

A note on density estimation for toroidal data

Marco Di Marzio ¹, Agnese Panzera
DMQTE, Università G. d'Annunzio, dimarzio@dmqte.unich.it,
agnesepanzera@yahoo.it

Charles C. Taylor
Department of Statistics, University of Leeds, charles@maths.leeds.ac.uk

Abstract: We study kernel estimation of densities defined on d -fold ($d \geq 1$) cartesian products of circles. Specifically, we introduce a class of circular kernels, and derive the AMISE-optimal smoothing degree.

Keywords: circular data; concentration parameter; nonparametric density estimation; von Mises kernel.

1. Introduction

A *circular* observation can be regarded as a point on the unit circle, and 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. 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). Concerning circular density estimation we observe that almost all of the related statistics appears to have a parametric nature, with the exception of a few contributions on kernel density estimation (Bai *et al.* (1988), Beran (1979), Hall *et al.* (1987), Klemelä (2000), Taylor (2008)). In this paper we consider kernel density estimation when a support point θ is a vector of angles $(\theta_1, \dots, \theta_d)^T \in [0, 2\pi)^d$. Our strategy is twofold. We *i*) introduce a class of circular kernels, and *ii*) locally approximate the unknown population density by a Taylor expansion where the increments of the angles are measured by their sine. We also obtain an AMISE-optimal bandwidth selector for the case in which the kernel is a d -fold product of von Mises densities.

2. Circular kernels

We introduce our kernels in the one-dimensional setting because we will use 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 $K_\kappa : [0, 2\pi) \rightarrow \mathbb{R}$ such that
i) it admits a convergent Fourier series representation $1/(2\pi)\{1 + 2 \sum_{j=1}^{\infty} \gamma_j(\kappa) \cos(j\theta)\}$;
ii) denoting $\eta_j(K_\kappa) := \int_0^{2\pi} \sin^j(\theta) K_\kappa(\theta) d\theta$, then $\eta_0(K_\kappa) = 1$, $\eta_j(K_\kappa) = 0$ for $0 < j < r$, and $\eta_r(K_\kappa) \neq 0$;

¹Address of correspondence: Viale Pindaro 42, 65127 Pescara, Italy.

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

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

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

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. Moreover, 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 fulfil the conditions of Definition 1.

Definition 2. (Multivariate kernel circular density estimator) Let $\{\Theta_i, i = 1, \dots, n\}$, with $\Theta_i := (\Theta_{i1}, \dots, \Theta_{id}) \in [0, 2\pi)^d$, be a random sample from a continuous density f defined on a d -dimensional torus. The kernel estimator of f at $\theta \in [0, 2\pi)^d$ is defined as

$$\hat{f}(\theta; \mathbf{C}) := \frac{1}{n} \sum_{i=1}^n K_{\mathbf{C}}(\theta - \Theta_i), \quad (1)$$

where $K_{\mathbf{C}}(\theta) := \prod_{j=1}^d K_{\kappa}(\theta_j)$, with $\mathbf{C} := \kappa \mathbf{I}$, and $\mathbf{I}$ denoting the identity matrix of order d .

The efficiency theory of euclidean kernels is based on the fact that the bandwidth and the kernel have separable contributions to the mean integrated squared error $\text{MISE}[\hat{g}] := \int \mathbb{E}[(\hat{g} - g)^2] \equiv \int (\mathbb{E}[\hat{g}] - g)^2 + \int \text{Var}[\hat{g}]$, 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 (1), which is given by

Theorem 1. Given the random sample $\{\Theta_i, i = 1, \dots, n\}$, consider the estimator $\hat{f}(\cdot; \mathbf{C})$ having as the kernel $K_{\mathbf{C}}$ the d -fold product of second-order circular kernels K_{κ} . If

- i) $\lim_{n \rightarrow \infty} \gamma_j(\kappa) = 1$, where $\gamma_j(\kappa)$ is the j th Fourier coefficient of K_{κ} ;
 - ii) $\lim_{n \rightarrow \infty} n^{-1} \{1 + 2 \sum_{i=1}^{\infty} \gamma_j^2(\kappa)\}^d = 0$;
 - iii) the hessian matrix of f at θ , $\mathbf{H}_f(\theta)$, has entries piecewise continuous and square integrable;
- then

$$\text{MISE}[\hat{f}(\cdot; \mathbf{C})] = \frac{1}{16} \{1 - \gamma_2(\kappa)\}^2 \int \text{tr}^2\{\mathbf{H}_f(\theta)\} d\theta + \frac{1}{n} \left\{ \frac{1 + 2 \sum_{i=1}^{\infty} \gamma_j^2(\kappa)}{2\pi} \right\}^d + o(1)$$

