Article:

Di Marzio,, M, Fensore, S, Panzera, A et al. (1 more author) (2016) A note on nonparametric estimation of circular conditional densities. Journal of Statistical Computation and Simulation, 86 (13). pp. 2573-2582. ISSN 0094-9655

https://doi.org/10.1080/00949655.2016.1146279

To appear in the *Journal of Statistical Computation and Simulation*
Vol. 00, No. 00, Month 20XX, 1–12

"Special Issue GRASPA"
A note on nonparametric estimation of circular conditional densities

M. Di Marzio[a*], S. Fensore[a], A. Panzera[b] and C.C. Taylor[c]

[a]*DSFPEQ, Università di Chieti-Pescara, Pescara, Italy;* [b]*DiSIA, Università di Firenze, Firenze, Italy;* [c]*Department of Statistics, University of Leeds, Leeds, UK*

(*Received XXX; final version received XXX*)

The conditional density offers the most informative summary of the relationship between explanatory and response variables. We need to estimate it in place of the simple conditional mean when its shape is not well-behaved. A motivation for estimating conditional densities, specific to the circular setting, lies in the fact that a natural alternative of it, like quantile regression, could be considered problematic because circular quantiles are not rotationally equivariant. We treat conditional density estimation as a local polynomial fitting problem as proposed by [1] in the euclidean setting, and discuss a class of estimators in the cases when the conditioning variable is either circular or linear. Asymptotic properties for some members of the proposed class are derived. The effectiveness of the methods for finite sample sizes is illustrated by simulation experiments and an example using real data.

Keywords: Circular data; conditional densities; local polynomials; optimal smoothing; von Mises kernel.

AMS Subject Classification: 62G07; 62G20; 65G60

1. Introduction

A circular observation can be regarded as a point on the circumference of the unit circle and, after both an origin and an orientation have been chosen, can be measured, in radians, by an angle $\theta \in [-\pi, \pi)$. Flight directions of birds from a point of release, wind and ocean current directions constitute classic examples of circular data. For a circular observation $\theta \in [-\pi, \pi)$, it holds that $\theta = 2m\pi + \theta$, for each $m \in \mathbb{Z}$: this makes standard real-line methods unsuited for circular data analysis.

The interest in predicting a circular variable given another one, of whatever nature, arises in many scientific fields. For example, in meteorology, it could be of interest to study the relation of wind direction on linear variables, such as the wind speed and the amount of rain, or to predict the wind direction, given measurements at previous hours. In atmospheric pollution studies, it is often of interest to investigate the dependency of a pollutant concentration on the wind direction. Further, in studying animals migration direction we often need to relate it to the distance moved.

Conditional densities are natural targets in prediction problems, where, for a given value of the explanatory variable, we wish to estimate the density of the response. Standard estimation of the conditional mean is customary, but in some cases is not proper

*Corresponding author. Email: mdimarzio@unich.it

due to the fact that the expectation is not very representative in the presence of multi-modality, significant asymmetry, and/or heavy tails. An alternative, robust approach, lies in quantile estimation. In particular, estimating some key quantiles could be a valid remedy when regression is not well-suited.

Nonparametric methods for estimating circular, conditional mean and quantiles, when the explanatory variable is either linear or circular, has been recently studied by [2] and [3], respectively. To the best of the authors' knowledge, nothing specific has been made for conditional density estimation. Compared to the euclidean setting, a strong additional motivation in favour of conditional density estimation is specific to the circular setting due to the problematic definition of quantiles for angular data. In fact, angle measurement relies on the *arbitrary* choice of the origin within the circle. Such arbitrariness has an impact on quantiles because the cumulative distribution function is calculated starting from the origin. As a consequence, the basic property of rotational equivariance is lost. This issue is recognized as serious, see [4], especially in problems when we need to compare datasets.

When it is difficult to motivate a parametric model, a sensible approach to conditional density estimation is the nonparametric one. Specifically, fully nonparametric estimation of conditional densities does not require either parametric assumptions for the target density, nor any parametric assumption on the link between response and explanatory variables. Surely, in the circular setting a nonparametric estimator fulfils the important property of rotational equivariance because of the local nature of weights, which depend on the difference between the estimation point and the sample observations.

The present paper considers the problem of estimating the density of $\Theta \mid U$, where Θ is a random angle, and U is either angular or linear. In particular, we follow the idea in [1], and formulate the problem of estimating the density of $\Theta \mid U$ as a nonparametric regression problem, discussing local polynomial estimators of circular conditional densities. The same approach, designed for distribution functions and quantiles estimation of circular random variable has been recently studied by [3].

In Section 2 we introduce our class of estimators. In Section 3, asymptotic properties of some members in this class, along with optimal smoothing, are derived. In Section 4, we show some results from a small simulation study, and illustrate the method in a real data example. Technical proofs are briefly outlined in the Appendix.

