n1} - \theta_1) & \cdots & \sin(\Theta_{nd} - \theta_d) \end{bmatrix}$$

the design matrix, and

$$\boldsymbol{W} := \mathrm{diag}\,\{K_{\boldsymbol{C}}(\boldsymbol{\Theta}_1 - \boldsymbol{\theta}), \cdots, K_{\boldsymbol{C}}(\boldsymbol{\Theta}_n - \boldsymbol{\theta})\}$$

the weight matrix, where $\boldsymbol{C} := \kappa \boldsymbol{I}$, $\boldsymbol{I}$ denoting the identity matrix of order d, and

$$K_{\boldsymbol{C}}(\boldsymbol{\Theta}_i - \boldsymbol{\theta}) := \prod_{j=1}^{d} K_\kappa(\Theta_{ij} - \theta_j), \quad i = 1, \ldots, n. \tag{3}$$

The local linear kernel estimator of $m(\boldsymbol{\theta})$ is given by the first entry of the vector

$$\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (Y_i - \boldsymbol{\beta}^\mathsf{T}\boldsymbol{\Theta})^2 K_{\boldsymbol{C}}(\boldsymbol{\Theta}_i - \boldsymbol{\theta}),$$

where $\boldsymbol{\beta} := (\beta_0, \beta_1, \ldots, \beta_d)^\mathsf{T}$. Assuming the non-singularity of $\boldsymbol{\Theta}^\mathsf{T}\boldsymbol{W}\boldsymbol{\Theta}$, standard weighted least squares theory yields $\hat{\boldsymbol{\beta}} = (\boldsymbol{\Theta}^\mathsf{T}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^\mathsf{T}\boldsymbol{W}\boldsymbol{y}$, and

$$\hat{m}(\boldsymbol{\theta}; \boldsymbol{C}) = \boldsymbol{e}_j^\mathsf{T}(\boldsymbol{\Theta}^\mathsf{T}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^\mathsf{T}\boldsymbol{W}\boldsymbol{y}, \tag{4}$$

where $\boldsymbol{e}_j$ is a $(d+1) \times 1$ vector having 1 as the jth entry and 0 elsewhere.

Given its efficiency, as well as its prevalence in kernel smoothing of circular data, we firstly give results when the von Mises kernel $V_\kappa(\cdot) := \exp\{\kappa \cos(\cdot)\}/\{2\pi \mathcal{I}_0(\kappa)\}$ is used to define the d-dimensional weight function.

Theorem 2. *Given the dataset $\{(\boldsymbol{\Theta}_i, Y_i), i = 1, \ldots, n\}$, where $\boldsymbol{\Theta}_i$'s are i.i.d. observations from the circular design density f, and Y_i's are i.i.d. real-valued random variables, take the local linear kernel regression estimator $\hat{m}(\cdot; \boldsymbol{C})$ equipped with the weight function $V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i - \boldsymbol{\theta}) := \prod_{j=1}^{d} V_\kappa(\Theta_{ij} - \theta_j)$. Assume that*

(i) $\lim_{n\to\infty}\kappa^{-1} = 0$;
(ii) $\lim_{n\to\infty} n^{-1}\kappa^{d/2} = 0$;
(iii) *the conditional variance σ^2 is continuous, and the density f is continuously differentiable;*
(iv) *all second-order derivatives of the regression function m are continuous.*

Then at $\boldsymbol{\theta} \in [0, 2\pi)^d$ the conditional mean squared error of $\hat{m}(\boldsymbol{\theta}; \boldsymbol{C})$ is given by

$$\mathrm{E}[\{\hat{m}(\cdot; \boldsymbol{C}) - m(\boldsymbol{\theta})\}^2 \mid \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_n] = \frac{1}{4}\left\{\frac{\mathcal{I}_1(\kappa)}{\kappa \mathcal{I}_0(\kappa)}\right\}^2 \mathrm{tr}^2\{\boldsymbol{H}_m(\boldsymbol{\theta})\} + \left[\frac{\mathcal{I}_0(2\kappa)}{2\pi\{\mathcal{I}_0(\kappa)\}^2}\right]^d \frac{\sigma^2(\boldsymbol{\theta})}{nf(\boldsymbol{\theta})} + o_p\left(\kappa^{-2} + n^{-1}\kappa^{d/2}\right), \tag{5}$$

where $\boldsymbol{H}_m(\boldsymbol{\theta})$ denotes the Hessian matrix of m at $\boldsymbol{\theta}$.

Proof. See Appendix. □

Once more, in the proof of the above theorem a major technical issue is that the concentration parameter κ cannot be "separated" from the kernel.

Remark 5. Since κ corresponds to the inverse of the squared bandwidth of the euclidean smoother, the remainder term in (5) is consistent with that obtained by Ruppert and Wand (1994).

