



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/94262/>

Version: Accepted Version

Article:

Di Marzio, M, Panzera, A and Taylor, CC (2016) Nonparametric circular quantile regression. *Journal of Statistical Planning and Inference*, 170. pp. 1-14. ISSN: 0378-3758

<https://doi.org/10.1016/j.jspi.2015.08.004>

© 2015. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Nonparametric circular quantile regression

Marco Di Marzio^a, Agnese Panzera^b, Charles C. Taylor^{c,1}

^a*DMQTE, Università di Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy.*

^b*DiSIA, Università di Firenze, Viale Morgagni 59, 50134 Firenze, Italy.*

^c*Department of Statistics, University of Leeds, Leeds LS2 9JT, UK.*

Abstract

We discuss nonparametric estimation of conditional quantiles of a circular distribution when the conditioning variable is either linear or circular. Two different approaches are pursued: inversion of a conditional distribution function estimator, and minimization of a smoothed check function. Local constant and local linear versions of both estimators are discussed.

Simulation experiments and a real data case study are used to illustrate the usefulness of the methods.

Key words: Check Function, Circular Quantile, Circular Distribution Function, Predictive Interval, Wind Turbine
2000 MSC: 62G07 - MSC 62G08

1. Introduction

Quantile regression focuses on estimating either the conditional median or other quantiles of the response variable. When compared with ordinary least squares regression, we can say that quantile estimates are: a) more robust when the distributions of the covariates and/or error terms are heavy tailed; b) more easily interpretable if the conditional distributions are asymmetric. However, quantile regression is typically used to gain insights on the whole underlying conditional distribution through the estimation of various percentiles. Similarly, predictive intervals are often determined by estimating pairs of extreme conditional quantiles.

Methods used for quantile regression range from fully parametric — in the simplest case, we have a normal distributed response, along with linear quantile curves — to fully nonparametric, where local smoothing of observed quantiles is carried out, see, for example, Jones and Hall (1990). Usually, either the quantile curves or response distribution is estimated nonparametrically. Jones and Noufaily (2013) provide a critical account of the literature.

In practice, as it happens for the majority of circular statistics indices, a set of quantiles can be estimated for each possible choice of the origin. In some situations there may be external information, or data could be confined to a small arc of the circle (reproducing an euclidean-like scenario), but more generally (as in our example) we could choose the origin according to a *minimum width* criterion. That is, for any specified quantile, the origin is chosen so as to minimise the width corresponding to the estimated interval. In any case, we observe that a change of the origin simply generates a linear shift on the cumulative distribution function (CDF) values, i.e. in the quantile order. Such an equivariance form links estimates based on same data but with different origin.

Also, the definition, and then the estimation, of the circular CDF is not an obvious extension of the standard theory. It is perhaps for these reasons that quantile regression seems unexplored in the circular setting, even from a parametric perspective. The CDF of a random angle Θ having density f is defined as $F(\theta) := \int_{-\pi}^{\theta} f(u)du$, $\theta \in [-\pi, \pi)$; this implies that $F(-\pi) = 0$ and $F(\pi) = 1$. When f is seen as a periodic function having \mathbb{R} as its support, then

$$\lim_{a \rightarrow \pm\infty} F(a) = \pm\infty,$$

Email addresses: mdimarzio@unich.it (Marco Di Marzio), a.panzera@disia.unifi.it (Agnese Panzera), charles@maths.leeds.ac.uk (Charles C. Taylor)

¹Corresponding author

and for a quantile of order $c \in \mathbb{R}$, say q_c , and an integer k , we have

$$F(q_c + k2\pi) = F(q_c) + k.$$

Clearly, this real-valued representation is pretty unfamiliar and practically un-necessary because all of the information is contained by its restriction over the interval of angles $[-\pi, \pi)$. Although this circle representation introduces an apparent discontinuity at the origin (F is not periodic), it implies a CDF assigning probabilities, and quantiles order ranging in $[0, 1]$.

In this paper, we propose estimators for circular *conditional* quantiles. Other than the usual pros of the nonparametric approach, notice that, because circular quantities are defined on a circle, boundary problems do not arise when the predictor is circular. Surely, we could still conceive circular populations with densities that are either discontinuous or whose support is a subset of the circle, but after a little thought we would conclude that such data are well represented also on the line, and conveniently studied using euclidean methods. For example, we could estimate the CDF using the methods proposed by Berg and Politis (2009). In general, consider that if the baseline of euclidean theory is a linear quantile function, a straight line is inadequate to interpolate quantiles whereas periodic functions are the natural tool.

Specifically, we investigate two strategies for the above task: a) inversion of *double-kernel* estimators of the conditional distribution function; b) minimization of a smoothed *circular check function*. We consider local constant and local linear versions for both estimators, and derive their asymptotic properties.

Section 2 deals with local constant and local linear versions of double-kernel estimator of a conditional circular distribution function, for both circular and linear conditioning cases. In Section 3 we define conditional quantile estimators as inverses of conditional distribution function estimators, while, in Section 4, we follow the approach of minimising a smoothed check function. In Section 5 we include a small simulation study to investigate and compare the finite sample performance of our estimators. Finally, in Section 6 we illustrate some of the methods using an application to aspects of wind turbine installation, by considering data on wind speed and direction at a particular location.

2. Conditional circular distribution function estimator

Let (U, Θ) be a $\mathbb{U} \times [-\pi, \pi)$ -valued random vector, where U is a random variable taking values over a generic domain, and Θ is a random angle. Let $f(\theta | u)$ denote the conditional density of Θ given $U = u$, at $\theta \in [-\pi, \pi)$. We address the problem of nonparametrically estimating $F(\theta | u) = \int_{-\pi}^{\theta} f(t | u) dt$, for $u \in \mathbb{U}$, in the cases where \mathbb{U} is the unit circle or $\mathbb{U} = [0, 1]$, and F is absolutely continuous. Notice that setting \mathbb{U} equal to the interval $[0, 1]$, which is common practice in nonparametric modeling, does not lead to any loss of generality.

In order to construct our estimator we start from the kernel estimator of $f(\theta | u)$, which can be conceived as the ratio between the kernel estimator of the joint density and the kernel estimator of the marginal one. In particular, letting $(U_1, \Theta_1), \dots, (U_n, \Theta_n)$ be independent copies of (U, Θ) , for $\alpha \in [-\pi, \pi)$, we write

$$\hat{f}_{\lambda, \kappa}(\alpha | u) = \frac{\hat{f}_{\lambda, \kappa}(u, \alpha)}{\hat{f}_{\lambda}(u)} = \frac{\sum_{i=1}^n Q_{\lambda}(u - U_i) K_{\kappa}(\alpha - \Theta_i)}{\sum_{i=1}^n Q_{\lambda}(u - U_i)}$$

with K_{κ} being a *circular kernel* with concentration parameter $\kappa \in (0, +\infty)$, i.e. a non-negative function defined on the unit circle, admitting an uniformly convergent Fourier series representation $K_{\kappa}(\theta) = (2\pi)^{-1} \{1 + \sum_{j=1}^{\infty} \gamma_j(\kappa) \cos(j\theta)\}$, where $\gamma_j(\kappa)$ is a strictly monotonic function of κ . Moreover, when \mathbb{U} is the unit circle ($[0, 1]$, resp.), Q_{λ} is a circular kernel (euclidean kernel supported on $[-1, 1]$, resp.) with (scaled by a, resp.) smoothing factor $\lambda \in (0, +\infty)$.

Then, a smooth estimator for $F(\theta | u)$ can be defined as

$$\hat{F}_{\lambda, \kappa}(\theta | u) = \int_{-\pi}^{\theta} \hat{f}_{\lambda, \kappa}(\alpha | u) = \frac{\sum_{i=1}^n Q_{\lambda}(u - U_i) \{W_{\kappa}(\theta - \Theta_i) - W_{\kappa}(-\pi - \Theta_i)\}}{\sum_{i=1}^n Q_{\lambda}(u - U_i)}, \quad (1)$$

where $W_{\kappa}(a) = \int_{-\pi}^a K_{\kappa}(t) dt$, for $a \in \mathbb{R}$.

