MDPI

Article

Robust Minimum Divergence Estimation for the Multinomial Circular Logistic Regression Model

Elena Castilla [1,*,†] and **Abhik Ghosh [2,†]**

[1] Departamento de Matematica Aplicada, Rey Juan Carlos University, Mostoles Campus, 28933 Madrid, Spain
[2] Indian Statistical Institute, Kolkata 700108, India; abhik.ghosh@isical.ac.in
[*] Correspondence: elena.castilla@urjc.es
[†] These authors contributed equally to this work.

Abstract: Circular data are extremely important in many different contexts of natural and social science, from forestry to sociology, among many others. Since the usual inference procedures based on the maximum likelihood principle are known to be extremely non-robust in the presence of possible data contamination, in this paper, we develop robust estimators for the general class of multinomial circular logistic regression models involving multiple circular covariates. Particularly, we extend the popular density-power-divergence-based estimation approach for this particular set-up and study the asymptotic properties of the resulting estimators. The robustness of the proposed estimators is illustrated through extensive simulation studies and few important real data examples from forest science and meteorology.

1. Introduction

One of the most representative examples in the area of directional statistics are circular data, which are characterized as points in the unit circle. Once the origin and the direction of rotation have been fixed, the observations are measured by their direction represented as an angle θ (from 0 to 2π if measured in radians), a unit vector $(\cos\theta, \sin\theta)$, or even a complex number with unit modulus, $e^{i\theta}$. These data represent periodic phenomena, such as directions or measurements over time (time of day, month, lunar cycle...). For a detailed survey of directional statistics we refer to reference [1].

Circular statistics is fast becoming a key instrument in many fields because of its applicability. In biology, for example, it has been applied to the study of avian migration routes [2–4] or animal movement [5,6], and to the analysis of mammalian circadian timekeeping [7]. Other examples include leaf inclination angles [8,9]. In atmospheric sciences, circular statistics has been widely applied to the analysis of wind directions [8–11]. Following this idea, ref. [12] used circular data in the study of wind and solar energy. Other examples include medicine [13] or astrophysics [14]. Circular data are also found in social sciences, such as policy making [15], sociology [16], economics [17] or criminology [9,18].

There are many approaches to model circular data that model the data using a circular distribution. The simplest one is the circular uniform distribution, which assigns equal probabilities to all the points of the circumference. The von Misses (vM) distribution is among the oldest and most used circular distributions. Originally introduced in reference [19] and deeply analyzed in reference [20], the vM distribution has been applied in numerous contexts [15,21,22]. In recent years, some extensions or alternatives to vM distributions have been developed, among which the spherical normal (SN) distribution [23–25] has become quite popular.

When we have additional covariates to explain the variations or predict the response data, appropriate regression models involving circular variables have been developed in recent years. The circular–circular regression model relates a circular response variable

with a circular explanatory variable [26], while circular–linear regression relates a circular response variable linearly with a vector of given covariates; see, e.g., reference [27] and references therein. On the other hand, the linear–circular regression [28] assumes a linear response variable and circular explanatory variables. The circular logistic regression, introduced by reference [29] and analyzed by references [8,30,31] among others, relates a binary response variable and one circular covariate. Recently, Ref. [9] extended this model to a multinomial response, in the so-called multinomial circular logistic regression (MCLR). The MCLR is an adaptation of the classical multinomial logistic regression (MLR) model, where covariates are not assumed to be circular [32]. Usually the maximum likelihood estimator (MLE) is used to perform inferences for such circular regression models, which is known to be asymptotically most efficient but highly non-robust under possible data contamination [33].

In this paper, we develop robust inference for the general class of MCLR models allowing multiple circular covariates. In particular, a new family of estimators, the minimum density power divergence (DPD) estimators, are defined as a generalization of the classical MLE and its asymptotic properties are studied. An extensive simulation study illustrates the improved robust performance of the proposed method over the existing MLE. The applicability and importance of our proposal is further justified through a few interesting real data examples. Finally, the paper ends with some concluding remarks and future direction for work.

2. Model Description

Let us consider that our response variable η has $d + 1$ categories, $\eta \in \{1, \ldots, d + 1\}$, and that it depends on k circular explanatory variables $\boldsymbol{u} = (u_1, \ldots, u_k)^T$. If $\pi_j(\boldsymbol{\beta})$ denotes the probability that η belongs to the j-th category, the multinomial circular logistic regression model is given by

$$
\begin{aligned}
\pi_j(\boldsymbol{\beta}|\boldsymbol{u}) &= \frac{\exp\left\{\beta_{j0} + \sum_{l=1}^k \left[\beta_{jl}^{(1)} \cos u_l + \beta_{jl}^{(2)} \sin u_l\right]\right\}}{1 + \sum_{s=1}^d \exp\left\{\beta_{s0} + \sum_{l=1}^k \left[\beta_{sl}^{(1)} \cos u_l + \beta_{sl}^{(2)} \sin u_l\right]\right\}} \\
&= \frac{\exp\{\beta_{j0} + \beta_{j1}^{(1)} \cos u_1 + \beta_{j1}^{(2)} \sin u_1 + \cdots + \beta_{jk}^{(1)} \cos u_k + \beta_{jk}^{(2)} \sin u_k\}}{1 + \sum_{s=1}^d \exp\{\beta_{s0} + \beta_{s1}^{(1)} \cos u_1 + \beta_{s1}^{(2)} \sin u_1 + \cdots + \beta_{sk}^{(1)} \cos u_k + \beta_{sk}^{(2)} \sin u_k\}},
\end{aligned}
\tag{1}
$$

for $j = 1, \ldots, d$, and $\pi_{d+1}(\boldsymbol{\beta}|\boldsymbol{u}) = 1 - \sum_{j=1}^d \pi_j(\boldsymbol{\beta}|\boldsymbol{u})$. Here, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}^{(1)}, \beta_{j1}^{(2)}, \ldots, \beta_{jk}^{(1)}, \beta_{jk}^{(2)})^T$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_d^T)^T \in \mathbb{R}^{d(2k+1)}$ is the model parameter vector representing the association between the circular predictors and the multinomial response variable.

Suppose that we observe n independent observations divided in I covariate patterns, each one with n_i observations ($\sum_{i=1}^I n_i = n$), and an associated vector of explanatory variables $\boldsymbol{u}_i = (u_{i1}, \ldots, u_{ik})^T$. For the i-th covariate pattern, let us consider that the number of response variables that fall in the j-th category is denoted by v_{ij}. For simplicity, we will also denote $\pi_{ij}(\boldsymbol{\beta}) \equiv \pi_j(\boldsymbol{\beta}|\boldsymbol{u}_i)$. Then, the likelihood associated with the model is given by $\mathcal{L}(\boldsymbol{\beta}) \propto \prod_{i=1}^I \prod_{j=1}^{d+1} \pi_{ij}(\boldsymbol{\beta})^{v_{ij}}$, and the log-likelihood

$$
\ell(\boldsymbol{\beta}) = \log \mathcal{L}(\boldsymbol{\beta}) \equiv c + \sum_{i=1}^I \sum_{j=1}^{d+1} v_{ij} \log \pi_{ij}(\boldsymbol{\beta}),
\tag{2}
$$

for any positive constant c; see reference [9] for details.

Definition 1. *Given the multinomial circular logistic regression model (1), the maximum likelihood estimator (MLE) of model parameters $\boldsymbol{\beta}$ is defined as*

$$
\widehat{\boldsymbol{\beta}}_{MLE} = \arg \max_{\boldsymbol{\beta} \in \mathbb{R}^{d(2k+1)}} \ell(\boldsymbol{\beta}),
$$

where $\ell(\boldsymbol{\beta})$ is the log-likelihood based on the observed data as defined in Equation (2).