Finally, the optimal smoothing degree is given by

Corollary 1. *The concentration parameter which minimizes the asymptotic mean squared error, i.e. the first two summands in the RHS of formula (5), is*

$$\left[\frac{\mathrm{tr}^4\{\boldsymbol{H}_m(\boldsymbol{\theta})\}\{nf(\boldsymbol{\theta})\}^2 2^{2d}\pi^d}{d^2\sigma^4(\boldsymbol{\theta})}\right]^{1/(4+d)}.$$

Proof. See Appendix. □

3.2. Generalizations and extensions

The results of Theorem 2 can be generalized to the class of second-order circular kernels K_κ. Given the square-integrable function g, define $R(g) := \int g^2$, then

Theorem 3. *Given the dataset $\{(\boldsymbol{\Theta}_i, Y_i), i = 1, \ldots, n\}$, where $\boldsymbol{\Theta}_i$'s are i.i.d. observations from the circular design density f, and Y_i's are i.i.d. real-valued random variables, take the local linear kernel regression estimator $\hat{m}(\cdot; \boldsymbol{C})$ equipped with the weight function in (3) with K_κ being a second-order circular kernel. Assume conditions* (i) *of Definition 1, and* (iii) *of Theorem 2, together with*

(i) $\lim_{n\to\infty} n^{-1}R(K_{\boldsymbol{C}}) = 0$.

Then, at $\boldsymbol{\theta} \in [0, 2\pi)^d$,

$$\mathrm{E}[\{\hat{m}(\cdot; \boldsymbol{C}) - m(\boldsymbol{\theta})\}^2 \mid \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_n] = \frac{1}{16}\{1 - \gamma_2(\kappa)\}^2 \mathrm{tr}^2\{\boldsymbol{H}_m(\boldsymbol{\theta})\} + \frac{R(K_{\boldsymbol{C}})\sigma^2(\boldsymbol{\theta})}{nf(\boldsymbol{\theta})} + o_p(1).$$

Proof. See Appendix. □

It would be of interest to determine the optimal smoothing degree in this case, but since the coefficients γ_j's depend on κ in a specific way for each kernel, the result in Corollary 1 is hard to generalize. Concerning the extension to higher-degree polynomials and whatever second-order circular kernel, we have

Theorem 4. *Given the dataset $\{(\Theta_i, Y_i), i = 1, \ldots, n\}$, where Θ_i's are i.i.d. observations from the circular one-dimensional density f, and Y_i's are i.i.d. real-valued random variables, take the local pth degree polynomial regression estimator $\hat{m}(\cdot; \kappa)$ equipped with a second-order circular kernel K_κ. Assume conditions* (i) *of Definition 1,* (iii) *and* (iv) *of Theorem 2. Moreover, assume that*

(i) *for the kernel $\mathcal{K}_{(p)}$ in Remark 2, $\lim_{n\to\infty} n^{-1}R(\mathcal{K}_{(p)}) = 0$;*

(ii) *$m^{(p+2)}$ is continuous in a neighborhood of θ.*

Then, for any $\theta \in [0, 2\pi)$,

$$\mathrm{E}[\hat{m}(\theta; \kappa) - m(\theta) \mid \Theta_1, \ldots, \Theta_n] = \begin{cases} \eta_{p+1}(\mathcal{K}_{(p)})\dfrac{m^{(p+1)}(\theta)}{(p+1)!} + o_p(1), & \text{if } p \text{ is odd;} \\ \eta_{p+2}(\mathcal{K}_{(p)})\left\{\dfrac{m^{(p+1)}(\theta)f'(\theta)}{f(\theta)(p+1)!} + \dfrac{m^{(p+2)}(\theta)}{(p+2)!}\right\} + o_p(1), & \text{otherwise;} \end{cases}$$

and

$$\mathrm{Var}[\hat{m}(\theta : \kappa) \mid \Theta_1, \ldots, \Theta_n] = R(\mathcal{K}_{(p)})\frac{\sigma^2(\theta)}{nf(\theta)}\{1 + o_p(1)\}.$$

