Journal of the Korean Statistical Society (2025) 54:314–341
https://doi.org/10.1007/s42952-024-00296-x

Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

A refreshing take on the inverted Dirichlet via a mode parameterization with some statistical illustrations

A. F. Otto[1] · **J. T. Ferreira**[1] · **A. Bekker**[1,2] · **A. Punzo**[3] · **S. D. Tomarchio**[3]

Received: 7 December 2023 / Accepted: 8 November 2024 / Published online: 23 November 2024

Abstract

The inverted Dirichlet (IDir) distribution is a popular choice for modeling multivariate data with positive support; however, its conventional parameterization can be challenging to interpret. In this paper, we propose a refreshing take on the IDir distribution through a convenient mode-based parameterization, resulting in the mode-reparameterized IDir (mIDir). This new parameterization aims to enhance the use of the IDir in various contexts. We provide relevant statistical illustrations in robust and nonparametric statistics, model-based clustering, and semiparametric density estimation, all benefiting from this novel perspective on the IDir for computation and implementation. First, we define finite mIDir mixtures for clustering and semiparametric density estimation. Secondly, we introduce a smoother based on mIDir kernels, which, by design, avoids allocating probability mass to unrealistic negative values, thereby addressing the boundary bias issue. Thirdly, we introduce a heavy-tailed generalization of the mIDir distribution, referred to as the contaminated mIDir (cmIDir), which effectively handles and detects mild outliers, making it suitable for robust statistics. Maximum likelihood estimates of the parameters for the parametric models are obtained using a developed EM algorithm as well as direct numerical optimization. A parameter recovery analysis demonstrates the successful application of the estimation method, while a sensitivity analysis examines the impact of mild outliers on both the mIDir and cmIDir models. The flexibility and advantages of the proposed mIDir-based models are showcased through several real data analyses and illustrations.

Keywords Kernel smoother · Mode · Mixture · Multivariate · Positive support

✉ A. F. Otto
arno.otto@up.ac.za

[1] Department of Statistics, University of Pretoria, Pretoria, South Africa

[2] Department of Geography, Geoinformatics and Meteorology, Centre for Environmental Studies, University of Pretoira, Pretoria, South Africa

[3] Department of Economics and Business, University of Catania, Catania, Italy

1 Introduction

It is argued by Kneib et al. (2021) that “anyone seriously conducting empirical data analysis will agree that statistics offers more than means.” Examples of the limitation of the mean and median as measures of location are comprehensively described by Chacón (2020). This limitation arises from the behaviour of real data scenarios, where symmetric distributions are inadequate for modelling datasets with vastly different underlying characteristics. In particular, when dealing with large datasets, it is essential to know where the bulk of the data lies. Consequently, there has been significant interest in the study of flexible and asymmetric models during the last three decades - here we briefly mention the work of Roberts (1966), Azzalini (1985, 2013), Adcock and Azzalini (2020), Genton (2004), Fang et al. (2022), and Tomarchio et al. (2023, 2024). The use of parametric models that are (i) positive, (ii) skewed, and (iii) multimodal have been well documented in the literature (Atkinson & Bourguignon, 2014; Vernic, 2006).

The mode is widely acknowledged to be a more “natural" measure of location than the mean or the median in the above-described circumstances. For that reason, there is interest in conducting research on mode statistics (Dalenius, 1965; Fukunaga & Hostetler, 1975; Cheng, 1995; Lee, 1989; Yao & Li, 2014; Chacón, 2019; Sando & Hino, 2020; Arias-Castro & Qiao, 2023; Yao et al., 2023). Moreover, the mode is easy to locate graphically and makes intuitive sense, as opposed to the mean, which may be far from the mode in skew distributions (Nolan, 1998). Several authors built flexible models by focusing on mode-parameterized versions of existing candidates, such as Punzo et al. (2018), Punzo (2019), Tomarchio and Punzo (2020), Bagnato and Punzo (2013), Liu et al. (2022), and Tomarchio et al. (2024). Another advantage of mode parameterization is the practical interpretation of the parameters. In a world where flexible models are more frequently needed for describing the complexities of data (Adcock & Azzalini, 2020), overparameterization and a loss of interpretable parameters are a big concern: meaningful and interpretable parameters are thus of important interest.

In addition to the above, the practitioner is challenged with modelling data with an inherent dependence structure in the sample; inferential statistics based on univariate distributions for such data might reflect inefficient estimates of the parameters and unreliable errors. This necessitates a multivariate model and specification of such models for modelling datasets with increased levels of heterogeneity is not a trivial exercise. Adding to the challenge is the issue of fitting distributions with real ($\mathbb{R}$) support when the data is bounded on the positive real space, denoted as $\mathbb{R}_+$, as this causes boundary bias (Punzo et al., 2018). Furthermore, nonparametric methods, usually employed for density (and hence modes) estimation, are known to be vulnerable to the presence of outliers, as much of the probability mass is led to flow to the tails of the density, possibly giving rise to the birth of spurious modes. This motivates the implementation of a parametric model based on a mode representation.

In this paper, we focus on the inverted Dirichlet (IDir) distribution—also known as the Dirichlet type II or multivariate inverted beta distribution—and we

introduce a mode-parametrization for it. The IDir distribution, which is a multivariate generalization of the beta prime distribution (Dubey, 1970), as described by Tiao and Cuttman (1965), is a positively skewed, unimodal distribution, which makes it a good choice for modelling data that exhibits non-Gaussian behaviour. The IDir has been successfully used in the calculation of confidence regions for variance ratios of random models for balanced data (Sahai & Anderson, 1973), as a conjugate prior for the negative multinomial distribution, and the modelling and classifying of positive data (Bdiri & Bouguila, 2012, 2013; Guo et al., 2023). It is essentially a special case of the multivariate F- and generalized multivariate Lomax (Pareto Type II) distribution (Lai & Balakrishnan, 2009), which has applications in reliability theory (Nayak, 1987) and modelling lifetime distributions in clinical trials (Lee & Gross, 1991). Other generalizations include the inverted beta Liouville distribution (Fang et al., 2018) and the generalized inverted Dirichlet distributions (Lingmnwah, 1976), which has significantly more number of parameters than the IDir. To the best of the authors' knowledge, the IDir has not been previously considered with this adopted mode parameterization. In this paper, we demonstrate how the mode parameterization of the IDir (denoted by mIDir) simplifies and motivates its use in various fields of statistics, namely in model-based clustering, nonparametric, and robust statistics.

The first considered use here of the mIDir distribution is in model-based clustering and semiparametric density estimation. To understand the significance of having mixture components parameterized in terms of the mode, consider that mixtures have often been used for their ability to model multimodal distributions, where each mode represents a distinct entity (Carreira-Perpinan, 2000). The underlying idea is that modes can be associated with important structures in an empirical distribution. Consequently, we propose and discuss finite mixtures of mIDir distribution for the intuitive use of clustering or classification of multivariate data with positive support.

Secondly, we propose a mIDir kernel smoother specifically tailored for nonparametric density estimation of positive data, as this popular statistical tool is essential in nonparametric statistics. This novel smoother addresses the issue of boundary bias that arises when smoothing occurs near the data boundaries. Unlike kernels with real support, which may allocate weight outside the theoretical range to unrealistic negative values, this mIDir kernel smoother provides a more accurate representation of the underlying distribution.

Lastly, we introduce the contaminated IDir (cmIDir) distribution, a heavy-tailed generalization of the mIDir distribution. The cmIDir distribution offers greater flexibility in tail behaviour, especially in the presence of atypical observations. In addition to the parameters of the mIDir distribution, the cmIDir distribution incorporates two additional parameters: δ controlling the proportion of atypical observations, and η specifying the degree of contamination. The cmIDir distribution is advantageous over the mIDir distribution in that it automatically down-weights atypical observations in the estimation of the mIDir parameters, making it a more robust method. Furthermore, once the cmIDir is fitted to a dataset, atypical observations ("mild" outliers) can be easily identified via maximum *a-posteriori* probabilities.

The paper is set out as follows: our proposed and adopted mode parameterization of the IDir is given in Sect. 2, while its simplified use in the various branches of

statistics is illustrated in Sect. 3. Finite mixtures consisting of the mIDir are defined in Sect. 3.1, along with the accompanying EM algorithm. Section 3.2 explores a kernel smoothing estimator that consists of the mIDir distribution, while the cmIDir is introduced in Sect. 3.3 to protect the mIDir against mild outliers. Accurate parameter recovery via a simulation study is presented in Sect. 4. The proposed models are applied to cantaloupe spectra data, and the physical activity of athletes in the Australian Institute of Sport dataset in Sect. 5. Final thoughts, key aspects, some limitations, and possible future extensions are discussed in Sect. 6.

2 Reparameterized inverted Dirichlet

2.1 The (conventional) inverted Dirichlet distribution

A random vector $\boldsymbol{X} = (X_1, \ldots, X_p) \in \mathbb{R}_+^p$ is said to have a p-dimensional IDir distribution with parameter vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p; \alpha_{p+1})$, $\alpha_i > 0$, $i = 1, \ldots, p+1$, if the probability density function (PDF) is given by

$$f(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{1}{B_{p+1}(\boldsymbol{\alpha})} \prod_{i=1}^{p} x_i^{\alpha_i - 1} \left(1 + \sum_{i=1}^{p} x_i\right)^{-\alpha_+}, \tag{1}$$

where $x_i \in \mathbb{R}_+$, $i = 1, \ldots, p$, $\alpha_+ = \sum_{i=1}^{p+1} \alpha_i$, and $B_{p+1}(\boldsymbol{\alpha}) = \dfrac{\prod_{i=1}^{p+1} \Gamma(\alpha_i)}{\Gamma(\alpha_+)}$, denoted as $\boldsymbol{X} \sim IDir_p(\boldsymbol{\alpha})$ (Ng et al., 2011). Here, $\Gamma(\cdot)$ denotes the gamma function. The PDF in (1) has mode

$$\frac{\alpha_i - 1}{p + \alpha_{p+1}}, \quad i = 1, \ldots, p,$$

if $\alpha_i > 1$, while the mode does not exist otherwise. An expression for the mixed moments is given by

$$\mathrm{E}\left(\prod_{i=1}^{p} X_i^{h_i}\right) = \frac{\Gamma\left(\alpha_{p+1} - h_+\right)}{\Gamma\left(\alpha_{p+1}\right)} \prod_{i=1}^{p} \frac{\Gamma\left(\alpha_i + h_i\right)}{\Gamma\left(\alpha_i\right)}, \tag{2}$$

if $\alpha_{p+1} > h_+ = \sum_{i=1}^{p} h_i$. See Fang et al. (2018) for other attractive statistical properties of the IDir distribution.

