Revista Colombiana de Estadística - Applied Statistics
January 2023, volume 46, issue 1, pp. 45–62
http://doi.org/10.15446/rce.v46n1.101517

Robust Circular Logistic Regression Model and Its Application to Life and Social Sciences

Modelo de regresión logística circular robusto y su aplicación a las ciencias de la vida y sociales

Elena Castilla[a]

Department of Applied Mathematics, Materials Science and Engineering and Electronic Technology, Rey Juan Carlos University, Madrid, Spain

Abstract

This paper presents robust estimators for binary and multinomial circular logistic regression, where a circular predictor is related to the response. An extensive Monte Carlo Simulation Study clearly shows the robustness of proposed methods. Finally, three numerical examples of Botany, Crime and Meteorology illustrate the application of these methods to Life and Social Sciences. Although in the Botany data the proposed method showed little improvement, in the Crime and Meteorological data an increment up to 5% and 4% of accuracy, respectively, is achieved.

***Key words*:** Circular data; Circular logistic regression; Maximum likelihood estimation; Multinomial circular logistic regression; Robustness.

Resumen

Este artículo presenta estimadores robustos para el modelo de regresión logística circular binomial y mutinomial. Un estudio de Monte Carlo muestra la robustez de los métodos propuestos. Finalmente, tres ejemplos numéricos en botánica, criminalística y meteorología muestran la aplicación de estos modelos a las Ciencias.

***Palabras clave*:** Datos circulares; Regresión logística circular; Regresión logística circular multinomial; Estimación de máxima verosimilitud; Robustez.

[a]Professor. E-mail: elena.castilla@urjc.es

1. Introduction

In the last years there has been renewed interest in the statistical treatment of circular data. Circular data can be represented in a circle, and can be expressed by their angle: in degrees from $0°$ to $360°$ or radians from 0 to 2π. Circular data is commonly used in Political Sciences. Gill & Hangartner (2010), for instance, applied circular data to study party preferences over policy issues for all Bundestag elections in post-World War II in Germany. Note that time data can be considered as circular data too. In case we have a 24 hour clock or a day of the year (DOY) calendar, for example, we only need to convert it to angular data. For example, Kibiak & Jonas (2007) used circular data to detect patterns in time through a monitoring study where mood and social interactions were assessed for 4 weeks. On the other hand, Jones & Pewsey (2012), used circular data to study the sudden infant death syndrome (SIDS) by taking the monthly totals of SIDS deaths in England and Wales, Scotland, and Northern Ireland for the years 1983-1998. Circular data have been also applied to Ecology (SenGupta & Ugwuowo, 2006), Medicine (Bell, 2008), Biology (Landler et al., 2018) or Meteorology (Abuzaid & Allahham, 2015).

Recently, Al-Daffaie & Khan (2017) defined the circular logistic regression model to relate a circular predictor to a binary response. Since then, this model has been widely applied in literature: see Uemura et al. (2021) or Wolpert and Tallon-Baudry Wolpert & Tallon-Baudry (2021), among many others. However, this approach assumes a binary response, which can be excesively simple in practice. On the other hand, the proposed method was based on maximum likelihood estimator (MLE) which is known for its lack of robustness. This means that the model may be seriously affected by the presence of outliers. However, it is very common to observe this kind of observations in practice. Therefore, we may be interested in the development of robust inference for the logistic regression model. In this line of research, Alshqaq et al. (2021) proposed some new robust estimators for the binomial model.

This paper is organized as follows. In Section 2 we extend the original model to the multinomial circular logistic regression model, where a categorical response is involved. In Section 3 we propose alternative robust divergence-based estimators for both the binomial and multinomial circular logistic regression models. In Section 4 an extensive Monte Carlo simulation study shows the robustness of proposed methods. Finally, in Section 5 three numerical examples illustrate the application of these models to Life and Social Sciences.

2. The Circular Logistic Regression model

2.1. The Binomial Circular Logistic Regression model

Let us consider that we have a binary outcome $\eta \in \{0, 1\}$ that depends on a circular explanatory variable $u \in [0, 2\pi]$. Let $\pi(\boldsymbol{\beta}, u)$ be the probability of success in the response variable, the (binomial) circular logistic regression model is given by

$$\pi(\boldsymbol{\beta}, u) = \frac{\exp\{\beta_0 + \beta_1 \cos u + \beta_2 \sin u\}}{1 + \exp\{\beta_0 + \beta_1 \cos u + \beta_2 \sin u\}}, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T \in \mathbb{R}^3$ is the model parameter vector. In this way, we take into account the circular nature of the predictor.

We assume that we have n independent observations divided into I groups, each one with n_i observations ($n = \sum_{i=1}^{I} n_i$) and associate covariate u_i. For the ith group, the number of successes is denoted by ν_i. As these observations come from a binomial distribution, the likelihood is given by

$$L(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) = \prod_{i=1}^{I} \binom{n_i}{\nu_i} \pi(\boldsymbol{\beta}, u_i)^{\nu_i} (1 - \pi(\boldsymbol{\beta}, u_i))^{n_i - \nu_i},$$

where $\boldsymbol{\nu} = (\nu_1, \dots, \nu_I)^T$ and $\boldsymbol{u} = (u_1, \dots, u_I)^T$. Thus, the log-likelihood is

$$\begin{aligned} \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) &= \log L(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) \\ &= \sum_{i=1}^{I} \left\{ \log \binom{n_i}{\nu_i} + \nu_i \log \pi(\boldsymbol{\beta}, u_i) + (n_i - \nu_i) \log(1 - \pi(\boldsymbol{\beta}, u_i)) \right\}. \end{aligned} \tag{2}$$

Definition 1. Given the circular logistic regression model in (1), the maximum likelihood estimator (MLE), $\widehat{\boldsymbol{\beta}}_{\text{MLE}}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_{\text{MLE}} = arg \max_{\boldsymbol{\beta} \in \mathbb{R}^3} \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}), \tag{3}$$

where $\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})$ is the log-likelihood based in the observed data (2).