2. Local polynomial estimators of circular conditional densities

Let $(U_1, \Theta_1), \ldots, (U_n, \Theta_n)$ be n independent copies of the $\mathbb{U} \times [-\pi, \pi)$-valued random variable (U, Θ), with unknown density $f_{U\Theta}$. In what follows, we discuss both the cases when U is a random angle, i.e. $\mathbb{U} = [-\pi, \pi)$, or U stands for a linear random variable, i.e. $\mathbb{U} = \mathbb{R}$.

Clearly, when $\mathbb{U} = [-\pi, \pi)$, $f_{U\Theta}$ is a toroidal density, and letting $f_{\Theta|U}$ denote the density of $\Theta \mid U$, for $(m, s) \in \mathbb{Z} \times \mathbb{Z}$,

$$f_{\Theta|U}(\theta, u) = f_{\Theta|U}(\theta + 2\pi m \mid u + 2\pi s)$$

while, when $\mathbb{U} = \mathbb{R}$, $f_{U\Theta}$ is a cylindrical density, and

$$f_{\Theta|U}(\theta, u) = f_{\Theta|U}(\theta + 2\pi m \mid u).$$

Now, letting f_U denote the marginal density of U, a kernel estimator for $f_{\Theta|U}(\theta \mid u)$,

$(\theta, u) \in [-\pi, \pi) \times \mathbb{U}$, can be easily obtained as

$$\hat{f}_{\Theta|U}(\theta \mid u) := \frac{\hat{f}_{U\Theta}(u,\theta)}{\hat{f}_U(u)} := \frac{n^{-1}\sum_{i=1}^n Q_\lambda(U_i - u)K_\kappa(\Theta_i - \theta)}{n^{-1}\sum_{i=1}^n Q_\lambda(U_i - u)}, \tag{1}$$

where, for $\mathbb{U} = [-\pi, \pi)$ ($\mathbb{U} = \mathbb{R}$, respectively) $\hat{f}_U(u)$ is a kernel estimator of f_U at $u \in \mathbb{U}$, with Q_λ being a circular (euclidean, resp.) kernel with (scaled by a, resp.) smoothing factor $\lambda \in (0, \infty)$ and $\hat{f}_{U\Theta}(u, \theta)$ is a kernel estimator of the toroidal (cylindrical, resp.) density $f_{U\Theta}$ at (u, θ), with weight function given by the product between Q_λ, and a circular kernel K_κ with concentration parameter $\kappa \in (0, \infty)$.

Clearly, a more general version of estimator (1) can be obtained by using a weight function different from Q_λ in estimating $f_U(u)$, and/or by using a multivariate kernel other than the product one in estimating $f_{U\Theta}(u, \theta)$.

As discussed in [1], estimating a conditional density can be regarded as a nonparametric regression problem. Specifically, once observed that, as $\kappa \to \infty$, for $(\theta, u) \in [-\pi, \pi) \times \mathbb{U}$,

$$\mathrm{E}[K_\kappa(\Theta - \theta) \mid U = u] \approx f_{\Theta|U}(\theta \mid u),$$

the estimation of $f_{\Theta|U}(\theta \mid u)$ can be treated as the estimation of a regression function with response $K_\kappa(\Theta - \theta)$ and predictor U.

Now, assuming that $f_{\Theta|U}(\theta \mid w)$ admits continuous derivatives up to order $p \in \mathbb{N}$ with respect to w, at $u \in \mathbb{U}$, a pth order series expansion of $f_{\Theta|U}(\theta \mid w)$, for w around u, yields

$$f_{\Theta|U}(\theta \mid w) = f_{\Theta|U}(\theta \mid u) + \sum_{j=1}^{p} \frac{f_{\Theta|U}^{(j,0)}(\theta \mid u)\Psi^j(w - u)}{j!} + o(\Psi^p(w - u)), \tag{2}$$

where, for $u \in \mathbb{U}$,

$$\Psi(u) := \begin{cases} u & \text{if } \mathbb{U} = \mathbb{R}, \\ \sin(u) & \text{if } \mathbb{U} = [-\pi, \pi), \end{cases}$$

and, for $(i, j) \in \mathbb{N} \times \mathbb{N}$,

$$f_{\Theta|U}^{(i,j)}(\theta \mid u) = \frac{\partial^{i+j}}{\partial a^i \partial b^j} f_{\Theta|U}(b \mid a)\big|_{\theta,u}.$$

Hence, a pth degree local polynomial estimator of $f_{\Theta|U}(\theta \mid u)$ can be defined as the solution for β_0 of the following least squares problem

$$\underset{\{\beta_0,\ldots,\beta_p\}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ K_\kappa(\Theta_i - \theta) - \sum_{j=0}^{p} \beta_j \Psi^j(U_i - u) \right\}^2 Q_\lambda(U_i - u). \tag{3}$$

Clearly, for $p = 0$ the solution for β_0 of (3) gives estimator (1), while for $p = 1$ such a solution defines a local linear estimator of $f_{\Theta|U}(\theta \mid u)$, i.e.

