Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2024;10(1): 111-123

Original Article

Bounded Multivariate Contaminated Normal Mixture Model with an Application in Skin Cancer Detection

Abbas Mahdavi*

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

ARTICLE INFO

Received 17.12.2023
Revised 11.01.2024
Accepted 28.02.2024
Published 15.03.2024

Key words:
ECME algorithm;
Mixture model;
Contaminated normal distribution;
Bounded distribution.

ABSTRACT

Introduction: In real-world datasets, outliers are a common occurrence that can have a significant impact on the accuracy and reliability of statistical analyses. Detecting these outliers and developing robust models to handle their presence is a crucial challenge in data analysis. For instance, natural images may have complex distributions of values due to environmental factors like noise and illumination, resulting in objects with overlapping regions and non-trivial contours that cannot be accurately described by Gaussian mixture models. In many real life applications, observed data always fall in bounded support regions. This leads to the idea of bounded support mixture models. Motivated by the aforementioned observations, we introduce a bounded multivariate cntaminated normal distribution for fitting data with non-Gaussian distributions, asymmetry, and bounded support which makes finite mixture models more robust to fitting, since rare observations are given less importance in calculations.

Methods: A family of finite mixtures of bounded multivariate contaminated normal distributions is introduced. The model is well-suited for computer vision and pattern recognition problems due to its heavily-tailed and bounded nature, providing flexibility in modeling data in the presence of outliers. A feasible expectation-maximization algorithm is developed to compute the maximum likelihood estimates of the model parameters using a selection mechanism.

Results: The proposed methodology is validated by conducting experiments on both simulated data and two real natural skin cancer images. We estimate the parameters by the proposed expectation-maximization algorithm. The obtained results shown that the proposed model has successfully enhanced accuracy in segmenting skin lesions.

Conclusion: The reliable model-based clustering using finite mixtures of bounded multivariate contaminated normal distributions is introduced. An expectation-maximization algorithm was created to estimate parameters, with closed-form expressions utilized at the E-step. Practical tests on images for skin cancer detection showed enhanced accuracy in delineating skin lesions.

Introduction

In real-world datasets, outliers are a common occurrence that can have a significant impact on the accuracy and reliability of statistical analyses. Detecting these outliers

*.Corresponding Author: a.mahdavi@vru.ac.ir

and developing robust models to handle their presence is a crucial challenge in data analysis. This is particularly true for model-based clustering methods, where even a small number of outliers can lead to biased estimates, incorrect classifications, and overfitting of the number of groups. For instance, natural images may have complex distributions of values due to environmental factors like noise and illumination, resulting in objects with overlapping regions and non-trivial contours that cannot be accurately described by Gaussian mixture models.

Tukey[1] introduced the concept of a multivariate contaminated normal (MCN) distribution, which consists of two normal mixture components. The first component has a high prior probability and represents the reference cluster distribution of good observations. The second component, with a small prior probability, has the same mean as the first component but an inflated covariance matrix, and represents the bad observations.[2] The probability density function (pdf) of MCN distribution, denoted by

$\mathbf{X} \sim \text{MCN}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2)$ is given by

$$f_{MCN}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2)=\nu_1\ \phi_p\ (\mathbf{x};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})+(1-\nu_1)\phi_p(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma}),\ x\in\mathbb{R}^p, \quad (1)$$

where ϕ_p (.; $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$) denotes the pdf of the p-variate normal random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Parameter $0<\nu_1<1$, can be interpreted as the proportion of outliers and $0<\nu_2<1$, is interpreted as a scale factor. Despite retaining its elliptical symmetry, the MCN distribution can exhibit long kurtosis and heavy tails through its shape parameters, ν_1 and ν_2, making it a suitable option for numerous modeling scenarios. Furthermore, if we set $\nu_2=1$, it reduces to the multivariate normal (MN) distribution.

It is worth mentioning that the MCN distribution can be expressed by a convolution stochastic representation of the MN distribution as

$$\mathbf{X} \overset{d}{=} \boldsymbol{\mu}+\tau^{-1/2}\ \boldsymbol{\Sigma}^{1/2}\ Z \sim \text{MCN}(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2) \quad (2)$$

where Z follows a p-variate standard normal distribution and τ is a discrete random variable taking one of two states with the following pdf:

$$f_\tau(\tau;\nu_1,\nu_2)=\nu_1\mathbb{I}_{(\tau=\nu_2)}+(1-\nu_1)\mathbb{I}_{(\tau=1)},$$

$$0<\nu_1<1, 0<\nu_2<1, \quad (3)$$

where $\mathbb{I}_{(.)}$ denotes the indicator function.