Proof. See Appendix. □

For the above indivisibility, when considering the (relative) efficiency of two circular kernels, the smoothing parameters do not “cancel” and so their equivalence needs first to be established. In our efficiency analysis we need MISE for the one-dimensional estimator, which, under the usual assumptions and by considering a circular density $f(\theta) = 1/(2\pi)[1 + 2 \sum_{j=1}^{\infty} \{\alpha_j \cos(j\theta) + \delta_j \sin(j\theta)\}]$, $\theta \in [0, 2\pi)$, is given by $\frac{1}{\pi} \sum_{j=1}^{\infty} \{\gamma_j(\kappa) - 1\}^2 (\alpha_j^2 + \delta_j^2) + \frac{1}{n\pi} \sum_{j=1}^{\infty} \gamma_j^2(\kappa) (1 - \alpha_j^2 - \delta_j^2)$. Without loss of generality we can suppose the

mean direction is 0, and we consider only densities and kernels which are determined by a single parameter, ρ and κ respectively. 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 ($\gamma_j(\kappa) = \mathbb{1}_{\{j \leq \kappa\}}$) is of higher order — and so expected to be asymptotically more efficient — we have measured the efficiency of other kernels relative to this one. In Figure 1 we show the relative efficiency of the von Mises ($\gamma_j(\kappa) = \mathcal{I}_j(\kappa)/\mathcal{I}_0(\kappa)$, where $\mathcal{I}_j(\cdot)$ is the modified Bessel function of the first kind and order j), wrapped normal ($\gamma_j(\kappa) = \kappa^{j^2}$), and Fejér ($\gamma_j(\kappa) = \mathbb{1}_{\{j \leq \kappa\}}(\kappa + 1 - j)/(\kappa + 1)$) 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 than the Dirichlet kernel; markedly so for the Cauchy distribution, or for data with low concentration.

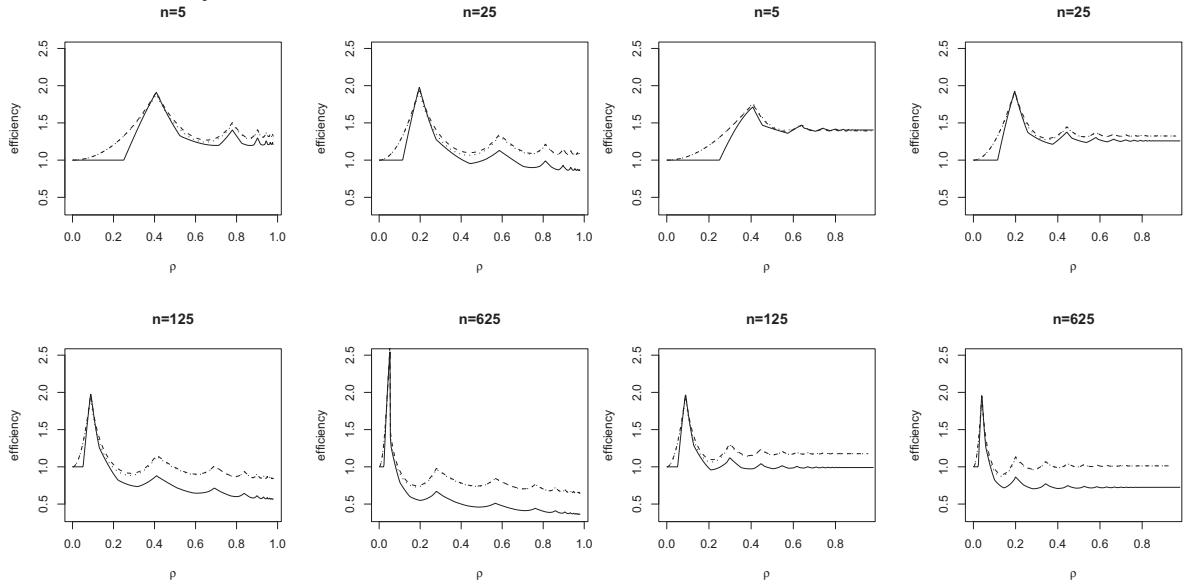


Figure 1: Relative efficiency of Fejér (—), wrapped normal (---), and von Mises (···) kernels to the Dirichlet kernel, for various values of n . Left group corresponds to the von Mises distribution with $\rho = \mathcal{I}_1(\nu)/\mathcal{I}_0(\nu)$, while right group corresponds to the wrapped Cauchy.