$$\hat{f}_{\Theta|U}(\theta \mid u) = \frac{\sum_{i=1}^n L_\lambda(U_i - u)K_\kappa(\Theta_i - \theta)}{\sum_{i=1}^n L_\lambda(U_i - u)}, \tag{4}$$

where

$$L_\lambda(U_i - u) := Q_\lambda(U_i - u)\Bigg\{ \sum_{j=1}^{n} Q_\lambda(U_j - u)\Psi^2(U_j - u) - \Psi(U_i - u)\sum_{j=1}^{n} Q_\lambda(U_j - u)\Psi(U_j - u)\Bigg\}.$$

3. Asymptotics

3.1. *Circular conditioning*

We consider the case where $\mathbb{U} = [-\pi, \pi)$, and denote the conditioning variable as Φ. Moreover, for any circular kernel with smoothing factor $q > 0$, say K_q, and $j \in \mathbb{N}$, we define

$$\eta_j(K_q) := \int_{-\pi}^{\pi} K_q(\alpha)\sin^j(\alpha)\mathrm{d}\alpha, \qquad \text{and} \qquad \nu(K_q) := \int_{-\pi}^{\pi} K_q^2(\alpha)\mathrm{d}\alpha.$$

Asymptotic properties, for both estimators (1) and (4), are collected in the following

THEOREM 3.1 *Given the $[-\pi,\pi)\times[-\pi,\pi)$-valued random sample $(\Phi_1,\Theta_1),\ldots,(\Phi_n,\Theta_n)$ from the unknown density $f_{\Phi\Theta}$, consider estimators* (1) *and* (4)*. If*

i) the marginal density f_Φ of Φ is strictly positive at $\phi \in [-\pi,\pi)$, $f_{\Theta|\Phi}$ and f_Φ are twice continuously differentiable in some neighborhood of $(\theta,\phi) \in [-\pi,\pi)\times[-\pi,\pi)$ and $\phi \in [-\pi,\pi)$, respectively;
ii) K_κ is a second-order circular kernel, which satisfies $\lim_{n\to\infty}\eta_2(K_\kappa) = 0$, $\eta_j(K_\kappa) = o(\eta_2(K_\kappa))$ for each $j > 2$, and $\lim_{n\to\infty} n^{-1}\nu(K_\kappa) = 0$;
iii) Q_λ is a second-order circular kernel which satisfies $\lim_{n\to\infty}\eta_2(Q_\lambda) = 0$, $\eta_j(Q_\lambda) = o(\eta_2(Q_\lambda))$ for each $j > 2$, and $\lim_{n\to\infty} n^{-1}\nu(Q_\lambda) = 0$;

then, for estimator (1)

$$\mathrm{E}[\hat f_{\Theta|\Phi}(\theta \mid \phi)] - f_{\Theta|\Phi}(\theta \mid \phi) = \frac{\eta_2(K_\kappa)}{2} f_{\Theta|\Phi}^{(0,2)}(\theta \mid \phi) + \frac{\eta_2(Q_\lambda)}{2}\left\{ f_{\Theta|\Phi}^{(2,0)}(\theta \mid \phi) + \frac{2f'_\Phi(\phi) f_{\Theta|\Phi}^{(1,0)}(\theta \mid \phi)}{f_\Phi(\phi)} \right\} + o(\eta_2(K_\kappa)) + o(\eta_2(Q_\lambda)),$$

while, for estimator (4)

$$\mathrm{E}[\hat f_{\Theta|\Phi}(\theta \mid \phi)] - f_{\Theta|\Phi}(\theta \mid \phi) = \frac{\eta_2(K_\kappa)}{2} f_{\Theta|\Phi}^{(0,2)}(\theta \mid \phi) + \frac{\eta_2(Q_\lambda)}{2} f_{\Theta|\Phi}^{(2,0)}(\theta \mid \phi) + o(\eta_2(K_\kappa)) + o(\eta_2(Q_\lambda)).$$

Moreover, for both estimators,

$$\mathrm{Var}[\hat f_{\Theta|\Phi}(\theta \mid \phi)] = \frac{\nu(Q_\lambda) f_{\Theta|\Phi}(\theta \mid \phi)}{n f_\Phi(\phi)}\left\{\nu(K_\kappa) - f_{\Theta|\Phi}(\theta \mid \phi)\right\} + O\left(\frac{\nu(Q_\lambda)}{n}\right) + O\left(\frac{\nu(Q_\lambda)\nu(K_\kappa)}{n}\right).$$

Proof. See Appendix. ■

Typical examples of densities which are second-order circular kernels and satisfy the condition on the η_js, stated in both assumption *ii*), and *iii*) are von Mises, wrapped normal and wrapped Cauchy; see [5].