2.2 The mode reparameterized inverted Dirichlet distribution

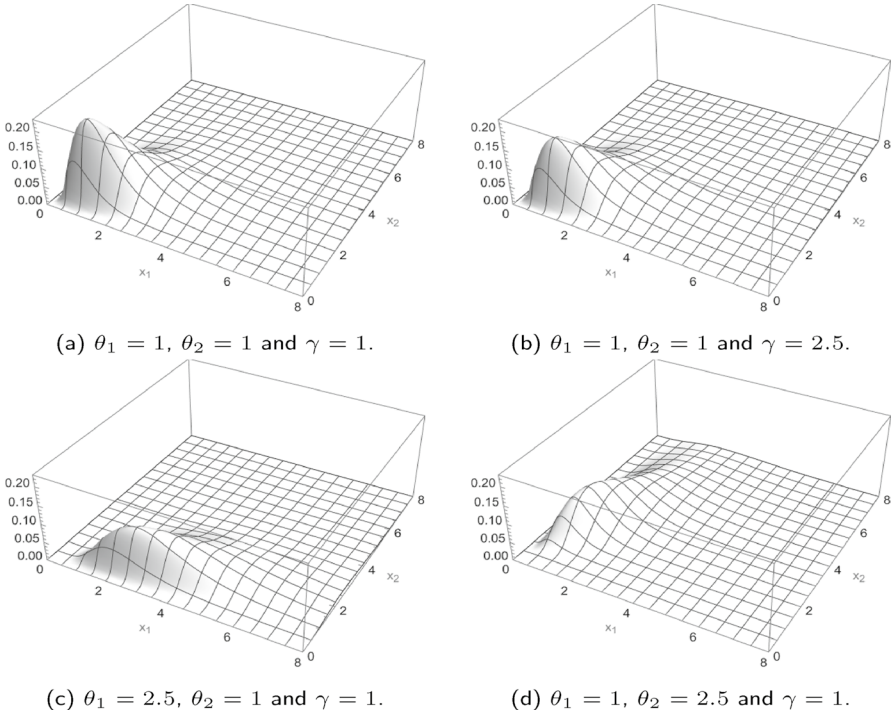
In this section, we introduce the mIDir distribution and describe its characteristics. Consider the following set of transformations on the parameter space of the IDir distribution (1):

$$\begin{cases} \alpha_i = 1 + \left(2 + p + \frac{1}{\gamma}\right)\theta_i,\ i = 1, \ldots, p \\ \alpha_{p+1} = 2 + \frac{1}{\gamma} \end{cases} \iff \begin{cases} \theta_i = \frac{\alpha_i - 1}{p + \alpha_{p+1}},\ i = 1, \ldots, p \\ \gamma = \frac{1}{\alpha_{p+1} - 2} \end{cases} \tag{3}$$

provided that $\alpha_i > 1, i = 1, \ldots, p$, and $\alpha_{p+1} > 2$. Because of these constraints on the parameters $\alpha_i, i = 1, \ldots, p+1$ for the mode and variance to exist, the mIDir distribution is a subclass of the IDir distribution in (1). Utilizing the link in (3), it follows that a random vector $\boldsymbol{X} \in \mathbb{R}_+^p$ is said to have a p-dimensional mIDir distribution if the PDF is

$$f(\boldsymbol{x};\boldsymbol{\theta},\gamma) = \frac{\Gamma\left(\left(2 + p + \frac{1}{\gamma}\right)\left(1 + \sum_{i=1}^{p} \theta_i\right)\right)}{\Gamma\left(2 + \frac{1}{\gamma}\right)\prod_{i=1}^{p}\Gamma\left(1 + \left(2 + p + \frac{1}{\gamma}\right)\theta_i\right)} \frac{\prod_{i=1}^{p} x_i^{(2+p+\frac{1}{\gamma})\theta_i}}{\left(1 + \sum_{i=1}^{p} x_i\right)^{(2+p+\frac{1}{\gamma})(1+\sum_{i=1}^{p}\theta_i)}}, \tag{4}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p), \theta_i > 0, i = 1, \ldots, p$, are the ith mode coordinates and $\gamma > 0$ is chosen to be related to the variability of $\boldsymbol{X}$ without making the form of the PDF more complex. If $\boldsymbol{X}$ has the PDF in (4), then we write $\boldsymbol{X} \sim mIDir_p(\boldsymbol{\theta}, \gamma)$. Figure 1



(a) $\theta_1 = 1$, $\theta_2 = 1$ and $\gamma = 1$. (b) $\theta_1 = 1$, $\theta_2 = 1$ and $\gamma = 2.5$.

(c) $\theta_1 = 2.5$, $\theta_2 = 1$ and $\gamma = 1$. (d) $\theta_1 = 1$, $\theta_2 = 2.5$ and $\gamma = 1$.

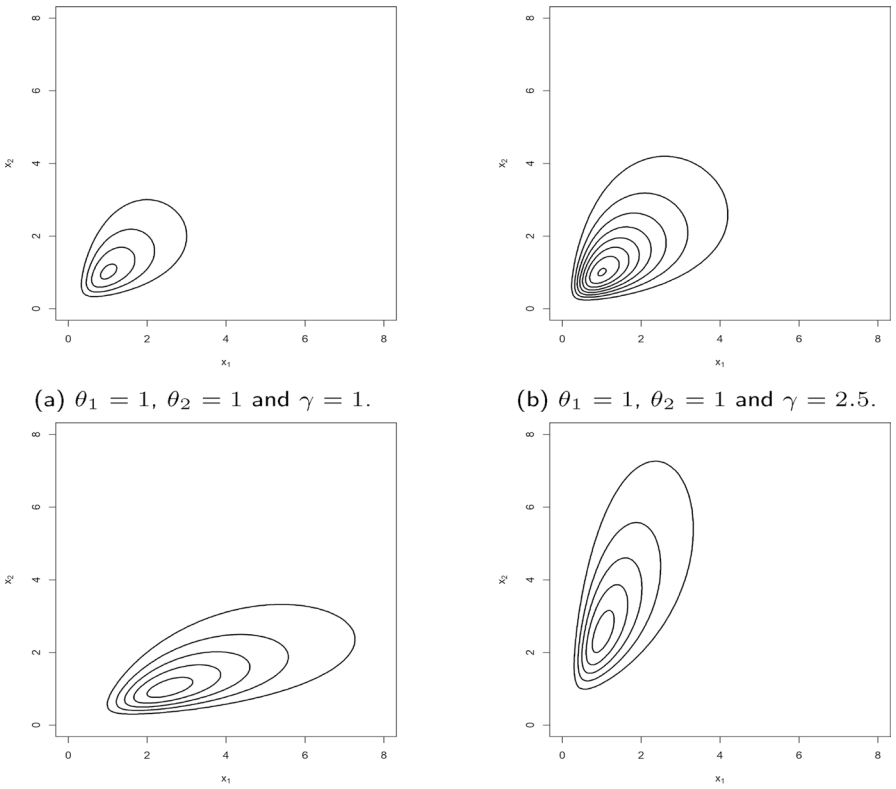
Fig. 1 Plots of the PDF of the mIDir distribution, in the case $p = 2$, for varying parameter values

illustrates some shapes the PDF in (4) can take for varying values of $\boldsymbol{\theta}$ and γ when $p = 2$, while Fig. 2 illustrates the corresponding contour plots.

From (2) and (3), an expression for the mixed moments of the mIDir distribution is given by

$$\mathrm{E}\left(\prod_{i=1}^{p} X_i^{h_i}\right) = \frac{\Gamma\left(2+\frac{1}{\gamma}-h_+\right)}{\Gamma\left(2+\frac{1}{\gamma}\right)} \prod_{i=1}^{p} \frac{\Gamma\left(1+\left(2+p+\frac{1}{\gamma}\right)\theta_i+h_i\right)}{\Gamma\left(1+\left(2+p+\frac{1}{\gamma}\right)\theta_i\right)}, \tag{5}$$

if $2+\frac{1}{\gamma} > h_+$. From (5), the expected value and variance of X_i follow for $\gamma > 0$ as



(a) $\theta_1 = 1$, $\theta_2 = 1$ and $\gamma = 1$. (b) $\theta_1 = 1$, $\theta_2 = 1$ and $\gamma = 2.5$.

(c) $\theta_1 = 2.5$, $\theta_2 = 1$ and $\gamma = 1$. (d) $\theta_1 = 1$, $\theta_2 = 2.5$ and $\gamma = 1$.

Fig. 2 Contour plots of the PDF of the mIDir distribution, in the case $p = 2$, for varying parameter values

$$\begin{aligned} \mathrm{E}(X_i) &= \frac{1+\left(2+p+\frac{1}{\gamma}\right)\theta_i}{1+\frac{1}{\gamma}}, \\ \mathrm{Var}(X_i) &= \frac{\gamma\left(1+\left(2+p+\frac{1}{\gamma}\right)\theta_i\right)\left(2+\frac{1}{\gamma}+\left(2+p+\frac{1}{\gamma}\right)\theta_i\right)}{\left(1+\frac{1}{\gamma}\right)^2}, \end{aligned} \tag{6}$$

while the covariance between X_i and X_j is

$$\mathrm{Cov}(X_i, X_j) = \frac{\gamma\left(1+\left(2+p+\frac{1}{\gamma}\right)\theta_i\right)\left(1+\left(2+p+\frac{1}{\gamma}\right)\theta_j\right)}{\left(1+\frac{1}{\gamma}\right)^2},$$

$i \neq j; i, j = 1, \ldots, p$. Convenient closed-form expressions for the skewness and kurtosis of X_i are given by

$$\mathrm{Skew}(X_i) = \frac{2\gamma\left(3\gamma + 2\theta_i(\gamma(p+2)+1)+1\right)}{\left(1-\gamma^2\right)\sqrt{\frac{\gamma\left(\gamma+\theta_i(\gamma(p+2)+1)\right)\left(2\gamma+\theta_i(\gamma(p+2)+1)+1\right)}{(\gamma+1)^2}}}, \quad \gamma > 0, \gamma \neq 1$$

$$\mathrm{Kurt}(X_i) = \frac{3\left(\gamma\left(\frac{2(\gamma+1)}{\gamma+\theta_i(\gamma(p+2)+1)} - \frac{2(\gamma+1)}{2\gamma+\theta_i(\gamma(p+2)+1)+1} + 7\right)+1\right)}{(\gamma-1)(2\gamma-1)}, \quad \gamma > 0, \gamma \notin \left\{\frac{1}{2}, 1\right\}.$$

2.3 Maximum likelihood estimation

To find estimates of the parameters for the mIDir model, we use the maximum likelihood (ML) approach. From a computational point of view, several methods may be considered to obtain ML estimates, including the popular expectation-maximization (EM) algorithm (Dempster et al., 1977). However, the computational ease and availability of optimization routines (for example in the `R` software) make direct numerical maximization (DNM) a practical choice to obtain ML estimates; see MacDonald (2014, 2021) for several examples depicting the advantages of DNM over the EM algorithm.