Proof. See Appendix. □

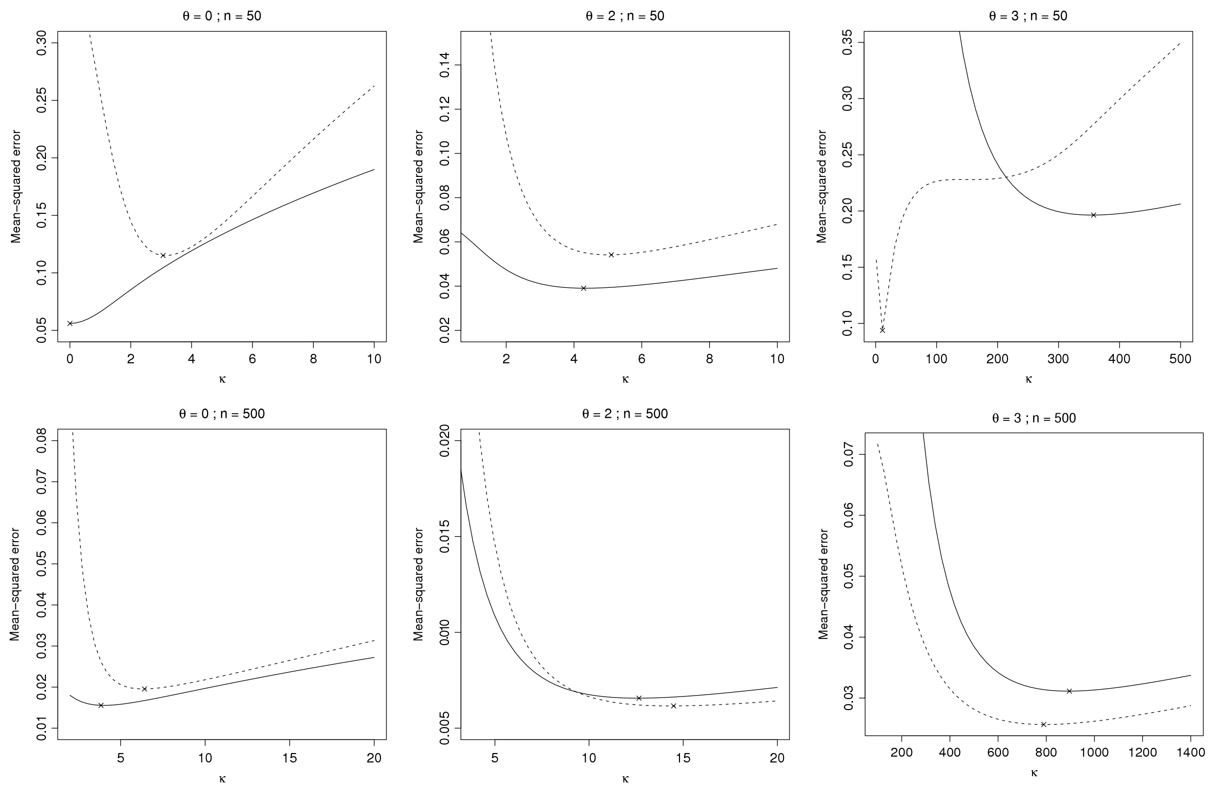

Fig. 2. Comparison of averaged squared error as a function of κ over 200 simulations (dashed line), and asymptotic mean squared error given by Theorem 2 (continuous line) with locations of minima. Top row: $n = 50$; lower row: $n = 500$, with m estimated at $\theta = 0$ (left), $\theta = 2$ (middle) and $\theta = 3$ (right).

4. Simulation results

We briefly explore the asymptotic result given by Theorem 2 in a simulation study. We first investigate the dependence of the mean squared error on θ, n and κ when $d = 1$ and choose a sharp-peaked response

$$m(\theta) = 2 + \sin(\theta - 1.2\pi) + 3\exp\left\{-10\left(15\frac{(\theta-\pi)}{2\pi}\right)^2\right\},$$

with $\varepsilon_i \sim N(0, 1)$, $\sigma^2(\Theta_i) = 1/2$, and Θ_i, $i = 1, \ldots, n$ coming from a von Mises density with mean π and concentration parameter 1. We estimate $m(\theta)$ at $\theta = 0, 2, 3$ and compare the average squared error of (4) with the asymptotic mean squared error given in Theorem 2 over κ for $n = 50$ and $n = 500$. The results are displayed in Fig. 2, and the asymptotic nature of the result is clear. Note that the values of the second derivative of m at $\theta = 0, 2, 3$ are $-0.59, 0.98, 140.89$, respectively, which explains the poorer performance at $\theta = 3$.

Secondly, we explore the dependence on d. In this case we use the model

$$m(\boldsymbol{\theta}) = \frac{1}{d}\sum_{i=1}^{d}\sin\theta_i + \frac{1}{d(d-1)}\sum_{i\neq j}\cos\theta_i\cos\theta_j \quad (d \geq 2)$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^\mathsf{T}$, $\sigma^2(\boldsymbol{\Theta}_i) = 1/2$, $i = 1, \ldots, n$, and f is a product of (independent) von Mises densities with mean zero and concentration parameter 1. We estimate $m(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = (0, \ldots, 0)^\mathsf{T}$ and $(\pi/2, \ldots, \pi/2)^\mathsf{T}$ for a range of κ, for $n = 500$. Fig. 3 shows good agreement for $d = 2$ between the average squared error and the asymptotic mean squared error. However, we note increasingly poor behaviour as d increases, indicating that the asymptotic nature of the result also depends on d, and again illustrating the well-known phenomenon of the *curse of dimensionality*.