Concerning optimal smoothing degrees, we assume K_κ and Q_λ to be von Mises densities both having zero mean direction, and concentration parameters $\kappa > 0$ and $\lambda > 0$, respectively. Now, for a von Mises density with concentration parameter $q > 0$, i.e. $K_q(\theta) := \{2\pi\mathcal{I}_0(q)\}^{-1}\exp(q\cos(\theta))$, where $\mathcal{I}_0(q)$ stands for the modified Bessel function of order 0, it holds that for sufficiently large q

$$\eta_2(K_q) \approx \frac{1}{q}, \quad \text{and} \quad \nu(K_q) \approx \sqrt{\frac{q}{4\pi}}.$$

Then, using the results of Theorem 3.1, the asymptotic mean squared error (AMSE) of $\hat{f}_{\Theta|U}(\theta \mid u)$, for both local constant and local linear estimators, has the form

$$\text{AMSE}[\hat{f}_{\Theta|\Phi}(\theta \mid \phi)] = \frac{C_1(\kappa\lambda)^{1/2}}{n} - \frac{C_2\lambda^{1/2}}{n} + \frac{C_3^p}{\kappa^2} + \frac{C_4^p}{\lambda^2} + \frac{C_5^p}{\kappa\lambda}, \tag{5}$$

where C_1 and C_2 denote the terms in the asymptotic variance expression which do not depend on κ, λ, and n, while C_j^p, $(p, j) \in \{0, 1\} \times \{3, 4, 5\}$, are the terms appearing in the asymptotic squared biases of local constant ($p = 0$), and local linear ($p = 1$) estimators which do not depend on κ and λ. The values of the smoothing parameters which minimize (5) are provided by the following

Corollary 3.2 *Given the $[-\pi, \pi) \times [-\pi, \pi)$-valued random sample $(\Phi_1, \Theta_1), \ldots, (\Phi_n, \Theta_n)$ from the unknown density $f_{\Phi\Theta}$, consider estimators* (1) *and* (4) *with Q_λ and K_κ being von Mises kernels with concentration parameters λ and κ, respectively. Then, the values of λ and κ minimizing the asymptotic mean squared errors of the estimators are, for $p \in \{0, 1\}$,*

$$\lambda_{\text{AMSE}} = C_1^{-1/3}\left\{4C_4^p\left(\frac{C_4^p}{C_3^p}\right)^{1/4} + 2C_5^p\left(\frac{C_4^p}{C_3^p}\right)^{3/4}\right\}^{1/3} n^{1/3},$$

and

$$\kappa_{\text{AMSE}} = \left(\frac{C_3^p}{C_4^p}\right)^{1/2}\lambda_{\text{AMSE}}.$$

After substituting κ_{AMSE} and λ_{AMSE} in (5), we see that the infimum, over (κ, λ), of $\text{AMSE}[\hat{f}_{\Theta|\Phi}(\theta \mid \phi)]$ has order $O(n^{-2/3})$. Then, from Corollary 3.2 it follows that, when von Mises kernels are employed for smoothing in both the θ-direction and ϕ-direction, the convergence rate for both estimators is of order $n^{-2/3}$, which corresponds to the convergence rate attained by their euclidean counterparts.

3.2. *Linear conditioning*

We now assume $\mathbb{U} = \mathbb{R}$, and denote the conditioning variable as X. Moreover, we set $Q_\lambda(\cdot) := \lambda^{-1}Q(\lambda^{-1}\cdot)$, $\lambda > 0$, with Q being a second order euclidean kernel. Then, letting

$$\mu_j(Q) := \int_{-\infty}^{\infty} x^j Q(x)\mathrm{d}x, \qquad \text{and} \qquad \nu(Q) := \int_{-\infty}^{\infty} Q^2(x)\mathrm{d}x,$$

with $j \in \mathbb{N}$, for both estimators (1) and (4), we get

THEOREM 3.3 *Given the $\mathbb{R}\times[-\pi,\pi)$-valued random sample $(X_1,\Theta_1),\ldots,(X_n,\Theta_n)$ from the unknown density $f_{X\Theta}$, if assumptions i) and ii) of Theorem 3.1 hold with $x \in \mathbb{R}$ in place of $\phi \in [-\pi,\pi)$, and*

i) λ is such that $\lim_{n\to\infty}\lambda = 0$ and $\lim_{n\to\infty} n\lambda = \infty$;

then, for estimator (1)

$$\begin{aligned}\mathrm{E}[\hat{f}_{\Theta|X}(\theta \mid x)] - f_{\Theta|X}(\theta \mid x) =& \frac{\eta_2(K_\kappa)}{2} f^{(0,2)}_{\Theta|X}(\theta \mid x)\\ &+ \frac{\lambda^2\mu_2(Q)}{2}\left\{ f^{(2,0)}_{\Theta|X}(\theta \mid x) + \frac{2f'_X(x)f^{(1,0)}_{\Theta|X}(\theta \mid x)}{f_X(x)}\right\}\\ &+ o(\eta_2(K_\kappa)) + o(\lambda^2),\end{aligned}$$

while, for estimator (4)

$$\mathrm{E}[\hat{f}_{\Theta|X}(\theta \mid x)] - f_{\Theta|X}(\theta \mid x) = \frac{\eta_2(K_\kappa)}{2} f^{(0,2)}_{\Theta|X}(\theta \mid x) + \frac{\lambda^2\mu_2(Q)}{2} f^{(2,0)}_{\Theta|X}(\theta \mid x) + o(\eta_2(K_\kappa)) + o(\lambda^2).$$