Given a random sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ from $\boldsymbol{X} \sim mIDir_p(\boldsymbol{\theta}, \gamma)$, the log-likelihood function of (4) is

$$l(\boldsymbol{\theta}, \gamma) = \sum_{l=1}^{n} \ln\left[f(\boldsymbol{x}_l; \boldsymbol{\theta}, \gamma)\right]$$

and is maximized using either the BFGS or Nelder-Mead algorithm implemented by the `optim()` function in `R` (R Core Team, 2013), which is available in the **stats** package. Since some of the parameters involved have constraints, a transformation/back-transformation approach is used to make the maximization unconstrained,

as required by both the BFGS and Nelder-Mead algorithms (Tomarchio & Punzo, 2020). The following transformations/back-transformations are employed:

$$\tilde{\theta}_i = \ln(\theta_i) \longleftrightarrow \theta_i = \mathrm{e}^{\tilde{\theta}_i}$$
$$\tilde{\gamma} = \ln(\tilde{\gamma}) \longleftrightarrow \gamma = \mathrm{e}^{\tilde{\gamma}},$$

where a 'tilde' denotes the unconstrained parameters.

3 Possible uses of the mode parameterized inverted Dirichlet in inferential statistics

Here we demonstrate how the mIDir facilitates the use and implementation of the IDir distribution in various statistical areas. Specifically, we define a finite mixture of mIDir distributions for clustering and semiparametric density estimation in Sect. 3.1. Additionally, we introduce a smoother consisting of mIDir kernels for nonparametric density estimation in Sect. 3.2, while a contaminated mIDir distribution for robustness against outliers is presented in Sect. 3.3.

3.1 Model-based clustering and semiparametric density estimation

Finite mixture models are mainly employed to represent datasets that are believed to consist of multiple subpopulations or clusters. These models have found applications in various fields such as clustering and pattern recognition (McLachlan & Peel, 2000), signal processing (Shih & Hero, 2003), and gene expression microarray data analysis (McLachlan et al., 2002). Despite the increasing recognition of the efficacy of mixtures involving asymmetric and, more generally, non-Gaussian distributions (Lee & McLachlan, 2013), the majority of the literature remains focused on Gaussian mixtures (An et al., 2022; Chen & Ludtke, 2021; Ni et al., 2020; Singhal et al., 2020). Another issue with using Gaussian components arises when the theoretical support of the data is positive. Although univariate Gaussian components are theoretically suitable for data supported on $\mathbb{R}$, they become inadequate for data supported on $\mathbb{R}_+$, leading to boundary bias. Because of these considerations, we suggest mixtures with mIDir components if the data is in the positive quadrant of the p-dimensional space and not normally distributed.

Mixtures of the inverted Dirichlet distributions have sporadically appeared in literature, generally in the context of image classification and segregation (Tirdad et al., 2015; Lai et al., 2018; Bdiri & Bouguila, 2011, 2012). The distribution of a random vector $\boldsymbol{X} \sim mIDir_p(\boldsymbol{\theta}, \gamma)$, taking values on $\mathbb{R}_+^p$, according to a K component parametric finite mixture model can be written as

$$p(\boldsymbol{x};\boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\gamma}) = \sum_{k=1}^{K} \pi_k f(\boldsymbol{x};\boldsymbol{\theta}_k, \gamma_k), \tag{7}$$

where $\boldsymbol{\pi} = \left(\pi_1, \ldots, \pi_K\right)^\top$ is the vector of mixing weights, with $\pi_k > 0$, and $\sum_{k=1}^{K} \pi_k = 1$, $\boldsymbol{\Theta} = \left(\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_K^\top\right)^\top$ is the vector of component modes $\boldsymbol{\theta}_k$, and $\boldsymbol{\gamma} = \left(\gamma_1, \ldots, \gamma_K\right)^\top$ is the vector of component "dispersion" parameters γ_k. The multimodal nature of (7) is illustrated by a two-component mixture example in Fig. 3.

3.1.1 EM algorithm

Let $\boldsymbol{z}_l = (z_{l1}, \ldots, z_{lK})$ be the latent variables where $z_{lk} = 1$ if $\boldsymbol{x}_l$ comes from component k and $z_{lk} = 0$ otherwise, i.e., $z_{lk} \in \{0, 1\}, l = 1, \ldots, n$ and $k = 1, \ldots, K$ (the source of incompleteness arises from the fact that we do not know to what group each observation belongs to). The complete-data are then given by $(\boldsymbol{x}_l, \boldsymbol{z}_l)$. Therefore we can find the ML estimator by maximizing the complete-data likelihood

$$L_c(\boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\gamma}) = \prod_{l=1}^{n} \prod_{k=1}^{K} \left[\pi_k f(\boldsymbol{x}_l; \boldsymbol{\theta}_k, \gamma_k)\right]^{z_{lk}}$$

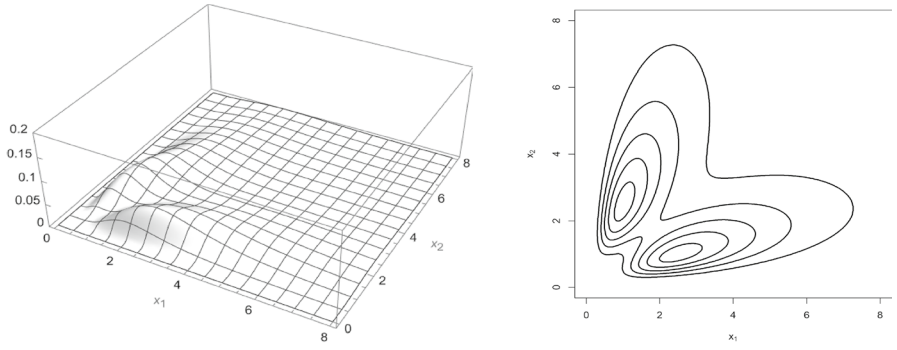
with the complete-data log-likelihood as

$$l_c(\boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\gamma}) = l_{1c}(\boldsymbol{\pi}) + l_{2c}(\boldsymbol{\Theta}, \boldsymbol{\gamma}), \tag{8}$$

where

$$l_{1c}(\boldsymbol{\pi}) = \sum_{l=1}^{n} \sum_{k=1}^{K} z_{lk} \ln \pi_k \tag{9}$$

and



(a) Density plot. (b) Contour plot.

Fig. 3 Two component finite mixture mIDir density and contour plot for $\boldsymbol{\theta}_1 = (2.5, 1)$, $\boldsymbol{\theta}_2 = (1, 2.5)$, $\gamma_1 = \gamma_2 = 1$ and $\pi = 0.5$

$$l_{2c}(\boldsymbol{\Theta}, \boldsymbol{\gamma}) = \sum_{l=1}^{n} \sum_{k=1}^{K} z_{lk} \ln\left[f(\boldsymbol{x}_l; \boldsymbol{\theta}_k, \gamma_k)\right]. \tag{10}$$

The EM algorithm iterates between two steps, the E-step and the M-step, until convergence. Each step is described below:

3.1.1.1 E-step The conditional expectation of the complete-data log-likelihood is denoted by

$$Q\left(\boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = Q_1\left(\boldsymbol{\pi} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) + Q_2\left(\boldsymbol{\Theta}, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right), \tag{11}$$

for the $(r+1)$th iteration and is obtained by substituting z_{ik} in (9) and (10) with

$$\mathrm{E}\left(Z_{lk} | \boldsymbol{x}_l; \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = \frac{\pi_k^{(r)} f\left(\boldsymbol{x}_l; \boldsymbol{\theta}_k^{(r)}, \boldsymbol{\gamma}^{(r)}\right)}{p\left(\boldsymbol{x}_l; \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right)} =: z_{lk}^{(r)}.$$

Note that the two right-hand terms of (11) are in the same order as (8).

3.1.1.2 M-step For the $(r+1)$-th iteration, updates for $\boldsymbol{\pi}$, $\boldsymbol{\Theta}$, and $\boldsymbol{\gamma}$ are calculated as the values that maximize $Q\left(\boldsymbol{\pi}, \boldsymbol{\Theta}, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right)$. Maximizing $Q_1\left(\boldsymbol{\pi} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right)$ with respect to $\boldsymbol{\pi}$ yields

$$\pi_k^{(r+1)} = \frac{1}{n} \sum_{l=1}^{n} z_{lk}^{(r)}, \quad k = 1, \ldots, K.$$

Similarly, updates for $\boldsymbol{\Theta}$ and $\boldsymbol{\gamma}$ are calculated for values that maximize $Q_2\left(\boldsymbol{\Theta}, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right)$, which is equivalent to independently maximizing each of the K components

$$Q_{2k}\left(\boldsymbol{\Theta}_k, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right) = \sum_{l=1}^{n} z_{lk}^{(r)} \ln\left[f\left(\boldsymbol{x}_l; \boldsymbol{\theta}_k^{(r)}, \boldsymbol{\gamma}_k^{(r)}\right)\right], \quad k = 1, \ldots, K.$$

The maximization of $Q_{2k}\left(\boldsymbol{\Theta}_k, \boldsymbol{\gamma} | \boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \boldsymbol{\gamma}^{(r)}\right)$ is simply a weighted log-likelihood, whose maximization is similar to that of the log-likelihood function discussed in Sect. 2.3, but with weights $z_{lk}^{(r)}$.

3.1.2 Initialization and convergence criterion

The starting values are a critical step in the EM algorithm and can greatly impact the accuracy and reliability of the model estimates (Biernacki et al., 2003; Bagnato & Punzo, 2013); their choice thus constitutes an important issue. If the starting values are chosen poorly, the algorithm may converge to a local maximum instead of the global maximum. Additionally, if the starting values are too far from the true values, the algorithm may converge slowly or not at all.