Taking into account

$$\frac{\partial \pi(\boldsymbol{\beta}, u_i)}{\partial \boldsymbol{\beta}} = (1, \cos u_i, \sin u_i)^T \pi(\boldsymbol{\beta}, u_i)(1 - \pi(\boldsymbol{\beta}, u_i)),$$

for $i = 1, \dots, I$, the maximum likelihood equations, which we want to set equal to zero, are given by

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial \beta_0} &= \sum_{i=1}^{I} (\nu_i - n_i \pi(\boldsymbol{\beta}, u_i)), \\ \frac{\partial \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial \beta_1} &= \sum_{i=1}^{I} (\nu_i - n_i \pi(\boldsymbol{\beta}, u_i)) \cos u_i, \\ \frac{\partial \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial \beta_2} &= \sum_{i=1}^{I} (\nu_i - n_i \pi(\boldsymbol{\beta}, u_i)) \sin u_i. \end{aligned}$$

Following proposition expresses these equations in a matricial form:

Proposition 1. *Given the circular logistic regression model in (1), the MLE, $\widehat{\boldsymbol{\beta}}_{MLE}$, is obtained by solving the following system of equations*

$$\boldsymbol{W}^T(\boldsymbol{\nu} - \boldsymbol{\mu}(\boldsymbol{\beta})) = \boldsymbol{0}_3, \tag{4}$$

where $\boldsymbol{0}_3$ is the null vector of dimension 3, $\boldsymbol{\nu}$ was the vector of observed successes,

$$\boldsymbol{\mu}(\boldsymbol{\beta}) = (n_1\pi(\boldsymbol{\beta}, \boldsymbol{u}_1), \dots, n_I\pi(\boldsymbol{\beta}, \boldsymbol{u}_I))^T$$

is the vector of expected successes and

$$\boldsymbol{W} = \begin{pmatrix} 1 & \cos u_1 & \sin u_1 \\ 1 & \cos u_2 & \sin u_2 \\ \vdots & & \\ 1 & \cos u_I & \sin u_I \end{pmatrix}.$$

Theorem 1. *Given the circular logistic regression model in (1), the asymptotic distribution of the MLE, $\widehat{\boldsymbol{\beta}}_{MLE}$, is given by*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{MLE} - \boldsymbol{\beta}^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}_3, \left(\boldsymbol{W}^T Diag\left(\delta_i\pi(\boldsymbol{\beta}^*, \boldsymbol{u}_i)(1-\pi(\boldsymbol{\beta}^*, \boldsymbol{u}_i)\right)_{i=1,\dots,I} \boldsymbol{W}\right)^{-1}\right),$$

where $\delta_i = \lim_{n\to\infty} \frac{n_i}{n}$ and $\boldsymbol{\beta}^*$ is the true value of the parameter vector.

***Proof*.** It is well known that the asymptotic distribution of the MLE is given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} - \boldsymbol{\beta}^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_F^{-1}(\boldsymbol{\beta}^*)\right),$$

where $\boldsymbol{I}_F(\boldsymbol{\beta})$ is the Fisher information matrix, which in this model is given by

$$n\boldsymbol{I}_F(\boldsymbol{\beta}) = -E\left[\frac{\partial^2\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial\boldsymbol{\beta}\boldsymbol{\beta}^T}\right] = \boldsymbol{W}^T \mathrm{Diag}\left(n_i\pi(\boldsymbol{\beta}^*, \boldsymbol{u}_i)(1-\pi(\boldsymbol{\beta}^*, \boldsymbol{u}_i)\right)_{i=1,\dots,I} \boldsymbol{W}.$$

Then, the result follows. □

MLE can be obtained by applying the Newton-Raphson method. The circular logistic regression model was first considered in Al-Daffaie & Khan (2017) and it is based on the classical logistic regression model for linear data, firstly used by Berkson (1944). However, this model was developed for a binomial response while in practice, we can have more than two response categories. Here we first introduce the multinomial circular logistic regression model to relate a circular predictor to a multinomial response.

2.2. The Multinomial Circular Logistic Regression model

Let us consider now that our response variable η has $d+1$ categories, $\eta \in \{0, 1, \dots, d+1\}$ and that this depends again on a circular explanatory variable. Let

$\pi_j(\boldsymbol{\beta}, u)$ denote the probability that η belongs to the jth category, the multinomial circular logistic regression model is given by

$$\pi_j(\boldsymbol{\beta}_j, u) = \frac{\exp\{\beta_{j0} + \beta_{j1}\cos u + \beta_{j2}\sin u\}}{1 + \sum_{j=1}^{d}\exp\{\beta_{j0} + \beta_{j1}\cos u + \beta_{j2}\sin u\}}, \quad j = 1, \dots, d, \tag{5}$$

and $\pi_{d+1}(\boldsymbol{\beta}, u) = 1 - \sum_{j=1}^{d}\pi_j(\boldsymbol{\beta}, u)$. Here $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j2}, \beta_{j3})^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_d^T)^T \in \mathbb{R}^{3d}$ is the model parameter vector. Considering a sample of n independent observations, the likelihood function is given by

$$L(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) = \prod_{i=1}^{I}\prod_{j=1}^{d+1}\frac{n_i!}{\nu_{11}! \dots \nu_{Id}!}\pi_j(\boldsymbol{\beta}_j, u_i)^{\nu_{ij}},$$

and

$$\begin{aligned}\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) &= \log L(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{d+1}\left\{\log\left(\frac{n_i!}{\nu_{11}! \dots \nu_{Id}!}\right) + \nu_{ij}\log\pi_j(\boldsymbol{\beta}_j, u_i)\right\}.\end{aligned} \tag{6}$$

Definition 2. Given the multinomial circular logistic regression model in (5), the MLE, $\widehat{\boldsymbol{\beta}}_{\text{MLE}}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}}_{\text{MLE}} = arg \max_{\boldsymbol{\beta}\in\mathbb{R}^{3d}} \ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u}), \tag{7}$$

where $\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})$ is the log-likelihood based in the observed data (6).

The maximum likelihood equations, which must be equal to zero for each $j = 1, \dots, d$, are given by

$$\begin{aligned}\frac{\partial\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial\beta_{j0}} &= \sum_{i=1}^{I}\left(\nu_{ij} - n_i\pi_j(\boldsymbol{\beta}_j, u_i)\right), \\ \frac{\partial\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial\beta_{j1}} &= \sum_{i=1}^{I}\left(\nu_{ij} - n_i\pi_j(\boldsymbol{\beta}_j, u_i)\right)\cos u_i, \\ \frac{\partial\ell(\boldsymbol{\beta}; \boldsymbol{\nu}, \boldsymbol{u})}{\partial\beta_{j2}} &= \sum_{i=1}^{I}\left(\nu_{ij} - n_i\pi_j(\boldsymbol{\beta}_j, u_i)\right)\sin u_i.\end{aligned}$$

A generalization of Proposition 1 applied to these equations gives us the following result.