Remark 1. Notice that estimates produced by (1) are bona fide circular CDFs values. Consider, in fact, that they are nondecreasing, right-continuous, and have these properties

- a) $\hat{F}_{\lambda,\kappa}(-\pi | u) = 0$, $\hat{F}_{\lambda,\kappa}(\pi | u) = 1$,
- b) $\lim_{a \rightarrow \pm\infty} \hat{F}_{\lambda,\kappa}(a | u) = \pm\infty$,
- c) $\hat{F}_{\lambda,\kappa}(q_c + k2\pi | u) = \hat{F}_{\lambda,\kappa}(q_c | u) + k$, $c \in \mathbb{R}$.

Also, estimator (1) is equivariant in the sense that, when we rotate the origin, its values remain shifted by a constant term. This property describes, obviously, also how circular quantiles estimates, defined as the inverse of (1), change when the origin is rotated.

A general class of *double-kernel local polynomial* estimators for $F(\theta | u)$, which includes estimator (1), can be derived as follows. Let

$$m(\theta, w) = E \left[\int_{-\pi}^{\theta} \hat{f}(u | w) du \right] = E[W_{\kappa}(\theta - \Theta_i) - W_{\kappa}(-\pi - \Theta_i) | U = w],$$

we know that, under the usual mild conditions, $\int_{-\pi}^{\theta} \hat{f}(u | w) du$ is consistent for $F(\theta | w)$, thus for a big enough κ we can write $m(\theta, w) \sim F(\theta | w)$. Now consider the following p th order series expansion of $F(\theta | w)$ for w around $u \in \mathbb{U}$,

$$F(\theta | w) = F(\theta | u) + \sum_{j=1}^p \frac{F^{(0,j)}(\theta | u) \Psi^j(w - u)}{j!} + o(\Psi^{p+1}(w - u)),$$

where $\Psi(u) = \sin(u)$ (u , respectively) when \mathbb{U} is the unit circle ($[0, 1]$, resp.), and, for $(i, j) \in \mathbb{N} \times \mathbb{N}$,

$$F^{(i,j)}(\theta | u) = \frac{\partial^{i+j}}{\partial a^i \partial b^j} F(a | b)|_{\theta, u},$$

Then, a p th degree local polynomial estimator of $F(\theta | u)$ can be defined as the solution for β_0 of

$$\operatorname{argmin}_{\beta_0, \dots, \beta_p} \sum_{i=1}^n \left\{ W_{\kappa}(\theta - \Theta_i) - W_{\kappa}(-\pi, \pi) - \sum_{j=0}^p \beta_j \Psi^j(u - U_i) \right\}^2 Q_{\lambda}(u - U_i).$$

Clearly, when $p = 0$, the solution for β_0 is given by estimator (1), whereas when $p = 1$, such a solution defines a local linear estimator of $F(\theta | u)$, i.e.

$$\hat{F}_{\lambda,\kappa}(\theta | u) = \frac{\sum_{i=1}^n L_{\lambda}(u - U_i) \{W_{\kappa}(\theta - \Theta_i) - W_{\kappa}(-\pi - \Theta_i)\}}{\sum_{i=1}^n L_{\lambda}(u - U_i)}, \quad (2)$$

where

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

It is possible, although rarely in practice, that a local linear estimate of $F(\theta | u)$ is non-monotone, yielding *non-bona-fide* estimates. Importantly, when $p = 1$ the solution for β_1 of the above minimization problem is an estimator of the conditional density. Although conditional density estimation in the circular-circular or linear-circular setting seems a promising research field, we will not pursue it further in the present paper.

Remark 2. Interestingly, observe that the structure of the estimators remains the same in the case of a multidimensional conditioning variable: it simply suffices to replace the one-dimensional kernels by multidimensional ones.

In what follows, we denote as g the density of U , and assume K_{κ} to be a von Mises density with zero mean direction and concentration parameter $0 < \kappa < \infty$, i.e. $K_{\kappa}(\theta) = \{2\pi \mathcal{I}_0(\kappa)\}^{-1} \exp(\kappa \cos(\theta))$, with $\mathcal{I}_0(\cdot)$ being the modified Bessel function of the first kind and order 0.

2.1. Conditioning on a circular variable

Here we consider the case in which \mathbb{U} is the unit circle, and denote the conditioning variable by Φ . Also, after defining $\eta_j(K_\kappa) = \int_{-\pi}^{\pi} K_\kappa(\theta) \sin^j(\theta) d\theta$, we say that K_κ is a s th sin-order circular kernel if and only if $\eta_0(K_\kappa) = 1$, $\eta_j(K_\kappa) = 0$ for each $0 < j < s$, and $\eta_s(K_\kappa) \neq 0$.

In particular, we assume that Q_λ is a second sin-order circular kernel, whose Fourier coefficients, say $\gamma_j(\lambda)$, for even $j > 0$, satisfy

$$\lim_{\lambda \rightarrow \infty} \{1 - \gamma_j(\lambda)\} / \{1 - \gamma_2(\lambda)\} = j^2/4. \quad (3)$$

This latter condition assures that, for even $j > 2$, $\eta_j(Q_\lambda) = o(\eta_2(Q_\lambda))$, as clarified in the following

Remark 3. For a circular kernel K_κ , and even $j > 0$

$$\eta_j(K_\kappa) = \frac{1}{2^{j-1}} \left\{ \binom{j-1}{j/2} + \sum_{s=1}^{j/2} (-1)^{j+s} \binom{j}{j/2+s} \gamma_{2s}(\kappa) \right\}.$$

If K_κ has second sin-order, then $\eta_j(K_\kappa) = o(\eta_2(K_\kappa))$, for each even $j > 2$, if and only if $\gamma_{2s}(K_\kappa) = 1$ for each $1 < s < j/2$. Now, the fact that $\lim_{\kappa \rightarrow \infty} \gamma_j(\kappa) = 1$ leads to condition (3).

Typical examples of densities which are second sin-order circular kernels and satisfy above condition are: von Mises, wrapped normal and wrapped Cauchy.

Now, letting $R(Q_\lambda) = \int_{-\pi}^{\pi} Q_\lambda^2(\theta) d\theta$, for both local constant and local linear estimators of $F(\theta | \phi)$, we get

Theorem 1. Given the $[-\pi, \pi) \times [-\pi, \pi)$ -valued random sample $(\Phi_1, \Theta_1), \dots, (\Phi_n, \Theta_n)$, if

- i) the density g of Φ is strictly positive, $F(\theta | \phi)$ and $g(\phi)$ are twice continuously differentiable in some neighborhoods of $\theta \in [-\pi, \pi)$ and $\phi \in [-\pi, \pi)$, respectively;
- ii) κ is such that $\lim_{n \rightarrow \infty} \kappa^{-1} = 0$;
- iii) λ increases with n , and $\lim_{n \rightarrow \infty} \eta_2(Q_\lambda) = 0$ and $\lim_{n \rightarrow \infty} n^{-1} R(Q_\lambda) = 0$;

then, for estimator (1) we have

$$\begin{aligned} E[\hat{F}_{\lambda, \kappa}(\theta | \phi)] - F(\theta | \phi) &= \frac{\eta_2(Q_\lambda)}{2} \left\{ F^{(0,2)}(\theta | \phi) + 2 \frac{g'(\phi)}{g(\phi)} F^{(0,1)}(\theta | \phi) \right\} \\ &+ \frac{1}{2\kappa} \{ F^{(2,0)}(\theta | \phi) - F^{(2,0)}(-\pi | \phi) \} + o(\eta_2(Q_\lambda) + \kappa^{-1}), \end{aligned}$$

while, for estimator (2) we have

$$E[\hat{F}_{\lambda, \kappa}(\theta | \phi)] - F(\theta | \phi) = \frac{\eta_2(Q_\lambda)}{2} F^{(0,2)}(\theta | \phi) + \frac{1}{2\kappa} \{ F^{(2,0)}(\theta | \phi) - F^{(2,0)}(-\pi | \phi) \} + o(\eta_2(Q_\lambda) + \kappa^{-1}),$$

and, in both cases,

$$\text{Var}[\hat{F}_{\lambda, \kappa}(\theta | \phi)] = \frac{R(Q_\lambda)}{ng(\phi)} \left\{ F(\theta | \phi)[1 - F(\theta | \phi)] - \frac{1}{\sqrt{\pi\kappa}} [F^{(1,0)}(\theta | \phi) + F^{(1,0)}(-\pi | \phi)] \right\} + o\left(\frac{R(Q_\lambda)}{n\sqrt{\kappa}}\right)$$

Proof. See appendix for Proof □

Because the asymptotic variances are the same, a simple index showing the relative efficiency of the local constant and local linear methods could be:

$$\text{EFF}(\phi) = \left| \frac{F^{(0,2)}(\theta | \phi)}{F^{(0,2)}(\theta | \phi) + \frac{2g'(\phi)}{g(\phi)} F^{(0,1)}(\theta | \phi)} \right|, \quad (4)$$

and we see that the quantity $g'(\phi)/g(\phi)$ is zero if the design density g is uniform, whereas it is proportional to $\sin(\phi)$ if g is von Mises or Wrapped Cauchy. In all of these cases, the behavior of local linear and local constant tend to be nearly similar in the tails and identical at the critical points or flat portions of the design density.