In many real life applications, observed data always fall in bounded support regions. This leads to the idea of bounded support mixture models. The bounded normal mixture (BNM) and bounded *t* mixture (BTM) models are widely used techniques for modeling observed data with bounded support regions. These models have been extensively used; for example, in signal processing,[3,4] climate data,[5] and image computer vision.[6-9] In all these cases, the observed data are usually in limited range and are not necessarily symmetric within their bounds. However, in many applications, the tail of the BNM model is shorter than required. Also, the BTM is not flexible enough to fit the shape of the data in the presence of the outliers. Outliers may also lead to overestimation or underestimation of the variability of the data, leading to a poor fit of the Gaussian distribution. One approach to address this issue involves is to consider generalizations of Gaussians with heavy tails such as t, power exponential

and contaminated normal distributions. The contaminated normal distribution has heavier tails compared to the normal distribution, which means it assigns more probability to extreme values. This makes it less sensitive to outliers, as extreme values have less impact on the overall shape of the distribution.[10,11] Therefore, the MCN distribution provides a more robust approach to the fitting of normal mixture models, as observations that are atypical of a component are given reduced weight in the calculation of its parameters. Also, the use of MCN components gives less extreme estimates of the posterior probabilities of component membership of the mixture model. Motivated by the aforementioned observations, we introduce a bounded MCN distribution (BMCN) for fitting data with non-Gaussian distributions, asymmetry, and bounded support which makes finite mixture models more robust to fitting, since rare observations are given less importance in calculations.

Dempster et al.[12] proposed the EM algorithm for the estimating maximum likelihood (ML) estimate of model parameters in the presence of censored and truncated data. Atkinson[13] introduced an EM algorithm for a finite mixture of two normal distributions with right-censored data, while McLachlan and Jones[14] developed an EM algorithm for univariate binned and truncated data. More recently, there have been many works on modeling bounded data. For instance, Nguyen et al.[15-16] proposed an EM algorithm to estimate the parameters of mixtures of bounded generalized normal models, while Azam and Bouguila[17] introduced a mixture of bounded Laplace distribution and suggested a maximum likelihood approach for parameter estimation, with parameter optimization performed by an EM algorithm. Similar studies investigating the effectiveness of EM algorithm for estimation of model parameters in the presence of truncated or bounded data can be found in Lee and Scott[18] and Yu et al.[19]

Inspired by the works of Mahdavi et al.,[20,21] we study now the theoretical framework of EM-type algorithm via selection mechanism for parameter estimation of the BMCN model. To demonstrate the performance of the proposed algorithm, we have successfully applied it to skin lesion images. The proposed model has successfully enhanced accuracy in segmenting skin lesions, showcasing the potential of the developed method in image analysis and image processing. Overall, the proposed algorithm offers a reliable and efficient way to model many real-world problems, especially in the field of image analysis.

Methods

The BMCN distribution

Let $\boldsymbol{\Omega} = (\mathbf{a}, \mathbf{b}) \subseteq \mathbb{R}^p$ be a measurable set. Then, the random vector $\mathbf{y} \in \boldsymbol{\Omega}$ is said to follow BMCN distribution, with support $\boldsymbol{\Omega}$ and parameters $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, $0 < \nu_1 < 1$ and $0 < \nu_2 < 1$, if its pdf is

$$f_{BMCN}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2) = \frac{f_{MCN}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2)}{\int_{\boldsymbol{\Omega}} f_{MCN}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_1, \nu_2) d\mathbf{y}} \quad (4)$$
$$= \frac{\nu_1 \phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma}) + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})}{\nu_1 \int_{\boldsymbol{\Omega}} \phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma}) d\mathbf{y} + (1-\nu_1) \int_{\boldsymbol{\Omega}} \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma}) d\mathbf{y}}.$$

Evidently, the MCN distribution is recovered if we set $\boldsymbol{\Omega} = \mathbb{R}^p$.

From (2), it is easy to show that the BMCN

distribution admits the following stochastic selection representation:

$$\boldsymbol{Y} \overset{d}{=} \boldsymbol{X} \mid (\boldsymbol{X} \in \boldsymbol{\Omega}), \tag{5}$$