Aknowledgements

The authors are grateful to two anonymous referees for their valuable comments which led to improvements in this article.

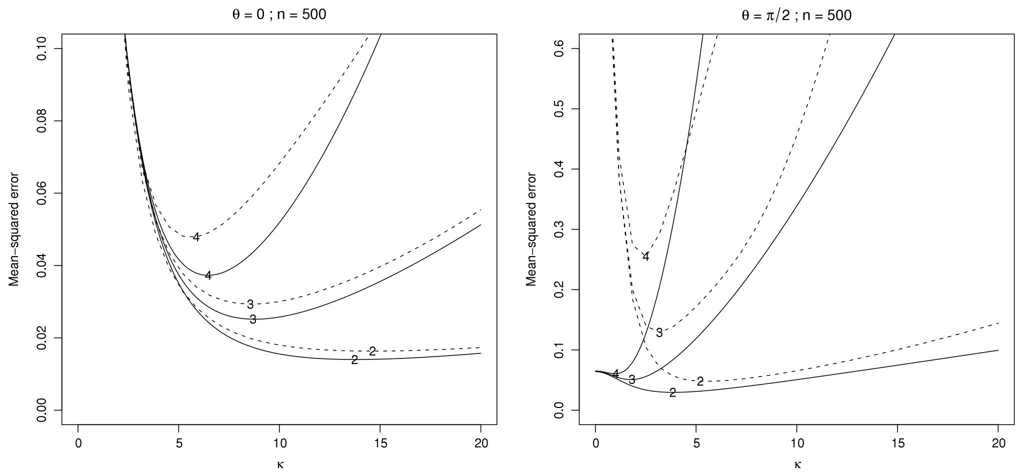

Fig. 3. Comparison of averaged squared error as a function of κ over 200 simulations (dashed line), and asymptotic mean squared error given by Theorem 2 (continuous line) with locations of minima shown by the integers 2, 3, 4 which corresponds to the dimension of the data. m is estimated at $\boldsymbol{\theta} = (0, \cdots, 0)^{\mathsf{T}}$ (left) and $\boldsymbol{\theta} = (\pi/2, \cdots, \pi/2)^{\mathsf{T}}$ (right).

Appendix

Proof of Theorem 1. Express $K_\kappa(\theta)$ in terms of a Fourier series, and, recalling that for very small values of u $\sin(u) \simeq u$, use the expansion $f(u+\theta) = f(\theta) + \sin(u)f'(\theta) + 1/2\sin^2(u)f''(\theta) + O\{\sin^3(u)\}$. Then, starting from (2), make a change of variable and use assumption (i) to get

$$\begin{aligned}\mathsf{E}[\hat{f}(\theta;\kappa)] &= \int_0^{2\pi} K_\kappa(\psi-\theta)f(\psi)\mathrm{d}\psi \\ &= \int_0^{2\pi} K_\kappa(u)f(u+\theta)\mathrm{d}u \\ &= f(\theta) + \frac{1}{4}\{1-\gamma_2(\kappa)\}f''(\theta) + o(1).\end{aligned}$$

Now, recalling assumptions (i) and (ii), we have

$$\begin{aligned}\mathsf{Var}[\hat{f}(\theta;\kappa)] &= \frac{1}{n}\int_0^{2\pi}\{K_\kappa(\psi-\theta)\}^2 f(\psi)\mathrm{d}\psi - \frac{1}{n}\left\{\mathsf{E}[\hat{f}(\theta;\kappa)]\right\}^2 \\ &= \frac{1}{n}\int_0^{2\pi}\{K_\kappa(u)\}^2\{f(\theta)+o(1)\}\mathrm{d}u - \frac{1}{n}\{f(\theta)+o(1)\}^2 \\ &= \frac{1}{2n\pi}\left\{1 + 2\sum_{j=1}^{\infty}\gamma_j^2(\kappa)\right\}f(\theta) + o(1). \quad \square\end{aligned}$$

Proof of Theorem 2. Put

$$\boldsymbol{S}_{\Theta_i-\theta} := \{\sin(\Theta_{i1}-\theta_1), \ldots, \sin(\Theta_{id}-\theta_d)\}^{\mathsf{T}}, \quad i = 1, \ldots, n$$

and use $\boldsymbol{D}_g(\boldsymbol{\theta})$ to denote the first-order partial derivatives vector of the function g at $\boldsymbol{\theta}$. To derive the conditional bias, we firstly note that (4) yields

$$\mathsf{E}[\hat{m}(\boldsymbol{\theta};\boldsymbol{C}) \mid \boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_n] = \boldsymbol{e}_1^{\mathsf{T}}(\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{m}, \tag{6}$$