2.2. Conditioning on a linear variable

We now assume $\mathbb{U} = [0, 1]$, and indicate 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 supported on $[-1, 1]$.

Then, letting $\mu_j(Q) = \int_{-1}^1 x^j Q(x) dx$, $j \in \mathbb{N}$, and $R(Q) = \int_{-1}^1 Q^2(x) dx$, for estimator (1), we get

Theorem 2. *Given the $[0, 1] \times [-\pi, \pi]$ -valued random sample $(X_1, \Theta_1), \dots, (X_n, \Theta_n)$, if assumptions i) and iii) of Theorem 1 hold with $\lambda < x < 1 - \lambda$ in place of ϕ , and*

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

then, for estimator (1) we have

$$E[\hat{F}_{\lambda, \kappa}(\theta | x)] - F(\theta | x) = \frac{\lambda^2 \mu_2(Q)}{2} \left\{ F^{(0,2)}(\theta | x) + 2 \frac{g'(x)}{g(x)} F^{(0,1)}(\theta | x) \right\} + \frac{1}{2\kappa} \{ F^{(2,0)}(\theta | x) - F^{(2,0)}(-\pi | x) \} + o(\lambda^2 + \kappa^{-1}),$$

while, for estimator (2) we have

$$E[\hat{F}_{\lambda, \kappa}(\theta | x)] - F(\theta | x) = \frac{\lambda^2 \mu_2(Q)}{2} F^{(0,2)}(\theta | x) + \frac{1}{2\kappa} \{ F^{(2,0)}(\theta | x) - F^{(2,0)}(-\pi | x) \} + o(\lambda^2 + \kappa^{-1}),$$

and, in both cases,

$$\text{Var}[\hat{F}_{\lambda, \kappa}(\theta | x)] = \frac{R(Q)}{n\lambda g(x)} \left\{ F(\theta | x)[1 - F(\theta | x)] - \frac{1}{\sqrt{\pi\kappa}} [F^{(1,0)}(\theta | x) + F^{(1,0)}(-\pi | x)] \right\} + o\left(\frac{1}{n\lambda\sqrt{\kappa}}\right)$$

Proof. See appendix for Proof □

Comparing the efficiency of local constant and local linear fits, we observe that here the situation could be quite different from the circular conditioning case in the tails of the design density. In fact, it can easily be seen that if g belongs to the exponential family, then the corresponding index (4) can be made arbitrarily small for sufficiently big x . For example, observe that $g'(x)/g(x) = -x$ for the standard normal density. When the tails are not exponentially bounded the situation improves for the local constant method. For example, in the standard Cauchy case we have $g'(x)/g(x) = -2x/(1 + x^2)$, such that $\lim_{|x| \rightarrow \infty} \text{EFF}(x) = 1$ making the behaviors identical.

The above theorem does not address the estimation at the *boundary* region, defined as $\{x \leq \lambda\} \cup \{x \geq 1 - \lambda\}$. Although the boundary region is asymptotically vanishing, it is known that the estimation is less accurate there. We do not present specific results on this because the usual theory applies with modifications exclusively due to the structure of the circular cumulative distribution estimator.

3. Double-kernel estimator of circular quantiles

Denote as $q_\alpha(u)$ the α th quantile of the conditional distribution of Θ given $U = u$, i.e. $F^{-1}(\alpha | u)$, $\alpha \in (0, 1)$. As the inverse of F , the definition of a quantile for circular distributions is awkward due to the need of an arbitrary choice of where to “cut”, the circle. Sensible choices include $\theta_0 = \arg \min f(\theta)$ or $\theta_0 = \pi + \arg \max f(\theta)$, but in general the density f will be unknown. Avoiding this issue requires the estimation of a family of quantiles corresponding to each possible θ_0 and interest may then lie in estimating an *interval* which maximizes some quantity of interest.

We address the problem of nonparametrically estimating $q_\alpha(u)$ for the case when U is either circular or linear.

A kernel estimator for the circular quantile $q_\alpha(u)$, $\alpha \in (0, 1)$, can be defined as

$$\hat{q}_{\alpha, \lambda, \kappa}(u) = \hat{F}_{\lambda, \kappa}^{-1}(\alpha | u) \tag{5}$$

Clearly the local constant (local linear, respectively) version of $\hat{F}_{\lambda, \kappa}$ in Equation (5) ((2), resp.) gives a local constant (local linear, resp.) version of the estimator $\hat{q}_{\alpha, \lambda, \kappa}(u)$. However recall that, very occasionally, local linear gives non-monotonic estimates of F , and this would give non-unique quantile estimates that may not increase with their order. As a consequence, confidence intervals would no longer be *nested*; this latter property assures that a confidence interval contains *any* other one defined in the same way but with smaller coverage. However, this problem is practically

solved by selecting the quantile solution which is the maximum solution of (5) if $\alpha > 0.5$ and the minimum solution if $\alpha < 0.5$.

Due to the angular nature of estimator (5), its performance can be assessed by the following accuracy measure

$$\mathcal{L}[\hat{q}_{\alpha,\lambda,\kappa}(u)] = E[2\{1 - \cos(\hat{q}_{\alpha,\lambda,\kappa}(u) - q_\alpha(u))\}], \quad (6)$$

which can be regarded as the circular counterpart of the mean squared error of a *linear* estimator, as clarified in the following

Remark 4. Let \mathbf{p}_1 and \mathbf{p}_2 respectively denote the cartesian coordinates of angles θ_1 and θ_2 , i.e. $\mathbf{p}_i = (\cos(\theta_i), \sin(\theta_i))$, for $i \in \{1, 2\}$. Then, using the fact that $\|\mathbf{p}_1\| = \|\mathbf{p}_2\| = 1$, one has $\|\mathbf{p}_1 - \mathbf{p}_2\|^2 = 2(1 - \cos(\theta_1 - \theta_2))$.

Specifically, letting

$$\xi(Q_\lambda) = \begin{cases} \eta_2(Q_\lambda) & \text{if } \mathbb{U} \text{ is the circle,} \\ \lambda^2 \mu_2(Q) & \text{if } \mathbb{U} = [0, 1], \end{cases}$$

and

$$\nu(Q_\lambda) = \begin{cases} R(Q_\lambda) & \text{if } \mathbb{U} \text{ is the circle,} \\ \lambda^{-1} R(Q) & \text{if } \mathbb{U} = [0, 1], \end{cases}$$

for the local constant version of our estimator, in both linear and circular conditioning cases we get

Theorem 3. Given the $\mathbb{U} \times [-\pi, \pi]$ -valued random sample $(U_1, \Theta_1), \dots, (U_n, \Theta_n)$, consider estimator (5) with $\hat{F}_{\lambda,\kappa}$ being defined in (1). If $f(q_\alpha(u) | u) \neq 0$, assumptions i) and iii) of Theorem 1 hold, and assumption ii) of Theorem 1 or assumption i) of Theorem 2 is satisfied when \mathbb{U} is the circle and $\mathbb{U} = [0, 1]$ respectively, then

$$\begin{aligned} \mathcal{L}[\hat{q}_{\alpha,\lambda,\kappa}(u)] = & \left\{ \frac{\xi(Q_\lambda)\{F^{(0,2)}(q_\alpha(u) | u) + 2\frac{g'(u)}{g(u)}F^{(0,1)}(q_\alpha(u) | u)\}}{2f(q_\alpha(u) | u)} + \frac{F^{(2,0)}(q_\alpha(u) | u) - F^{(2,0)}(-\pi | u)}{2\kappa f(q_\alpha(u) | u)} \right\}^2 \\ & + \frac{\nu(Q_\lambda)\{\alpha(1 - \alpha) - (\pi\kappa)^{-1/2}[f(q_\alpha(u) | u) + f(-\pi | u)]\}}{ng(u)f^2(q_\alpha(u) | u)} + o\left(\xi^2(Q_\lambda) + \kappa^{-2} + n^{-1}(\pi\kappa)^{-1/2}\nu(Q_\lambda)\right) \end{aligned}$$