Let us introduce the following notation. For the i-th covariate pattern, let $\boldsymbol{v}_i = (v_{i1}, \ldots, v_{i(d+1)})^T$ denote the vector of the number of responses within each category, and let $\boldsymbol{\pi}_i(\boldsymbol{\beta}) = (\pi_{i1}(\boldsymbol{\beta}), \ldots, \pi_{i(d+1)}(\boldsymbol{\beta}))^T$ denote the vector of model probabilities. With superscript * along with a vector (or matrix), we will denote the truncated vector (matrix) obtained by deleting the last value (row) from the initial vector (matrix). Thus, $\boldsymbol{v}_i^* = (v_{i1}, \ldots, v_{id})^T$, $\boldsymbol{\pi}_i(\boldsymbol{\beta})^* = (\pi_{i1}(\boldsymbol{\beta}), \ldots, \pi_{id}(\boldsymbol{\beta}))^T$ and so on. Let $\boldsymbol{p}$ be a vector and let us denote $\boldsymbol{\Delta}(\boldsymbol{p}) = diag(\boldsymbol{p}) - \boldsymbol{p}\boldsymbol{p}^T$, where $diag(\boldsymbol{p})$ denotes the matrix with the entries of $\boldsymbol{p}$ along the diagonal. Finally, let us denote

$$\boldsymbol{\omega}_i = (1, \cos u_{i1}, \sin u_{i1}, \ldots, \cos u_{ik}, \sin u_{ik})^T, \quad i = 1, \ldots, I.$$

Note that, for $j = 1, \ldots, d$, we have

$$\frac{\partial \pi_{ij}(\boldsymbol{\beta})}{\boldsymbol{\beta}_j} = \pi_{ij}(\boldsymbol{\beta})(1 - \pi_{ij}(\boldsymbol{\beta}))\boldsymbol{\omega}_i. \tag{3}$$

Computing the derivative of (2) with respect to $\boldsymbol{\beta}$ and using (3), the estimating equations associated to the MLE are then given, as in the following result. The subsequent result presents the asymptotic distribution of the MLE for the multinomial circular logistic regression model (1); see reference [9] for proofs of both results.

Proposition 1. *Given the multinomial circular logistic regression model (1), the MLE ($\widehat{\boldsymbol{\beta}}_{MLE}$) of the parameter vector $\boldsymbol{\beta}$ as defined in Definition 1 can be equivalently obtained by solving the following system of estimating equations:*

$$\sum_{i=1}^{I}(\boldsymbol{v}_i^* - n_i\boldsymbol{\pi}_i^*(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i = \boldsymbol{0}_{d(2k+1)}, \tag{4}$$

where $\otimes$ denotes the Kronecker product.

Proposition 2. *Given the multinomial circular logistic regression model (1), the asymptotic distribution of the MLE of the parameter vector $\boldsymbol{\beta}$, as defined in Definition 1, is given by*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_{MLE} - \boldsymbol{\beta}^0\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(\boldsymbol{0}_{d(2k+1)}, \boldsymbol{\Omega}^{-1}(\boldsymbol{\beta}^0)\right),$$

where $\boldsymbol{\beta}^0$ denotes the true value of $\boldsymbol{\beta}$ and

$$\boldsymbol{\Omega}(\boldsymbol{\beta}) = \lim_{n\to\infty} \sum_{i=1}^{I} \frac{n_i}{n} \boldsymbol{\Delta}(\boldsymbol{\pi}_i^*(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i\boldsymbol{\omega}_i^T.$$

Since the above MLE is known to be highly non-robust (see Sections 4 and 5 as well for numerical illustrations), we propose a general class of robust minimum density power divergence estimator under the MCLR models in the following section. Note that, in this paper, we refer to robust estimators to those insensitive to outliers. While outliers can provide meaningful information concerning potential model misspecification, these may also stem from measurement or input errors. These failed observations may make a large difference in the results of regression analysis when following the classical maximum likelihood procedures [33], and the development of alternative robust estimators becomes necessary. When outliers come from model misspecification, the problem of outliers may also be solved through an adequate modification on the model describing our data. Other types of robustness, for example to model misspecification, are not discussed here.

3. The Minimum DPD Estimators under the MCLR Models

The minimum density power divergence estimator (MDPDE), originally introduced by Basu et al. [34] for independent identically distributed (IID) data, has become extremely popular as a robust generalization of the classical MLE that produces desired robustness under possible data contamination without significant loss in efficiency under clean data. The MDPDE is obtained by minimizing a suitable estimate of a density-based statistical divergence, namely the DPD, between the observed data and the model density. In particular, if we observe IID data $X_1, \ldots, X_n$ from a population having true (unknown) density g, that we model by a family of parametric densities $f_{\boldsymbol{\theta}}$ with $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$, then the MDPDE of the unknown parameter $\boldsymbol{\theta}$ is to be obtained by minimizing an estimate of the DPD between g and $f_{\boldsymbol{\theta}}$, which is defined as [34]

$$d_\alpha(g, f_{\boldsymbol{\theta}}) = \int f_{\boldsymbol{\theta}}^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \int f_{\boldsymbol{\theta}}^\alpha g + \frac{1}{\alpha} \int g^{1+\alpha},$$

where $\alpha > 0$ is a robustness tuning parameter that controls the trade-off between robustness and efficiency. Since the third term in the form of the DPD does not depend on the parameter of interest ($\boldsymbol{\theta}$), estimating the second term using the empirical distribution function based on the observed data $X_1, \ldots, X_n$, we get the final MDPDE objective function to be minimized as given by

$$H_{n,\alpha}(\boldsymbol{\theta}) = \int f_{\boldsymbol{\theta}}^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) \frac{1}{n} \sum_{i=1}^n f_{\boldsymbol{\theta}}^\alpha(X_i) + \frac{1}{\alpha}. \tag{5}$$

As $\alpha \to 0$, the DPD measure coincides, in a limiting sense, to the Kullback–Leibler divergence and the associated minimizer (the MDPDE at $\alpha = 0$) is then nothing but the MLE, the most efficient but highly non-robust MLE. This can also be seen by noting that, as $\alpha \to 0$, $H_{n,\alpha}(\boldsymbol{\theta})$ converges to the negative log-likelihood (ensured by the addition of the constant $1/\alpha$). Thus, the MDPDEs provide increased robustness at a cost of slight loss in efficiency as α increases. Please refer to references [34,35] for more details and examples.

Due to various favorable properties of the MDPDE, it has now been extended to several important complex data structures and problem set-ups. In particular, the authors of reference [36] extended the definition and properties of the MDPDE for independent but non-homogeneous set-ups, which is later utilized to study the MDPDEs for many different regression models [36–38]. Castilla et al. [39,40] have used the same idea to study the MDPDE for MLR models based on data obtained using simple random sampling and complex survey sampling, respectively. In the present set-up of MCLR models also, the observed responses are independent but each follows a different multinomial distribution depending on the given (or observed) value of the circular predictor variables, and so we will define the MDPDE of the parameters of the MCLR model (1) in the line of references [36,39], which is formally presented in the following definition.

Definition 2. *Given the multinomial circular logistic regression model (1), the MDPDE of model parameters $\boldsymbol{\beta}$ with tuning parameter $\alpha > 0$ is defined as*

$$\widehat{\boldsymbol{\beta}}_{n,\alpha} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^{d(2k+1)}} H_{n,\alpha}(\boldsymbol{\beta}),$$

where we now have

$$H_{n,\alpha}(\boldsymbol{\beta}) = \frac{1}{n^{1+\alpha}} \sum_{i=1}^I \sum_{j=1}^{d+1} \left[n_i^{1+\alpha} \pi_{ij}(\boldsymbol{\beta})^{1+\alpha} - \left(1 + \frac{1}{\alpha}\right) v_{ij} n_i^\alpha \pi_{ij}(\boldsymbol{\beta})^\alpha + \frac{1}{\alpha} v_{ij}^{1+\alpha} \right]. \tag{6}$$

Note that, as $\alpha \to 0$, $H_{n,\alpha}(\boldsymbol{\beta})$ converges to the symmetric log-likelihood value defined in (2), plus a constant, so that $\widehat{\boldsymbol{\beta}}_{n,0}$ is nothing but the MLE defined in Definition 1. The

MDPDEs thus again provide a robust generalization of the MLE under the MCLR models with increased robustness under data contamination as α increases with only a little loss in efficiency under pure data; see Sections 4 and 5 for numerical illustrations justifying this claim. For the computation of the MDPDE under the MCLR models, we can either numerically minimize the objective function given in (6) or, alternatively, solve the associated estimating equations given in the following theorem.