Proposition 2. *Given the multinomial circular logistic regression model in (5), the MLE, $\widehat{\boldsymbol{\beta}}_{MLE}$, is obtained by solving the following system of equations*

$$\tilde{\boldsymbol{W}}^T(\tilde{\boldsymbol{\nu}} - \tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})) = \mathbf{0}_{3d}, \tag{8}$$

where $\tilde{\boldsymbol{\nu}} = (\nu_{11}, \ldots, \nu_{1d}, \ldots, \nu_{I1}, \ldots, \nu_{Id})^T$ *is the (truncated) vector of observed successes,*

$$\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta}) = (n_1 \tilde{\boldsymbol{\pi}}_1^T(\boldsymbol{\beta}), \ldots, n_I \tilde{\boldsymbol{\pi}}_I^T(\boldsymbol{\beta}))^T, \tag{9}$$

is the (truncated) vector of expected successes with $\tilde{\boldsymbol{\pi}}_i(\boldsymbol{\beta}) = (\pi_1(\boldsymbol{\beta}_1, \boldsymbol{u}_i), \ldots, \pi_d(\boldsymbol{\beta}_d, \boldsymbol{u}_i))^T$ *and*

$$\tilde{\boldsymbol{W}}^T = (\tilde{\boldsymbol{W}}_1^T, \ldots, \tilde{\boldsymbol{W}}_I^T)_{3d \times Id}, \tag{10}$$

$$\tilde{\boldsymbol{W}}_i = \boldsymbol{I}_d \otimes \boldsymbol{\omega}_i^T = \begin{pmatrix} \boldsymbol{\omega}_i^T & \boldsymbol{0}_3^T & \ldots & \boldsymbol{0}_3^T \\ \boldsymbol{0}_3^T & \boldsymbol{\omega}_i^T & \ldots & \boldsymbol{0}_3^T \\ \vdots & & & \\ \boldsymbol{0}_3^T & \boldsymbol{0}_3^T & \ldots & \boldsymbol{\omega}_i^T \end{pmatrix}_{d \times 3d},$$

where $\boldsymbol{I}_d$ *is the identity matrix of dimension* d, $\boldsymbol{\omega}_i^T = (1, \cos u_i, \sin u_i)$ *and* $\otimes$ *denotes the Kronecker product.*

As expected, when taking $d = 1$ (binomial case), we have the equations in Proposition 1. We can also compute the asymptotic distribution of the MLE:

Theorem 2. *Given the multinomial circular logistic regression model in (5), the asymptotic distribution of the MLE ,* $\widehat{\boldsymbol{\beta}}_{MLE}$*, is given by*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{MLE} - \boldsymbol{\beta}^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}_{3d}, \left(\tilde{\boldsymbol{W}}^T \boldsymbol{\Delta}(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta}^*))\tilde{\boldsymbol{W}}\right)^{-1}\right),$$

where $\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})$ *and* $\tilde{\boldsymbol{W}}$ *are given in (9) and (10), respectively, and*

$$\boldsymbol{\Delta}(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})) = \lim_{n\to\infty} \frac{1}{n} \boldsymbol{\Delta}^{(n)}(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})),$$
$$\boldsymbol{\Delta}^{(n)}(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})) = diag(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})) - \tilde{\boldsymbol{\mu}}(\boldsymbol{\beta})\tilde{\boldsymbol{\mu}}^T(\boldsymbol{\beta}),$$

and $\boldsymbol{\beta}^*$ *is the true value of the parameter vector.*

***Proof*.** It can be proved by generalizing the proof of Theorem to the multinomial case. □

3. Minimum phi-divergence estimators

Let us consider the vectors of empirical and predicted probabilities

$$\widehat{\boldsymbol{p}} = \frac{1}{n}\boldsymbol{\nu} = \frac{1}{n}(\nu_{11}, \ldots, \nu_{1(d+1)}, \ldots, \nu_{I1}, \ldots, \nu_{I(d+1)})^T,$$
$$\boldsymbol{p}(\boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{\mu}(\boldsymbol{\beta}) = \frac{1}{n}(n_1 \boldsymbol{\pi}_1^T(\boldsymbol{\beta}), \ldots, n_I \boldsymbol{\pi}_I^T(\boldsymbol{\beta}))^T,$$

respectively, with $\boldsymbol{\pi}_i(\boldsymbol{\beta}) = (\pi_1(\boldsymbol{\beta}_1, \boldsymbol{u}_i), \ldots, \pi_{d+1}(\boldsymbol{\beta}_{d+1}, \boldsymbol{u}_i))^T$ the Kullback-Leibler divergence between $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\boldsymbol{\beta})$ is given by

$$
\begin{aligned}
D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta})) &= \sum_{i=1}^{I} \sum_{j=1}^{d+1} \frac{\nu_{ij}}{n} \log \frac{\nu_{ij}}{n_i \pi_j(\boldsymbol{\beta}_j, u_i)} \\
&= K - \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{d+1} \nu_{ij} \log \pi_j(\boldsymbol{\beta}_j, u_i), \qquad (11)
\end{aligned}
$$

with K a constant that does not depend on $\boldsymbol{\beta}$. For more details about the Kullback-Leibler divergence one can refer to the pioneer paper by Kullback & Leibler (1951). One can observe that maximizing the log-likelihood in (6) is equivalent to minimizing the Kullback-Leibler divergence in (11). Therefore, we can give an alternative definition to the MLE:

Definition 3. Given the multinomial circular logistic regression model in (5), the MLE, $\widehat{\boldsymbol{\beta}}_{\text{MLE}}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$
\widehat{\boldsymbol{\beta}}_{\text{MLE}} = arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{3d}} D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta})), \qquad (12)
$$

where $D_{KL}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta}))$ is the Kullback-Leibler divergence between $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\boldsymbol{\beta})$ in (11).

Thus, the MLE can be obtained through the minimization of the Kullback-Leibler divergence between the observed and the model probability vectors. The main idea of the proposed approach is the following: *Why not to minimize other divergences, instead of Kullback-Leibler, between both probability vectors in order to obtain alternative (and maybe more suitable) estimators?*

Following this idea, we introduce the Cressie-Read family of minimum phi-divergences (Cressie & Read, 1984).

$$
d_{\phi_\lambda}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta})) = \frac{1}{n} \sum_{i=1}^{I} \sum_{j=1}^{d+1} \pi_j(\boldsymbol{\beta}_j, u_i) \phi_\lambda \left(\frac{\nu_{ij}}{n_i \pi_j(\boldsymbol{\beta}_j, u_i)} \right), \qquad (13)
$$

where

$$
\phi_\lambda(x) = \begin{cases} \frac{1}{\lambda(1+\lambda)} \left[x^{\lambda+1} - x - \lambda(x-1) \right], & \lambda \in \mathbb{R} \setminus \{-1, 0\} \\ \lim_{\upsilon \to \lambda} \frac{1}{\upsilon(1+\upsilon)} \left[x^{\upsilon+1} - x - \upsilon(x-1) \right], & \lambda \in \{-1, 0\} \end{cases}.
$$

This family of divergences depends on a tuning parameter λ. When $\lambda = 0$, we have the Kullback-Leibler divergence. Other well known divergences are obtained for $\lambda = -0.5$ (Hellinger distance), $\lambda = 2/3$ (Cressie-Read divergence) or $\lambda = 1$ (Chi-square divergence).

Definition 4. Given the multinomial circular logistic regression model in (5), the minimum Cressie Read estimator (MCRE), $\widehat{\boldsymbol{\beta}}_{\phi_\lambda}$ of the parameter vector $\boldsymbol{\beta}$ is given by

$$
\widehat{\boldsymbol{\beta}}_{\phi_\lambda} = arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{3d}} d_{\phi_\lambda}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta})), \qquad (14)
$$

where $d_{\phi_\lambda}(\widehat{\boldsymbol{p}}, \boldsymbol{p}(\boldsymbol{\beta}))$ is the Cressie-Read phi-divergence between $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}(\boldsymbol{\beta})$ in (13). For the particular case in which $\lambda = 0$, we have the MLE.