We utilized random soft and hard clustering for the initial values of $z_{lk}^{(0)}$. Due to the structural shape of the mIDir distribution, the k-means algorithm is not suitable for initial values as it assumes that the data is spherical. The initial values of $\theta_{lk}^{(0)}$ are calculated as the empirical mode of the marginal (argued that the empirical marginal modes might be close to the true value of θ_l, even though not exactly equal) while $\gamma_k^{(0)}$ is calculated as a function of (6).

There are several convergence criteria that can be used to determine whether the EM algorithm has converged or not. One common method is to track the change in the log-likelihood function, say l, between consecutive iterations. If the change falls below a predetermined threshold the algorithm can be considered to be converged, i.e., $l\left(\boldsymbol{\pi}^{(r+1)}, \boldsymbol{\Theta}^{(r+1)}, \gamma^{(r+1)}\right) - l\left(\boldsymbol{\pi}^{(r)}, \boldsymbol{\Theta}^{(r)}, \gamma^{(r)}\right) < \epsilon$. In this paper, we use a threshold of $\epsilon = 1 \times 10^{-10}$ or a maximum of 1000 iterations as the stopping criteria.

3.2 Nonparametric density estimation

Nonparametric density estimation serves as a crucial data analysis tool for effectively revealing the underlying structure within datasets. Kernel smoothing, one of the most widely used techniques for nonparametric density estimation, is particularly valued for its conceptual simplicity and its practical and theoretical advantages.

In kernel density estimation, a kernel (or bump) is centered on each observation, and the overall PDF is estimated by summing these kernels, each with equal weight. The smoothing parameter, or bandwidth, which controls the width of each kernel, must be chosen or conveniently selected. However, when dealing with positive variables, as Chen (2000) highlighted, commonly used Gaussian kernels are unsuitable as they assign weight to unrealistic negative values, leading to boundary bias. Chen (1999, 2000), Punzo (2010), and Mazza and Punzo (2014) addressed this issue by using kernels defined on the same support as the unknown density. In light of this, this section proposes a smoother based on mIDir kernels. The mIDir kernel estimator is non-negative and mitigates boundary bias, making it a viable alternative for positive multivariate data. Given that the choice of the smoothing parameter has a substantial impact on the resulting estimate, an automated method for selecting it is also discussed.

3.2.1 Reparameterized ID kernel density estimation

If we place a mIDir density over each $\boldsymbol{X}_l$, $l = 1, \ldots, n$, by fixing $\boldsymbol{\theta} = \boldsymbol{X}_l$, i.e., each $\boldsymbol{X}_l$ is considered as the mode, with equal weights, this leads to the following kernel density smoother

$$\widehat{f}(\boldsymbol{x};\gamma) = \frac{1}{n}\sum_{l=1}^{n} f(\boldsymbol{x};\boldsymbol{\theta} = \boldsymbol{X}_l, \gamma) = \frac{1}{n}\sum_{l=1}^{n} k_\gamma(\boldsymbol{x};\boldsymbol{X}_l),$$

which satisfies all the fundamental properties of a PDF.

We utilize likelihood cross-validation (LCV) to determine the optimal smoothing parameter. As described by Silverman (1986), the rationale behind LCV is as

follows: by taking the log of $\widehat{f}$ and fixing $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, $\ln\widehat{f}$ can be regarded as the log-likelihood of γ, since it only depends on γ. Since there is no specific $\boldsymbol{X}_l$ that should be omitted, the log-likelihood is averaged across all possible choices of omitted $\boldsymbol{X}_l$. The smoothing parameter γ is then chosen as the value that minimizes the score function

$$\mathrm{LCV}(\gamma) = \frac{1}{n}\sum_{l=1}^{n} \ln\left[\widehat{f}_{-l}(\boldsymbol{x} = \boldsymbol{X}_l;\gamma)\right]$$

over the possible values of γ, where $\widehat{f}_{-l}$ is the estimate constructed from all the observations excluding $\boldsymbol{X}_l$.

3.3 Robustness against mild outliers

Real data may often be contaminated by unusual observations that can affect the estimation of the model parameters. The identification of these unusual observations and the development of robust methods of parameter estimation that are not affected by their presence is an important problem (Hennig, 2004). Outliers are generally divided into two broad categories: (i) mild outliers, which are observations sampled from a population different or even far from the assumed model, or (ii) gross outliers, which are observations that cannot be modelled by any distribution as they are unpredictable. For mild outliers, which are also referred to as "bad" observations herein in analogy with Aitkin and Wilson (1980), Punzo et al. (2020), and Punzo and Tortora (2021), a model flexible enough to accommodate all observations, including the bad observations, is recommended (Ritter, 2014). To address this issue, we propose the contaminated mIDir (cmIDir) model. The PDF of the cmIDir model is given by

$$p(\boldsymbol{x};\boldsymbol{\theta},\gamma,\delta,\eta) = \delta f(\boldsymbol{x};\boldsymbol{\theta},\gamma) + (1-\delta) f(\boldsymbol{x};\boldsymbol{\theta},\eta\gamma), \tag{12}$$

where $\eta > 1$ and $\delta \in (0,1)$. As well-documented in Zhang et al. (2023) and Tortora et al. (2024), although not necessary, a restriction on δ can be imposed such that $\delta \in (0.5, 1)$. This will ensure that at least half of the sample points are considered "good", which is a general assumption within robust statistical inference. The additional contamination parameters δ and η have an interpretation of practical interest: δ is the proportion of points from the reference mIDir distribution, while η denotes the degree of contamination. Because of the assumption that $\eta > 1$, it can be viewed as an inflation parameter, i.e., the increase in variability due to the points that do not come from the reference distribution. For a discussion on the concept of a reference distribution, which is here assumed to be (4), see Davies and Gather (1993) and Hennig (2002). The cmIDir reduces to the mIDir when $\delta \to 1^-$ and $\eta \to 1^+$.

An advantage of model (12) is that, once the parameters are estimated, it is possible to determine whether a data point $\boldsymbol{x}$ comes from the reference mIDir distribution or not via the *a posteriori* probability

$$v(\boldsymbol{x};\boldsymbol{\theta},\gamma,\delta,\eta) := P\big(\boldsymbol{x} \text{ comes from the reference mIDir distribution}|\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta},\hat{\eta}\big) = \frac{\hat{\delta}f\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma}\big)}{p\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta},\hat{\eta}\big)}. \tag{13}$$

In detail, $\boldsymbol{x}$ is considered good if the resulting probability is greater than 0.5, i.e., if $v(\boldsymbol{x};\boldsymbol{\theta},\gamma,\delta,\eta) > 0.5$, while it is considered bad otherwise. This can equivalently be defined in terms of discriminant functions

$$D_{\text{good}}\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta}\big) = \hat{\delta}f\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma}\big) \tag{14}$$

and

$$D_{\text{bad}}\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta},\hat{\eta}\big) = \big(1-\hat{\delta}\big)f\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\eta}\hat{\gamma}\big) \tag{15}$$

such that $\boldsymbol{x}$ is considered good if

$$D_{\text{good}}\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta}\big) > D_{\text{bad}}\big(\boldsymbol{x};\hat{\boldsymbol{\theta}},\hat{\gamma},\hat{\delta},\hat{\eta}\big), \tag{16}$$

and bad otherwise (Duda & Hart, 1973). If (16) is solved as a function of $\boldsymbol{x}$, the positive p-dimensional real space, denoted by $\mathbb{R}_+^p$ is partitioned in two regions of good and bad points, delimited by the intersection points of (14) and (15). This is visualized in Sect. 5.1.

3.3.1 EM algorithm

The ML estimation procedure is briefly described for the proposed cmIDir model. Let $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ be an observed sample from the cmIDir model (12). In this case, the source of incompleteness stems from the fact that we do not know if the generic data point $\boldsymbol{x}_l$ is good or bad. To denote this incompleteness, we use an indicator vector $\boldsymbol{v} = (v_1,\ldots,v_n)$ so that $v_l = 1$ if x_l is good and $v_l = 0$ otherwise. From (12), the complete-data likelihood can be written as

$$L_c(\boldsymbol{\theta},\gamma,\delta,\eta) = \prod_{l=1}^{n}\big[\delta f(\boldsymbol{x}_l;\boldsymbol{\theta},\gamma)\big]^{v_l}\big[(1-\delta)f(\boldsymbol{x}_l;\boldsymbol{\theta},\eta\gamma)\big]^{1-v_l}.$$

Then, the complete-data log-likelihood can be written as

$$\begin{aligned} l_c(\boldsymbol{\theta},\gamma,\delta,\eta) &= \sum_{l=1}^{n}\big\{v_l\ln\delta + v_l\ln f(\boldsymbol{x}_l;\boldsymbol{\theta},\gamma) + (1-v_l)\ln(1-\delta) + (1-v_l)\ln f(\boldsymbol{x}_l;\boldsymbol{\theta},\eta\gamma)\big\} \\ &= l_{1c}(\delta) + l_{2c}(\boldsymbol{\theta},\gamma,\eta) \end{aligned}$$

where

$$l_{1c}(\delta) = \sum_{l=1}^{n}\big\{v_l\ln\delta + (1-v_l)\ln(1-\delta)\big\}$$

and

$$l_{2c}(\boldsymbol{\theta}, \gamma, \eta) = \sum_{l=1}^{n} \left\{ v_l \ln f(\boldsymbol{x}_l; \boldsymbol{\theta}, \gamma) + (1 - v_l) \ln f(\boldsymbol{x}_l; \boldsymbol{\theta}, \eta\gamma) \right\}. \tag{17}$$

3.3.2 E-step

In the E-step, the conditional expectation of the complete-data log-likelihood function is denoted by

$$Q\left(\boldsymbol{\theta}, \gamma, \delta, \eta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right) = Q_1\left(\delta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right) + Q_2\left(\boldsymbol{\theta}, \gamma, \eta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right)$$

for the $(r+1)$-th iteration, which is in the same order as (17). $Q\left(\boldsymbol{\theta}, \gamma, \delta, \eta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right)$ is obtained by substituting v_l by the expected *a posteriori* probability for a point to come from the reference distribution

$$\mathrm{E}\left(V_l | \boldsymbol{X}_l = \boldsymbol{x}_l; \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right) = \frac{\delta^{(r)} f\left(\boldsymbol{x}_l; \boldsymbol{\theta}^{(r)}, \gamma^{(r)}\right)}{p\left(\boldsymbol{x}_l; \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right)} := v_l^{(r)}$$

in (17).

