Nonparametric smoothing of circular data

Agnese Panzera and Charles C. Taylor

Abstract A local regression smoother is proposed for the case when the response is a circular variable. The method allows for both smoothing circular time series and circular quantiles estimation.

Key words: circular kernels, circular quantiles, circular regression, circular time series.

1 Introduction

We propose a nonparametric regression smoother for the case when the response is a circular random variable. Our smoother is defined by the arc-tangent of the ratio between the locally weighted components of the first sample trigonometric moment of the response variable. Simple adaptations of the weight function enable a unified formulation for both linear and circular predictors, whereas these cases have been tackled by quite distinct parametric methods. See, for example, [9], [7], [10], [5], [6], [4] for some contributions on circular-circular regression, and [7] and [8] for the linear predictor case. In Section 2 we introduce our regression smoother, whereas Section 3 and 4 are respectively devote to its adaptation for circular time series analysis and circular quantiles estimation.

Agnese Panzera
DMQTE, University of Chieti-Pescara, Viale Pindaro 42, Pescara, Italy
e-mail: agnesepanzera@yahoo.it

Charles C. Taylor
Department of Statistics, University of Leeds, Leeds LS2 9JT, UK
e-mail: charles@maths.leeds.ac.uk

2 Regression smoother for circular response

Let $\mathbb{T} := [-\pi, \pi)$. Consider a $\mathbb{A} \times \mathbb{T}-$ valued random vector (Δ, Θ), where $\mathbb{A}$ is a generic domain. The dependence of the *response* Θ on the *predictor* Δ is well described by a (measurable) function $m : \mathbb{A} \to \mathbb{T}$, such that the risk

$$E[1 - \cos(\Theta - m(\Delta))],$$

is as small as possible. For $\delta \in \mathbb{A}$, let $m_1(\delta) := E[\sin(\Theta)|\Delta = \delta]$, and $m_2(\delta) := E[\cos(\Theta)|\Delta = \delta]$. Then the minimizer of the above risk is given by $m(\delta) = \text{atan2}[m_1(\delta), m_2(\delta)]$, where the function atan2$[y,x]$ returns the angle between the x-axis and the vector from the origin to (x, y). This approach implies the model

$$\Theta_i = [m(\Delta_i) + \varepsilon_i](\text{mod}2\pi) \qquad (i = 1, \dots, n)$$

where the ε_is are *i.i.d.* random angles having zero mean direction, finite concentration, and are independent of the Δ_is. To introduce our smoother we preliminarily define the sample statistics

$$\hat{m}_1(\delta) = \frac{1}{n}\sum_{i=1}^{n} \sin(\Theta_i)W(d(\Delta_i, \delta)) \quad \text{and} \quad \hat{m}_2(\delta) = \frac{1}{n}\sum_{i=1}^{n} \cos(\Theta_i)W(d(\Delta_i, \delta)), \quad (1)$$

where $d(\cdot, \cdot)$ is a distance, and W denotes a local weight, conceived in such a way that the ratio $\hat{m}_1(\delta)/\hat{m}_2(\delta)$ is asymptotically unbiased for $m_1(\delta)/m_2(\delta)$. Then, an estimator for the regression function at $\delta \in \mathbb{A}$ is defined as

$$\hat{m}(\delta) = \text{atan2}[\hat{m}_1(\delta), \hat{m}_2(\delta)]. \quad (2)$$

Clearly, the weights have to be specific to the nature of the predictors. Here we consider the case where $\mathbb{A} = \mathbb{T}$, *i.e.* the circular predictor case, and the case where $\mathbb{A} = [0, 1]$, which corresponds, without loss of generality, to the linear predictor case. For both cases, we propose two kind of weights, one aimed at a local constant fit, the other one at a local linear fit.

When $\mathbb{A} = [0, 1]$, we use standard euclidean weights. When $\mathbb{A} = \mathbb{T}$, to take into account the circular nature of the predictor variable, we use periodic functions as weights. In particular, we propose as weights for local constant fit, *circular kernels*, say K_κ, introduced by [1] in the context of density estimation. Here $\kappa > 0$ is a concentration parameter which, although not being a scale factor, plays the rôle of the inverse of the bandwidth of euclidean kernels. A different version of our smoother can be obtained by setting

$$W(d(\delta,\Delta_i)) = n^{-1}K_\kappa(\delta-\Delta_i)\left\{\sum_{j=1}^{n} K_\kappa(\delta-\Delta_j)\sin(\delta-\Delta_j)\right.$$
$$\left. -\sin(\delta-\Delta_i)\sum_{j=1}^{n} K_\kappa(\delta-\Delta_j)\sin(\delta-\Delta_j)\right\},$$

which resembles the structure of the weight function of euclidean local linear fit. Accuracy measures for the estimator equipped with the two kinds of weights, in both circular and linear predictor case, have been derived, along with optimal smoothing, in [2] . Moreover, the derivation of a central limit theorem for the estimator allowed the construction of approximate confidence intervals for $m(\delta)$. As a further result, we found that, when data are assumed to be observations from a stationary process satisfying some mixing conditions, (2) shares essentially the same asymptotic behavior as in the *i.i.d.* case. Then, assuming $\Theta_i = \Theta_{i+a}$, with $a \in \mathbb{Z}$, we have that (2) can be used for autoregressive functions estimation.