Theorem 1. *Given the multinomial circular logistic regression model (1), the MDPDE $\widehat{\boldsymbol{\beta}}_{n,\alpha}$ of the parameter vector $\boldsymbol{\beta}$ with tuning parameter $\alpha \geq 0$, as defined in Definition 2, can be equivalently obtained by solving the following system of estimating equations:*

$$\sum_{i=1}^{I} n_i^{\alpha} \left[\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) diag^{\alpha-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))(\boldsymbol{\nu}_i - n_i \boldsymbol{\pi}_i(\boldsymbol{\beta})) \right] \otimes \boldsymbol{\omega}_i = \mathbf{0}_{d(2k+1)}, \tag{7}$$

where $\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) = (\boldsymbol{I}_d | \mathbf{0}_{d \times 1}) \boldsymbol{\Delta}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))$.

Proof. The proof follows by the standard differentiation of the objective function $H_{n,\alpha}$ given in (6) with respect to the parameter vector $\boldsymbol{\beta}$ and using Equation (3). Details are omitted for brevity. □

It may be noted that the estimating equations of the MDPDE given in (7) coincide exactly with the estimating equations of the MLE in (4) at $\alpha = 0$. So, the above theorem is a generalization of Proposition 1 covering both the cases of the existing MLE and our newly proposed MDPDEs under the MCLR models.

Next, we derive the asymptotic properties of the proposed MDPDE under the MCLR models using the general results from reference [36]. For this purpose, we will assume that Assumptions (A1)–(A7) of reference [36] hold true for our MCLR models as given in (1), which we will refer to as the Ghosh–Basu Conditions. Under these conditions, it can be seen that the MDPDE is asymptotically consistent as $n \to \infty$, having a nice asymptotic normal distribution as formally presented in the following theorem.

Theorem 2. *Suppose that the Ghosh–Basu Conditions hold true for the given multinomial circular logistic regression model (1) with the true value of parameter vector $\boldsymbol{\beta}$ being $\boldsymbol{\beta}^0$. Then, the asymptotic distribution of the MDPDE $\widehat{\boldsymbol{\beta}}_{n,\alpha}$ with tuning parameter $\alpha \geq 0$, as defined in Definition 2, is given by*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{n,\alpha} - \boldsymbol{\beta}^0) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_{d(2k+1)}, \boldsymbol{\Psi}_\alpha^{-1}\left(\boldsymbol{\beta}^0\right) \boldsymbol{\Omega}_\alpha\left(\boldsymbol{\beta}^0\right) \boldsymbol{\Psi}_\alpha^{-1}\left(\boldsymbol{\beta}^0\right) \right), \tag{8}$$

where $\boldsymbol{\Psi}_\alpha(\boldsymbol{\beta}) = \lim_{n \to \infty} \boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta})$ and $\boldsymbol{\Omega}_\alpha(\boldsymbol{\beta}) = \lim_{n \to \infty} \boldsymbol{\Omega}_{n,\alpha}(\boldsymbol{\beta})$ with

$$\boldsymbol{\Psi}_{n,\alpha}(\boldsymbol{\beta}) = \sum_{i=1}^{I} \left(\frac{n_i}{n} \right)^{1+\alpha} \widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) diag^{\alpha-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \widetilde{\boldsymbol{\Delta}}^T(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T,$$

$$\boldsymbol{\Omega}_{n,\alpha}(\boldsymbol{\beta}) = \sum_{i=1}^{I} \left(\frac{n_i}{n} \right)^{1+\alpha} \widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) diag^{\alpha-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \boldsymbol{\Delta}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) diag^{\alpha-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \widetilde{\boldsymbol{\Delta}}^T(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T.$$

Proof. The proof follows from the general result of reference [36] in a manner similar to the proof of Theorem 1 in reference [39]. Details are omitted for brevity. □

Remark 1. *After some algebraic manipulations detailed in Appendix A, it can be seen that $\boldsymbol{\Psi}_0(\boldsymbol{\beta}) = \boldsymbol{\Omega}_0(\boldsymbol{\beta}) = \boldsymbol{\Omega}(\boldsymbol{\beta})$, so that Proposition 2 can be recovered directly from Theorem 2 at $\alpha = 0$.*

Theorem 2 is thus applicable, again, for the whole class of MDPDE, including the MLE at $\alpha = 0$, and can be used to obtain the standard error of these estimates under the MCLR models. For this purpose, the asymptotic variance matrix of $\widehat{\boldsymbol{\beta}}_{n,\alpha}$ can be directly estimated

by $n^{-1}\mathbf{\Psi}_{n,\alpha}^{-1}\left(\widehat{\boldsymbol{\beta}}_{n,\alpha}\right)\mathbf{\Omega}_{n,\alpha}\left(\widehat{\boldsymbol{\beta}}_{n,\alpha}\right)\mathbf{\Psi}_{n,\alpha}^{-1}\left(\boldsymbol{\beta}^{0}\right)$, from which the standard errors of the estimates of each component parameter can be obtained by taking the square root of the corresponding diagonal entries of this estimated (asymptotic) variance matrix. This asymptotic variance estimates of the MDPDEs, along with the estimators themselves, can also be used to develop a robust test of statistical hypothesis under the MCLR models in a routine manner.

4. Simulation Studies

We consider a simulation scheme with $d + 1 = 3$ categories and one circular explanatory variable ($k = 1$). The true value of the parameters is taken to be $\boldsymbol{\beta} = (0, 2, 2, 0.2, 2.5, 1.5)^{T}$. To generate the data, we consider different numbers of covariate patterns (I), different sample sizes (n_i, $i = 1, \ldots, I$), and three different generating distributions: the uniform distribution, the vM distribution with mean $60°$ and concentration parameter $\kappa = 2$, and the SN distribution with mean $60°$ and concentration parameter $\kappa = 6$. These samples are generated with libraries `circular` and `Riemann` in R (see Appendix B for more details on the circular distributions used here).

The robustness of the proposed estimators is evaluated by introducing 10% outliers in the data. These are artificially classified in the first category, independent from the value of their explanatory variables.

For 500 replications, the vector of parameters $\boldsymbol{\beta}$ is estimated for different tuning parameters α, and the mean absolute error (MAE) of the estimated probabilities (obtained based on the estimated vector of parameters) is computed for each one of the following scenarios:

- Scenario 1: $I \in \{10, 20, 50\}$, $n_i \in \{1, \ldots, 20\}$ for $i = 1, \ldots, I$. Explanatory variables generated from uniform distribution (Figure 1).
- Scenario 2: $I \in \{20, \ldots, 100\}$, $n_i \in \{1, 5, 10\}$ for $i = 1, \ldots, I$. Explanatory variables generated from uniform distribution (Figure 2).
- Scenario 3: $I = 50$, $n_i \in \{1, \ldots, 20\}$ for $i = 1, \ldots, I$. Explanatory variables generated from vM distribution (top of Figure 3).
- Scenario 4: $I = 50$, $n_i \in \{1, \ldots, 20\}$ for $i = 1, \ldots, I$. Explanatory variables generated from SN distribution (bottom of Figure 3).
- Scenario 5: $I \in \{20, \ldots, 100\}$, $n_i = 5$ for $i = 1, \ldots, I$. Explanatory variables generated from vM distribution (top of Figure 4).
- Scenario 6: $I \in \{20, \ldots, 100\}$, $n_i = 5$ for $i = 1, \ldots, I$. Explanatory variables generated from SN distribution (bottom of Figure 4).
- Scenario 7: $I \in \{20, \ldots, 100\}$, $n_i = 10$ for $i = 1, \ldots, I$. Explanatory variables generated from vM distribution (top of Figure 5).
- Scenario 8: $I \in \{20, \ldots, 100\}$, $n_i = 10$ for $i = 1, \ldots, I$. Explanatory variables generated from SN distribution (bottom of Figure 5).

The resulting values of the MAEs are plotted in Figures 1–5 for each of the above simulation set-ups. An increment on the sample size increases the confidence levels of our estimates, as observed in Figures 1–5, where MAE decreases, generally, as I or n_i increases. It can be seen that MDPDEs generally outperform the MLE when data are contaminated for all the generating distributions and the sample sizes considered. In particular, when vM and SN distributions generate the explanatory variables, the MDPDEs even outperform the MLE when no contamination is introduced in our simulated data.