where $\boldsymbol{X}=\boldsymbol{\mu}+\tau^{-1/2}\,\boldsymbol{\Sigma}^{1/2}\,\boldsymbol{Z}$, $\boldsymbol{Z} \sim N_p(\mathbf{0},\mathbf{I}_p)$ and $\tau \sim f_\tau(\tau;\nu_1,\nu_2)$ defined in (3). The expression (5) will be useful in generating random numbers from this distribution and also in studying its theoretical properties. Moreover, it is easy to see $\mathbf{X}|\tau \sim N_p(\boldsymbol{\mu},\tau^{-1}\boldsymbol{\Sigma})$.
To capture some features of selection variable **Y**, we introduce an additional latent variable $\boldsymbol{\gamma} \overset{d}{=} \boldsymbol{\tau} \mid (\boldsymbol{X} \in \boldsymbol{\Omega})$. Combining **Y** and the latent variable γ together, we get complete variable $(\mathbf{Y},\gamma)^\top \overset{d}{=} (\mathbf{X},\tau)^\top \mid (\mathbf{X} \in \boldsymbol{\Omega})$ with the following joint density:

$$\begin{aligned} f_{\mathbf{Y},\gamma}(\mathbf{y},\gamma) &= \frac{f_{\mathbf{X},\tau}(\mathbf{y},\gamma)}{Pr(\mathbf{X} \in \boldsymbol{\Omega})} \\ &= \frac{f_\tau(\gamma) f_{\mathbf{X}|\tau}(\mathbf{y})}{\int_{\boldsymbol{\Omega}} f_{MCN}(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2)d\mathbf{y}} \\ &= \frac{\phi_p(\mathbf{y};\boldsymbol{\mu},\gamma^{-1}\boldsymbol{\Sigma}) f_\tau(\gamma;\nu_1,\nu_2)}{\nu_1 \int_{\boldsymbol{\Omega}} \phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})d\mathbf{y} + (1-\nu_1)\int_{\boldsymbol{\Omega}} \phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})d\mathbf{y}}, \\ & \qquad \boldsymbol{y} \in \boldsymbol{\Omega}, \gamma > 0. \end{aligned} \tag{6}$$

From (4) and (6), it is easy to show that

$$f_\gamma(\gamma|\mathbf{y}) = \frac{\phi_p(\mathbf{y};\boldsymbol{\mu},\gamma^{-1}\boldsymbol{\Sigma}) f_\tau(\gamma;\nu_1,\nu_2)}{\nu_1 \phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma}) + (1-\nu_1)\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})} \tag{7}$$

In the following subsection, we propose an EM-type algorithm for determining the ML estimates of the parameters of *BMCN*($\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2$) based on the joint density in (6).

Parameter estimation via the EM-type algorithm

The EM algorithm is a widely used iterative algorithm for ML estimation in the presence of missing, censored or latent variables. The algorithm alternates between an E-step, where the expected value of the complete log-likelihood function is computed with respect to the conditional distribution of the latent variables, given the observed data and the current estimates of the model parameters, and a M-step, where the estimates of the model parameters are updated by maximizing the so called Q-function obtained from the E-step. The EM algorithm guarantees convergence to a maxima, of the likelihood function, but it can be slow to converge or can get stuck in local optima.
To address some of these issues, several variants of the EM algorithm have been proposed, such as the Expectation Conditional Maximization[22] (ECM) algorithm and the Expectation Conditional Maximization Either[23] (ECME) algorithm. The ECM algorithm replaces the M-step with a series of conditional maximization steps (CM) of the Q-function with respect to each parameter, while holding the others fixed. The ECME algorithm extends the ECM algorithm by replacing some CM-steps of ECM with the CML-step that maximizes the corresponding constrained actual-likelihood function, which can lead to faster convergence and improved estimation accuracy. These algorithms have been successfully applied in many problems in statistics, machine learning, and signal processing.
Let $\mathbf{y}=(\mathbf{y}_1,\ldots,\mathbf{y}_n)^\top$ denote the observed data, $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_n)^\top$ be the latent variable

and $\mathbf{y}_c=(\mathbf{y}^\top,\boldsymbol{\gamma}^\top)^\top$ be the complete data. From (6), the log-likelihood function of $\boldsymbol{\theta}=(\boldsymbol{\mu},\boldsymbol{\Sigma},\nu_1,\nu_2)^\top$ corresponding to the complete-data $\mathbf{y}_c$, excluding additive constants and terms that do not involve the parameters of the model, is given by

$$\ell_c(\boldsymbol{\theta}|\mathbf{y}_c) = \sum_{i=1}^{n}\{-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{\gamma_i}{2}(\mathbf{y}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu}) + \log f_\tau(\gamma_i;\nu_1,\nu_2) - \log(\nu_1\int_{\boldsymbol{\Omega}}\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})d\mathbf{y} + (1-\nu_1)\int_{\boldsymbol{\Omega}}\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})d\mathbf{y})\}. \quad (8)$$