3.3.3 M-step

An update $\delta^{(r+1)}$ for δ is calculated by independently maximizing

$$Q_1\left(\delta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}\right) = \sum_{l=1}^{n} \left\{ v_l^{(r)} \ln \delta^{(r)} + \left(1 - v_l^{(r)}\right) \ln\left(1 - \delta^{(r)}\right) \right\}$$

with respect to δ and subjects to constraints. It follows that an update for the $(r+1)$ -th iteration is given as

$$\delta^{(r+1)} = \max\left\{ 0.5, \frac{1}{n} \sum_{l=1}^{n} v_l^{(r)} \right\}.$$

Updates for $\boldsymbol{\theta}, \gamma$ and η are obtained by maximizing

$$Q_2(\boldsymbol{\theta}, \gamma, \eta | \boldsymbol{\theta}^{(r)}, \gamma^{(r)}, \delta^{(r)}, \eta^{(r)}) = \sum_{l=1}^{n} \left\{ v_l^{(r)} \ln f\left(\boldsymbol{x}_l; \boldsymbol{\theta}^{(r)}, \gamma^{(r)}\right) + \left(1 - v_l^{(r)}\right) \ln f\left(\boldsymbol{x}_l; \boldsymbol{\theta}^{(r)}, \eta^{(r)}\gamma^{(r)}\right) \right\}.$$

This can be achieved in `R` using the `optim` function as described in Sect. 2.3. Similarly, since some of the parameters involved have constraints, the following transformations/back-transformations are used

$$\tilde{\theta}_l = \ln(\theta_l) \longleftrightarrow \theta_l = \mathrm{e}^{\tilde{\theta}_l}$$
$$\tilde{\gamma} = \ln(\tilde{\gamma}) \longleftrightarrow \gamma = \mathrm{e}^{\tilde{\gamma}}$$
$$\tilde{\eta} = \ln(\eta - 1) \longleftrightarrow \eta = \mathrm{e}^{\tilde{\eta}} + 1$$

where parameters marked with a 'tilde' denote the unconstrained parameters.

4 Simulated data analyses

In this section, various aspects related to the mIDir model are investigated, including the ability of the estimation methods described in Sects. 2.3 for successful parameter recovery. A sensitivity analysis is also presented that evaluates the influence of mild outliers on the parameter estimates of the mIDir model. Additionally, the proposed cmIDir model is fit to the data in the sensitivity analysis to demonstrate its effectiveness in modelling mild outliers and identifying whether an observation is a mild outlier or not.

4.1 Parameter recovery

Parameter recovery is a critical aspect of algorithmic performance. When the empirical mean of estimates consistently diverges from actual parameters across numerous replications, the estimator is described as biased. The extent of variability in the estimates across the replications is also a matter of concern. To investigate the performance of the DNM method described in Sect. 2.3, we simulate datasets of size $n = 100, 500, 1000$ from the mIDir distribution in (4), with $p = 2$, for the following scenarios:

1. mIDir with $\boldsymbol{\theta} = (1, 1)$ and $\gamma = 0.01$.
2. mIDir with $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 0.1$.
3. mIDir with $\boldsymbol{\theta} = (1, 1)$ and $\gamma = 1$.
4. mIDir with $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 1$.

In each scenario, we report the empirical mean, bias, standard deviation (SD), and the mean squared error (MSE) of the estimates in Table 1 and Table 2.

In the various scenarios investigated, we observe accurate parameter recovery and note that an increase in sample size leads to reduced bias, variability, and consequently, MSE within these estimates. The estimation of the parameters is nearly perfect across all scenarios.

4.2 Sensitivity analysis

In this analysis, we investigate the impact of outliers on both the mIDir and the cmIDir, while also examining the cmIDir's capability to accurately classify whether an observation can be considered as good or bad via the *a posteriori* probability

Table 1 Parameter recovery results of Scenario 1 and 2, based on 500 replications of different sample sizes

	Scenario 1: $\boldsymbol{\theta} = (1, 1)$ and $\gamma = 0.01$			Scenario 2: $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 0.1$		
	100	500	1000	100	500	1000
Mean						
θ_1	1.00036	1.00017	0.99998	2.51040	2.50110	2.50045
θ_2	0.99980	0.99934	1.00012	1.00690	0.99923	1.00029
γ	0.00993	0.01001	0.00999	0.09855	0.09954	0.10003
Bias						
θ_1	0.00036	0.00017	− 0.00002	0.01040	0.00110	0.00045
θ_2	− 0.00020	− 0.00066	0.00012	0.00690	− 0.00077	0.00029
γ	− 0.00007	0.00001	− 0.00001	− 0.00145	− 0.00046	0.00003
SD						
θ_1	0.01417	0.00649	0.00442	0.08829	0.04185	0.02829
θ_2	0.01395	0.00640	0.00431	0.04347	0.01985	0.01305
γ	0.00100	0.00047	0.00033	0.01170	0.00541	0.00382
MSE						
θ_1	0.00020	0.00004	0.00002	0.00790	0.00175	0.00080
θ_2	0.00019	0.00004	0.00002	0.00194	0.00039	0.00017
γ	0.00000	0.00000	0.00000	0.00014	0.00003	0.00001

Table 2 Parameter recovery results of Scenario 3 and 4, based on 500 replications of different sample sizes

	Scenario 3: $\boldsymbol{\theta} = (1, 1)$ and $\gamma = 1$			Scenario 4: $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 1$		
	100	500	1000	100	500	1000
Mean						
θ_1	1.01525	1.00107	1.00035	2.51890	2.50360	2.49902
θ_2	1.01123	1.00175	1.00189	1.00787	1.00110	1.00011
γ	1.03415	1.01674	1.00770	1.05862	1.01126	1.00949
Bias						
θ_1	0.01525	0.00107	0.00035	0.01890	0.00360	− 0.00098
θ_2	0.01123	0.00175	0.00189	0.00787	0.00110	0.00011
γ	0.03415	0.01674	0.00770	0.05862	0.01126	0.00949
SD						
θ_1	0.08747	0.03639	0.02823	0.17536	0.07489	0.05740
θ_2	0.08467	0.03618	0.02837	0.08424	0.03672	0.02886
γ	0.36720	0.14188	0.09727	0.46926	0.14626	0.09757
MSE						
θ_1	0.00788	0.00133	0.00080	0.03111	0.00562	0.00330
θ_2	0.00729	0.00131	0.00081	0.00716	0.00135	0.00083
γ	0.13600	0.02041	0.00952	0.22364	0.02152	0.00961

given in (13). To explore this, we generate data where the majority of the observations are from the reference mIDir distribution, but a small proportion are from a uniform distribution. Consider the following two data generating processes with $p = 2$:

1. mIDir with $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 0.01$ where 1% of the points are randomly substituted by data generated from a uniform distribution over the interval (0, 5) for both dimensions.
2. mIDir with $\boldsymbol{\theta} = (2.5, 1)$ and $\gamma = 0.01$ where 5% of the points are randomly substituted by data generated from a uniform distribution over the interval (0, 5) for both dimensions.

For each configuration, 300 datasets are generated with sample size $n = 500$. The mIDir and cmIDir are then fit to the data to see what the impact of the outlier is on the estimates. The mean, bias, SD, and MSE are reported in Table 3.

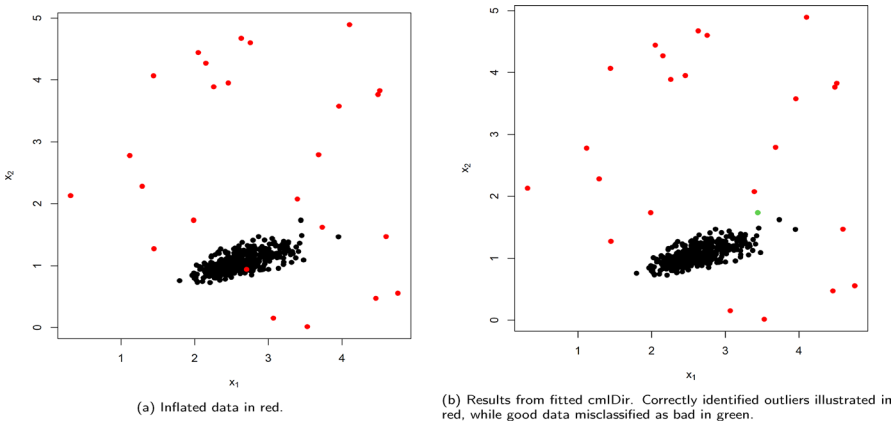
In Table 3, we observe that in both scenarios, the cmIDir model demonstrates noticeably reduced bias and MSE compared to the mIDir distribution. Additionally, the bias and MSE are slightly higher for Scenario 2 than for Scenario 1. This is expected given the increased presence of bad points in Scenario 2, which influences the estimates of the mIDir distribution. We also note that the variability parameter, γ is consistently greater for the mIDir distribution compared to the cmIDir distribution. This makes intuitive sense as γ tries to compensate for the increased variance caused by the bad points.

Table 3 Results of sensitivity analysis for outliers added to mIDir data according to Scenario's 1 and 2

	Scenario 1		Scenario 2	
	mIDir	cmIDir	mIDir	cmIDir
Mean				
θ_1	2.42132	2.49875	2.12474	2.49061
θ_2	0.97330	0.99952	0.87661	0.99838
γ	0.02195	0.00997	0.07766	0.01000
Bias				
θ_1	− 0.07868	− 0.00125	− 0.37526	− 0.00940
θ_2	− 0.02670	− 0.00048	− 0.12339	− 0.00162
γ	0.01195	− 0.00003	0.06766	0.00000
SD				
θ_1	0.05330	0.01290	0.10092	0.01287
θ_2	0.02061	0.00638	0.03805	0.00657
γ	0.00768	0.00046	0.02067	0.00050
MSE				
θ_1	0.00903	0.00017	0.15101	0.00025
θ_2	0.00114	0.00004	0.01667	0.00005
γ	0.00020	0.00000	0.00501	0.00000

Table 4 Values of TPRs and FPRs over 300 replications

	Scenario 1 (%)	Scenario 2 (%)
TPR	91.533	92.893
FPR	0.024	0.078



(a) Inflated data in red.

(b) Results from fitted cmIDir. Correctly identified outliers illustrated in red, while good data misclassified as bad in green.

Fig. 4 Example of simulated mIDir data described in Scenario 2

To assess the ability of the cmIDir to automatically classify whether a point is good or bad, as given in (13) we report: (i) the true positive rate (TPR), which measures the proportion of bad observations that are correctly identified as bad; and (ii) the false positive rate (FPR), which corresponds to the proportion of good points incorrectly classified as bad. The results are reported in Table 4.