When data are generated under the uniform distribution, an increment on the tuning parameter α increases the robustness but also decreases the efficiency when pure data are considered. A high value of α may also lead to greater errors when data are generated under vM and SN distributions, both for pure and contaminated data. If we suspect our data are not highly contaminated, a value of $\alpha \in (0, 0.4]$ may present a good trade-off between efficiency and robustness. If we detect important outliers in our data, a higher value of α may be preferable. In Section 5, we present some numerical examples which confirm this idea.
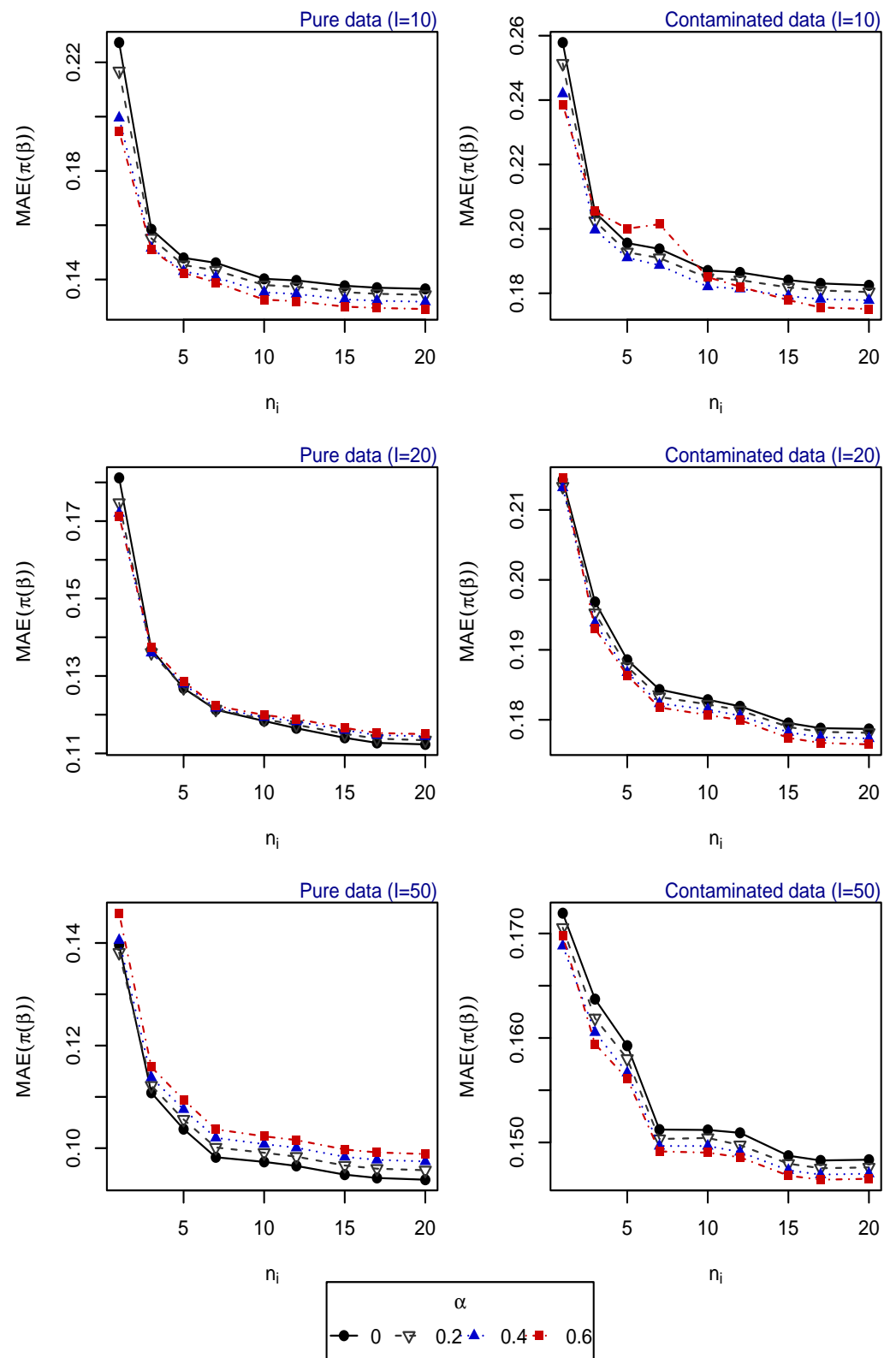
Figure 1. MAE of estimated probabilities when data are generated using uniform distribution for different values of I and n_i, $i = 1, \ldots, I$.
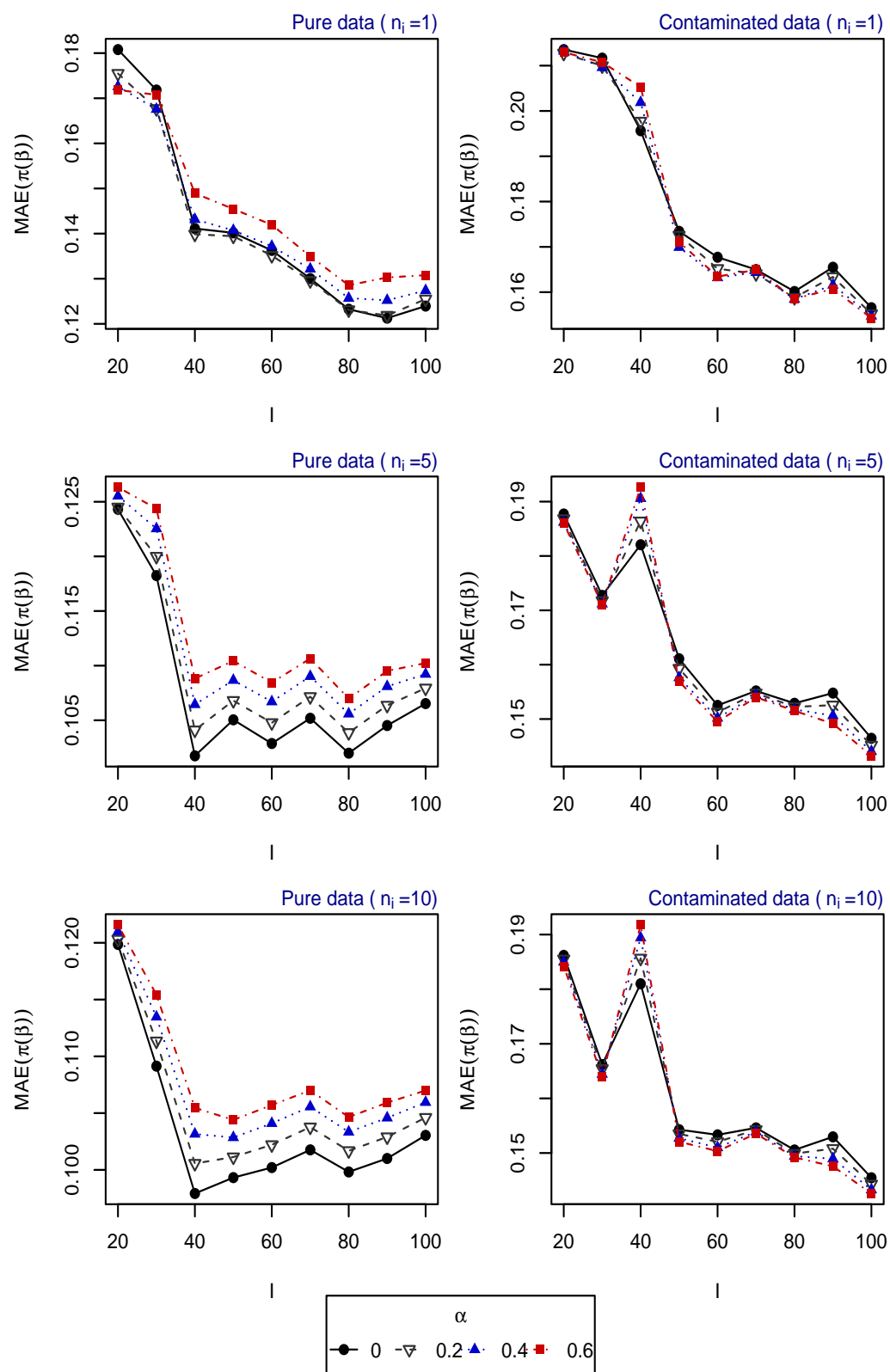
Figure 2. MAE of estimated probabilities when data are generated using uniform distribution for different values of I and n_i, $i = 1, \ldots, I$.
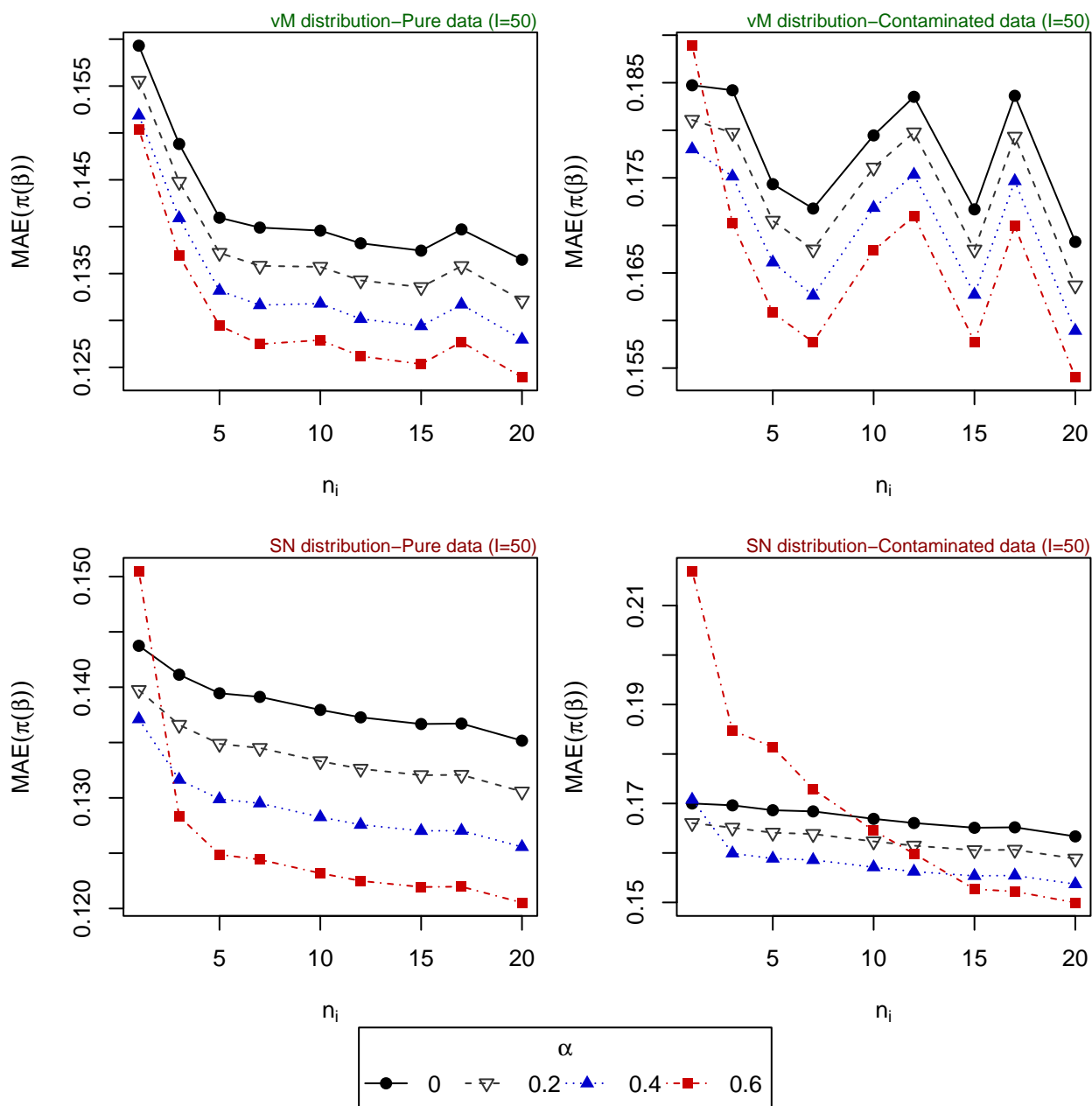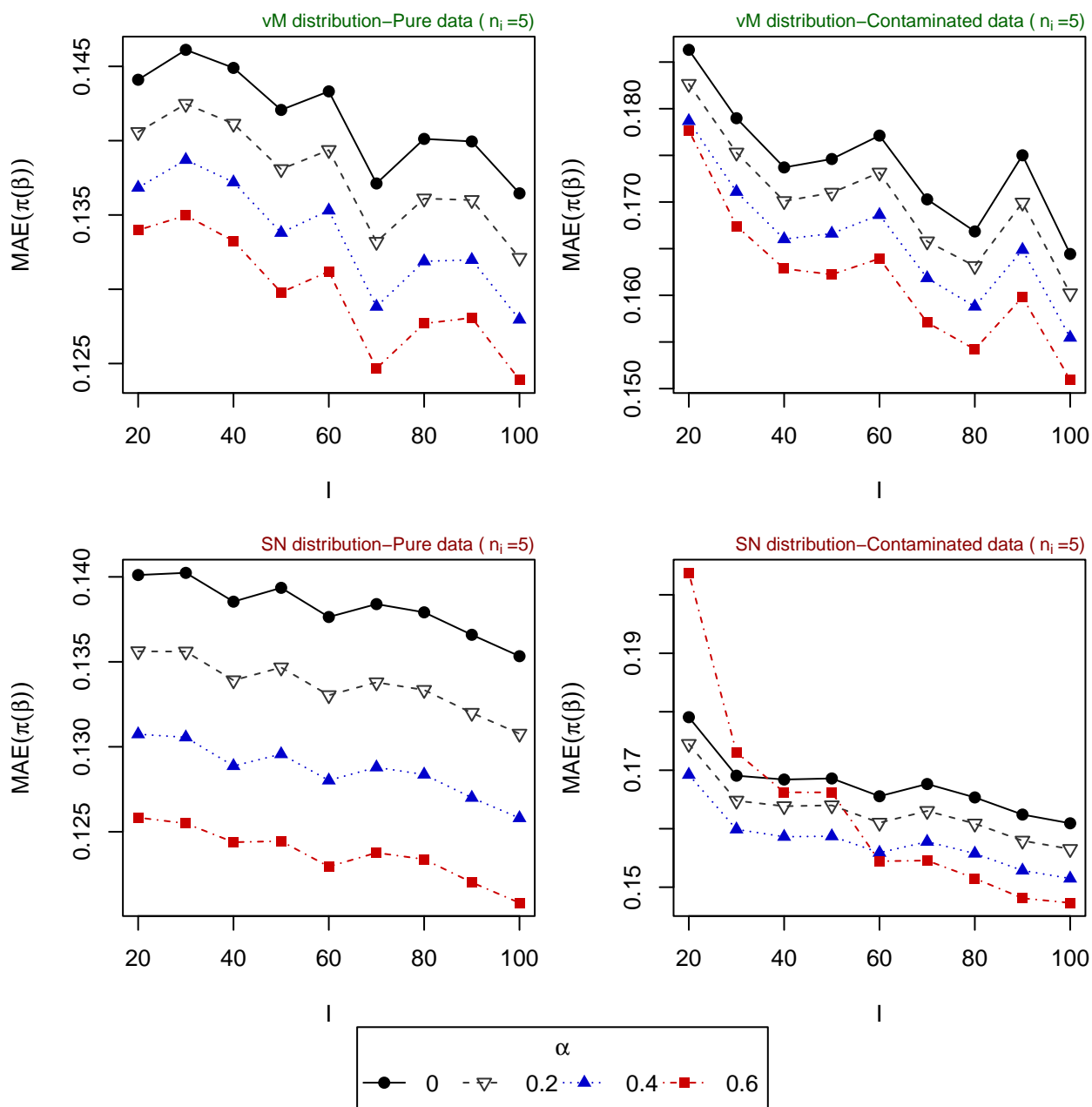
Figure 3. MAE of estimated probabilities when data are generated using vM distribution (above) or SN distribution (below) for $I = 50$ and different values of n_i, $i = 1, \ldots, I$.