Theorem 3. *Given the multinomial circular logistic regression model in (5), the asymptotic distribution of the MCRE, $\widehat{\boldsymbol{\beta}}_{\phi_\lambda}$, is given by*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{\phi_\lambda} - \boldsymbol{\beta}^*\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_{3d}, \left(\tilde{\boldsymbol{W}}^T \boldsymbol{\Delta}(\tilde{\boldsymbol{\mu}}(\boldsymbol{\beta}^*))\tilde{\boldsymbol{W}}\right)^{-1}\right),$$

i.e., the asymptotic distribution of the MCRE is independent to the tuning parameter λ, particularly, it has the same asymptotic distribution as the MLE ($\lambda = 0$).

***Proof*.** The result is straightforward following the theory given in Lindsay (1994). □

The main idea to understand this result is that all the estimators have the same first order approximation of the residual adjustment function (RAF). See the cited paper by Lindsay (1994). This result suggests that the MCRE is asymptotically fully efficient at the model, so the method provides an efficient estimator of the model parameters when the model is true. However, it is also known that negative values of the parameter λ usually yield to more robust estimators, with an unavoidable loss in efficiency. In particular, the Hellinger distance ($\lambda = -0.5$) is known in the statistical literature for its robustness properties. Although this robustness can not be proved through the Influence Function (Rousseeuw et al., 2011), as it is based again in the first-order approximation, we can empirically show it through an extensive simulation study.

4. Monte Carlo Simulation Study

We develop a simulation study to evaluate the behaviour of the proposed estimators. We consider both binomial and multinomial responses (three categories, $d = 2$). To generate the data we consider three different scenarios: von Misses distribution (Mardia & Zemroch, 1975) with mean 60° and concentration parameters $\kappa = 4, 8$ and Spherical Normal distribution (Hauberg, 2018; Castilla, 2022) with concentration parameter $\kappa = 6$.

For the binomial case, we consider $\boldsymbol{\beta}^T = (0, 2, 2)$ and $n \in \{20, \ldots, 100\}$ different responses, without repeated covariates. The outliers are generated by interchanging a 10% of the responses, selected randomly. In the case of the multinomial response, we consider $\boldsymbol{\beta}^T = (0, 2, 2, 0.2, 2.5, 1.5)$ and $I \in \{10, \ldots, 50\}$ categories with $n_i = 5$ samples in each one. The outliers are obtained by assigning all the responses to the third category, the one with less probability. The vector of parameters $\boldsymbol{\beta}$ is estimated for each one of 1000 replications.

The mean absolute errors (MAE) of the estimated probabilities (computed with the estimated vector of parameters) are obtained for different MCRE and presented in Figure 1 (binomial response) and Figure 2 (multinomial response).

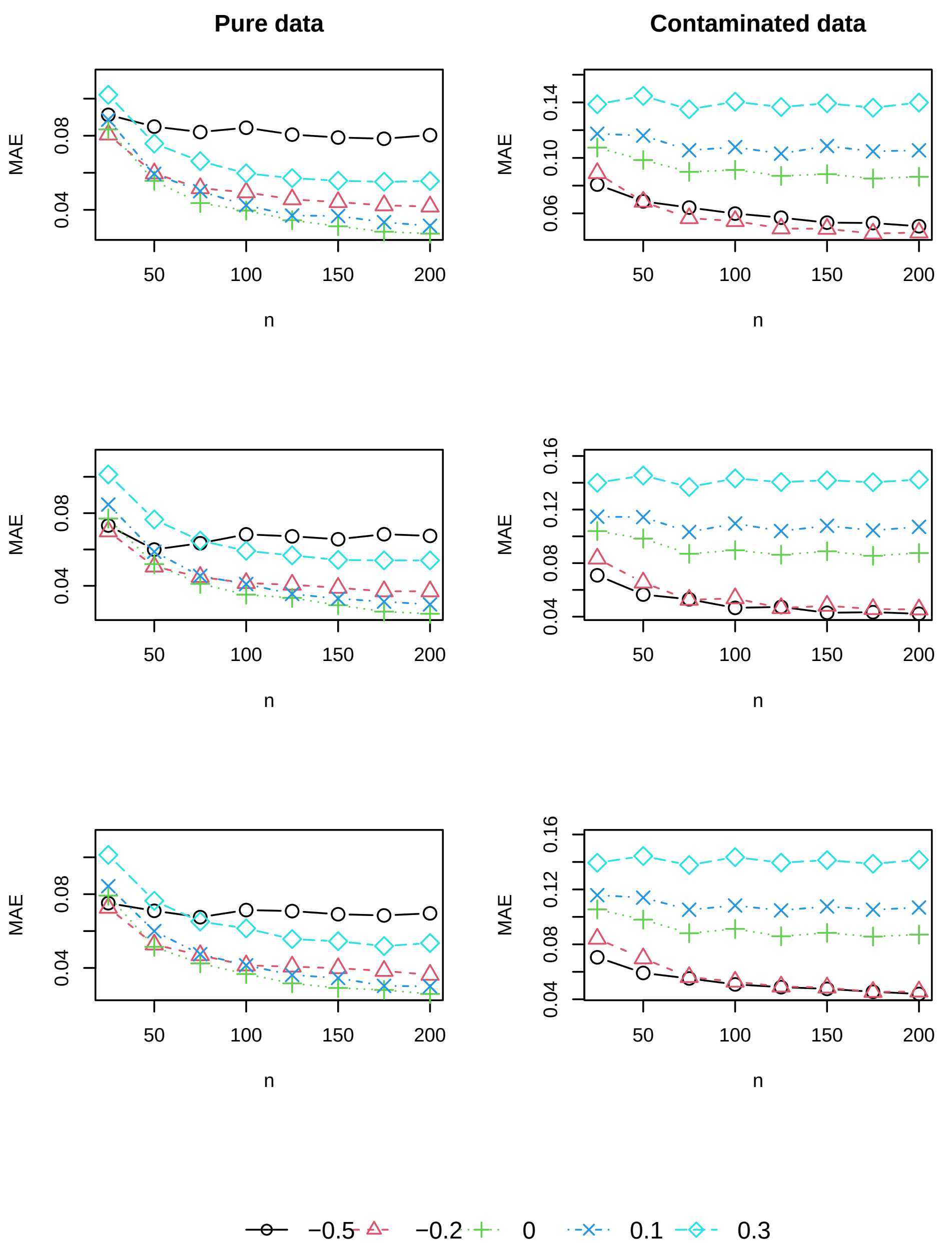

FIGURE 1: Binomial case. Results from the Monte Carlo study. Von Misses distribution with $\kappa = 4$ (above), Von Misses distribution with $\kappa = 6$ (middle) and Spherical Normal distribution with $\kappa = 8$ (below).