Proof. See appendix for Proof □

When local linear weights are employed in Equation (5), we obtain

Theorem 4. Given the $\mathbb{U} \times [-\pi, \pi]$ -valued random sample $(U_1, \Theta_1), \dots, (U_n, \Theta_n)$, consider estimator (5) with $\hat{F}_{\lambda,\kappa}$ being defined in (2). Under the assumptions of Theorem 3, then

$$\begin{aligned} \mathcal{L}[\hat{q}_{\alpha,\lambda,\kappa}(u)] = & \left\{ \frac{\xi(Q_\lambda)F^{(0,2)}(q_\alpha(u) | u)}{2f(q_\alpha(u) | u)} + \frac{F^{(2,0)}(q_\alpha(u) | u) - F^{(2,0)}(-\pi | u)}{2\kappa f(q_\alpha(u) | u)} \right\}^2 \\ & + \frac{\nu(Q_\lambda)\{\alpha(1 - \alpha) - (\pi\kappa)^{-1/2}[f(q_\alpha(u) | u) + f(-\pi | u)]\}}{ng(u)f^2(q_\alpha(u) | u)} + o\left(\xi^2(Q_\lambda) + \kappa^{-2} + n^{-1}(\pi\kappa)^{-1/2}\nu(Q_\lambda)\right) \end{aligned}$$

Proof. See appendix for Proof □

As for links to previous work, Di Marzio et al. (2012) defined an estimator of a (unconditional) circular quantile as the inverse of a kernel estimator of a circular distribution function. Here, the definition of a circular distribution function implies the choice of a (finite) lower limit of integration, which yields a kernel estimator with an ‘extra term’ which depends on it. Clearly, accuracy (and optimal smoothing) for both circular distribution and circular quantile estimators depend on this integration limit.

Concerning the optimal smoothing we treat separately the cases where U is circular or linear. In particular, in the first case we have that, since for a circular kernel the smoothing parameter is not a scale factor, the optimal smoothing theory requires a specific form for Q_λ . Here, we assume that Q_λ is a von Mises kernel with mean direction 0, and concentration parameter $\lambda > 0$. Hence, recalling that for this kernel, when λ increases $\eta_2(Q_\lambda) \approx \lambda^{-1}$, and

$R(Q_\lambda) \approx \{\lambda/(4\pi)\}^{1/2}$, when we choose $\lambda \ll \kappa$, we have that the leading term of $\mathcal{L}[\hat{q}_{\alpha,\lambda,\kappa}(\phi)]$ essentially depends only on λ , whose optimal value turns out to be

$$\lambda_0 = \left\{ \frac{2B^2(q_\alpha(\phi) | \phi)g(\phi)n\pi^{1/2}}{\alpha(1-\alpha)} \right\}^{2/5}, \quad (7)$$

where, for $u \in \mathbb{U}$, $B(q_\alpha(u) | u) = F^{(0,2)}(q_\alpha(u) | u) + 2\frac{g'(u)}{g(u)}F^{(0,1)}(q_\alpha(u) | u)$ in the local constant case, and $B(q_\alpha(u) | u) = F^{(0,2)}(q_\alpha(u) | u)$ in the local linear one.

Conversely, when $\mathbb{U} = [0, 1]$, choosing $\lambda \gg \kappa^{-1/2}$ leads to the following optimal value of the (interior) bandwidth

$$\lambda_0 = \left\{ \frac{\alpha(1-\alpha)R(Q)}{\mu_2^2(Q)B^2(q_\alpha(x) | x)g(x)n} \right\}^{1/5},$$

which corresponds to the optimal bandwidth of local double-kernel estimators of *euclidean* conditional quantiles when the bandwidth in the dependent variable direction is assumed to be smaller than that in the direction of the conditioning one.

Notice also that, as n increases, while the smoothing parameter in the Θ -direction goes to infinity, the smoothing parameter in the U -direction goes to zero when U is *linear* and to infinity when U is a circular random variable. This is the reason why we write $\lambda \gg \kappa^{-1/2}$ and $\lambda \ll \kappa$ to indicate that in both cases there is asymptotically lower smoothing degree on the side of κ .

The idea behind the above approach to optimal smoothing is that it is well known (see, for example, Yu and Jones (1998)) that the value of κ has not a very big impact in the quantile estimate (see also results in Figure 3) because it concerns smoothing only at the level of distribution function, whereas the choice of λ is the crucial one. Therefore, we can rely on a few degrees of freedom in setting the value of κ , in order to get an optimal (and simple) setting of λ . In particular, the choice of less smoothing within the CDF function estimate is theoretically argued by Li and Racine (2008) in a very similar context.

4. Circular check function minimization

The α -th quantile, $\alpha \in (0, 1)$, of the distribution of Θ given $U = u$ can be also defined as the solution for $\beta \in (-\pi, \pi)$ of the following minimization problem

$$\operatorname{argmin}_{\beta} E[\rho_\alpha(\Theta - \beta) | U = u] \quad (8)$$

with

$$\rho_\alpha(\Theta - \beta) = \alpha(\Theta - \beta)\mathbf{1}_{\{0 < \Theta - \beta < \pi - \beta\}} - (1 - \alpha)(\Theta - \beta)\mathbf{1}_{\{-\pi - \beta < \Theta - \beta < 0\}}, \quad (9)$$

where $\mathbf{1}_{\{A\}}$ is the indicator function of event A .

Notice that the loss in Equation (8) is the expected value of the *check function* for a circular quantile. This check function depends on the finite lower (upper) limit of integration of a circular CDF throughout the event A in $\mathbf{1}_A$.

Hence, a kernel estimator for $q_\alpha(u)$, $u \in \mathbb{U}$, can be defined as

$$\hat{q}_{\alpha,\lambda}(u) = \operatorname{argmin}_{\beta} \frac{1}{n} \sum_{i=1}^n Q_\lambda(U_i - u) \rho_\alpha(\Theta_i - \beta), \quad (10)$$

where, for \mathbb{U} being the circle ($[0, 1]$, respectively) Q_λ stands for a circular (euclidean, resp.) kernel with smoothing factor $\lambda > 0$. The accuracy measure (6) for the above estimator is derived in the following

Theorem 5. *Given the $\mathbb{U} \times [-\pi, \pi]$ -valued random sample $(U_1, \Theta_1), \dots, (U_n, \Theta_n)$, consider estimator (10). If $f(q_\alpha(u) | u) \neq 0$, assumptions i) and iii) of Theorem 1 hold, and assumption ii) of Theorem 1 or assumption i) of Theorem 2 is satisfied when \mathbb{U} is the circle and $\mathbb{U} = [0, 1]$ respectively, then*

$$\mathcal{L}[\hat{q}_{\alpha,\lambda}(u)] = \frac{\xi^2(Q_\lambda)}{4} \left\{ \frac{F^{(0,2)}(q_\alpha(u) | u)}{f(q_\alpha(u) | u)} + \frac{2F^{(0,1)}(q_\alpha(u) | u)g'(u)}{f(q_\alpha(u) | u)g(u)} \right\}^2 + \frac{\alpha(1-\alpha)\nu(Q_\lambda)}{ng(u)f^2(q_\alpha(u) | u)} + o(\eta(Q_\lambda) + n^{-1}\nu(Q_\lambda)) \quad (11)$$

Proof. See appendix for Proof □

Now, considering the following series expansion of $q_\alpha(U)$ for U around u ,

$$q_\alpha(U) \simeq q_\alpha(u) + \Psi(U - u)q'_\alpha(u),$$

a local linear version of estimator (10) can be obtained as the solution for β of the following minimization problem

$$\operatorname{argmin}_{(\beta, \omega)} \frac{1}{n} \sum_{i=1}^n Q_\lambda(U_i - u) \rho_\alpha(\Theta_i - \beta - \Psi(U_i - u)\omega) \quad (12)$$