where $\boldsymbol{m} := \{m(\boldsymbol{\Theta}_1), \ldots, m(\boldsymbol{\Theta}_n)\}^{\mathsf{T}}$, and $\boldsymbol{W} := \mathsf{diag}\,\{V_{\boldsymbol{C}}(\boldsymbol{\Theta}_1-\boldsymbol{\theta}), \cdots, V_{\boldsymbol{C}}(\boldsymbol{\Theta}_n-\boldsymbol{\theta})\}$. Using the expansion

$$\boldsymbol{m} = \boldsymbol{\Theta}\begin{bmatrix} m(\boldsymbol{\theta}) \\ \boldsymbol{D}_m(\boldsymbol{\theta})\end{bmatrix} + \frac{1}{2}\begin{bmatrix}\boldsymbol{S}_{\Theta_1-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_1-\theta} \\ \vdots \\ \boldsymbol{S}_{\Theta_n-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_n-\theta}\end{bmatrix} + \boldsymbol{R}_m(\boldsymbol{\theta}),$$

where $\boldsymbol{R}_m(\boldsymbol{\theta})$ denotes the remainder, we have that the first term in the expansion of (6) is $m(\boldsymbol{\theta})$. Thus

$$\mathsf{E}[\hat{m}(\boldsymbol{\theta};\boldsymbol{C})-m(\boldsymbol{\theta})\mid\boldsymbol{\Theta}_1,\cdots,\boldsymbol{\Theta}_n]=\frac{1}{2}\boldsymbol{e}_1^{\mathsf{T}}(\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\left\{\begin{bmatrix}\boldsymbol{S}_{\Theta_1-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_1-\theta}\\ \vdots\\ \boldsymbol{S}_{\Theta_n-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_n-\theta}\end{bmatrix}+\boldsymbol{R}_m(\boldsymbol{\theta})\right\}.$$

Observe that

$$\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta}=\begin{bmatrix}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta}) & \sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\\ \sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta} & \sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\end{bmatrix} \tag{7}$$

and

$$\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\begin{bmatrix}\boldsymbol{S}_{\Theta_1-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_1-\theta}\\ \vdots\\ \boldsymbol{S}_{\Theta_n-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\theta)\boldsymbol{S}_{\Theta_n-\theta}\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}\\ \sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\left\{\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}\right\}\boldsymbol{S}_{\Theta_i-\theta}\end{bmatrix}, \tag{8}$$

then, using the expansion

$$f(\boldsymbol{u}+\boldsymbol{\theta})=f(\boldsymbol{\theta})+\boldsymbol{S}_u^{\mathsf{T}}\boldsymbol{D}_f(\boldsymbol{\theta})+O(\boldsymbol{S}_u^{\mathsf{T}}\boldsymbol{S}_u),$$

and recalling assumption (i), a change of variables leads to these approximations

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta}) &= \int_{[0,2\pi)^d}V_{\boldsymbol{C}}(\boldsymbol{\alpha}-\boldsymbol{\theta})f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}+o_p(1)\\
&= f(\boldsymbol{\theta})+o_p(1);\\
\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta} &= \int_{[0,2\pi)^d}V_{\boldsymbol{C}}(\boldsymbol{\alpha}-\boldsymbol{\theta})\boldsymbol{S}_{\alpha-\theta}f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}+o_p(\mathbf{1})\\
&= \frac{\mathcal{I}_1(\kappa)}{\mathcal{I}_0(\kappa)}\boldsymbol{C}^{-1}\boldsymbol{D}_f(\boldsymbol{\theta})+o_p(\boldsymbol{C}^{-1}\boldsymbol{1});\\
\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}} &= \int_{[0,2\pi)^d}V_{\boldsymbol{C}}(\boldsymbol{\alpha}-\boldsymbol{\theta})\boldsymbol{S}_{\boldsymbol{\alpha}-\boldsymbol{\theta}}\boldsymbol{S}_{\boldsymbol{\alpha}-\boldsymbol{\theta}}^{\mathsf{T}}f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}+o_p(\boldsymbol{I})\\
&= \frac{\mathcal{I}_1(\kappa)}{\mathcal{I}_0(\kappa)}\boldsymbol{C}^{-1}f(\boldsymbol{\theta})+o_p(\boldsymbol{C}^{-1});\\
\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta} &= \int_{[0,2\pi)^d}V_{\boldsymbol{C}}(\boldsymbol{\alpha}-\boldsymbol{\theta})\boldsymbol{S}_{\boldsymbol{\alpha}-\boldsymbol{\theta}}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\boldsymbol{\alpha}-\boldsymbol{\theta}}f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}+o_p(1)\\
&= \frac{\mathcal{I}_1(\kappa)}{\kappa\,\mathcal{I}_0(\kappa)}\,\mathrm{tr}\,\{\boldsymbol{H}_m(\boldsymbol{\theta})\}f(\boldsymbol{\theta})+o_p\left(\kappa^{-1}\right);\\
\frac{1}{n}\sum_{i=1}^{n}V_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\left\{\boldsymbol{S}_{\Theta_i-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_i-\theta}\right\}\boldsymbol{S}_{\Theta_i-\theta} &= \int_{[0,2\pi)^d}V_{\boldsymbol{C}}(\boldsymbol{\alpha}-\boldsymbol{\theta})\left\{\boldsymbol{S}_{\alpha-\theta}^{\mathsf{T}}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\alpha-\theta}\right\}\boldsymbol{S}_{\alpha-\theta}f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha}+o_p(\mathbf{1})\\
&= O_p(\boldsymbol{C}^{-2}\boldsymbol{1});
\end{aligned}$$