Finally, for both estimators

$$\mathrm{Var}[\hat{f}_{\Theta|X}(\theta \mid x)] = \frac{\nu(Q)f_{\Theta|X}(\theta \mid x)}{\lambda n f_X(x)}\{\nu(K_\kappa) - f_{\Theta|X}(\theta \mid x)\} + O\left(\frac{1}{n\lambda}\right) + O\left(\frac{\nu(K_\kappa)}{n\lambda}\right).$$

Proof. See Appendix. ∎

Theorem 3.3 implies that the resulting asymptotic mean squared error, for both $p = 0$ and $p = 1$, takes the form

$$\mathrm{AMSE}[\hat{f}_{\Theta|X}(\theta \mid x)] = \frac{D_1\kappa^{1/2}}{n\lambda} - \frac{D_2}{n\lambda} + \frac{D_3^p}{\kappa^2} + \lambda^4 D_4^p + \frac{\lambda^2 D_5^p}{\kappa}, \tag{6}$$

where D_1 and D_2 stand for the terms not depending on n, λ and κ in the asymptotic variance expression, and D_3^p, D_4^p and D_5^p are the terms not depending on λ and κ in the asymptotic squared bias of local constant ($p = 0$) and local linear ($p = 1$) estimators. The pair (κ, λ) minimizing (6) is given by the following

COROLLARY 3.4 *Given the random sample $(X_1,\Theta_1),\ldots,(X_n,\Theta_n)$ from the unknown density $f_{X\Theta}$ defined on $\mathbb{R}\times[-\pi,\pi)$, consider estimators* (1) *and* (4) *with K_κ being a von*

Mises kernel with concentration parameter κ. Then, for $p \in \{0,1\}$,

$$\lambda_{\text{AMSE}} = D_1^{1/6}\left\{4D_3^p\left(\frac{D_4^p}{D_3^p}\right)^{5/4} + 2D_5^p\left(\frac{D_4^p}{D_3^p}\right)^{3/4}\right\}^{-1/6} n^{-1/6},$$

and

$$\kappa_{\text{AMSE}} = \left(\frac{D_3^p}{D_4^p}\right)^{1/2}\lambda_{\text{AMSE}}^{-2}.$$

Using results of Corollary 3.4 we have that both local constant and local linear estimators attain a convergence rate of order $n^{-2/3}$.

4. Numerical results

4.1. *Simulations*

We present results from a brief simulation study which indicates the relative merits of the local constant estimator (1) and local linear one (4) for both the cases of a real-valued explanatory variable, and a circular explanatory variable, including different sample sizes.

Various authors (for example, [6], [7], and [8]) have proposed variations of cross-validation methods to choose smoothing parameters for conditional density estimates. Here we adopt the following strategy:

(1) Given the sample $u_1, \ldots, u_n$ we can select λ by likelihood cross-validation to maximize

$$\sum_i \log \hat{f}_U^{(-i)}(u_i)$$

where $\hat{f}_U^{(-i)}$ is the kernel estimator of f_U using all the data except the ith observation.

(2) Using this value, say λ_0, we then select κ by likelihood cross-validation to maximize

$$\sum_i \log\left\{\sum_{j\neq i} Q_{\lambda_0}(u_j - u)K_\kappa(\theta_j - \theta_i)\right\}$$

for the local constant case, and

$$\sum_i \log\left\{\sum_{j\neq i} L_{\lambda_0}(u_j - u)K_\kappa(\theta_j - \theta_i)\right\}$$

for the local linear one.

(3) Using these values, say (λ_0, κ_0) as starting values, we can maximize the likelihood cross-validation function for the joint likelihoods, given by the above equations, for both λ and κ.

The simulation settings are as follows. For the linear case, we took $X \sim N(0.5, 0.2^2)$ and then, conditional on this $\Theta \mid X = x \sim vM(2\tan^{-1}(x - 0.5), 4)$. Here we considered

Table 1. Averages (standard deviations) of the integrated squared errors of local constant ($p = 0$) and local linear ($p = 1$) versions of $\hat{f}_{\Theta|U}(\theta|u)$ obtained using 1000 samples of various size, different u and λ and κ selected from a range of values. Left panel: linear conditioning case; right panel: circular conditioning case.