Now, for the resulting estimator, we get

Theorem 6. *Given the $\mathbb{U} \times [-\pi, \pi]$ -valued random sample $(U_1, \Theta_1), \dots, (U_n, \Theta_n)$, let $\tilde{q}_{\alpha, \lambda}(u)$, $u \in \mathbb{U}$, be the solution for β of the minimization problem (12). If the assumptions of Theorem 5 hold, and $q_\alpha(\cdot)$ has continuous second derivative at $u \in \mathbb{U}$, then*

$$\mathcal{L}[\tilde{q}_{\alpha, \lambda}(u)] = \frac{\xi^2(Q_\lambda)}{4} \{q''_\alpha(u)\}^2 + \frac{\nu(Q_\lambda)\alpha(1-\alpha)}{ng(u)f^2(q_\alpha(u) | u)} + o(\eta(Q_\lambda) + n^{-1}\nu(Q_\lambda))$$

Proof. See appendix for Proof □

Notice that the asymptotic versions of loss (6) for local constant and local linear fits differ only in the bias part. Here, as it happens in the standard setting, the asymptotic biases differ even when the design density is uniform. They are the same when, for example, $q'_\alpha(u) = 0$, or when both $g(u)$ and $f(q_\alpha(u) | u)$ are uniform on \mathbb{U} , this can be seen using the formulation of $F^{(0,2)}(q_\alpha(u) | u)/f(q_\alpha(u) | u)$.

As a consequence of Theorems 5 and 6, when \mathbb{U} is the circle and Q_λ is a von Mises kernel, the value of λ which minimises the leading term of Equation (11) is essentially the same as the value in Equation (7) obtained for the corresponding local constant and local linear versions of inverse-cumulative distribution estimator with $\lambda \ll \kappa$. The same result, with due modifications, holds for the case $\mathbb{U} = [0, 1]$. This is because under the said conditions the respective losses are asymptotically equivalent.

5. Simulations

In this section, we explore and compare the performance of our estimators through simulations. Specifically, in Section 5.1 we compare the double-kernel estimator with the estimator based on the circular check function, and then in Section 5.2 we compare the local linear estimator with the local constant in the case that the explanatory variable may have finite support. In both cases we suppose that the explanatory variable is circular.

5.1. Double-kernel and check function estimators

We compare the double-kernel estimator given by (5) and the estimator based on the circular check function, given by (10) in the case that \mathbb{U} is the unit circle, and $f(\theta | u)$ is a von Mises density with concentration $\kappa = 10$ and conditional mean

$$\operatorname{atan2}[\sin(u - 1), 1 + 0.8 \cos(u - 1)] \quad (13)$$

where $\operatorname{atan2}[y, x]$ is the angle between the x -axis and the vector from the origin to (x, y) . We will estimate quantiles of orders 0.01, 0.05, 0.95, 0.99 in correspondence of these three values of the conditioning variable: $-2.2, -1.5, 1.5$. These latter values are chosen to consider a variety of slopes. A single example is given in Figure 1 for sample size $n = 100$, which shows the estimates at the given design points together with the theoretical quantiles, which are of constant width.

We note that the estimator (5) depends on a pair of smoothing parameters (say (κ, λ_1)), whereas (10) depends only on one (say λ_2) which would suggest that the double-kernel estimator will provide more flexibility, at the expense of potential difficulty in using cross-validation to select appropriate values.

As a second simulation experiment, for each $n \in \{50, 100, 500\}$ we simulated 100 samples. In each case we obtained quantile estimates, corresponding to the same orders and conditioning variable values as the first experiment,

using (5) for a grid of values of (κ, λ_1) and (10) for a grid of values of λ_2 . The results from the simulations were combined by using a sample version of the accuracy measure (6).

Three sets of results are shown. In the first set we fix $n = 100$ and $u = 1.5$ and consider $\alpha \in \{0.01, 0.05, 0.95, 0.99\}$, In the second set, we take $n = 100$ and $\alpha = .05$ and consider $u \in \{-2.2, -1.5, 1.5\}$. In the third set, we take $\alpha = 0.05$ and $u = 1.5$ and consider $n \in \{50, 100, 250, 500\}$. In each set we report: the values of (κ, λ_1) which maximize the accuracy for estimator (5), together with the corresponding loss, and the value of λ_2 which maximizes the accuracy for (10), with its corresponding loss. The results are presented in Table 1.

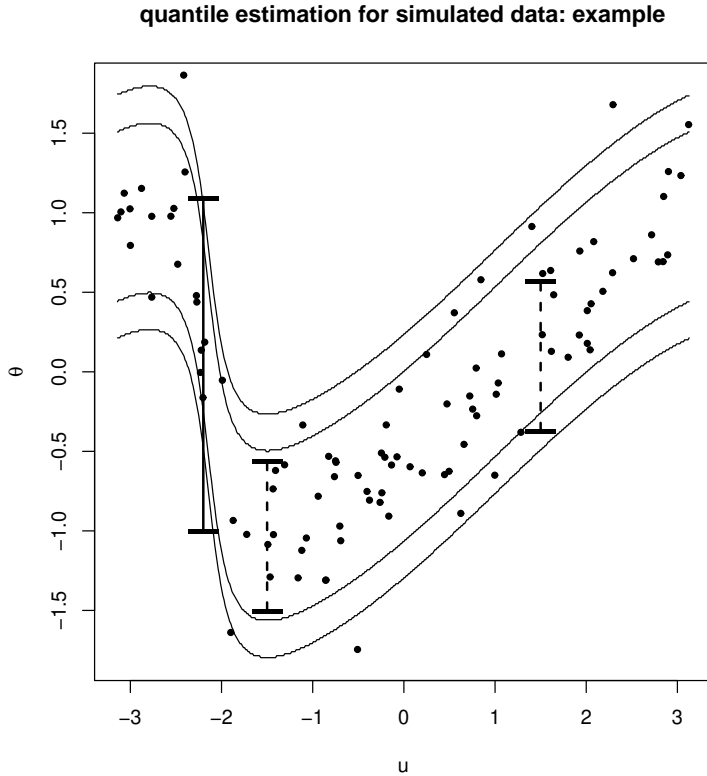


Figure 1: Theoretical quantiles (lines) for $\alpha \in \{0.01, 0.05, 0.95, 0.99\}$ for the von Mises model with conditional mean given by (13) and $\kappa = 10$. $n = 100$ points are generated from this model, in which u is uniform on the circle. As an illustration, the check quantile estimates (10) are shown for $\alpha \in \{0.01, 0.99\}$ (continuous) and $\alpha \in \{0.05, 0.95\}$ (dashed) at locations $u \in \{-2.2, -1.5, 1.5\}$. In this illustration we have chosen $\lambda = 100$.

Firstly we note that, as expected, estimator (5) is a little better than (10) in all settings. However, the improvement comes at a computational price. The first set of results shows a similar performance for each of the quantile estimates. In the second set, we can see that estimating quantiles at $u = -2.2$ is more difficult than either of the other locations. At this position there is a steep gradient over a small interval, and so little information is available. Correspondingly, much larger values of λ_1 and λ_2 are selected at this position. Finally, the dependence on sample size is also much as expected; there is improving accuracy with n and a general increase in concentration parameters.

5.2. Local constant and local linear estimators

We use similar settings as in the third block of Table 1 in that we consider $u = 1.5$ and $\alpha = 0.05$ for various values of n , and continuing to use the same conditional mean function (eq:simmean). Here, we also consider the same local constant estimator (5) but now we compare this to the corresponding local linear estimator, based on (2), in the case that the explanatory variable is uniform on the circle (as before), and in the case when it is uniform on $(-1.5, 1.5)$ so that we are estimating the quantile at the boundary of the sample space. The results are shown in Table 2.

position $u = 1.5$ and sample size $n = 100$				
quantile α	estimator (5)		estimator (10)	
	(κ, λ_1)	loss	λ_2	loss
0.01	(33,31)	0.028	16	0.049
0.05	(48,20)	0.025	10	0.032
0.95	(116,11)	0.022	10	0.026
0.99	(11,18)	0.022	7	0.027
quantile $\alpha = 0.05$ and sample size $n = 100$				
position u	estimator (5)		estimator (10)	
	(κ, λ_1)	loss	λ_2	loss
-2.2	(43,300)	0.082	194	0.094
-1.5	(29,17)	0.025	6	0.037
1.5	(57,17)	0.030	11	0.037
quantile $\alpha = 0.05$ and position $u = 1.5$				
size n	estimator (5)		estimator (10)	
	(κ, λ_1)	loss	λ_2	loss
50	(34,13)	0.030	9	0.039
100	(175,10)	0.026	9	0.030
250	(210,13)	0.011	12	0.011
500	(246,19)	0.009	13	0.011

Table 1: Results of three simulation sets from model (13), for various settings of sample size n , location u and quantiles α (see Figure 1 for an example). In each setting, we report the optimal pair (κ, λ_1) or λ_2 for estimators (5) & (10), respectively with corresponding minimized loss (6). Results were obtained from 100 samples in each setting.

