proceedings

Proceedings

Nonparametric Regression Estimation for Circular Data †

Andrea Meilán-Vila [1,*], Mario Francisco-Fernández [1], Rosa M. Crujeiras [2] and Agnese Panzera [3]

1 Department of Mathematics, Universidade da Coruña, 15071 A Coruña , Spain
2 Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, 15705 Santiago de Compostela, Spain
3 Department of Statistics, Computer Science, Applications "Giuseppe Parenti", Università degli Studi di Firenze, 50134 Florence, Italy
* Correspondence: andrea.meilan@udc.es; Tel.: +34-981-167000
† Presented at the 2nd XoveTIC Conference, A Coruña, Spain, 5–6 September 2019.

Published: 31 July 2019

Abstract: Non-parametric regression with a circular response variable and a unidimensional linear regressor is a topic which was discussed in the literature. In this work, we extend the results to the case of multivariate linear explanatory variables. Nonparametric procedures to estimate the circular regression function are formulated. A simulation study is carried out to study the sample performance of the proposed estimators.

Keywords: circular data; nonparametric methods; multidimensional data analysis

1. Introduction

Circular data can be regarded as points whose support is on a circle (with unit radius) measured in degrees or radians and with periodic nature. Examples of circular data arise in many applied fields such as biology (orientation of animals), meteorology (wind direction), oceanography (ocean currents), among others. Comprehensive reviews on circular statistics are provided by [1] and [2].

Nonparametric regression with circular response and univariate linear predictor was studied by [3]. They defined local estimators for the circular regression function. The aim of this work is to propose and study nonparametric procedures to estimate the circular regression function, assuming a multivariate linear-circular regression model (circular responses and multivariate linear predictors). Simulation studies are carried out to check the finite sample performance of the considered estimators.

2. The Model

Let $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample of $(\mathbf{X}, \Theta)$, where Θ is an angular or circular random variable taking values on $\mathbb{T} = [0, 2\pi)$, and $\mathbf{X}$ is a random variable taking values in $D \subseteq \mathbb{R}^d$. The following regression model is assumed:

$$\Theta_i = (m(\mathbf{X}_i) + \varepsilon_i)(\text{mod} 2\pi), \quad i = 1, \ldots, n, \tag{1}$$

where the random angles ε_i have zero mean direction, finite concentration, and are independent of the $\mathbf{X}_i$. This implies that $\mathbb{E}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] = 0$ and $\mathbb{E}[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}] < \infty$. Additionally, it is supposed that $\text{Var}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] < \infty$, $\text{Var}[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}] < \infty$ and $\text{Cov}[\sin(\varepsilon), \cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}] < \infty$.

3. Kernel-Type Estimators

In order to construct an estimator for the regression circular function m, the risk measure $\mathbb{E}(1 - \cos(\Theta - m(\mathbf{x})))$ can be used. Defining $m_1(\mathbf{x}) = E[\sin(\Theta) \mid \mathbf{X} = \mathbf{x}]$ and $m_2(\mathbf{x}) = E[\cos(\Theta) \mid \mathbf{X} = \mathbf{x}]$, then the minimizer of the risk function is $m(\mathbf{x}) = \text{atan2}(m_1(\mathbf{x}), m_2(\mathbf{x}))$, where the function atan2(y, x) returns the angle between the x-axis and the vector from the origin to (x, y). Therefore, the proposed estimator for the circular regression function $m(\mathbf{x})$ is given by:

$$\hat{m}_{\mathbf{H}}(\mathbf{x}; p) = \text{atan2}(\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)), \quad (2)$$

where $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)$ denote the p-th order local polynomial estimators of $m_1(x)$ and $m_2(x)$, respectively. Although theoretical results are omitted here, the asymptotic properties of the estimator given in (2) were derived in terms of bias and variance.