Figure 4. MAE of estimated probabilities when data are generated using vM distribution (above) or SN distribution (below) for $n_i = 10$, $i = 5, \ldots, I$, and different values of I.
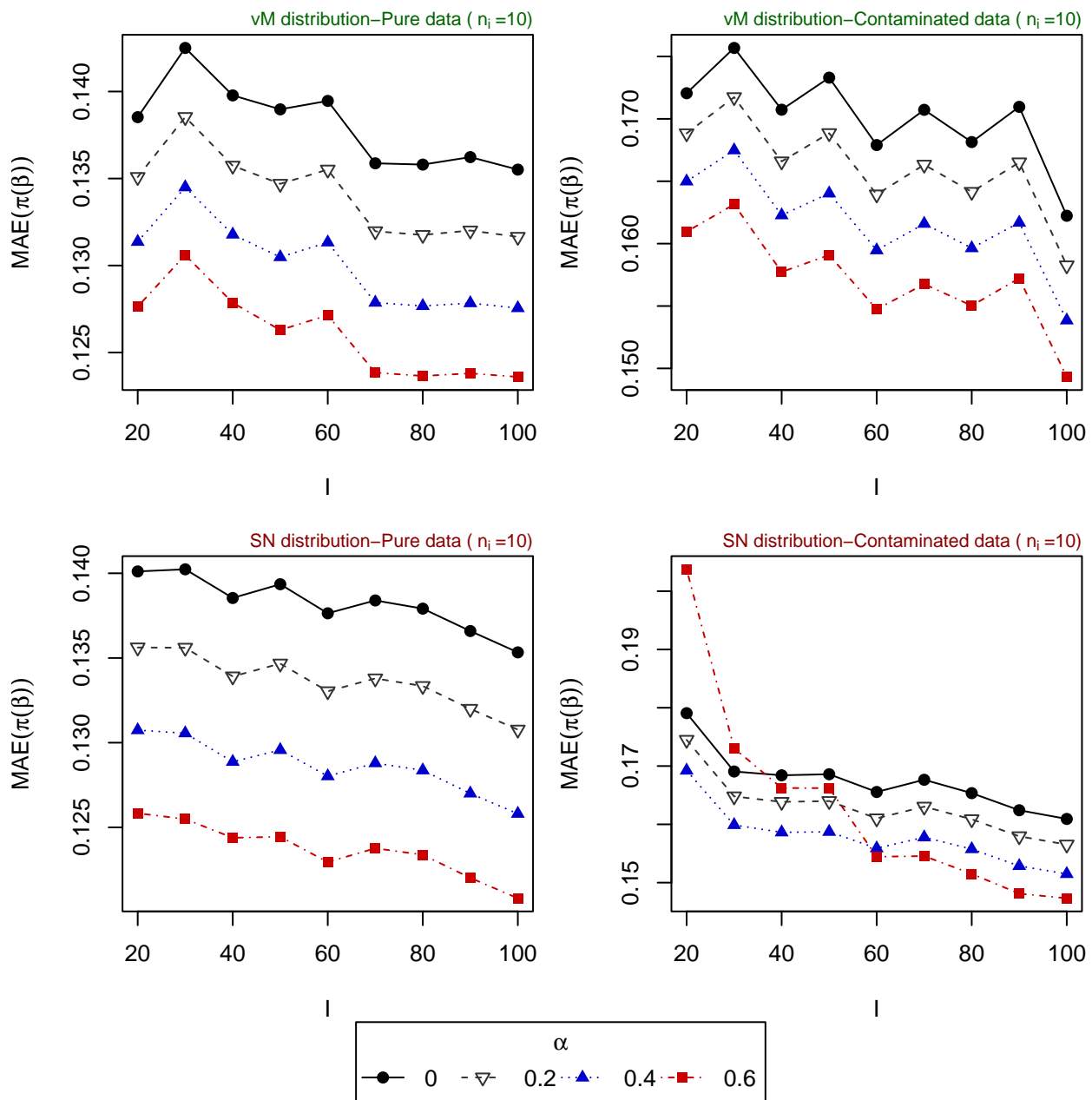
Figure 5. MAE of estimated probabilities when data are generated using vM distribution (above) or SN distribution (below) for $n_i = 10$, $i = 1, \ldots, I$, and different values of I.