In Scenario 1, of the $5 \times 300 = 1500$ outliers, 1373 were correctly classified as outliers, resulting in a TPR of 39.867%. Similarly, out of the $495 \times 300 = 148500$ good data points, 36 were incorrectly classified as bad. The cmIDir provides almost optimal results in terms of the FPR being very close to 0, indicate an extremely low rate of misclassifying good points as bad. The fact that the TPRs do not approach one is not necessarily an indication of the ineffectiveness of the cmIDir's outlier detection capabilities. The way the outliers are inserted into the data may lead to an overlap with values related to good points and, as such, these points will be detected as good points by our model.

This is exemplified in Fig. 4, for a replication of Scenario 2. The 5% replaced points are denoted in red in Fig. 4a, while the reference mIDir data is shown in black. The classification result of the cmIDir is displayed in Fig. 4b, with correctly identified outliers illustrated in red, while good data misclassified as bad is depicted in green. In this example, 20 of the 25 noise points were correctly identified as outliers and 1 good observation was wrongly misclassified as bad.

5 Real data analyses

To demonstrate the relevance of the approaches proposed in Sect. 3, several examples in diverse fields of the mIDir mixture (Sect. 3.1), mIDir kernel smoother (Sect. 3.2), and cmIDir distribution (Sect. 3.3), are illustrated in this section. We applied the proposed mIDir and cmIDir model to the spectra of different cultivars of cantaloupe (Sect. 5.1). Additionally, we illustrate how the mIDir kernel smoother can be successfully applied to uncover structural features within the Australian Institute of Sport data set (Sect. 5.2), followed by fitting mixtures of the mIDir to the data.

The results of these applications provide valuable insights into the models' performance, generalizability, and relevance across different fields of application. Furthermore, by comparing the mIDir and cmIDir to other positively skewed distributions like a bivariate log-normal (Nadarajah & Lyu, 2022) and two different multivariate generalizations of the inverted beta distributions i.e., $GB2_I$ and $GB2_I I$ as given by Sarabia et al. (2020), it becomes evident that our proposed models can be considered as viable alternatives with the added advantage of interpretable parameter estimates. The model comparison is conducted using the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978).

5.1 Cantaloupe spectra

The spectra of different cultivars of cantaloupe (*Cucumis melo L. Cantaloupensis* group) are available in the fruit dataset (Hubert & Van Driessen, 2004; Hubert et al., 2012), which can be accessed in the `R` package `rrcov` (Todorov, 2009). The dataset contains $n = 1096$ observations, consisting of D, M, and HA cultivates, sized 490, 106, and 500 respectively. The spectra were measured in 256 wavelengths.

In our analysis, we examined all the pairs of the 3rd to 6th wavelengths, denoted as V3, V4, V5, and V6, respectively, across various cantaloupes, and fitted the proposed mIDir models to these pairs. We used these pairs to illustrate the effectiveness of the cmIDir model. The marginal summary statistics for the considered wavelengths are presented in Table 5. We note that the data consistently exhibit positive skewness, with the mean always exceeding the mode, and show excess kurtosis, regardless of the wavelength considered. The "empirical" mode is determined as the value with the highest density based on the kernel density estimator provided by the `density()` function in `R`, using its default arguments (Punzo, 2019).

Table 5 Cantaloupe spectra dataset. Marginal summary statistics of cantaloupe spectra wavelengths

	V3	V4	V5	V6
Mode	5.49	5.57	5.23	5.61
Mean	7.33	7.19	7.27	7.35
Variance	12.11	10.36	12.72	10.95
Skewness	1.22	1.15	1.27	1.34
Kurtosis	4.28	4.26	4.28	5.05

Table 6 presents the model comparison separately for each pair of cantaloupe spectra wavelengths. Here, we see that the cmIDir constantly ranked first according to the AIC and BIC (represented by bold values). The cmIDir is thus a viable alternative to other existing competitors, as validated by the AIC and BIC. The added benefit of the cmIDir is that the parameter estimates are of interpretable interest. For example, the estimates of the cmIDir for V4 vs V5 are $\widehat{\theta}_1 = 4.584, \widehat{\theta}_2 = 4.605, \widehat{\gamma} = 0.067, \widehat{\delta} = 0.501$, and $\widehat{\eta} = 8.056$. Approximately 50.1% of the observations are considered "good", and from (13), it is possible to determine whether each observation is a mild outlier. This is illustrated in Fig. 5, where points

Table 6 Cantaloupe spectra dataset The number of parameters (# par), log-likelihood (Log-lik), AIC, and BIC for the competing models fitted to each pair of cantaloupe spectra wavelengths, along with rankings from these criteria

Wavelengths	Model	# par	Log-lik	AIC	Rank	BIC	Rank
V3 vs V4	mIDir	3	− 4595.81	9197.62	5	9212.62	5
	cmIDir	5	− 4552.91	**9115.81**	1	**9140.81**	1
	Log-Normal	5	− 4575.06	9160.12	4	9185.12	4
	$GB2_I$	7	− 4563.59	9141.17	2	9176.17	2
	$GB2_{II}$	7	− 4565.34	9144.68	3	9179.68	3
V3 vs V5	mIDir	3	− 4477.52	8961.05	5	8976.05	5
	cmIDir	5	− 4403.94	**8817.88**	1	**8842.88**	1
	Log-Normal	5	− 4423.13	8856.26	2	8881.26	2
	$GB2_I$	7	− 4424.69	8863.38	4	8898.38	4
	$GB2_{II}$	7	− 4421.75	8857.50	3	8892.50	3
V3 vs V6	mIDir	3	− 4738.68	9483.36	5	9498.36	5
	cmIDir	5	− 4676.27	**9362.54**	1	**9387.54**	1
	Log-Normal	5	− 4712.50	9435.00	3	9460.00	2
	$GB2_I$	7	− 4707.57	9429.14	2	9464.14	3
	$GB2_{II}$	7	− 4714.15	9442.31	4	9477.31	4
V4 vs V5	mIDir	3	− 4772.63	9551.26	5	9566.26	5
	cmIDir	5	− 4723.85	**9457.70**	1	**9482.69**	1
	Log-Normal	5	− 4754.65	9519.31	4	9544.30	4
	$GB2_I$	7	− 4744.23	9502.46	2	9537.46	2
	$GB2_{II}$	7	− 4747.41	9508.81	3	9543.81	3
V4 vs V6	mIDir	3	− 4461.19	8928.37	5	8943.37	5
	cmIDir	5	− 4404.85	**8819.70**	1	**8844.70**	1
	Log-Normal	5	− 4424.86	8859.73	2	8884.72	2
	$GB2_I$	7	− 4423.48	8860.97	3	8895.96	3
	$GB2_{II}$	7	− 4425.83	8865.66	4	8900.66	4
V5 vs V6	mIDir	3	− 4717.89	9441.79	5	9456.79	4
	cmIDir	5	− 4680.72	**9371.45**	1	**9396.44**	1
	Log-Normal	5	− 4703.96	9417.91	2	9442.91	2
	$GB2_I$	7	− 4705.08	9424.16	4	9459.16	5
	$GB2_{II}$	7	− 4703.11	9420.21	3	9455.21	3

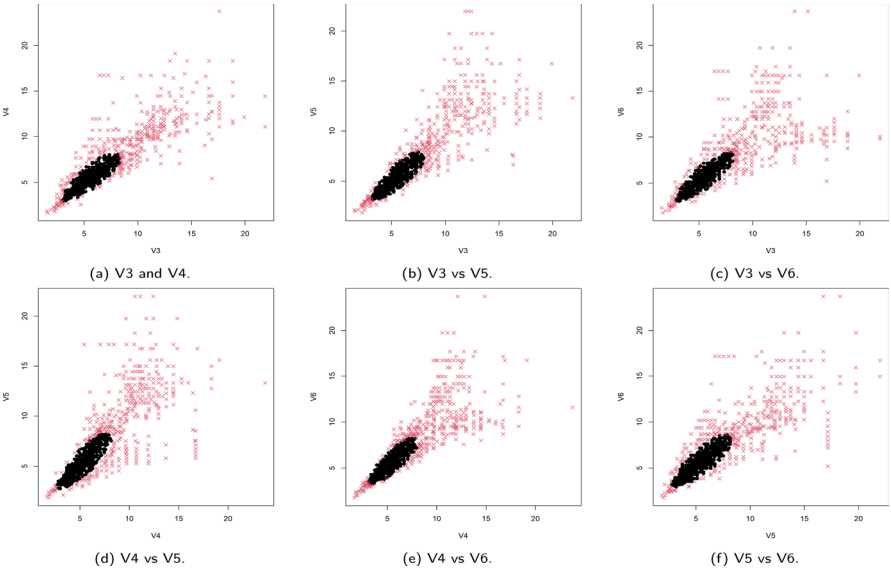


Fig. 5 Scatter plots of pairs of cantaloupe spectra wavelengths, where points considered "good" are colored in black, and "bad" points in red

considered "good" are colored in black, while the "bad" points are red, for all the pairs of wavelengths considered. The plots also show a clear deviation of the data from the reference mIDir model, regardless of the pair considered, which corroborates the need for a more flexible approach like our cmIDir model.

5.2 Australian Institute of Sport data

In this example, we analyze the dataset from the Australian Institute of Sport (AIS), which contains measures of physical activity for 102 male and 100 female athletes. This dataset has been analyzed by several authors (see, e.g., Li et al., 2016; Tomarchio et al., 2022; Loperfido, 2024), and is freely available in the `sn` package in R (Azzalini, 2015). We focus on five particular variables: lean body mass (LBM), weight in kg (Wt), body mass index (BMI), white cell count (WCC), and body fat percentage (BFP). Our objective is to cluster the data using the mIDir mixture as defined in (7) and to assess whether this mixture can effectively distinguish clusters of male and female athletes, assuming that this information is not known beforehand. The marginal summary statistics for each sex are provided in Table 7.