	n	$u = 0.5$	$u = 0.9$		n	$u = 0$	$u = \pi/2$
$p = 0$	100	0.0201 (0.0142)	0.0572 (0.0465)	$p = 0$	100	0.0160 (0.0111)	0.1110 (0.0827)
	500	0.0080 (0.0052)	0.0272 (0.0189)		500	0.0060 (0.0038)	0.0604 (0.0467)
$p = 1$	100	0.0200 (0.0143)	0.0685 (0.0629)	$p = 1$	100	0.0159 (0.0111)	0.0962 (0.0847)
	500	0.0080 (0.0052)	0.0247 (0.0187)		500	0.0060 (0.0037)	0.0336 (0.0263)

Table 2. Averages (standard deviations) of the integrated squared errors of local constant ($p = 0$) and local linear ($p = 1$) versions of $\hat{f}_{\Theta|U}(\theta|u)$ obtained using 1000 samples of various size, different u and λ and κ selected by the cross-validation strategy. Left panel: linear conditioning case; right panel: circular conditioning case.

	n	$u = 0.5$	$u = 0.9$		n	$u = 0$	$u = \pi/2$
$p = 0$	100	0.0330 (0.0160)	0.1060 (0.0365)	$p = 0$	100	0.0227 (0.0139)	0.1591 (0.0999)
	500	0.0230 (0.0062)	0.0960 (0.0180)		500	0.0130 (0.0066)	0.1403 (0.0838)
$p = 1$	100	0.0346 (0.0162)	0.0486 (0.0414)	$p = 1$	100	0.0260 (0.0137)	0.2340 (0.2831)
	500	0.0241 (0.0065)	0.0191 (0.0126)		500	0.0165 (0.0061)	0.0466 (0.0408)

two different values for the predictor X being equal to 0.5 and 0.9. In the circular case we took $\Phi \sim vM(0, 4)$ and then $\Theta \mid \Phi = \phi \sim vM(\text{atan2}(\sin(\phi - 1), 1 + 0.8\cos(\phi - 1)), 4)$ where atan2(a, b) returns the angle between the x-axis and the (non-zero) vector (a, b). Here we considered the conditioning values of ϕ being equal to 0 and $\pi/2$. We took 1000 samples of size $n = 100$ and $n = 500$, and for each sample we computed the estimates through (1) and (4) using the optimal combination of (κ, λ) from a grid search, as well as for (λ_0, κ_0) obtained by cross-validation according to the above strategy. The results are presented in Table 1 and Table 2. As expected, the estimation is better when there are more data available; both in the sample size n, and where the conditioning variable is denser. It can be seen that the cross-validation strategy performs well, particularly for small samples, by comparison with the benchmark minimum ISE in which λ and κ are optimally determined over the entire simulation. However it should be also noted that the standard deviations of the errors have moderately big magnitudes if compared to the respective averages. This reflects the well known high variability of cross-validation selectors.

4.2. *Example*

We illustrate the methods using some data on wind speed and wind directions. The objective is to estimate the density of wind directions, conditional on some specific wind speeds. This could be useful when designing buildings, or planning wind turbine locations. The data, which is taken from an observation station close to the Florida coastline, was collected by the National Data Buoy Center of the NOAC. The wind speeds are recorded to one decimal place, and the directions are recorded as degrees (integers). In order to reduce the serial correlation, we used only every tenth available observation for data covering a full year, which resulted in 1748 observations at intervals of 5 hours. We have chosen the smoothing parameters by cross-validation. The conditional density estimates for wind direction, corresponding to three wind speeds (5, 10, 12) are shown in Figure 1.