$\alpha = 0.05$ and position $u = 1.5$. $U(-\pi, \pi)$				
size n	local constant		local linear	
	(κ_1, λ_1)	loss	(κ_2, λ_2)	loss
50	(76,13)	0.032	(164,11)	0.032
100	(112,13)	0.026	(74,14)	0.026
250	(148,13)	0.013	(217,14)	0.012
500	(434,17)	0.010	(325,21)	0.010
$\alpha = 0.05$ and position $u = 1.5$. $U(-1.5, 1.5)$				
size n	local constant		local linear	
	(κ_1, λ_1)	loss	(κ_2, λ_2)	loss
50	(199,22)	0.034	(68,0.7)	0.010
100	(231,26)	0.024	(52,1.2)	0.005
250	(199,38)	0.023	(68,1.0)	0.002
500	(515,38)	0.016	(84,1.1)	0.001

Table 2: Results of two simulation sets from model (13), for various settings of sample size n . In each row, we report the optimal pair (κ, λ) for the estimator (5) based on (1) and (2), respectively with corresponding minimized loss (6). Results were obtained from 100 samples in each setting. In the top block the explanatory variable is uniform on the circle (and a repeat of the last block of Table 1) and in the second block, U is uniform on $(-1.5, 1.5)$.

We can see that the performance is quite similar between the local constant and the local linear in the case that the quantile is estimated at an interior point. However, in the case that the explanatory variable has limited support and the quantile is estimated at a boundary, then the local linear estimator performs much better, and the improvement increases with sample size. This is as expected.

We note that in all double-kernel experiments, the choice of λ was much more important than the choice of κ .

6. Application

We illustrate the estimation of a conditional circular distribution function using some data on wind speed and wind directions. The data, which is taken from an observation station close to the Florida coastline and was collected by the National Data Buoy Center of the NOAC, is intended to illustrate the potential uses of the methodology.

The first part of our analysis concerns finding an “optimal” direction corresponding to wind speeds near to a specified value. Then we consider conditional quantile estimation, and assess its coverage. Note that any concern about where to “cut” the circle is not an issue in the questions we address here.

6.1. Estimation of optimal direction

A wind rose is often used to display information about the direction of particular wind speeds, and Figure 2 shows some sample rose diagrams revealing the distribution of wind directions corresponding to wind speeds 9 m/s and 12 m/s . We initially note that our method is not hampered by an arbitrary, equi-spaced partition of the circle as rose diagrams are.

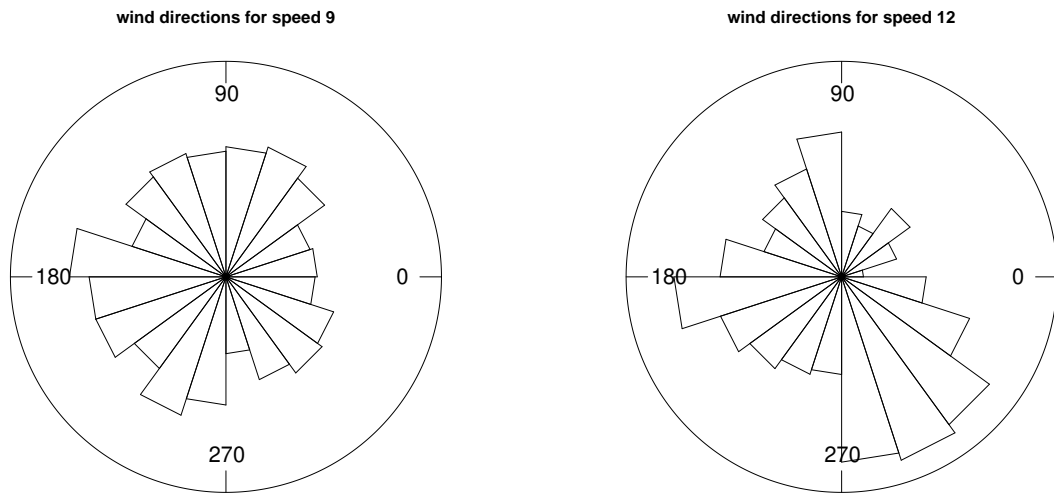


Figure 2: Rose diagrams for wind directions corresponding to the observed wind speeds (m/s) in the range (8, 10) (left) and (11, 13) (right).

This type of data is highly relevant to the analysis of proposed wind turbine installations. The energy (and revenue) generated by a wind turbine is crucially dependent on the wind speed; different turbines may have different (energy–wind speed) response rates. Locations with little wind will often be uneconomic, and locations with very high wind speeds will cause different problems. By considering the wind speed and wind direction distributions, the anticipated energy may be calculated. Our dataset consists of wind speed and wind directions at a specific location so in this illustration U , which represents the wind speed, is linear, and $\theta \in [-\pi, \pi)$ is the direction.

We will suppose that the energy output for a proposed turbine is highest for wind speeds close to some w m/s and we are interested to know the *direction* of the turbine which will “optimize” the energy. If we suppose that the turbine

will function within an angle of $\pi/4$ radians of the wind, then we need to find a direction $\theta = \theta(w)$ to maximize

$$F(\theta(w) + \pi/4 | u = w) - F(\theta(w) - \pi/4 | u = w).$$

We can estimate this using Equation (1), in which the weights L can be based on either a local constant, or local linear estimate. One way to select the smoothing parameters (λ and κ) is to consider the mean squared error of the estimates, and then use the results of Theorems 3 and 4. This would require suitable plug-in methods, similar to those of Yu and Jones (1998), who deal with analogous linear methods. Bandwidth selection for simple directional density estimation as, for example, in García-Portugués (2013), would be inappropriate in this setting. However, in our application the choice of λ is practically determined by consideration of the fact that we actually want to estimate $F(\theta(w) + \pi/4 | u \approx w) - F(\theta(w) - \pi/4 | u \approx w)$ and so (this will often depend on the operating characteristics of the specific turbine) we have considered $\lambda = 0.6$ and $\lambda = 1.2$ in our calculations with a normal kernel for K_λ (so λ is the standard deviation). Given the choice of normal kernel, these choices of λ were chosen to focus on a range of wind speeds of about $\pm 2\lambda$. Our illustrative data consists of 17,480 observations, with wind speeds which range from 0.1 m/s to 18 m/s (mean 5.9).