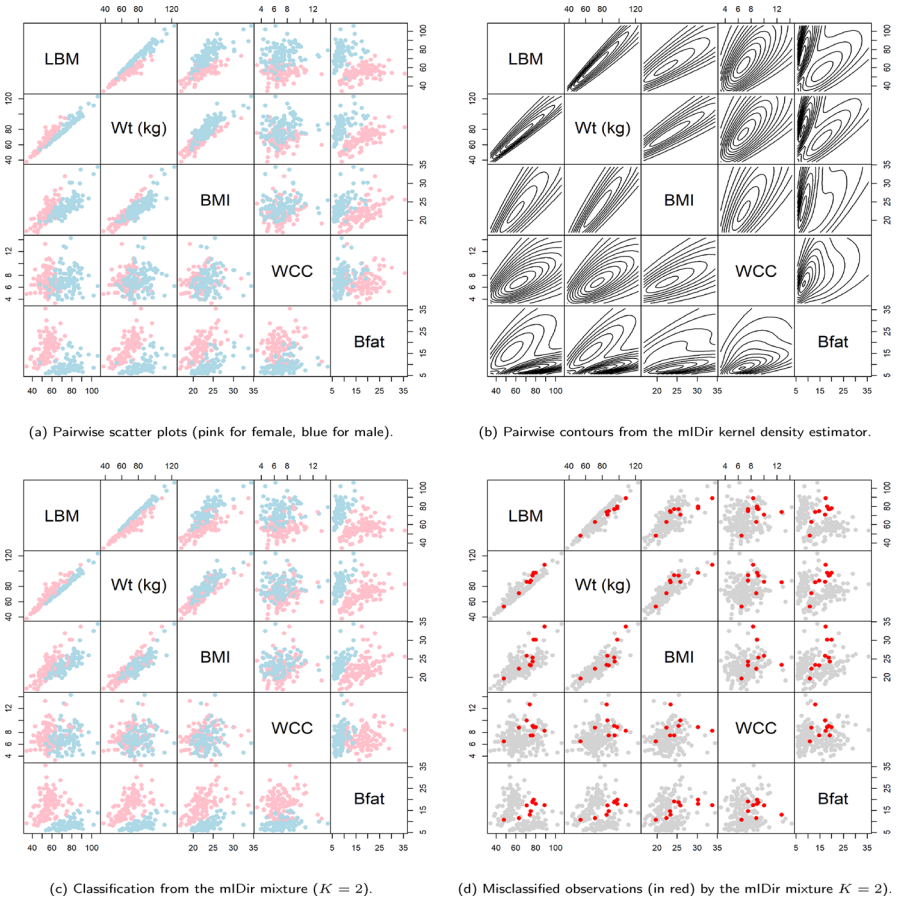
The matrix of pairwise scatter plots of the subset of the AIS data is depicted in Fig. 6a, where the female observations are shown in pink and the male observations in blue, and when considered separately for each sex, appear non-Gaussian and the data often exhibit asymmetry.

As a preliminary step, we apply the mIDir kernel density smoother to uncover the data's structural features. The resulting pairwise contours in Fig. 6b, where the

Table 7 AIS dataset

	Male					Female				
	LBM	Wt	BMI	WCC	Bfat	LBM	Wt	BMI	WCC	Bfat
Mode	77.54	75.69	23.25	6.68	8.83	54.71	71.12	21.20	6.39	18.84
Mean	74.66	82.52	23.90	7.22	9.25	54.89	67.34	21.99	6.99	17.85
Variance	97.75	153.92	7.66	3.61	10.14	47.92	119.15	6.97	2.87	29.73
Skewness	0.27	0.39	1.41	0.86	1.53	− 0.31	− 0.17	0.69	0.75	0.35
Kurtosis	3.62	3.41	5.99	4.58	5.08	3.45	3.13	4.18	4.00	2.91

Marginal summary statistics of AIS data for males and females



(a) Pairwise scatter plots (pink for female, blue for male).

(b) Pairwise contours from the mlDir kernel density estimator.

(c) Classification from the mlDir mixture ($K = 2$).

(d) Misclassified observations (in red) by the mlDir mixture $K = 2$).

Fig. 6 Matrices of pairwise behaviors for the AIS data

smoothing parameter is determined according to the LCV approach illustrated in Sect. 3.2, highlight bimodality, particularly the pairs involving the Bfat variable.

Motivated by these preliminary findings, we fit a 2-component mIDir mixture to the data. The clusters from the fitted mIDir mixture are depicted in Fig. 6c. Since the sex of each athlete is known, we can evaluate the misclassification rate of the binary classification from the fitted mIDir mixture (refer to Fig. 6d). We compared our model-based clustering approach to other clustering methods, including *k*-means, hierarchical clustering, and density-based methods such as GB2*I*, GB2*II*, and Gaussian finite mixtures. The comparison was based on the misclassification rate and the adjusted Rand index (ARI; Hubert & Arabie, 1985), as presented in Table 8, which measures the agreement between the true and predicted classifications. An ARI of 1 indicates perfect clustering, while an ARI of 0 suggests the results are no better than random chance. Among the models evaluated, mIDir mixtures achieved a misclassification rate of 4.95%, corresponding to 10 misclassified observations, and a high ARI of 0.81. As shown in Fig. 6d, most of the misclassified observations are located in the overlapping region between the two clusters, demonstrating that mIDir mixtures are a viable alternative.

The parameter estimates for the mIDir mixture model are $\widehat{\boldsymbol{\pi}} = (0.54, 0.46)$, $\widehat{\boldsymbol{\theta}}_1 = (18.81, 22.93, 7.43, 2.29, 5.67)$, $\widehat{\boldsymbol{\theta}}_2 = (36.60, 40.05, 11.62, 3.33, 4.07)$, and $\widehat{\boldsymbol{\gamma}} = (0.77, 0.26)$, where $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_2$ represent the estimated mode-coordinates for males and females, respectively. Thus, $\widehat{\boldsymbol{\theta}}$ offers valuable insights into the approximate locations of the clusterwise modes. Furthermore, the estimates for $\boldsymbol{\gamma}$ suggest greater variability between male athletes than females, which agrees with the statistics in Table 7.

6 Discussion and conclusion

In this paper, we proposed a mode-based reparameterization of the inverted Dirichlet (IDir) distribution. From the application of the model in various branches of statistics including model-based clustering, nonparametric statistics as well as robust statistics, it becomes clear from several examples that this new parameterization simplifies the use of the IDir distribution in various statistical branches. In

Table 8 AIS dataset

Model	# Misclassified	Misclassification rate (%)	ARI
$GB2_I$	9	4.46	0.83
$GB2_{II}$	9	4.46	0.83
mIDir	10	4.95	0.81
k-means	34	16.83	0.44
Gaussian	45	22.28	0.30
Hierarchical	53	26.24	0.22

The number of misclassified observations (# Misclassified), Misclassification rate, and ARI for the competing clustering models

particular, we first proposed the use of the mode-parameterized inverted Dirichlet (mIDir) distribution for model-based clustering since a modal approach allows for a natural and convenient interpretation of the detected clusters. Secondly, a mIDir kernel smoother for nonparametric density estimation is introduced to effectively uncover structural features of the data. Thirdly, we proposed the contaminated inverted Dirichlet (cmIDir) distribution as a heavy-tailed generalization of the mIDir, which conveniently allows for the modelling of mild outliers in the data. An added benefit of the cmIDir is its ability to automatically identify mild outliers.

We employed direct numerical maximization for the maximum likelihood (ML) estimation to estimate the parameters of the mIDir distribution. A parameter recovery analysis was completed to demonstrate the successful parameter recovery of the proposed ML estimation approach. An expectation maximization (EM) algorithm was outlined for ML estimation of the mIDir mixtures and the cmIDir distribution, and a sensitivity analysis was also undertaken to illustrate the effectiveness of the cmIDir distribution in handling mild outliers and its outlier detection capabilities.

As an illustration, we fitted the mIDir to the spectra of different cultivars of cantaloupe, and data from the Australian Institute of Sport. The considered illustrative applications highlight the relevance of this model, but its applicability extends beyond these areas. The mIDir model may be further explored in statistical areas such as (multivariate) Bayesian analysis. The cmIDir is also identified as a viable alternative model, as it outperformed many of the considered distributions.

Due to the formulation of the mIDir distribution in (4), the covariance structure of the model is governed by only one parameter, γ, regardless of the dimension of the data. This leads to challenges in higher dimensions as the flexibility of accommodating this high-dimensional variability behaviour is potentially limiting (Hu et al., 2019). Particular mixture constructions, as considered in Ascari et al. (2021), Ferreira et al. (2022), and Ling et al. (2024), could be investigated and adapted for the inverse Dirichlet choice to facilitate the modelling of more general correlation behaviour that data may exhibit. Other generalized IDir distributions include the generalized IDir distribution (Lingmnwah, 1976) and the inverse beta Liouville distributions (Fang et al., 2018), which contains the IDir as a special case, which may be considered in the future for further extension and implementation.

Acknowledgements The authors express their gratitude towards the associate editor and two anonymous reviewers whose comments led to an improved paper.

Funding Open access funding provided by University of Pretoria. Otto, Ferreira, and Bekker have been partially supported by the National Research Foundation (NRF) of South Africa (SA), grant RA201125576565 & RA211204653274, nr 145681 & 151035; NRF ref. SRUG2204203865 & PMDS22063029311, and the Centre of Excellence in Mathematical and Statistical Sciences, based at the University of the Witwatersrand, Johannesburg (SA). The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

Data availability All datasets considered in this paper are freely available online.

Declarations

Conflict of interest The authors report there are no conflict of interest or competing interests to declare.

References

Adcock, C., & Azzalini, A. (2020). A selective overview of skew-elliptical and related distributions and of their applications. *Symmetry, 12*, 118.

Aitkin, M., & Wilson, G. T. (1980). Mixture models, outliers, and the EM algorithm. *Technometrics, 22*, 325–331.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

An, P., Wang, Z., & Zhang, C. (2022). Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection. *Information Processing & Management, 59*, 102844.

Arias-Castro, E., & Qiao, W. (2023). A unifying view of modal clustering. *Information and Inference: A Journal of the IMA, 12*, 897–920.

Ascari, R., Migliorati, S., & Ongaro, A. (2021). The double flexible Dirichlet: A structured mixture model for compositional data. *Applied Modeling Techniques and Data Analysis 2: Financial Demographic, Stochastic and Statistical Models and Methods, 8*, 135–152.

Atkinson, A. B., & Bourguignon, F. (2014). *Handbook of income distribution* (Vol. 2). Elsevier.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics, 12*, 171–178.

Azzalini, A. (2013). *The skew-normal and related families* (Vol. 3). Cambridge University Press.

Azzalini, A. (2015). Package 'sn'. *The skew-normal and skew-t distributions* (pp. 1–3). https://mirror.linux.duke.edu/cran/web/packages/sn/sn.pdf

Bagnato, L., & Punzo, A. (2013). Finite mixtures of unimodal beta and gamma densities and the k-bumps algorithm. *Computational Statistics, 28*, 1571–1597.

Bdiri, T., & Bouguila, N. (2011). An infinite mixture of inverted Dirichlet distributions. In *Neural Information Processing: 18th International Conference, ICONIP 2011, Shanghai, China, November 13–17, 2011, Proceedings, Part II* (Vol. 18, pp. 71–78). Springer.