We can see that there is a clear dependence of the density on the wind speed. As

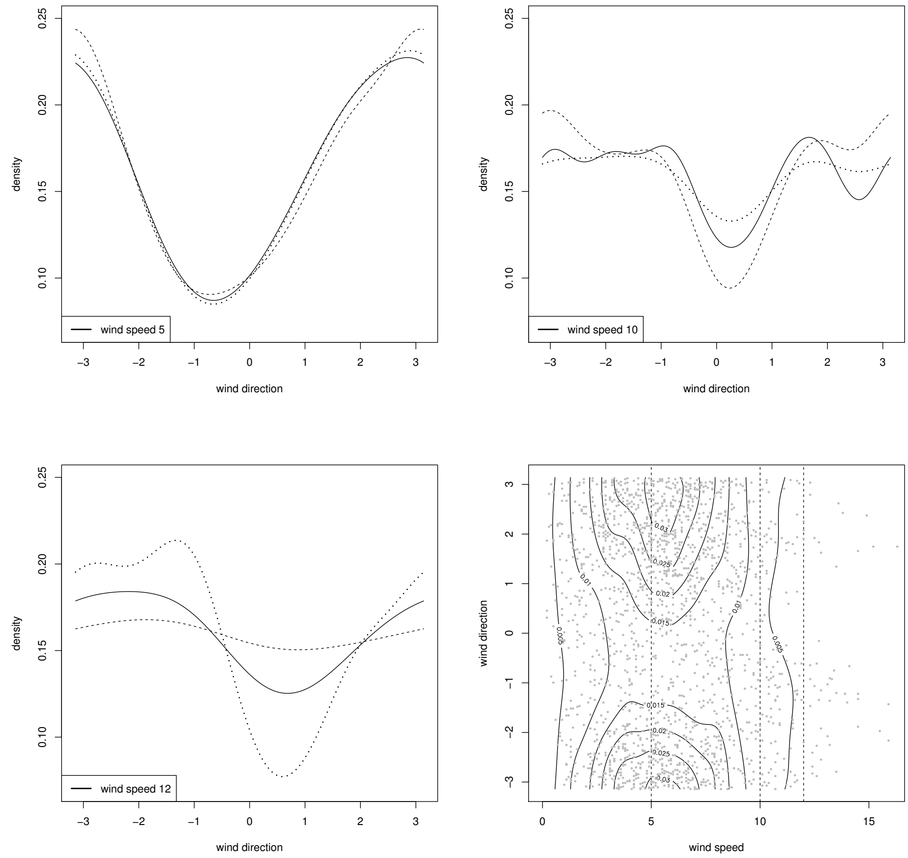

Figure 1. Density estimates for wind directions conditional on wind speeds 5, 10, and 12, and contour plot for the joint density estimate, with observations shown. The continuous lines correspond to estimator (1), the dashed lines to (4), and the dotted lines correspond to the normalized joint density, conditional on the wind speed. Smoothing parameters were selected by stage-wise cross-validation with values $\lambda_0 = 0.41$ and $\kappa_0 = 1.8, 4.4, 0.72$, respectively, corresponding to (1) and $\lambda_0 = 4.1$ and $\kappa_0 = 4.3, 4.2, 0.3$, respectively, corresponding to (4). For the joint density estimate, the smoothing parameters (also selected by cross-validation) were $\lambda = 0.53$ and $\kappa = 2.1$.

a benchmark, we have also computed a standard kernel estimate of the joint density, with smoothing parameters chosen by cross-validation. This is then used to compute a conditional density estimate for comparison. The two estimators perform somewhat differently, with the local linear tending to reveal more structure. Although this is a moderately-sized dataset, the big differences between the estimators at wind speeds 10 and 12 is due to there being less data at these speeds (see the bottom right panel of Figure 1), so all the estimates will have some degree of uncertainty.

Appendix

Proof of Theorem 3.1. Letting $m(\theta, \phi) := \mathrm{E}[K_\kappa(\Theta - \theta) \mid \Phi = \phi]$, assumptions $i) - iii)$ along with results in Theorem 4 of [9], yield

$$\mathrm{E}[\hat{f}_{\Theta|\Phi}(\theta \mid \phi)] - m(\theta, \phi) = \begin{cases} \frac{\eta_2(Q_\lambda)}{2}\left\{m^{(2,0)}(\theta, \phi) + 2\frac{f'_\Phi(\phi) m^{(1,0)}(\theta,\phi)}{f_\Phi(\phi)}\right\} + o(\eta_2(Q_\lambda)) & \text{if } p = 0, \\ \frac{\eta_2(Q_\lambda)}{2} m^{(2,0)}(\theta, \phi) + o(\eta_2(Q_\lambda)) & \text{if } p = 1, \end{cases}$$

and letting $s^2(\theta,\phi) := \mathrm{Var}[K_\kappa(\Theta_i - \theta) \mid \Phi = \phi]$, for both estimators

$$\mathrm{Var}[\hat{f}_{\Theta|\Phi}(\theta \mid \phi)] = \frac{\nu(Q_\lambda)s^2(\theta,\phi)}{nf_\Phi(\phi)} + o\left(\frac{\nu(Q_\lambda)}{n}\right).$$

Now, in virtue of assumption i) and ii), using a Taylor-like expansion yields

$$\begin{aligned}
m(\theta,\phi) &= \int_{-\pi}^{\pi} K_\kappa(\omega-\theta) f_{\Theta|\Phi}(\omega \mid \phi)\mathrm{d}\omega \\
&= \int_{-\pi}^{\pi} K_\kappa(\alpha)\{f_{\Theta|\Phi}(\theta \mid \phi) + \sin(\alpha) f^{(0,1)}_{\Theta|\Phi}(\theta \mid \phi) + 2^{-1}\sin^2(\alpha) f^{(0,2)}_{\Theta|\Phi}(\theta \mid \phi) + o(\sin^2(\alpha))\}\mathrm{d}\alpha \\
&= f_{\Theta|\Phi}(\theta \mid \phi) + 2^{-1}\eta_2(K_\kappa) f^{(0,2)}_{\Theta|\Phi}(\theta \mid \phi) + o(\eta_2(K_\kappa)),
\end{aligned}$$

and

$$\begin{aligned}
s^2(\theta,\phi) &= \int_{-\pi}^{\pi} K^2_\kappa(\omega-\theta) f_{\Theta|\Phi}(\omega \mid \phi)\mathrm{d}\omega - \left\{\int_{-\pi}^{\pi} K_\kappa(\omega-\theta) f_{\Theta|\Phi}(\omega-\theta)\mathrm{d}\omega\right\}^2 \\
&= f_{\Theta|\Phi}(\theta,\phi)\{\nu(K_\kappa) - f_{\Theta|\Phi}(\theta,\phi)\} + o(\nu(K_\kappa)) + O(\eta_2^2(K_\kappa)).
\end{aligned}$$

Then, substituting the above approximations for $m(\theta,\phi)$ and $s^2(\theta,\phi)$ into the asymptotic biases and variance expressions, respectively, leads to the results.