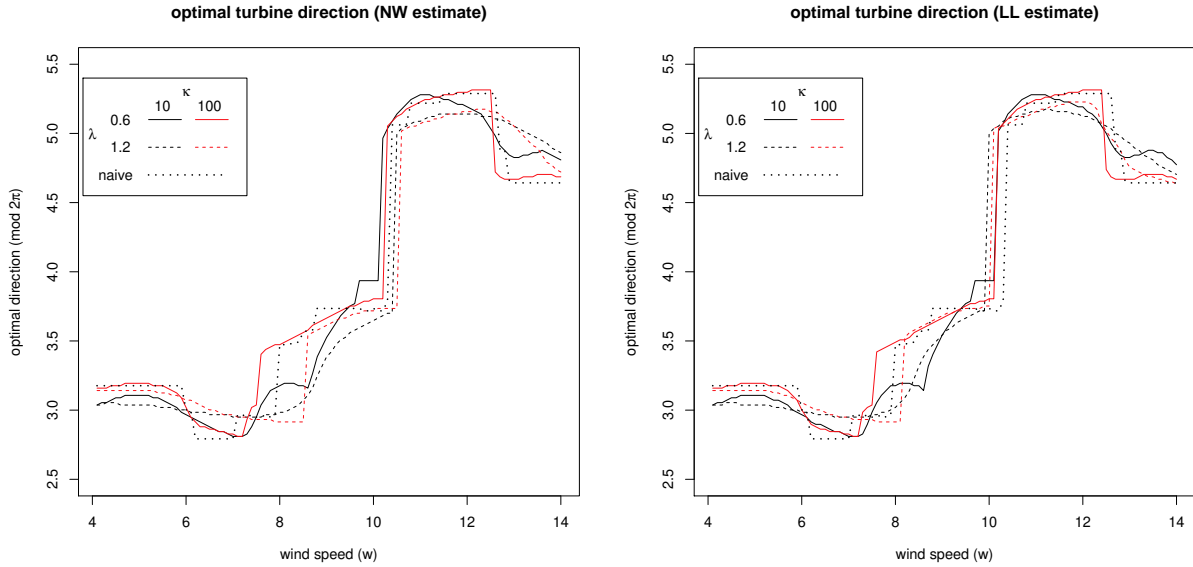


Figure 3: Wind turbine direction to maximize output using an estimate of the conditional CDF based on Equation (1). Four smoothed estimates (dependent on λ and κ) are shown, as well as a naive estimate, for each of local constant (left) and local linear (right) estimation.

Some results are presented in Figure 3, where the dependence of the optimal direction on the wind speed can be clearly seen. The dependence on κ is small, as is the effect of λ on the result. For ease of interpretation, we have used angles $\text{mod}(2\pi)$ to avoid “wrapping”. We have also considered a naive approach for comparison. In this case, we chose the angle ($\theta_0 = \theta_0(w)$) to maximize

$$\sum_{\{i:w_i \in (w-1, w+1)\}} \mathbf{1}_{\{\cos(\theta_0 - \theta_i) > 1/\sqrt{2}\}},$$

and the result is shown on both plots. This naive estimate is more akin to a histogram, though the resulting values of θ_0 are still quite smooth, with all estimates showing a similar overall structure. Figure 3 shows a marked difference in the optimal direction between wind speeds 9 m/s (where the optimal direction is about 3.5 radians) and 12 m/s (with optimal direction about 5.3 radians).

6.2. Conditional quantile estimation

Estimation of specific quantiles will depend on the choice of “cut point” on the circle, so we avoid this arbitrary selection by considering the estimation of confidence intervals conditional on wind speed, and investigate the potential

to choose λ and κ by cross-validation. Ideally, this should be considered with respect to a loss function which measures the coverage accuracy for a specific set of quantiles. However, since the coverage is discrete, this typically results in a cross-validation function which is piecewise constant, and harder to optimize. Instead, we again consider estimation of the conditional CDF as a means to get suitable smoothing parameters, and we use an adaptation of Equation (2) in Di Marzio et al. (2012), so that λ and κ are chosen to minimise

$$CV(\kappa, \lambda | u) = \sum_{i: U_i \approx u} \int_{-\pi}^{\pi} \{ \mathbf{1}_{\{\theta \geq \theta_i\}} - \hat{F}_{\lambda, \kappa}^{(i)}(\theta | u) \}^2 d\theta \quad (14)$$

where $\hat{F}_{\lambda, \kappa}^{(i)}(\cdot)$ is the smoothed estimate (1) of the conditional CDF using all the data except the i th observation, with smoothing parameters κ and λ . There are two points to note here. Firstly, the smoothing parameter selection does not take any account of the quantile(s) to be estimated, although the parameters do depend on the wind speed (u). Secondly, there is a need to make precise what “ \approx ” means in the limit of the summation in Equation (14). In our example, the wind speeds are recorded to one decimal place, but in general one will require a sufficient number of observations in order to achieve stable results.

For given wind speeds, we first find the smoothing parameters by using Equation (14) for both local linear and local constant estimators. Then, using these smoothing parameters we compute (1). Finally, for given α we choose an “origin” to minimise the width of the interval, which leads to estimates of the “lower” and “upper” endpoints. This final step is used to avoid the somewhat arbitrary choice of origin, or “cut point” for data which lie on a circle. In Figure 4 we show the 80% confidence intervals corresponding to the distribution of the direction conditional on wind speed. Also on this plot, we give the actual coverage, as a percentage, for both the local constant, and local linear estimates. Although the coverages are mostly close to 80% (as desired) it should be remembered that the data is not independent of the cross-validation data. The intervals (and coverage) are quite similar for the two methods, except for wind speed 13 m/s; the estimates of the conditional CDFs for this wind speed are also shown.

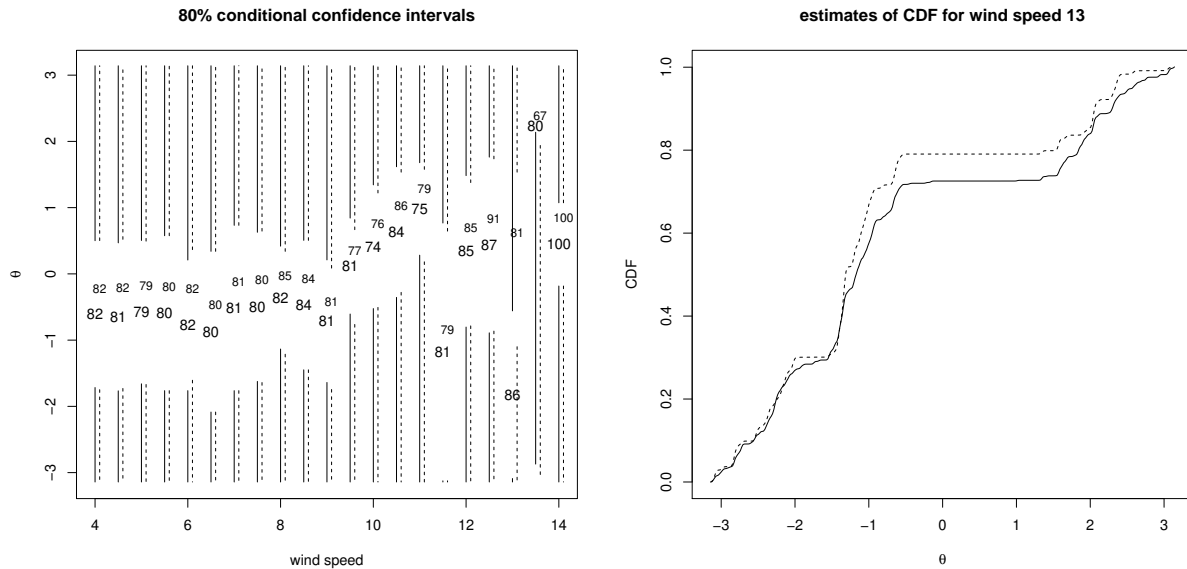


Figure 4: Left: 80% confidence interval estimates of direction conditional on wind speed for local linear (lines) and local constant (dashed). The numbers give actual coverage for each method; the larger font is used for local linear, and the smaller for local constant. Note that the intervals are shown by vertical lines (not by the gaps). Right: the estimated conditional CDF for wind speed 13 m/s for local linear (lines) and local constant (dashed).

7. Discussion

We should stress that in any practical application the choice of origin (“cut-point”) should be chosen dependent on u to minimise the width of the interval to be estimated. This adjustment is important to obtain meaningful interpretations, especially if the conditional mean can take values close to π . We note that the double-kernel estimator has two tuning parameters (λ and κ) whereas the circular check function estimator has only one. Maybe, it is for this reason that, in some simulation experiments with small samples, we have observed a slightly better performance for the double-kernel estimator in most settings. However, this comes with the cost of more computational effort, as well as the need to select good smoothing choices. This effort can be reduced by first selecting a good value of λ in the check function, and then using this value, together with an appropriate κ , in the double-kernel estimator. Cross-validation can be used in this process, though we have found a joint selection of both κ and λ to be problematic.