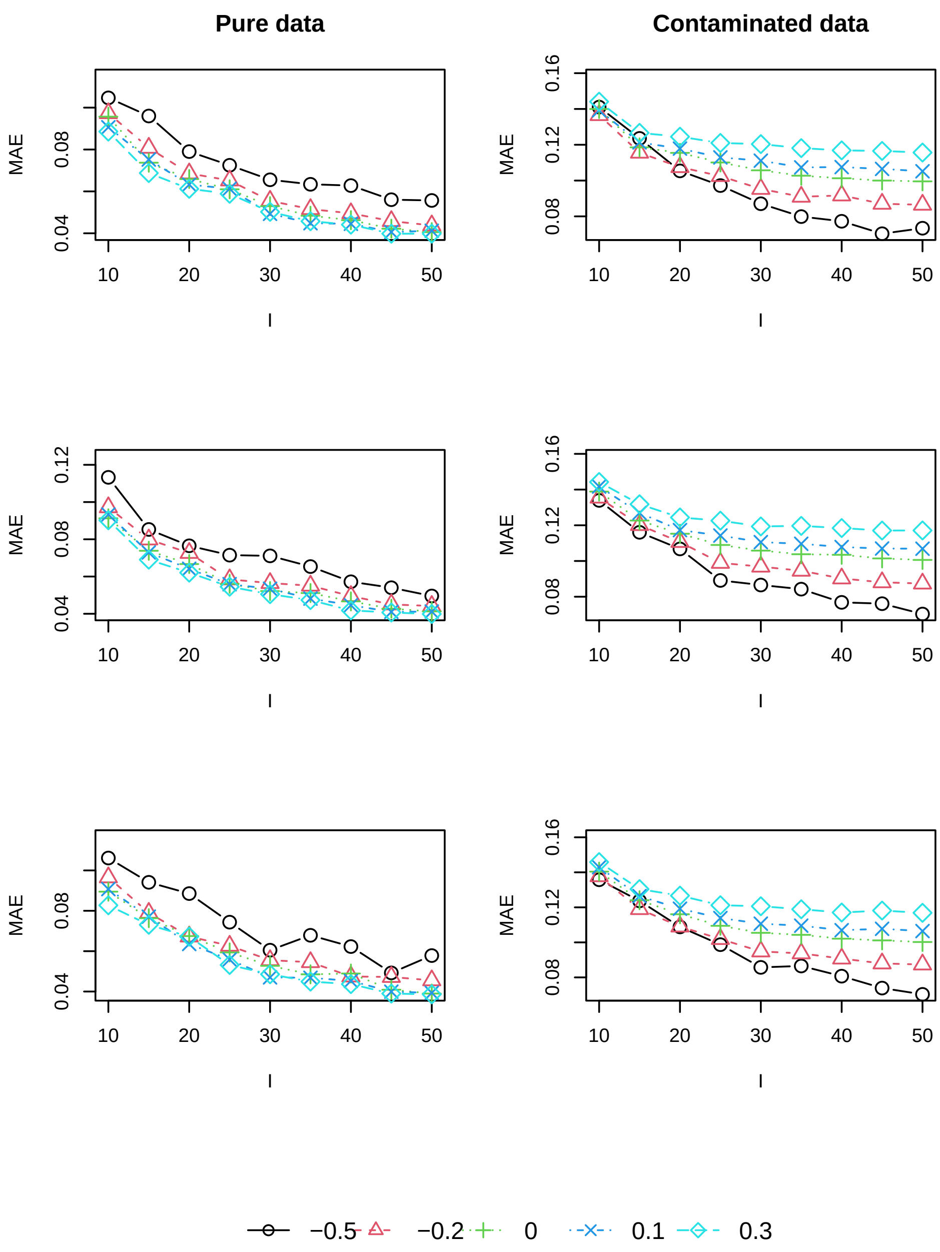

FIGURE 2: Multinomial case. Results from the Monte Carlo study. Von Misses distribution with $\kappa = 4$ (above), Von Misses distribution with $\kappa = 6$ (middle) and Spherical Normal distribution with $\kappa = 8$ (below).

$$\text{MAE}_\lambda(\boldsymbol{\pi}) = \frac{1}{Id}\sum_{i=1}^{I}\sum_{j=1}^{d}|\pi_{ij}(\widehat{\boldsymbol{\beta}}_{\phi_\lambda}) - \pi_{ij}(\boldsymbol{\beta})|.$$

In both cases, minimum phi-divergence estimators with negative tuning parameter outperforms the MLE ($\lambda = 0$) and minimum phi-divergence estimators with positive tuning parameter when considering a contaminated scenario. MCREs with negative tuning parameter are sometimes a more efficient alternative to MLE in case of pure data.

5. Application to Life and Social Sciences

In this section, we present three numerical examples to illustrate the applicability of the proposed methods. In order to evaluate the performance of them, we compute the accuracy of the estimations, i.e., the proportion of events that are classified correctly.

Example 1 (**Botany data**). Let us consider the dataset of leaf inclination angle measurement recorded in Chianucci et al. (2018) and also analyzed in Alshqaq et al. (2021). This dataset contains the leaf inclination angles of 138 plant species. In particular, we want to classify the species *Betula pendula* and *Aesculus hippocastanum* by the angle inclination of their bottom canopy. We first split the dataset into training and testing sets assigning a random 70% of data points to the former and the remaining 30% to the latter. We apply the circular logistic regression to the training set (see Figure 3, top left), and evaluate the fitted model with the test set. Results for different values of the tuning parameter λ are shown in Table 1. Although there is not a huge difference among MCREs, estimators with $\lambda < 0$ outperform the classical MLE. This suggests presence of outliers in our data, which is in concordance with Figure 3.

***Note* 1.** The use of divergence-based estimators in our data may have two different utilities: (1) develop a more robust inference than that based on MLE and (2) detect the presence of outliers in our data, as happened in Example 1.

Example 2 (**Crime data**). Let us consider a data set obtained from The Crime Open Database (CODE) (Ashby, 2019), a service that records crime data from multiple US cities. In particular, we randomly select 250 motor vehicle thefts and 250 fraud offenses (except counterfeiting/forgery and bad checks) commited in the city of Chicago during 2020. We split again the dataset into training and testing sets (70% and 30% of data points, respectively) and we wonder if the time of crime is able to predict the crime commited. See top right part of Figure 3. We apply a circular logistic model taking as explanatory variable the time of crime, illustrating how time can be treated as circular data. In this case, the proportion of crimes that are estimated correctly is not excessively high, not exceeding a 70%, as it can be seen in Table 2. However, MCRE with $\lambda < 0$ may improve a 5% the prediction via MLE.

Example 3 (Meteorological data). Finally, let us illustrate the multinomial circular logistic regression model here presented. We take data from the "Portale Open Data della Regione Siciliana" which contains meteorological data from Sicilia (Italy) (Open Data, 2019). In particular, we take the temperature of wind at two metters of height in June 2016 in the region of Palermo (see Figure 3, below). We divide the temperature in three different groups (less than 20, between 20 and 27, and more than 27 which correspond to the terciles of the variable temperature). MCREs can improve MLE a 4% with good performance results, as seen in Table 3.