where $\boldsymbol{1}$ is the unit vector of length d. Hence, recalling assumption (i) we have

$$\boldsymbol{e}_1^{\mathsf{T}}\left(n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta}\right)^{-1}\simeq\left[\{f(\boldsymbol{\theta})\}^{-1}+o_p(1)\quad -\boldsymbol{D}_f(\boldsymbol{\theta})^{\mathsf{T}}\{f(\boldsymbol{\theta})\}^{-2}+o_p(\mathbf{1})\right], \tag{9}$$

thus

$$\mathsf{E}[\hat{m}(\boldsymbol{\theta};\boldsymbol{C})-m(\boldsymbol{\theta})\mid\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_n]=\frac{1}{2}\frac{\mathcal{I}_1(\kappa)}{\kappa\,\mathcal{I}_0(\kappa)}\mathrm{tr}\,\{\boldsymbol{H}_m(\boldsymbol{\theta})\}+o_p(\kappa^{-1}).$$

For the conditional variance, according to multivariate local linear regression theory

$$\mathsf{Var}[\hat{m}(\boldsymbol{\theta};\boldsymbol{C})\mid\boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_n]=\boldsymbol{e}_1^{\mathsf{T}}(\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{\Theta}(\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1}\boldsymbol{e}_1,$$

where $\boldsymbol{\Sigma} := \text{diag}\{\sigma^2(\boldsymbol{\Theta}_1), \ldots, \sigma^2(\boldsymbol{\Theta}_n)\}$. Consider that

$$n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}\boldsymbol{\Theta} = \begin{bmatrix} n^{-1}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\sigma^2(\boldsymbol{\Theta}_i) & n^{-1}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}^{\mathsf{T}}_{\Theta_i-\theta}\sigma^2(\boldsymbol{\Theta}_i) \\ n^{-1}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}_{\Theta_i-\theta}\sigma^2(\boldsymbol{\Theta}_i) & n^{-1}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}_{\Theta_i-\theta}\boldsymbol{S}^{\mathsf{T}}_{\Theta_i-\theta}\sigma^2(\boldsymbol{\Theta}_i) \end{bmatrix}, \tag{10}$$

and approximate the components of the above matrix using the following relationships

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\sigma^2(\boldsymbol{\Theta}_i) &= \int_{[0,2\pi)^d}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\sigma^2(\boldsymbol{\alpha})f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} + o_p(1) \\
&= \left[\frac{\mathcal{I}_0(2\kappa)}{2\pi\{\mathcal{I}_0(\kappa)\}^2}\right]^d \sigma^2(\boldsymbol{\theta})f(\boldsymbol{\theta})\{1+o_p(1)\}; \\
\frac{1}{n}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}^{\mathsf{T}}_{\Theta_i-\theta}\sigma^2(\boldsymbol{\Theta}_i) &= \int_{[0,2\pi)^d}\{V_{\mathbf{C}}(\boldsymbol{\alpha}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}^{\mathsf{T}}_{\alpha-\theta}\sigma^2(\boldsymbol{\alpha})f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} + o_p(\mathbf{1}) \\
&= o_p(\mathbf{1}); \\
\frac{1}{n}\sum_{i=1}^{n}\{V_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}_{\Theta_i-\theta}\boldsymbol{S}^{\mathsf{T}}_{\Theta_i-\theta}\sigma^2(\boldsymbol{\Theta}_i) &= \int_{[0,2\pi)^d}\{V_{\mathbf{C}}(\boldsymbol{\alpha}_i-\boldsymbol{\theta})\}^2\boldsymbol{S}_{\alpha-\theta}\boldsymbol{S}^{\mathsf{T}}_{\alpha-\theta}\sigma^2(\boldsymbol{\alpha})f(\boldsymbol{\alpha})\mathrm{d}\boldsymbol{\alpha} + o_p(\boldsymbol{I}) \\
&= \frac{\widetilde{\mathcal{F}}(2,\kappa^2)}{4\pi\{\mathcal{I}_0(\kappa)\}^2}\left[\frac{\mathcal{I}_0(2\kappa)}{2\pi\{\mathcal{I}_0(\kappa)\}^2}\right]^{d-1}\sigma^2(\boldsymbol{\theta})f(\boldsymbol{\theta})\{\boldsymbol{I}+o_p(\boldsymbol{I})\},
\end{aligned}$$