Bdiri, T., & Bouguila, N. (2012). Positive vectors clustering using inverted Dirichlet finite mixture models. *Expert Systems with Applications, 39*, 1869–1882.

Bdiri, T., & Bouguila, N. (2013). Bayesian learning of inverted Dirichlet mixtures for SVM kernels generation. *Neural Computing and Applications, 23*, 1443–1458.

Biernacki, C., Celeux, G., & Govaert, G. (2003). Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. *Computational Statistics & Data Analysis, 41*, 561–575.

Carreira-Perpinan, M. A. (2000). Mode-finding for mixtures of Gaussian distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 1318–1323.

Chacón, J. E. (2019). Mixture model modal clustering. *Advances in Data Analysis and Classification, 13*, 379–404.

Chacón, J. E. (2020). The modal age of statistics. *International Statistical Review, 88*, 122–141.

Chen, M., & Ludtke, S. J. (2021). Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods, 18*, 930–936.

Chen, S. X. (1999). Beta kernel estimators for density functions. *Computational Statistics & Data Analysis, 31*, 131–145.

Chen, S. X. (2000). Probability density function estimation using gamma kernels. *Annals of the Institute of Statistical Mathematics, 52*, 471–480.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*, 790–799.
Dalenius, T. (1965). The mode-a neglected statistical parameter. *Journal of the Royal Statistical Society: Series A (General), 128*, 110–117.
Davies, L., & Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association, 88*, 782–792.
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 39*, 1–22.
Dubey, S. D. (1970). Compound gamma, beta and F distributions. *Metrika, 16*, 27–31.
Duda, R. O., Hart, P. E., et al. (1973). *Pattern classification and scene analysis* (Vol. 3). Wiley.
Fang, K. W., Kotz, S., & Ng, K. W. (2018). *Symmetric multivariate and related distributions*. CRC Press.
Fang, Y., Karlis, D., & Subedi, S. (2022). Infinite mixtures of multivariate normal-inverse Gaussian distributions for clustering of skewed data. *Journal of Classification, 39*, 510–552.
Ferreira, J. T., Botha, T., & Bekker, A. (2022). Tsallis and other generalised entropy forms subject to Dirichlet mixture priors. *Symmetry, 14*, 1110.
Fukunaga, K., & Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory, 21*, 32–40.
Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A journey beyond normality*. CRC Press.
Guo, J., Amayri, M., Najar, F., Fan, W., & Bouguila, N. (2023). Occupancy estimation in smart buildings using predictive modeling in imbalanced domains. *Journal of Ambient Intelligence and Humanized Computing, 14*, 10917–10929.
Hennig, C. (2002). Fixed point clusters for linear regression: Computation and comparison. *Journal of Classification, 19*, 249.
Hennig, C. (2004). Breakdown points for maximum likelihood estimators of location-scale mixtures. *The Annals of Statistics, 32*, 1313–1340.
Hu, C., Fan, W., Du, J. X., & Bouguila, N. (2019). A novel statistical approach for clustering positive data based on finite inverted Beta-Liouville mixture models. *Neurocomputing, 333*, 110–123.
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*, 193–218.
Hubert, M., Rousseeuw, P. J., & Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics, 21*, 618–637.
Hubert, M., & Van Driessen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics & Data Analysis, 45*, 301–320.
Kneib, T., Silbersdorff, A., & Säfken, B. (2021). Rage against the mean—A review of distributional regression approaches. *Econometrics and Statistics, 26*, 99–123.
Lai, C. D., & Balakrishnan, N. (2009). *Continuous bivariate distributions*. Springer.
Lai, Y., Ping, Y., He, W., Wang, B., Wang, J., & Zhang, X. (2018). Variational Bayesian inference for finite inverted Dirichlet mixture model and its application to object detection. *Chinese Journal of Electronics, 27*, 603–610.
Lee, M. J. (1989). Mode regression. *Journal of Econometrics, 42*, 337–349.
Lee, M. L. T., & Gross, A. J. (1991). Lifetime distributions under unknown environment. *Journal of Statistical Planning and Inference, 29*, 137–143.
Lee, S. X., & McLachlan, G. J. (2013). Model-based clustering and classification with non-normal mixture distributions. *Statistical Methods & Applications, 22*, 427–454.
Li, M., Xiang, S., & Yao, W. (2016). Robust estimation of the number of components for mixtures of linear regression models. *Computational Statistics, 31*, 1539–1555.
Ling, M. H., Balakrishnan, N., & Bae, S. J. (2024). On the application of inverted Dirichlet distribution for reliability inference of completely censored components with dependence structure. *Computers & Industrial Engineering, 196*, 110452.
Lingmnwah, G. (1976). On the generalised inverted Dirichlet distribution. *Demonstratio Mathematica, 9*, 119–130.
Liu, Q., Huang, X., & Zhou, H. (2022). The flexible Gumbel distribution: A new model for inference about the mode. arXiv preprint arXiv:2212.01832
Loperfido, N. (2024). The skewness of mean-variance normal mixtures. *Journal of Multivariate Analysis, 199*, 105242.
MacDonald, I. L. (2014). Numerical maximisation of likelihood: A neglected alternative to EM? *International Statistical Review, 82*, 296–308.

MacDonald, I. L. (2021). Is EM really necessary here? Examples where it seems simpler not to use EM. *AStA Advances in Statistical Analysis, 105*, 629–647.
Mazza, A., & Punzo, A. (2014). DBKGrad: An R package for mortality rates graduation by discrete beta kernel techniques. *Journal of Statistical Software, 57*, 1–18.
McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley.
McLachlan, G. J., Bean, R. W., & Peel, D. (2002). A mixture model-based approach to the clustering of microarray expression data. *Bioinformatics, 18*, 413–422.
Nadarajah, S., & Lyu, J. (2022). New bivariate and multivariate log-normal distributions as models for insurance data. *Results in Applied Mathematics, 14*, 100246.
Nayak, T. K. (1987). Multivariate Lomax distribution: Properties and usefulness in reliability theory. *Journal of Applied Probability, 24*, 170–177.
Ng, K. W., Tian, G. L., & Tang, M. L. (2011). *Dirichlet and related distributions: Theory, methods and applications*. Wiley.
Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., & Liu, J. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology, 586*, 124901.
Nolan, J. P. (1998). Parameterizations and modes of stable distributions. *Statistics & Probability Letters, 38*, 187–195.
Punzo, A. (2010). Discrete beta-type models. In H. Locarek-Junge & C. Weihs (Eds.), *Classification as a tool for research* (pp. 253–261). Springer.
Punzo, A. (2019). A new look at the inverse Gaussian distribution with applications to insurance and economic data. *Journal of Applied Statistics, 46*, 1260–1287.
Punzo, A., Bagnato, L., & Maruotti, A. (2018). Compound unimodal distributions for insurance losses. *Insurance: Mathematics and Economics, 81*, 95–107.
Punzo, A., Blostein, M., & McNicholas, P. D. (2020). High-dimensional unsupervised classification via parsimonious contaminated mixtures. *Pattern Recognition, 98*, 107031.
Punzo, A., Mazza, A., & Maruotti, A. (2018). Fitting insurance and economic data with outliers: A flexible approach based on finite mixtures of contaminated gamma distributions. *Journal of Applied Statistics, 45*, 2563–2584.
Punzo, A., & Tortora, C. (2021). Multiple scaled contaminated normal distribution and its application in clustering. *Statistical Modelling, 21*, 332–358.
R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
Ritter, G. (2014). *Robust cluster analysis and variable selection*. CRC Press.
Roberts, C. (1966). A correlation model useful in the study of twins. *Journal of the American Statistical Association, 61*, 1184–1190.
Sahai, H., & Anderson, R. (1973). Confidence regions for variance ratios of random models for balanced data. *Journal of the American Statistical Association, 68*, 951–952.
Sando, K., & Hino, H. (2020). Modal principal component analysis. *Neural Computation, 32*, 1901–1935.
Sarabia, J. M., Jordá, V., Prieto, F., & Guillén, M. (2020). Multivariate classes of GB2 distributions with applications. *Mathematics, 9*, 72.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.
Shih, M. F., & Hero, A. O. (2003). Unicast-based inference of network link delay distributions with finite mixture models. *IEEE Transactions on Signal Processing, 51*, 2219–2228.
Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC Press.
Singhal, A., Singh, P., Lall, B., & Joshi, S. D. (2020). Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. *Chaos, Solitons & Fractals, 138*, 110023.
Tiao, G. G., & Cuttman, I. (1965). The inverted Dirichlet distribution with applications. *Journal of the American Statistical Association, 60*, 793–805.
Tirdad, P., Bouguila, N., & Ziou, D. (2015). *Variational learning of finite inverted Dirichlet mixture models and applications*. Springer.
Todorov, V. (2009). *rrcov: Scalable robust estimators with high breakdown point. R package version 0.5-03*. http://CRAN.R-project.org/package=rrcov
Tomarchio, S. D., Bagnato, L., & Punzo, A. (2022). Model-based clustering via new parsimonious mixtures of heavy-tailed distributions. *AStA Advances in Statistical Analysis, 2022*, 1–33.
Tomarchio, S. D., Bagnato, L., & Punzo, A. (2023). Model-based clustering using a new multivariate skew distribution. *Advances in Data Analysis and Classification, 18*, 61–83.

Tomarchio, S. D., & Punzo, A. (2020). Dichotomous unimodal compound models: Application to the distribution of insurance losses. *Journal of Applied Statistics, 47*, 2328–2353.

Tomarchio, S. D., Punzo, A., Ferreira, J. T., & Bekker, A. (2024). A new look at the Dirichlet distribution: Robustness, clustering, and both together. *Journal of Classification, 2024*, 1–23.

Tortora, C., Franczak, B. C., Bagnato, L., & Punzo, A. (2024). A Laplace-based model with flexible tail behavior. *Computational Statistics and Data Analysis, 192*, 107909.

Vernic, R. (2006). Multivariate skew-normal distributions with applications in insurance. *Insurance:Mathematics and Economics, 38*, 413–426.

Yao, S., Kitahara, D., Kuroda, H., & Hirabayashi, A. (2023). Modal interval regression based on spline quantile regression. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 106*, 106–123.

Yao, W., & Li, L. (2014). A new regression model: Modal linear regression. *Scandinavian Journal of Statistics, 41*, 656–671.

Zhang, Y., Melnykov, V., & Melnykov, I. (2023). On model-based clustering of directional data with heavy tails. *Journal of Classification, 40*, 527–551.