TABLE 1: Botany data: proportion of plants that are classified correctly.

				Accuracy	
λ	$\widehat{\beta}_{0,\phi_\lambda}$	$\widehat{\beta}_{1,\phi_\lambda}$	$\widehat{\beta}_{2,\phi_\lambda}$	Trainig set	Test set
0 (MLE)	9.3925	-9.9489	-2.8964	0.7274	0.7454
-0.7	38.2534	-39.5737	-13.0746	0.7290	0.7491
-0.5	24.1549	-24.8110	-8.4341	0.7290	0.7491
-0.2	13.6915	-14.1959	-4.6150	0.7274	0.7491
0.2	68.169	-7.3941	-1.8710	0.7227	0.7380
0.3	59.480	-6.5211	-1.5381	0.7211	0.7343
0.5	47.246	-5.2731	-1.0936	0.7242	0.7343

TABLE 2: Crime data: proportion of crimes that are classified correctly.

				Accuracy	
λ	$\widehat{\beta}_{0,\phi_\lambda}$	$\widehat{\beta}_{1,\phi_\lambda}$	$\widehat{\beta}_{2,\phi_\lambda}$	Trainig set	Test set
0 (MLE)	0.0737	0.6786	-0.2708	0.6314	0.6200
-0.7	2.9454	6.7962	0.3494	0.6686	0.6733
-0.5	0.3267	1.5944	-0.3707	0.6400	0.6333
-0.2	0.1063	0.8718	-0.3254	0.6314	0.6333
0.2	0.0583	0.5671	-0.2328	0.6314	0.6200
0.3	0.0527	0.5223	-0.2164	0.6314	0.6200
0.5	0.0444	0.4513	-0.1894	0.6314	0.6200

TABLE 3: Meteorological: proportion of temperatures that are classified correctly.

							Accuracy	
λ	$\widehat{\beta}_{10,\phi_\lambda}$	$\widehat{\beta}_{11,\phi_\lambda}$	$\widehat{\beta}_{12,\phi_\lambda}$	$\widehat{\beta}_{20,\phi_\lambda}$	$\widehat{\beta}_{21,\phi_\lambda}$	$\widehat{\beta}_{22,\phi_\lambda}$	Trainig set	Test set
0 (MLE)	0.3782	5.5019	1.4715	21.061	2.7386	0.4274	0.6302	0.7037
-0.7	-0.8461	14.1766	4.3646	41.744	5.8311	0.7724	0.6362	0.7407
-0.5	0.0320	9.4926	2.7134	32.253	4.2988	0.6020	0.6362	0.7407
-0.2	0.3155	6.6638	1.7906	24.592	3.2197	0.4747	0.6362	0.7407
0.2	0.4067	4.7189	1.2791	18.583	2.4037	0.3977	0.6302	0.7037
0.3	0.4139	4.3822	1.2009	17.484	2.2563	0.3850	0.6362	0.7407
0.5	0.4165	3.8294	1.0765	15.620	2.0087	0.3634	0.6322	0.7269

6. Conclusions and future lines of research

In this paper, we develop robust inference for the circular logistic regression, introducing the multinomial circular logistic regression as well. The simulation

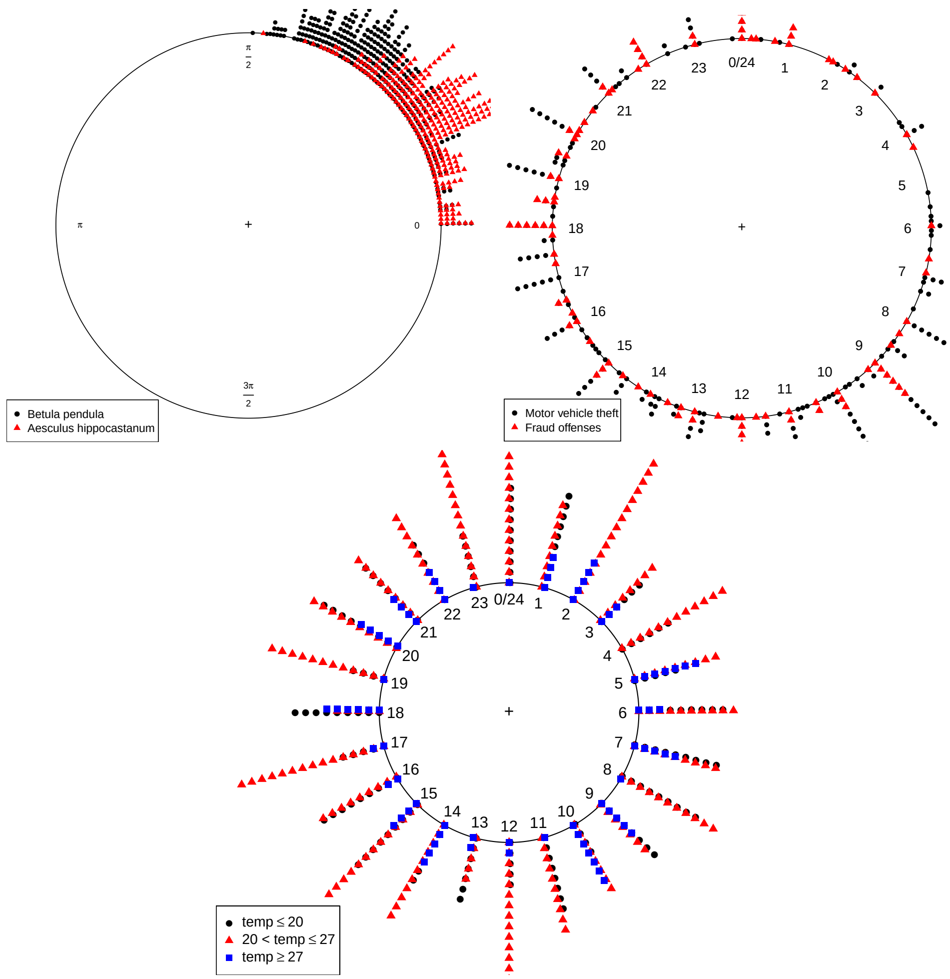

Figure 3: Botany data (top left), Crime data (top right) and Meteorological data (bottom).

studies show the robustness of proposed estimators. Finally, three numerical examples of Botany, Crime and Meteorology illustrate the application of these methods to Life and Social Sciences. Although in the Botany data the proposed method showed very little improvement, in the Crime and Meteorological data an increment up to 5% and 4% of accuracy, respectively, is achieved. Results also suggest the presence of outliers in our data sets.