References

- Berg, A. and Politis, D. CDF and survival function estimation with infinite-order kernels. *Electronic Journal of Statistics*, 3, 1436–1454 (2009).
- Di Marzio, M., Panzera, A. and Taylor, C. C. Local polynomial regression for circular predictors. *Statistics & Probability Letters*, 79, 2066–2075 (2009)
- Di Marzio, M., Panzera, A. and Taylor, C. C. Density estimation on the torus. *Journal of Statistical Planning and Inference*, 141, 2156–2173 (2011)
- Di Marzio, M., Panzera, A. and Taylor, C. C. Smooth estimation of circular distribution functions and quantiles. *Journal of Nonparametric Statistics*, 24, 935–949 (2012)
- Fan, J., Hu, T. and Truong, Y. K. Robust Non-parametric Function Estimation. *Scandinavian Journal of Statistics*, 21, 433–446 (1994)
- García-Portugués, E. Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7, 1655–1685 (2013)
- Jones, M.C. and Hall, P. Mean squared error properties of kernel estimates of regression quantiles. *Statistics & Probability Letters*, 10, 283–289 (1990)
- Jones, M.C. and Noufaily, A. Parametric quantile regression based on the generalised gamma distribution. *Journal of the Royal Statistical Society, Series C*, 62, 723–740 (2013)
- Li, Q. and Racine, J. S. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26, 423–434 (2008)
- Yu, K. and Jones, M.C. Local linear quantile regression. *Journal of the American Statistical Association*, 441, 228–237 (1998)

Appendix

Proof of Theorem 1 Observing that estimators (1) and (2) can be respectively regarded as a local constant and local linear estimators of

$$m(\theta, \phi) = E[W_\kappa(\theta - \Theta) - W_\kappa(-\pi - \Theta) \mid \Phi = \phi],$$

in virtue of assumptions i) and iii), Theorem 4 of Di Marzio et al. (2009) gives

$$E[\hat{F}_{\lambda, \kappa}(\theta \mid \phi)] - m(\theta, \phi) = \begin{cases} \frac{\eta_2(Q_\lambda)}{2} \{m^{(0,2)}(\theta, \phi) + 2 \frac{g'(\phi)}{g(\phi)} m^{(0,1)}(\theta, \phi)\} + o(\eta_2(Q_\lambda)), & \text{for estimator (1);} \\ \frac{\eta_2(Q_\lambda)}{2} m^{(0,2)}(\theta, \phi) + o(\eta_2(Q_\lambda)), & \text{for estimator (2);} \end{cases}$$

and, for both estimators,

$$\text{Var}[\hat{F}_{\lambda, \kappa}(\theta \mid \phi)] = \frac{R(Q_\lambda) s^2(\phi, \theta)}{ng(\phi)} + o(\eta_2(Q_\lambda)),$$

where $s^2(\theta, \phi) = \text{Var}[W_\kappa(\theta - \Theta) - W_\kappa(-\pi - \Theta) \mid \Phi = \phi]$.

Now, notice that $m(\theta, \phi)$ and $s^2(\theta, \phi)$ respectively amount to expectation and variance of the circular CDF estimator provided by Di Marzio et al. (2012), conditional on $\Phi = \phi$. Thus, an adaptation of the results stated in Theorem 2.1 of Di Marzio et al. (2012), for the unconditional case, along with assumption ii), yield

$$m(\theta, \phi) = F(\theta \mid \phi) + \frac{1}{2k} \{F^{(2,0)}(\theta \mid \phi) - F^{(2,0)}(-\pi \mid \phi)\} + o(\kappa^{-1}), \quad (15)$$

and

$$s^2(\theta, \phi) = F(\theta \mid \phi) \{1 - F(\theta \mid \phi)\} - \frac{1}{\sqrt{\pi k}} \{F^{(1,0)}(\theta \mid \phi) - F^{(1,0)}(-\pi \mid \phi)\} + o(\kappa^{-1/2}). \quad (16)$$

Hence, by plugging (15) and (16) in the expectations and variance expressions, the results follow. \square

Proof of Theorem 2 The results follow by using the same arguments as in the proof of Theorem 1, once observed that, in virtue of assumption i), euclidean local polynomial fitting theory gives

$$E[\hat{F}_{\lambda,\kappa}(\theta | x)] - m(\theta, x) = \begin{cases} \frac{\lambda^2 \mu_2(Q)}{2} \{m^{(0,2)}(\theta, x) + 2 \frac{g'(x)}{g(x)} m^{(0,1)}(\theta, x)\} + o(\lambda^2), & \text{for estimator (1);} \\ \frac{\lambda^2 \mu_2(Q)}{2} m^{(0,2)}(\theta, x) + o(\lambda^2), & \text{for estimator (2);} \end{cases}$$

and, for both estimators,

$$\text{Var}[\hat{F}_{\lambda,\kappa}(\theta | x)] = \frac{R(Q)s^2(x, \theta)}{n\lambda g(x)} + o\left(\frac{1}{n\lambda}\right).$$

Proof of Theorem 3 First observe that, by standard Taylor series arguments, we can write

$$\begin{aligned} E[2\{1 - \cos(\hat{q}_{\alpha,\lambda,\kappa}(u) - q_\alpha(u))\}] &\approx E[\{\hat{q}_{\alpha,\lambda,\kappa}(u) - q_\alpha(u)\}^2] \\ &= \{E[\hat{q}_{\alpha,\lambda,\kappa}(u) - q_\alpha(u)]\}^2 + E[\{\hat{q}_{\alpha,\lambda,\kappa}(u) - E[\hat{q}_{\alpha,\lambda,\kappa}(u)]\}^2] \end{aligned}$$

Now, using the approximation

$$\hat{q}_{\alpha,\lambda,\kappa}(u) - q_\alpha(u) \simeq -\frac{\hat{F}_{\lambda,\kappa}(q_\alpha(u) | u) - F(q_\alpha(u) | u)}{f(q_\alpha(u) | u)},$$

Theorem 1 when $u = \phi$ and Theorem 2 when $u = x$, along with $F(q_\alpha(u) | u) = \alpha$, yield the result. \square

Proof of Theorem 4 The result directly follows by using the same arguments as in the proof of Theorem 3, along with results of Theorem 1 when $u = \phi$, and Theorem 2 when $u = x$. \square

Proof of Theorem 5 Reasoning as in Jones and Hall (1990), define

$$H_\alpha(u) = \frac{1}{n} \sum_{i=1}^n Q_\lambda(U_i - u) \{\alpha \mathbf{1}_{\{\beta < \Theta_i < \pi\}} - (1 - \alpha) \mathbf{1}_{\{-\pi < \Theta_i < \beta\}}\},$$

obtaining that

$$E[H_\alpha(u) | U_1, \dots, U_n] = \frac{1}{n} \sum_{i=1}^n Q_\lambda(u - U_i) \{\alpha - F(\beta | U_i)\}.$$

Now, expanding $F(\beta | U_i)$ in Taylor series, for (β, U_i) around $(q_\alpha(u), u)$, it results

$$F(\beta | U_i) \approx F(q_\alpha(u) | u) + \{\beta - q_\alpha(u)\} f(q_\alpha(u) | u) + \Psi(U_i - u) F^{(0,1)}(q_\alpha(u) | u) + \frac{1}{2} \Psi^2(U_i - u) F^{(0,2)}(q_\alpha(u) | u),$$

and using similar approximations as those used in the proof of Theorem 3, we finally get

$$E[H_\alpha(u)] = -\{\beta - q_\alpha(u)\} f(q_\alpha(u) | u) g(u) - \frac{\xi(Q_\lambda)}{2} \{2F^{(0,1)}(q_\alpha(u) | u) g'(u) + F^{(0,2)}(q_\alpha(u) | u) g(u)\} + o(\xi(Q_\lambda))$$

Moreover, reasoning again as in Jones and Hall (1990), we have

$$\text{Var}[H_\alpha(u)] = \frac{\nu(Q_\lambda) g(u)}{n} \alpha(1 - \alpha) + o\left(\frac{\nu(Q_\lambda)}{n}\right).$$

Now, use the fact that $\{\text{Var}[H_\alpha(u)]\}^{-1/2} \{H_\alpha(u) - E[H_\alpha(u)]\}$ converges to a standard normal distribution, and that estimator (9) is the solution for β of $H_\alpha(\theta) - E[H_\alpha(u)] = -E[H_\alpha(u)]$. \square

Proof of Theorem 6 When $u \in [0, 1]$, the result directly follows by using Theorem 3 in Fan et al. (1994), with λ in place of h . When u is an angle, the same result holds, with due modifications, by using the assumptions in Theorem 5. \square