Given its efficiency, we consider the d -variate von Mises kernel $V_C(\boldsymbol{\theta}) := \prod_{j=1}^d V_\kappa(\theta_j)$, where $V_\kappa(\cdot) := \exp\{\kappa \cos(\cdot)\}/\{2\pi\mathcal{I}_0(\kappa)\}$ and $\kappa > 0$ is the smoothing parameter, to get

Theorem 2. Given the random sample $\{\boldsymbol{\Theta}_i, i = 1, \dots, n\}$, consider the estimator $\hat{f}(\cdot; C)$ using the kernel $V_C(\boldsymbol{\theta})$. Assume that condition iii) of Theorem 1 holds, and

i) $\lim_{n \rightarrow \infty} \kappa^{-1} = 0$;

ii) $\lim_{n \rightarrow \infty} n^{-1} \kappa^{d/2} = 0$;

then the AMISE optimal concentration parameter for $\hat{f}(\cdot; C)$ is

$$\left[\frac{2^d \pi^{d/2} n \int \text{tr}^2\{\mathbf{H}_f(\boldsymbol{\theta})\} d\boldsymbol{\theta}}{d} \right]^{2/(4+d)}.$$

Proof. See Appendix. □

Appendix

Proof of Theorem 1. Put $\mathbf{S}_{\Theta_i - \theta} := \{\sin(\Theta_{i1} - \theta_1), \dots, \sin(\Theta_{id} - \theta_d)\}^\top$, $i = 1, \dots, n$, and use $\mathbf{D}_f$ to denote the first-order partial derivatives vector of the function f at θ . Using the expansion $f(\mathbf{u} + \theta) = f(\theta) + \mathbf{S}_u^\top \mathbf{D}_f(\theta) + \frac{1}{2} \mathbf{S}_u^\top \mathbf{H}_f(\theta) \mathbf{S}_u + o(\mathbf{S}_u^\top \mathbf{S}_u)$, and recalling assumptions i) and iii), a change of variables leads to

$$\begin{aligned} \mathbb{E}[\hat{f}(\theta; \mathbf{C})] &= \int_{[0, 2\pi)^d} K_{\mathbf{C}}(\psi - \theta) f(\psi) d\psi \\ &= f(\theta) + \frac{1}{4} \{1 - \gamma_2(\kappa)\} \text{tr}\{\mathbf{H}_f(\theta)\} + o(1). \end{aligned}$$

Now, recalling assumption ii), we obtain

$$\begin{aligned} \text{Var}[\hat{f}(\theta; \mathbf{C})] &= \frac{1}{n} \int_{[0, 2\pi)^d} \{K_{\mathbf{C}}(\psi - \theta)\}^2 f(\psi) d\psi - \frac{1}{n} \left\{ \mathbb{E}[\hat{f}(\theta; \mathbf{C})] \right\}^2 \\ &= \frac{1}{n} \int_{[0, 2\pi)^d} \{K_{\mathbf{C}}(\mathbf{u})\}^2 \{f(\theta) + o(1)\} d\mathbf{u} - \frac{1}{n} \{f(\theta) + o(1)\}^2 \\ &= \frac{f(\theta)}{n} \left\{ \frac{1 + 2 \sum_{j=1}^{\infty} \gamma_j^2(\kappa)}{2\pi} \right\}^d + o(1). \end{aligned}$$

□

Proof of Theorem 2. Recall that for the von Mises kernel $\gamma_j(\kappa) = \mathcal{I}_j(\kappa)/\mathcal{I}_0(\kappa)$, and follow the proof of Theorem 1 to get

$$\text{MISE}[\hat{f}(\cdot; \mathbf{C})] = \frac{1}{4} \left\{ \frac{\mathcal{I}_1(\kappa)}{\kappa \mathcal{I}_0(\kappa)} \right\}^2 \int \text{tr}^2\{\mathbf{H}_f(\theta)\} d\theta + \frac{1}{n} \left[\frac{\mathcal{I}_0(2\kappa)}{2\pi \{\mathcal{I}_0(\kappa)\}^2} \right]^d + o_p(\kappa^{-2} + n^{-1} \kappa^{d/2}). \quad (2)$$

Now, replace $\mathcal{I}_1(\kappa)/\mathcal{I}_0(\kappa)$ by 1 with an error of magnitude $O(\kappa^{-1})$, use $\lim_{\kappa \rightarrow \infty} \left[\frac{\mathcal{I}_0(2\kappa)}{2\pi \{\mathcal{I}_0(\kappa)\}^2} \right]^d = \left(\frac{\kappa}{4\pi} \right)^{d/2}$, then minimize the leading term of (2). □

References

- Bai Z.D., Rao R.C. and 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.
- Hall P., Watson G. and Cabrera J. (1987) Kernel density estimation with spherical data, *Biometrika*, 74, 751–762.
- Jammalamadaka S.R. and 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.
- Mardia K.V. and Jupp P.E. (1999) *Directional Statistics*, John Wiley, New York.
- Taylor C.C. (2008) Automatic bandwidth selection for circular density estimation, *Computational Statistics & Data Analysis*, 52, 3493–3500.