A more extended model may consider more than one predictor variables, and may also combine linear and circular predictors. Without loss of generality, let us consider a binary response variable η, $\eta \in \{0,1\}$ and that this depends on R circular explanatory variables $u_1, \dots, u_R$ and S linear explanatory variables $x_1, \dots, x_S$. The (binomial) linear-circular logistic regression model would be given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{x}) = \frac{\exp\left\{\beta_0 + \sum_{r=1}^{R} (\beta_1^{(r)} \cos u_r + \beta_2^{(r)} \sin u_r) + \sum_{s=1}^{S} \beta_3^{(s)} x_s\right\}}{1 + \exp\left\{\beta_0 + \sum_{r=1}^{R} (\beta_1^{(r)} \cos u_r + \beta_2^{(r)} \sin u_r) + \sum_{s=1}^{S} \beta_3^{(s)} x_s\right\}}, \tag{15}$$

where $\boldsymbol{\beta}^T = (\beta_0, \beta_1^{(1)}, \beta_2^{(1)}, \dots, \beta_1^{(R)}, \beta_2^{(R)}, \beta_3^{(1)}, \dots, \beta_3^{(S)}) \in \mathbb{R}^{1+2R+S}$ is the model parameter vector. In a similar manner as done in Section 2.2, we can extend model (15) to the multinomial response case. We may also consider a complex set-up, see Morel (1989), Skinner et al. (1992) and Castilla & Chocano (2022) for more details.

[Received: April 2022 — Accepted: November 2022]

References

Abuzaid, A. H. & Allahham, N. R. (2015), 'Simple circular regression model assuming wrapped cauchy error', *Pakistan Journal of Statistics* **31**(4).

Al-Daffaie, K. & Khan, S. (2017), Logistic regression for circular data, *in* 'AIP Conference Proceedings', Vol. 1842, AIP Publishing LLC, p. 030022.

Alshqaq, S. S., Ahmadini, A. A. & Abuzaid, A. H. (2021), 'Some new robust estimators for circular logistic regression model with applications on meteorological and ecological data', *Mathematical Problems in Engineering* **2021**.

Ashby, M. P. (2019), 'Studying crime and place with the crime open database: Social and behavioural scienes', *Research Data Journal for the Humanities and Social Sciences* **4**(1), 65–80.

Bell, K. (2008), *Analysing Cycles in Biology and Medicine–a practical introduction to circular variables & periodic regression*, Razorbill Press, St. John's.

Berkson, J. (1944), 'Application of the logistic function to bio-assay', *Journal of the American statistical association* **39**(227), 357–365.

Castilla, E. (2022), 'Robust estimation of the spherical normal distribution', *Mathematica Applicanda* **50**(1), 43–63.

Castilla, E. & Chocano, P. J. (2022), 'A new robust approach for multinomial logistic regression with complex design model', *IEEE Transactions on Information Theory* **68**(11), 7379–7395.

Chianucci, F., Pisek, J., Raabe, K., Marchino, L., Ferrara, C. & Corona, P. (2018), 'A dataset of leaf inclination angles for temperate and boreal broadleaf woody species', *Annals of Forest Science* **75**(2), 1–7.

Cressie, N. & Read, T. R. (1984), 'Multinomial goodness-of-fit tests', *Journal of the Royal Statistical Society: Series B (Methodological)* **46**(3), 440–464.

Gill, J. & Hangartner, D. (2010), 'Circular data in political science and how to handle it', *Political Analysis* **18**(3), 316–336.

Hauberg, S. (2018), Directional statistics with the spherical normal distribution, *in* '2018 21st International Conference on Information Fusion (FUSION)', IEEE, pp. 704–711.

Jones, M. & Pewsey, A. (2012), 'Inverse Batschelet distributions for circular data', *Biometrics* **68**(1), 183–193.

Kibiak, T. & Jonas, C. (2007), 'Applying circular statistics to the analysis of monitoring data', *European Journal of Psychological Assessment* **23**(4), 227–237.

Kullback, S. & Leibler, R. A. (1951), 'On information and sufficiency', *The Annals of Mathematical Statistics* **22**(1), 79–86.

Landler, L., Ruxton, G. D. & Malkemper, E. P. (2018), 'Circular data in biology: advice for effectively implementing statistical procedures', *Behavioral ecology and sociobiology* **72**(8), 1–10.

Lindsay, B. G. (1994), 'Efficiency versus robustness: the case for minimum Hellinger distance and related methods', *The annals of statistics* **22**(2), 1081–1114.

Mardia, K. & Zemroch, P. (1975), 'Algorithm AS 86: The von Mises distribution function', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **24**(2), 268–272.

Morel, J. G. (1989), 'Logistic regression under complex survey designs', *Survey Methodology* **15**(2), 203–223.

Open Data (2019), 'Portale open data della regione siciliana'. https://dati.regione.sicilia.it/

Rousseeuw, P. J., Hampel, F. R., Ronchetti, E. M. & Stahel, W. A. (2011), *Robust statistics: the approach based on influence functions*, John Wiley & Sons.

SenGupta, A. & Ugwuowo, F. I. (2006), 'Asymmetric circular-linear multivariate regression models with applications to environmental data', *Environmental and Ecological Statistics* **13**(3), 299–309.

Skinner, C. J. et al. (1992), 'Pseudo-likelihood and quasi-likelihood estimation for complex sampling schemes', *Computational statistics & data analysis* **13**(4), 395–405.

Uemura, M., Meglič, A., Zalucki, M. P., Battisti, A. & Belušič, G. (2021), 'Spatial orientation of social caterpillars is influenced by polarized light', *Biology Letters* **17**(2), 20200736.

Wolpert, N. & Tallon-Baudry, C. (2021), 'Coupling between the phase of a neural oscillation or bodily rhythm with behavior: Evaluation of different statistical procedures', *NeuroImage* **236**, 118050.

Appendix. Tables of Results of the Monte Carlo Simulation Study

In this Appendix, we present the tables of results of the simulation study for each case, which can also be found in Figures 1 and 2.

Table A1: Binomial case and Pure data. Results from the Monte Carlo study.

	MAEs							
λ	$n_i = 25$	50	75	100	125	150	175	200
	Von Misses distribution ($\kappa = 4$)							
-0.5	0.0913	0.0849	0.082	0.0843	0.0806	0.0791	0.0784	0.0803
-0.2	0.0806	0.0596	0.0517	0.0494	0.0456	0.0442	0.0424	0.0418
0(MLE)	0.0834	0.0556	0.0437	0.0396	0.0344	0.0311	0.0282	0.0272
0.1	0.0886	0.0594	0.0501	0.0425	0.0369	0.0367	0.0333	0.0314
0.3	0.1021	0.0758	0.0663	0.0597	0.0571	0.0557	0.0551	0.0556
	Von Misses distribution ($\kappa = 6$)							
-0.5	0.0732	0.0600	0.0634	0.0683	0.0672	0.0656	0.0684	0.0675
-0.2	0.0700	0.0506	0.0449	0.0415	0.0406	0.0389	0.0370	0.0369
0(MLE)	0.0770	0.0520	0.0411	0.0351	0.0333	0.0293	0.0258	0.0246
0.1	0.0847	0.0587	0.0454	0.0409	0.0355	0.0329	0.0313	0.0298
0.3	0.1013	0.0765	0.0649	0.0593	0.0567	0.0542	0.0540	0.0540
	Spherical Normal distribution ($\kappa = 8$)							
-0.5	0.0751	0.0710	0.0676	0.0714	0.0708	0.0691	0.0685	0.0696
-0.2	0.0726	0.0527	0.0469	0.0414	0.0407	0.0399	0.0384	0.0363
0(MLE)	0.0792	0.0514	0.0425	0.0367	0.0316	0.0292	0.0281	0.0259
0.1	0.0843	0.0600	0.0475	0.0415	0.0362	0.0346	0.0303	0.0300
0.3	0.1013	0.0764	0.0652	0.0615	0.0556	0.0545	0.0519	0.0535

Table A2: Binomial case and Contaminated data. Results from the Monte Carlo study.