5. Applications to Real Data

5.1. Application to Forest Science

For this application, we use the leaf inclination dataset recorded in reference [41] and analyzed in reference [8,9]. These data contain the leaf inclination angles of 138 plant species. With only this circular variable and applying the circular logistic regression model presented here, we want to classify these species of plants. In particular, we focus on the following examples that have a binary response variable. For each one, we compute the accuracy of the model under different tuning parameters. Contamination in the data is also induced in order to see the effect on our model.

5.1.1. First Example [*Alnus incana* vs. *Alnus glutinosa*]

These data contain 160 observations of the leaf inclination of *Alnus incana* and *Alnus glutinosa* (80 of each type), which are illustrated in the top left of Figure 6. The circular logistic model with only this covariate has more than 76% of accuracy for the MLE, which is increased by 2% for the proposed MDPDEs (see Table 1). Further, we artificially introduce 16 extreme samples (a 10% of total observations) of *Alnus incana* with null inclination angle, as can be seen in Figure 6; then, the accuracy of all estimations decays but our MDPDEs with $\alpha > 0$ provide significantly higher accuracy than the MLE (Table 1), illustrating their claimed robustness.
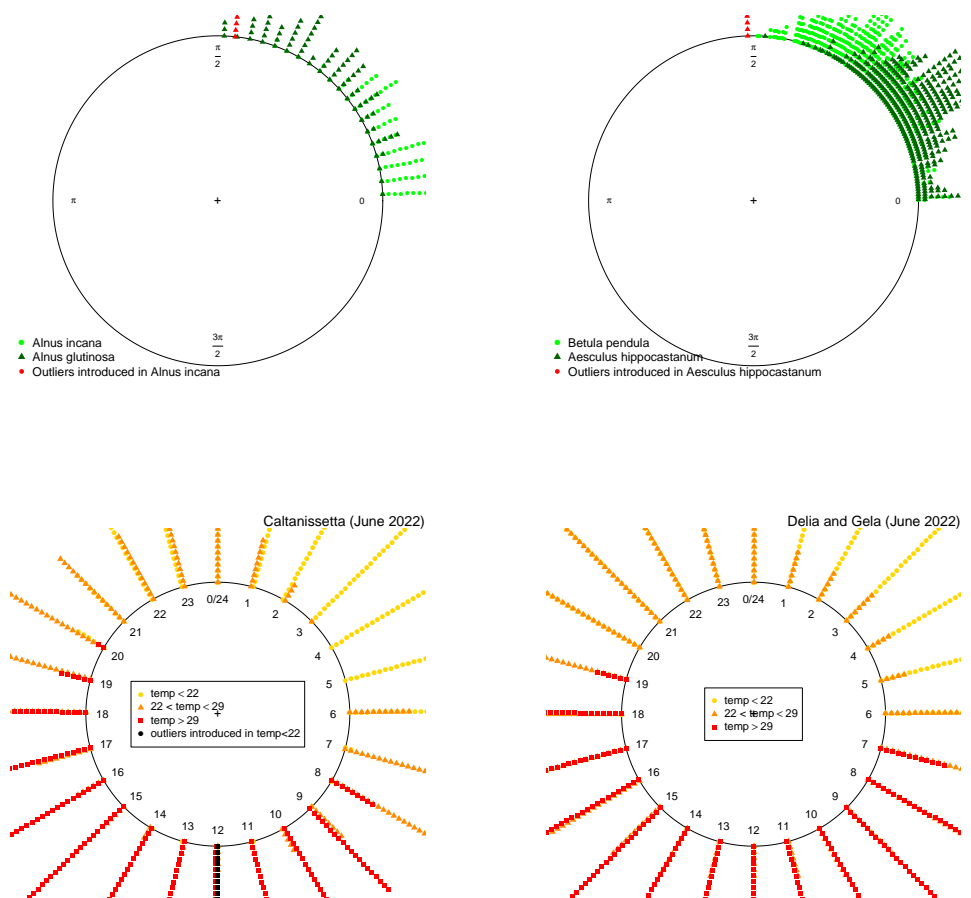

Figure 6. Above: Application to forest science (leaf inclinations). **Below**: Application to meteorological science (hours) for original (**left**) and new data (**right**).

Table 1. Accuracy of the multinomial circular regression model for forest and meteorological data.