Proof of Corollary 3.2. We need similar arguments as those used in [10] for deriving optimal smoothing of euclidean kernel conditional density estimator. Specifically, after setting the partial derivatives of (5) to zero, some simplifications give

$$\frac{C_1}{2n} - \frac{2C_3^p}{\kappa^{5/2}\lambda^{1/2}} - \frac{C_5^p}{\kappa^{3/2}\lambda^{3/2}} = 0, \tag{7}$$

and

$$\frac{C_1}{2n} - \frac{C_2}{2n\kappa^{1/2}} - \frac{2C_4^p}{\kappa^{1/2}\lambda^{5/2}} - \frac{C_5^p}{\kappa^{3/2}\lambda^{3/2}} = 0. \tag{8}$$

Finally, subtracting (8) to (7), and solving for κ yields

$$\tilde{\kappa} = \left\{\frac{4C_3^p\lambda^2 n}{C_2\lambda^{5/2} + 4C_4^p n}\right\}^{1/2}.$$

Now, in order to obtain λ_{AMSE}, replace κ by $\tilde{\kappa}$ in (7), then use a first order Taylor series approximation for $n^{-1}\lambda^{1/2}$ around 0, and finally solve the resulting equation. Then, substituting λ_{AMSE} in $\tilde{\kappa}$, and retaining the dominant term of the first order Taylor series approximation lead to κ_{AMSE}.

Proof of Theorem 3.3. The results easily follow by using the same arguments as in the proof of Theorem 3.1, with x in place of ϕ. Starting from the fact that, in virtue of assumption i), one has

$$\mathrm{E}[\hat{f}_{\Theta|X}(\theta \mid x)] - m(\theta, x) = \begin{cases} \frac{\lambda^2\mu_2(Q)}{2}\left\{m^{(2,0)}(\theta,x) + 2\frac{f'_X(x)m^{(1,0)}(\theta,x)}{f_X(x)}\right\} + o(\lambda^2) & \text{if } p = 0, \\ \frac{\lambda^2\mu_2(Q)}{2}m^{(2,0)}(\theta,x) + o(\lambda^2) & \text{if } p = 1, \end{cases}$$

and, for both $p = 0$ and $p = 1$

$$\mathrm{Var}[\hat{f}_{\Theta|X}(\theta \mid x)] = \frac{\nu(Q)s^2(\theta,x)}{n\lambda f_X(x)} + o\left(\frac{1}{n\lambda}\right).$$

Now, asymptotic approximations of $m(\theta, x)$ and $s^2(\theta, x)$ follow by adapting their corresponding approximations in the proof of Theorem 3.1 to the linear setting.

Proof of Corollary 3.4. The result follows by reasoning as in the proof of Corollary 3.2, and then modifying the asymptotic approximations according to the assumptions in Theorem 3.3.

Aknowledgements

The authors thank a Referee for the constructive comments which led to an improved presentation.

References

[1] Fan J, Yao Q, Tong H. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika. 1996; 83:189–206.
[2] Di Marzio M, Panzera A, Taylor CC. Non-parametric Regression for Circular Responses. Scandinavian Journal of Statistics. 2013; 40:238–255.
[3] Di Marzio M, Panzera A, Taylor CC. Nonparametric circular quantile regression. Journal of Statistical Planning and Inference. 2016; 170:1–14.
[4] Mardia KV, Jupp EP. Directional statistics. New York: Wiley; 1999; pp. 16–17.
[5] Di Marzio M, Panzera A, Taylor CC. Kernel density estimation on the torus. Journal of Statistical Planning and Inference. 2011; 141:2156–2173.
[6] Bashtannyk DM, Hyndman RJ. Bandwidth selection for kernel conditional density estimation. Computational Statistics & Data Analysis. 2001; 36:279–298.
[7] Fan J, Yim HT. A crossvalidation method for estimating conditional densities. Biometrika. 2004; 91:819–834.
[8] Hall P, Racine J, Li Q. Cross-Validation and the Estimation of Conditional Probability Densities. Journal of the American Statistical Association. 2004; 99:1015–1026.
[9] Di Marzio M, Panzera A, Taylor CC. Local polynomial regression for circular predictors. Statistics & Probability Letters. 2009; 79:2066–2075.
[10] Hyndman RJ, Bashtannyk DM, Grunwald GK. Estimating and Visualizing Conditional Densities. Journal of Computational and Graphical Statistics. 1996; 5:315–336.