where $\widetilde{\mathcal{F}}(2,\kappa^2) := \{\mathcal{I}_0(k)\}^2 + \{\mathcal{I}_1(k)\}^2 + 2\sum_{j=2}^{\infty}\mathcal{I}_j(\kappa)\{\mathcal{I}_j(\kappa) - \mathcal{I}_{j-2}(\kappa)\}$ is the regularized confluent hypergeometric function of the first kind. Combining the previous results with the approximations in (9), and recalling assumption (ii), we finally obtain

$$\text{Var}[\hat{m}(\boldsymbol{\theta};\mathbf{C}) \mid \boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_n] = \left[\frac{\mathcal{I}_0(2\kappa)}{2\pi\{\mathcal{I}_0(\kappa)\}^2}\right]^d \frac{\sigma^2(\boldsymbol{\theta})}{nf(\boldsymbol{\theta})} + o_p(n^{-1}\kappa^{d/2}). \quad \square$$

Proof of Corollary 1. Replace $\mathcal{I}_1(\kappa)/\mathcal{I}_0(\kappa)$ by 1 with an error of magnitude $O(\kappa^{-1})$, and use

$$\lim_{\kappa\to\infty}\left[\frac{\mathcal{I}_0(2\kappa)}{2\pi\{\mathcal{I}_0(\kappa)\}^2}\right]^d = \left(\frac{\kappa}{4\pi}\right)^{d/2},$$

then minimize the asymptotic MSE. □

Proof of Theorem 3. Follow the proof of Theorem 2, with $K_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})$ as ith entry of the weight matrix, $i = 1,\ldots,n$. In particular, to derive the conditional bias firstly note that

$$n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta} \simeq \begin{bmatrix} f(\boldsymbol{\theta}) + o_p(1) & 1/2\{1-\gamma_2(\kappa)\}\boldsymbol{D}_f^{\mathsf{T}}(\boldsymbol{\theta}) + o_p(\mathbf{1}) \\ 1/2\{1-\gamma_2(\kappa)\}\boldsymbol{D}_f(\boldsymbol{\theta}) + o_p(\mathbf{1}) & 1/2\{1-\gamma_2(\kappa)\}f(\boldsymbol{\theta})\boldsymbol{I} + o_p(\boldsymbol{I}) \end{bmatrix},$$

and, in virtue of condition (i) of Definition 1,

$$\boldsymbol{e}_1^{\mathsf{T}}(n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1} \simeq \left[\{f(\boldsymbol{\theta})\}^{-1} + o_p(1) \quad -D_f^{\mathsf{T}}(\boldsymbol{\theta})\{f(\boldsymbol{\theta})\}^{-2} + o_p(\mathbf{1})\right].$$

Moreover, observe that

$$n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\begin{bmatrix} \boldsymbol{S}^{\mathsf{T}}_{\Theta_1-\theta}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_1-\theta} \\ \vdots \\ \boldsymbol{S}^{\mathsf{T}}_{\Theta_n-\theta}\boldsymbol{H}_m(\boldsymbol{\theta})\boldsymbol{S}_{\Theta_n-\theta} \end{bmatrix} \simeq \begin{bmatrix} 1/2\{1-\gamma_2(\kappa)\}\text{tr}\{\boldsymbol{H}_m(\boldsymbol{\theta})\}f(\boldsymbol{\theta}) + o_p(1) \\ O_p(\mathbf{1}) \end{bmatrix},$$

to get

$$\text{E}[\hat{m}(\boldsymbol{\theta};\mathbf{C}) - m(\boldsymbol{\theta}) \mid \boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_n] = \frac{1}{4}\{1-\gamma_2(\kappa)\}\text{tr}\{\boldsymbol{H}_m(\boldsymbol{\theta})\} + o_p(1).$$

To derive the conditional variance, observe that the upper-left entry of the matrix (10) generalizes as

$$\frac{1}{n}\sum_{i=1}^{n}\{K_{\mathbf{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2\sigma^2(\boldsymbol{\Theta}_i) \simeq R(K_{\mathbf{C}})\sigma^2(\boldsymbol{\theta})f(\boldsymbol{\theta})\{1+o_p(1)\},$$

where $R(K_{\boldsymbol{C}}) = \{R(K_\kappa)\}^d = \{(2\pi)^{-1}(1+2\sum_{j=1}^{\infty}\gamma_j^2(\kappa))\}^d$, the diagonal blocks are $o_p(\mathbf{1})$, whereas letting