Dataset→	First Forest Example		Second Forest Example		Meteorological Example		
$\alpha \downarrow$	Original	Contaminated	Original	Contaminated	Original	Contaminated	New
0 (MLE)	0.7625	0.7486	0.7339	0.6705	0.7135	0.6120	0.7208
0.2	0.7812	0.7486	0.7350	0.6727	0.7135	0.6218	0.7208
0.4	0.7812	0.7600	0.7350	0.7223	0.7135	0.6387	0.7208
0.6	0.7812	0.7771	0.7350	0.7307	0.7135	0.6387	0.7208
0.8	0.7812	0.7771	0.7350	0.7381	0.7135	0.6639	0.7208

5.1.2. Second Example [*Betula pendula* vs. *Aesculus hippocastanum*]

These data, presented in the top right of Figure 6, contain 902 observations of the species *Betula pendula* and *Aesculus hippocastanum*, and were also analyzed in references [8,9]. The circular logistic model with only this covariate has more than 73% of the accuracy with the MLE, which is increased again for the MDPDEs with $\alpha > 0$ (see Table 1). When we manually introduced 45 extreme observations (5% of total observations) of *Aesculus hippocastanum* with null inclination angle, the higher robustness of the proposed MDPDEs are, again, clearly evident from the results presented in Table 1.

5.2. Application to Meteorological Science

We now apply the MCLR model to a meteorological dataset, obtained from the "Portale Open Data della Regione Siciliana" (https://dati.regione.sicilia.it/) (accessed on 1 April 2023) containing the temperature of wind at two meters of height in June 2022 in the region of Caltanissetta, a commune capital of the Province of Caltanissetta (Sicilia, Italy), during the whole day with a total of 684 observations. Our response variable is the temperature associated to the hour, divided into three categories: (i) lower than 22°, (ii) between 22° and 29° and (iii) greater than 29°. Then, we have balanced data with approximately one third of the observations in each group (See Figure 6). The results of the accuracy of the model are presented in Table 1. As observed, more than 70% of the cases are correctly classified, without a significant difference between estimates. Let us now introduce some outliers to our data. In particular, let us introduce 5% of the low temperatures at 12 noon. The accuracy of MLE and minimum DPD with low values of the tuning parameter decreases significantly, but remains stable for higher values of the tuning parameter.

Finally, we also consider the wind temperatures in the same dates of two other communes in Caltanissetta: Delia and Gela (see Figure 6). If we predict the wind temperature of these 1368 new observations with the model estimated for the first region, we obtain more than 72% accuracy.

6. Conclusions

In this paper, we have studied a class of robust minimum divergence estimators, based on the popular density power divergence measure, for the multinational circular logistic regression models, which is extremely important for the circular data frequently occurring in ecological and environmental sciences, among other domains. We have defined the minimum DPD estimators for the multinomial circular logistic regression models as a robust generalization of the classical MLE and derived its asymptotic distribution. The improved performances of the proposed MDPDE under the multinational circular logistic regression models are illustrated numerically with extensive simulation studies and important real-life applications. Our proposed estimators present an increase of up to 6% in accuracy compared to MLE when artificial outliers are introduced in our data sets.

We have also discussed the estimation of the asymptotic variance of the MDPDEs, and hence their standard errors, which can further be used to conduct statistical hypothesis testing robustly under the multinational circular logistic regression models. Along with

further study on such hypothesis testing problems, we would also like to extend the MDPDE further for more complex important model set-ups involving circular and general directional data in our future works.

Abbreviations

The following abbreviations are used in this manuscript:

DPD	Density Power Divergence
IID	Independent Identically Distributed
MAE	Mean Absolute Error
MCLR	Multinomial Circular Logistic Regression
MDPDE	Minimum Density Power Divergence Estimator
MLE	Maximum Likelihood Estimator
MLR	Multinomial Logistic Regression
vM	Von Mises (distribution)
SN	Spherical Normal (distribution)

Appendix A. Proof of Remark 1

For each $i = 1, \ldots, I$, we have

$$\widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\widetilde{\boldsymbol{\Delta}}^T(\boldsymbol{\pi}_i(\boldsymbol{\beta}))$$
$$= (\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})\boldsymbol{\Delta}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\boldsymbol{\Delta}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))(\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T$$
$$= (\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})[\mathrm{diag}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) - \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})]\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))$$
$$\times [\mathrm{diag}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) - \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})](\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T$$
$$= (\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})[\boldsymbol{I}_{d+1} - \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))]$$
$$\times [\mathrm{diag}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) - \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})](\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T$$
$$= (\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})[\boldsymbol{I}_{d+1} - 2\boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta}) + \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})](\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T$$
$$= (\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})[\mathrm{diag}(\boldsymbol{\pi}_i(\boldsymbol{\beta})) - \boldsymbol{\pi}_i(\boldsymbol{\beta})\boldsymbol{\pi}_i^T(\boldsymbol{\beta})](\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T = \widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))(\boldsymbol{I}_d|\boldsymbol{0}_{d\times 1})^T = \boldsymbol{\Delta}(\boldsymbol{\pi}_i^*(\boldsymbol{\beta})).$$

Therefore,

$$\boldsymbol{\Psi}_0(\boldsymbol{\beta}) = \lim_{n\to\infty} \sum_{i=1}^{I} \left(\frac{n_i}{n}\right) \widetilde{\boldsymbol{\Delta}}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\mathrm{diag}^{-1}(\boldsymbol{\pi}_i(\boldsymbol{\beta}))\widetilde{\boldsymbol{\Delta}}^T(\boldsymbol{\pi}_i(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T$$

$$= \lim_{n\to\infty} \sum_{i=1}^{I} \left(\frac{n_i}{n}\right) \boldsymbol{\Delta}(\boldsymbol{\pi}_i^*(\boldsymbol{\beta})) \otimes \boldsymbol{\omega}_i \boldsymbol{\omega}_i^T = \boldsymbol{\Omega}(\boldsymbol{\beta}).$$

In a similar manner, $\boldsymbol{\Omega}_0(\boldsymbol{\beta}) = \boldsymbol{\Omega}(\boldsymbol{\beta})$.

Appendix B. Details on Circular Distributions

Let us consider data distribution $x \in \mathbb{S}^1$ (i.e., $x \in \mathbb{R}^2$ and $||x||_2 = 1$). The unifrom distribution assigns equal probability to all points of the circle, with probability density function

$$f_{\text{Uniform}_{\mathbb{S}^1}}(x) = \frac{1}{2\pi}.$$

On the other hand, for the unit circle, the vM distribution with mean $\mu \in \mathbb{S}^1$ and concentration $\kappa \in \mathbb{R}^+$ has probability density function

$$f_{\text{vM}}(x; \mu, \kappa) = \frac{1}{2\pi \mathcal{I}_0(\kappa)} \exp(\mu^T x),$$

where $\mathcal{I}_0(\kappa)$ is the modified Bessel function of the first kind of order 0. The vM distribution is a particular case of the vM–Fisher distribution on the *n*-sphere [20].

Finally, the SN distribution [24] with mean $\mu \in \mathbb{S}^1$ and concentration $\kappa \in \mathbb{R}^+$ is defined in the unit circle by

$$f_{\text{SN}}(x; \mu, \kappa) = \frac{1}{\mathcal{Z}(\kappa)} \exp\left(-\frac{\kappa}{2} d^2(x, \mu)\right),$$

where $d^2(\cdot, \cdot)$ is the geodesic distance and $\mathcal{Z}(\kappa)$ is a normalizing parameter

$$\mathcal{Z}(\kappa) = \int_{\mathbb{S}^1} \exp\left(-\frac{\kappa}{2} d^2(x, \mu)\right) dx.$$

Note that, for both vM and SN distribution, the mean μ and concentration κ are measures of direction and concentration, respectively. For small κ both distributions are close to the uniform and, in particular,

$$\text{vM}_{\mathbb{S}^1}(\mu, \kappa = 0) = \text{SN}_{\mathbb{S}^1}(\mu, \kappa = 0) = \text{Uniform}_{\mathbb{S}^1}.$$

For more details on these distributions, one may refer to reference [25].