	MAEs							
λ	$n_i = 25$	50	75	100	125	150	175	200
	Von Misses distribution ($\kappa = 4$)							
-0.5	0.0808	0.0686	0.0642	0.0598	0.0569	0.0533	0.0530	0.0507
-0.2	0.0890	0.0684	0.0565	0.0544	0.0491	0.0488	0.0454	0.0462
0(MLE)	0.1074	0.0985	0.0899	0.0913	0.0870	0.0882	0.0851	0.0864
0.1	0.1174	0.1161	0.1055	0.1078	0.1031	0.1086	0.1047	0.1055
0.3	0.1387	0.1447	0.1351	0.1405	0.1367	0.1394	0.1362	0.1399
	Von Misses distribution ($\kappa = 6$)							
-0.5	0.0709	0.0565	0.0530	0.0466	0.0473	0.0429	0.0434	0.0422
-0.2	0.0834	0.0655	0.0526	0.0537	0.0464	0.0482	0.0457	0.0454
0(MLE)	0.1039	0.0983	0.0870	0.0896	0.0862	0.0888	0.0855	0.0875
0.1	0.1147	0.1143	0.1032	0.1096	0.1039	0.1079	0.1044	0.1071
0.3	0.1400	0.1454	0.1368	0.1432	0.1405	0.1419	0.1404	0.1423
	Spherical Normal distribution ($\kappa = 8$)							
-0.5	0.0706	0.0592	0.0552	0.0508	0.0488	0.0476	0.0455	0.0439
-0.2	0.0841	0.0698	0.0562	0.0528	0.0493	0.0485	0.0453	0.0456
0(MLE)	0.1054	0.0980	0.0881	0.0912	0.0859	0.0884	0.0857	0.0871
0.1	0.1158	0.114	0.1052	0.1083	0.1047	0.1076	0.1052	0.1068
0.3	0.1394	0.1443	0.1377	0.1435	0.1395	0.1413	0.1387	0.1415

Table A3: Multinomial case and Pure data. Results from the Monte Carlo study.

	MAEs								
λ	I =10	15	20	25	30	35	40	45	50
	Von Misses distribution ($\kappa = 4$)								
-0.5	0.1047	0.096	0.079	0.0725	0.0656	0.0634	0.0628	0.0561	0.0557
-0.2	0.0973	0.0808	0.0686	0.0652	0.0555	0.0515	0.0495	0.0458	0.0437
0(MLE)	0.0957	0.0737	0.0659	0.0610	0.0530	0.0484	0.0463	0.0422	0.0405
0.1	0.0906	0.0752	0.0633	0.0613	0.0493	0.0448	0.0445	0.0406	0.0412
0.3	0.0886	0.0687	0.0612	0.0589	0.0502	0.0458	0.0441	0.0398	0.0398
	Von Misses distribution ($\kappa = 6$)								
-0.5	0.1133	0.0853	0.0765	0.0715	0.0711	0.0654	0.0572	0.054	0.0496
-0.2	0.0972	0.0799	0.0727	0.0585	0.0565	0.0551	0.0495	0.0450	0.0442
0(MLE)	0.0913	0.0738	0.0667	0.0561	0.0522	0.0511	0.0466	0.0427	0.0413
0.1	0.0933	0.0729	0.0642	0.0559	0.0534	0.0482	0.0446	0.0410	0.0412
0.3	0.0902	0.069	0.0621	0.0545	0.0505	0.0474	0.0418	0.0407	0.0399
	Spherical Normal distribution ($\kappa = 8$)								
-0.5	0.1062	0.0941	0.0885	0.0744	0.0605	0.0679	0.0623	0.0492	0.0578
-0.2	0.0966	0.0790	0.0673	0.0627	0.0555	0.0545	0.0475	0.0472	0.0455
0(MLE)	0.0895	0.0763	0.0675	0.0594	0.0527	0.0485	0.0489	0.0409	0.0390
0.1	0.0907	0.0771	0.0636	0.0559	0.0470	0.0468	0.0453	0.0399	0.0393
0.3	0.0827	0.0731	0.0672	0.0532	0.0486	0.0449	0.0437	0.0390	0.0386

Table A4: Multinomial case and Contaminated data. Results from the Monte Carlo study.

λ	MAEs								
	I =10	15	20	25	30	35	40	45	50
	Von Misses distribution ($\kappa = 4$)								
-0.5	0.1411	0.1236	0.1054	0.0972	0.0871	0.0799	0.0773	0.0703	0.0734
-0.2	0.1366	0.1156	0.1075	0.1025	0.0953	0.0910	0.0917	0.0870	0.0865
0(MLE)	0.1400	0.1185	0.1155	0.1101	0.1057	0.1027	0.1013	0.1000	0.0996
0.1	0.1392	0.1213	0.1178	0.1129	0.1111	0.1074	0.1075	0.1065	0.1052
0.3	0.1441	0.1266	0.1245	0.1210	0.1204	0.1181	0.1169	0.1165	0.1156
	Von Misses distribution ($\kappa = 6$)								
-0.5	0.1339	0.1161	0.1068	0.0891	0.0866	0.0842	0.0768	0.0761	0.0703
-0.2	0.1355	0.1201	0.1107	0.0989	0.0968	0.0945	0.0902	0.0881	0.0874
0(MLE)	0.1389	0.1227	0.1153	0.1090	0.1058	0.1038	0.1035	0.1014	0.1006
0.1	0.1417	0.1269	0.1173	0.1142	0.1106	0.1096	0.1078	0.1070	0.1069
0.3	0.1442	0.1318	0.1244	0.1226	0.1194	0.1197	0.1184	0.1171	0.1171
	Spherical Normal distribution ($\kappa = 8$)								
-0.5	0.1358	0.1237	0.1088	0.0987	0.0857	0.0865	0.0807	0.0739	0.0703
-0.2	0.1378	0.1190	0.1090	0.1019	0.0948	0.0937	0.0908	0.0880	0.0874
0(MLE)	0.1404	0.1241	0.1161	0.1095	0.1054	0.1043	0.1021	0.1011	0.1002
0.1	0.1429	0.1264	0.1192	0.1139	0.1107	0.1095	0.1069	0.1077	0.1064
0.3	0.1458	0.1304	0.1267	0.1212	0.1206	0.1189	0.1171	0.1182	0.1169