4. Simulation Study

In order to explore the performance of the estimators proposed in Section 3, a simulation study considering different scenarios is carried out. Only the case $d = 2$ is analyzed in the study. For each scenario, 500 samples of size n ($n = 64, 100, 225$ and 400) are generated on a bidimensional regular grid in the unit square considering the following regression models:

R1. $\Theta = (\text{atan2}(6X_1^5 - 2X_1^3 - 1, -2X_2^5 - 3X_2 - 1) + \varepsilon)(\text{mod} 2\pi)$

R2. $\Theta = (\text{acos}(X_1^5 - 1) + \frac{3}{2}\text{asin}(X_2^3 - X_2 + 1) + \varepsilon)(\text{mod} 2\pi)$

where $\mathbf{X} = (X_1, X_2)$ denote the bidimensional covariate and the circular errors, ε_i, are drawn from a von Mises distribution $vM(0, \kappa)$ with different values of κ (5, 10 and 15). For each sample the estimator (2) is computed, assuming $p = 0$ (Nadaraya-Watson estimator) and $p = 1$ (local linear estimator). In both cases, the smoothing parameter $\mathbf{H}$ is chosen by cross-validation, minimizing $\sum_{i=1}^{n}\{1 - \cos[\Theta_i - \hat{m}^{(i)}(\mathbf{X}_i)]\}$, where $\hat{m}^{(i)}(\mathbf{X}_i)$ represents the Nadaraya-Watson or the local linear estimator, computed using all observations except $(\mathbf{X}_i, \Theta_i)$, and evaluated at $\mathbf{X}_i$.The results of errors $\frac{1}{n}\sum_{i=1}^{n}\{1 - \cos[m(\mathbf{X}_i) - \hat{m}(\mathbf{X}_i)]\}$ are shown in Table 1 for models R1 and R2. It can be observed that the local linear estimator gives a smaller error than the Nadaraya-Watson estimator in most cases. On the other hand, as expected, considering a larger sample size, the error is smaller. Further, smaller errors are also obtained if the concentration parameter κ is larger (less *variance*).

Table 1. Means of errors $\frac{1}{n}\sum_{i=1}^{n}\{1 - \cos[m(\mathbf{X}_i) - \hat{m}(\mathbf{X}_i)]\}$ over 500 simulations, for models R1 and R2, using Nadaraya-Watson and local linear fits.

		Model R1 κ			**Model R2** κ		
Estimator	n	5	10	15	5	10	15
Nadaraya-Watson	64	0.0226	0.0121	0.0089	0.0367	0.0213	0.0165
	100	0.0171	0.0108	0.0080	0.0388	0.0024	0.0152
	225	0.0058	0.0049	0.0038	0.0185	0.0125	0.0108
	400	0.0056	0.0035	0.0026	0.0129	0.0080	0.0062
Local linear	64	0.0234	0.0125	0.0089	0.0283	0.0144	0.0107
	100	0.0165	0.0086	0.0061	0.0209	0.0013	0.0083
	225	0.0050	0.0039	0.0029	0.0103	0.0061	0.0047
	400	0.0050	0.0026	0.0018	0.0074	0.0043	0.0033

5. Conclusions

Nonparametric procedures to estimate the the circular regression function were proposed and studied. The new estimators were based on local polynomial fits as part of their construction. Simulation studies were carried out justifying the correct performance of the proposed estimators.

Funding: This research has received financial support from the Xunta de Galicia and the European Union (European Social Fund—ESF). This research has been partially supported by MINECO grant MTM2016-76969-P and MTM2017-82724-R and by the Xunta de Galicia (ED481A-2017/361, Grupos de Referencia Competitiva ED431C-2016-015 and Centro Singular de Investigación de Galicia ED431G/01), all of them through the ERDF.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fisher, N.I. *Statistical Analysis of Circular Data*; Cambridge University Press: Cambridge, UK, 1995.
2. Mardia, K.V.; Jupp, P.E. *Directional Statistics*; John Wiley & Sons: New York, NY, USA, 2009.
3. Di Marzio, M.; Panzera, A.; Taylor, C.C. Non-parametric Regression for Circular Responses. *Scand. J. Stat.* **2013**, *40*, 142–149.