3 Smoothing and prediction for circular time series

The estimator introduced in Section 2 can be extended to nonparametric time series analysis, by smoothing on the time domain. This obviously implies the introduction of a different model. Specifically, letting $\{\Theta_t\}_{t=1}^T$ be a time series of angles, we assume the model

$$\Theta_t = [m(t/T) + \varepsilon_t](\mathrm{mod} 2\pi) \tag{3}$$

where $m : [0,1] \to \mathbb{T}$ is an unknown smooth function of time representing the trend, and $\{\varepsilon_t\}$ is a $\mathbb{T}$-valued stationary stochastic process, with $E[\sin(\varepsilon_t)] = 0$ and with autocovariance function regularly varying at infinity with exponent $\alpha > 0$, *i.e.*, as ℓ goes to infinity,

a1)$Cov[\cos(\varepsilon_t), \cos(\varepsilon_{t+\ell})] \sim L_1|\ell|^{-\alpha}$;
a2)$E[\cos(\varepsilon_t)\sin(\varepsilon_{t+\ell})] \sim L_2|\ell|^{-\alpha}$;
a3)$E[\sin(\varepsilon_t)\sin(\varepsilon_{t+\ell})] \sim L_3|\ell|^{-\alpha}$;

where $L_i \in \mathbb{R}\setminus\{0\}$, for $i \in \{1,2,3\}$, $\ell \in \mathbb{Z}$, and $|\ell|^\alpha := 1$ if $\ell = 0$. We say that the case $0 < \alpha < 1$ indicates a *long-range dependence*, whereas the case $\alpha > 1$ implies so-called *short-range dependence*. Then an estimator for the trend function at t/T can be constructed by adapting the smoother (2), by setting $\delta = t/T$ and $\Delta_i = i/T$, and assuming the weight in (1) to be a euclidean kernel supported on $[0,1]$, or to have the formulation of the local linear fitting weight. We obtain asymptotic behavior of the resulting estimators, in the different settings of α. For the task of prediction–for which we mean the estimation of m at a point $t/T \in (1,+\infty)$– we still use model (3), but now the domain of m is assumed to be $[0,+\infty)$.

4 Kernel quantile estimation

Letting $\Theta_1, \cdots, \Theta_n$ be a random sample from an absolutely continuous circular distribution function, an estimator of the circular population quantile of order $p \in (0,1)$ can be defined as $\hat{Q}_\kappa(p) := \mathrm{atan2}(\hat{q}_1(p), \hat{q}_2(p))$, where

$$\hat{q}_1(p) := \frac{1}{n}\sum_{i=1}^{n} K_\kappa(2\pi i/n - 2\pi p)\sin(\Theta_{(i)}), \text{ and } \hat{q}_2 := \frac{1}{n}\sum_{i=1}^{n} K_\kappa(2\pi i/n - 2\pi p)\cos(\Theta_{(i)}),$$

where $\Theta_{(i)}$ denotes the order statistics of circular rank $2\pi i/n - \pi$. A different estimator, in the same spirit, can be defined by setting

$$\hat{q}_1(p) := \int_0^1 2\pi K_\kappa(2\pi p - 2\pi u)\sin(F_n^{-1}(u))du,$$

and

$$\hat{q}_2(p) := \int_0^1 2\pi K_\kappa(2\pi p - 2\pi u)\cos(F_n^{-1}(u))du,$$

where $F_n(\theta) := n^{-1}\sum_1^n \mathbb{1}_{\{\Theta_i \leq \theta\}}$, and $F_n^{-1}(u) := \inf\{\theta : F_n(\theta) \geq u\}$ is the empirical circular quantile of order u.

References

1. Di Marzio, M., Panzera, A. & Taylor, C. C.: Density estimation on the torus. Journal of Statistical Planning and Inference, **141**, 2156-2173. (2011).
2. Di Marzio, M., Panzera A., & Taylor, C. C.: Nonparametric regression for circular responses. Submitted. (2011).
3. Downs, T. D. & Mardia, K. V.: Circular regression, Biometrika, **89**, 683–697. (2002).
4. Kato, S. & Jones, M.: A family of distributions on the circle with links to, and applications arising from, möbius transformation. Journal of the American Statistical Association, **105**, 249–262. (2010).
5. Kato, S. & Shimizu, K.: Dependent models for observations which include angular ones. Journal of Statistical Planning and Inference, **138**, 3538–3549. (2008).
6. Kato, S., Shimizu, K. & Shieh, G.: A circular-circular regression model. Statistica Sinica, **18**, 633–645. (2008).
7. Fisher, N. I. & Lee, A. J.: Regression models for angular responses. Biometrics, **48**, 665–677. (1992).
8. Presnell, B., Morrison, S.P. & Littel R.C.: Projected multivariate linear models for directional data. Journal of the American Statistical Association, **93**, 1068–1077. (1998).
9. Rivest, L.-P.: Some statistical methods for bivariate circular data. Journal of the Royal Statistical Society. Series B, **44**, 81–90. (1982).
10. Rivest, L.-P.: A decentred predictor for circular-circular regression. Biometrika, **84**, 717–726. (1997).