$$A(K_{\boldsymbol{C}}) := \frac{\left[\gamma_0^2(\kappa)+\gamma_1^2(\kappa)+2\sum_{j=2}^{\infty}\gamma_j(\kappa)\{\gamma_j(\kappa)-\gamma_{j-2}(\kappa)\}\right]\{R(K_\kappa)\}^{d-1}}{4\pi},$$

where $\gamma_0(\kappa) := \int_0^{2\pi} K_\kappa(\theta)\cos(0)\mathrm{d}\theta = 1$, the lower-right entry is

$$\frac{1}{n}\sum_{i=1}^{n}\{K_{\boldsymbol{C}}(\boldsymbol{\Theta}_i-\boldsymbol{\theta})\}^2 \boldsymbol{S}_{\boldsymbol{\Theta}_i-\boldsymbol{\theta}}\boldsymbol{S}_{\boldsymbol{\Theta}_i-\boldsymbol{\theta}}^{\mathsf{T}}\sigma^2(\boldsymbol{\Theta}_i) \simeq A(K_{\boldsymbol{C}})\sigma^2(\boldsymbol{\theta})f(\boldsymbol{\theta})\{\boldsymbol{I}+o_p(\boldsymbol{I})\}.$$

Hence, it finally results

$$\mathsf{Var}[\hat{m}(\boldsymbol{\theta};\boldsymbol{C}) \mid \boldsymbol{\Theta}_1,\ldots,\boldsymbol{\Theta}_n] = \frac{R(K_{\boldsymbol{C}})\sigma^2(\boldsymbol{\theta})}{nf(\boldsymbol{\theta})}\{1+o_p(1)\}. \quad \square$$

Proof of Theorem 4. Follow the proof of Theorem 4.1 of Ruppert and Wand (1994) with these two recommendations: in the design matrix replace $(X_i - x)^j$, with $\sin^j(\Theta_i-\theta)$, and use the expansion $f(u+\theta) = f(\theta)+\sin(u)f'(\theta)+O\{\sin^2(u)\}$. In particular, to derive the conditional bias, let $\boldsymbol{Q}_p$ be the matrix of order $p+1$ having as (i,j) entry $\eta_{i+j-1}(K_\kappa)$, and observe that, in virtue of assumption (i) of Definition 1, $n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta} = f(\theta)\boldsymbol{E}_p + f'(\theta)\boldsymbol{Q}_p + o_p(1)$, with $\boldsymbol{E}_p$ being the matrix defined in Remark 2, to get

$$\boldsymbol{r}_1^{\mathsf{T}}(n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}\boldsymbol{\Theta})^{-1} = f(\theta)^{-1}\{\boldsymbol{r}_1^{\mathsf{T}}\boldsymbol{E}_p^{-1} - f'(\theta)f(\theta)^{-1}\boldsymbol{r}_1^{\mathsf{T}}\boldsymbol{E}_p^{-1}\boldsymbol{Q}_p\boldsymbol{E}_p^{-1}\} + o_p(\mathbf{1}),$$

where $\boldsymbol{r}_1$ is a $(p+1)\times 1$ vector having 1 as first entry and 0 elsewhere. For the conditional variance, denoting as $\boldsymbol{T}_p$ the matrix of order $p+1$ having $\int \sin^{i+j-2}(u)\{K_\kappa(u)\}^2\mathrm{d}u$ as (i,j) entry, and recalling condition (i), it follows that $n^{-1}\boldsymbol{\Theta}^{\mathsf{T}}\boldsymbol{W}^2\boldsymbol{\Theta} = f(\theta)\boldsymbol{T}_p + o_p(\boldsymbol{I})$. $\square$

References

Bai, Z.D., Rao, R.C., Zhao, L.C., 1988. Kernel estimators of density function of directional data. Journal of Multivariate Analysis 27, 24–39.
Beran, R., 1979. Exponential models for directional data. The Annals of Statistics 7, 1162–1178.
Cogburn, I., Davis, H.T., 1974. Periodic splines and spectral estimation. The Annals of Statistics 2, 1108–1126.
Hall, P., Watson, G., Cabrera, J., 1987. Kernel density estimation with spherical data. Biometrika 74, 751–762.
Jammalamadaka, S.R., SenGupta, A., 2001. Topics in Circular Statistics. World Scientific, Singapore.
Klemelä, J., 2000. Estimation of densities and derivatives of densities with directional data. Journal of Multivariate Analysis 73, 18–40.
Lejeune, M., Sarda, P., 1992. Smooth estimators of distribution and density functions. Computational Statistics & Data Analysis 14, 457–471.
Mardia, K.V., Jupp, P.E., 1999. Directional Statistics. John Wiley, New York.
Ruppert, D., Wand, M.P., 1994. Multivariate locally weighted least squares regression. The Annals of Statistics 22, 1346–1370.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.