References

1. Pewsey, A.; García-Portugués, E. Recent advances in directional statistics. *Test* **2021**, *30*, 1–58.
2. Ożarowska, A.; Ilieva, M.; Zehtindjiev, P.; Akesson, V.; Muś, K. A new approach to evaluate multimodal orientation behaviour of migratory passerine birds recorded in circular orientation cages. *J. Exp. Biol.* **2013**, *216*, 4038–4046. [CrossRef] [PubMed]
3. Akesson, S.; Klaassen, R.; Holmgren, J.; Fox, J.W.; Hedenstrom, A. Migration routes and strategies in a highly aerial migrant, the common swift Apus apus, revealed by light-level geolocators. *PLoS ONE* **2012**, *7*, e41195. [CrossRef] [PubMed]
4. Guilford, T.; Freeman, R.; Boyle, D.; Dean, B.; Kirk, H.; Phillips, R.; Perrins, C. A dispersive migration in the Atlantic puffin and its implications for migratory navigation. *PLoS ONE* **2011**, *6*, e21336. [CrossRef]
5. Kilic, M.B.; Kalaylioglu, Z.; Sengupta, A. A flexible Bayesian mixture approach for multi-modal circular data. *Hacet. J. Math. Stat.* **2022**, *51*, 1160–1173. [CrossRef]
6. Ramkilawon, G.D.; Ferreira, J.; Nakhaeirad, N. Sine-skewed von Mises-and Lindley/Gumbel models as candidates for direction and distance in modelling animal movement. *Braz. J. Biom.* **2023**, *41*, 175–190. [CrossRef]
7. Gustafson, C.L.; Partch, C.L. Emerging models for the molecular basis of mammalian circadian timing. *Biochemistry* **2014**, *54*, 134–149. [CrossRef]
8. Alshqaq, S.S.; Ahmadini, A.A.; Abuzaid, A.H. Some new robust estimators for circular logistic regression model with applications on meteorological and ecological data. *Math. Probl. Eng.* **2021**, *2021*, 9944363. [CrossRef]
9. Castilla, E. Robust circular logistic regression model and its application to life and social sciences. *Rev. Colomb. Estad.* **2023**, *46*, 45–62.
10. Agostinelli; C. Robust estimation for circular data. *Comput. Stat. Data Anal.* **2007**, *51*, 5867–5875. [CrossRef]
11. Gaumond M.; Réthoré, P.E.; Ott, S.; Peña, A.; Bechmann, A.; Hansen, K.S. Evaluation of the wind direction uncertainty and its impact on wake modeling at the Horns Rev offshore wind farm. *Wind Energy* **2014**, *17*, 1169–1178. [CrossRef]
12. SenGupta, A.; Ugwuowo, F.I. Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environ. Ecol. Stat.* **2006**, *13*, 299–309. [CrossRef]
13. Jones, M.; Pewsey, A. Inverse Batschelet distributions for circular data. *Biometrics* **2012**, *68*, 183–193. [CrossRef]
14. Archibald, A.M.; Bogdanov, S.; Patruno, A.; Hessels, J.W.T.; Deller, A.T.; Bassa, C.; Janssen, G.H.; Kaspi, V.M.; Lyne, A.G.; Stappers, B.W.; et al. Accretion-powered pulsations in an apparently quiescent neutron star binary. *Astrophys. J.* **2015**, *807*, 62. [CrossRef]

15. Gill, J.; Hangartner, D. Circular data in political science and how to handle it. *Political Anal.* **2010**, *18*, 316–336. [CrossRef]
16. Kibiak, T.; Jonas, C. Applying circular statistics to the analysis of monitoring data. *Eur. J. Psychol. Assess.* **2007**, *23*, 227–237. [CrossRef]
17. Lourenço, N.; Rua, A. Business cycle clocks: Time to get circular. *Empir. Econ.* **2023**, *65*, 1513–1541. [CrossRef]
18. Ashby, M.P. Studying crime and place with the crime open database: Social and behavioural scienes. *Res. Data J. Humanit. Social Sci.* **2019**, *4*, 65–80. [CrossRef]
19. von Mises, R. Uber Die Ganzzahligkeit der Atomgewichte und verwandte Fragen. *Phys. Z.* **1918**, *19*, 490–500.
20. Khatri, C.G.; Mardia, K.V. The von Mises-Fisher matrix distribution in orientation statistics. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 95–106. [CrossRef]
21. Banerjee, A.; Dhillon, I.S.; Ghosh, J.; Sra, S.; Ridgeway, G. Clustering on the unit hypersphere using von Mises-Fisher distributions. *J. Mach. Learn. Res.* **2005**, *6*, 1345–1382.
22. Bangert, M.; Hennig, P.; Oelfke, U. Using an infinite von mises-fisher mixture model to cluster treatment beam directions in external radiation therapy. In Proceedings of the 2010 Ninth International Conference on Machine Learning and Applications, Washington, DC, USA, 12–14 December 2010; pp. 746–751.
23. Hauberg, S. Directional statistics with the spherical normal distribution. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 704–711.
24. You, K.; Suh, C. Parameter estimation and model-based clustering with spherical normal distribution on the unit hypersphere. *Comput. Stat. Data Anal.* **2022**, *171*, 107457. [CrossRef]
25. Castilla, E. Robust estimation of the spherical normal distribution. *Math. Appl.* **2022**, *50*, 43–63. [CrossRef]
26. Polsen, O.; Taylor, C.C. Parametric circular-circular regression and diagnostic analysis. In *Geometry Driven Statistics*; Wiley Series in Probability and Statistics; Dryden, I.L., Kent, J.T., Eds.; Wiley: Chichester, UK, 2015; pp. 115–128.
27. Presnell, B.; Morrison, S.P.; Littell, R.C. Projected multivariate linear models for directional data. *J. Am. Stat. Assoc.* **1998**, *93*, 1068–1077. [CrossRef]
28. Johnson, R.A.; Wehrly, T.E. Some angular-linear distributions and related regression models. *J. Am. Stat. Assoc.* **1978**, *73*, 602–606. [CrossRef]
29. Al-Daffaie, K.; Khan, S. Logistic regression for circular data. *AIP Conf. Proc.* **2017**, *1842*, 030022.
30. Uemura, M.; Meglic, A.; Zalucki, M.P.; Battisti, A.; Belusic, G. Spatial orientation of social caterpillars is influenced by polarized light. *Biol. Lett.* **2021**, *17*, 20200736. [CrossRef]
31. Wolpert, N.; Tallon-Baudry, C. Coupling between the phase of a neural oscillation or bodily rhythm with behavior: Evaluation of different statistical procedures. *NeuroImage* **2021**, *236*, 118050. [CrossRef]
32. Castilla, E.; Chocano, P.J. A new robust approach for multinomial logistic regression with complex design model. *IEEE Trans. Inf. Theory* **2022**, *68*, 7379–7395. [CrossRef]
33. Abuzaid, A.H.; Ahmed, H.I.E. On outliers detection in circular logistic regression. *J. Appl. Probab. Stat.* **2021**, *16*, 95–110.
34. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559. [CrossRef]
35. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2011.
36. Ghosh, A.; Basu, A. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **2013**, *7*, 2420–2456. [CrossRef]
37. Ghosh, A.; Basu, A. Robust Estimation for Non-Homogeneous Data and the Selection of the Optimal Tuning Parameter: The DPD Approach. *J. Appl. Stat.* **2015**, *42*, 2056–2072. [CrossRef]
38. Ghosh, A.; Majumdar, S. Ultrahigh-dimensional Robust and Efficient Sparse Regression using Non-Concave Penalized Density Power Divergence. *IEEE Trans. Inf. Theory* **2020**, *66*, 7812–7827. [CrossRef]
39. Castilla, E.; Ghosh, A.; Martin, N.; Pardo, L. New robust statistical procedures for the polytomous logistic regression models. *Biometrics* **2018**, *74*, 1282–1291. [CrossRef] [PubMed]
40. Castilla, E.; Ghosh, A.; Martin, N.; Pardo, L. Robust semiparametric inference for polytomous logistic regression with complex survey design. *Adv. Data Anal. Classif.* **2021**, *15*, 701–734. [CrossRef]
41. Chianucci, F.; Pisek, J.; Raabe, K.; Marchino, L.; Ferrara, C.; Corona, P. A dataset of leaf inclination angles for temperate and boreal broadleaf woody species. *Ann. For. Sci.* **2018**, *75*, 50. [CrossRef]