This function can be maximized to obtain the estimate of $\boldsymbol{\theta}$. However, a problem arises because the estimate is dependent on the latent variables, rendering them unusable. For this reason, in the E-step of the EM algorithm, we evaluate the so-called Q-function, which is the conditional expectation of (8), given the observed data using the curr ent estimates of the model parameters, $\hat{\theta}^{(k)}$ where the superscript (k) denotes the estimate of $\boldsymbol{\theta}$ at k-th iteration. To evaluate the Q-function, we require the following conditional expectations:

$$\hat{\gamma}_i^{(k)} = E(\gamma_i|\mathbf{Y}_i=\mathbf{y}_i,\widehat{\boldsymbol{\theta}}^{(k)}) = \frac{1-\nu_1+\nu_1\nu_2^{1+p/2}\exp\{(1-\nu_2)(\mathbf{y}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})/2\}}{1-\nu_1+\nu_1\nu_2^{p/2}\exp\{(1-\nu_2)(\mathbf{y}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu})/2\}}, \quad (9)$$

which is resulting from (7), along with

$$\hat{s}_i^{(k)} = E(\log h_\tau(\gamma_i;\nu_1,\nu_2)|\mathbf{Y}_i=\mathbf{y}_i,\widehat{\boldsymbol{\theta}}^{(k)}). \quad (10)$$

We can then form the objective function $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$ as:

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = \sum_{i=1}^{n}\{-\frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\hat{\gamma}_i^{(k)}(\mathbf{y}_i-\boldsymbol{\mu})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i-\boldsymbol{\mu}) + \hat{s}_i^{(k)} - \log(\nu_1\int_{\boldsymbol{\Omega}}\phi_p(\mathbf{y};\boldsymbol{\mu},\nu_2^{-1}\boldsymbol{\Sigma})d\mathbf{y} + (1-\nu_1)\int_{\boldsymbol{\Omega}}\phi_p(\mathbf{y};\boldsymbol{\mu},\boldsymbol{\Sigma})d\mathbf{y})\} \quad (11)$$

After performing some algebraic manipulations, the CM-steps can be implemented as follows:

CMQ-step 1: Fix $\boldsymbol{\Sigma}=\widehat{\boldsymbol{\Sigma}}^{(k)}$, we update $\widehat{\boldsymbol{\mu}}^{(k)}$ by maximizing (11) with respect to $\boldsymbol{\mu}$ which leads to

$$\widehat{\boldsymbol{\mu}}^{(k+1)} = \frac{1}{\sum_{i=1}^n\hat{\gamma}_i^{(k)}}(\sum_{i=1}^n\hat{\gamma}_i^{(k)}\mathbf{y}_i - n\kappa_1(\hat{\theta}^{(k)};\boldsymbol{\Omega}^*)), \quad (12)$$

where

$$\kappa_1(\hat{\theta}^{(k)},\boldsymbol{\Omega}^*) = \frac{\alpha_1\nu_1\nu_2M^1(\mathbf{0},\nu_2^{-1}\boldsymbol{\Sigma};\boldsymbol{\Omega}^*)+\alpha_2(1-\nu_1)M^1(\mathbf{0},\boldsymbol{\Sigma};\boldsymbol{\Omega}^*)}{\nu_1\alpha_1+(1-\nu_1)\alpha_2}, \quad (13)$$

here we introduce the notation $M^1(\boldsymbol{\mu},\boldsymbol{\Sigma};\mathbb{A})$ to denote the first moment of a truncated multivariate normal distribution on $\mathbb{A}$ with location parameter $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\alpha_1=P(\mathbf{X}_1\in\boldsymbol{\Omega}^*)$, $\alpha_2=P(\mathbf{X}_2\in\boldsymbol{\Omega}^*)$ where $\mathbf{X}_1\sim N_p(\mathbf{0},\nu_2^{-1}\boldsymbol{\Sigma})$, $\mathbf{X}_2\sim N_p(\mathbf{0},\boldsymbol{\Sigma})$ and $\boldsymbol{\Omega}^*=(\mathbf{a}-\boldsymbol{\mu},\mathbf{b}-\boldsymbol{\mu})$.

CMQ-step 2: Fixing $\boldsymbol{\mu}=\widehat{\boldsymbol{\mu}}^{(k+1)}$, we can then update $\widehat{\boldsymbol{\Sigma}}^{(k)}$ by maximizing (11) over $\boldsymbol{\Sigma}$. This yields

$$\widehat{\boldsymbol{\Sigma}}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\gamma}_i^{(k)}(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k+1)})(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}^{(k+1)})^{\top} + \kappa_2(\hat{\theta}^{(k)}, \boldsymbol{\Omega}^*),$$

where

$$\kappa_2(\hat{\theta}^{(k)}, \boldsymbol{\Omega}^*) = \boldsymbol{\Sigma}^{(k)} - \frac{\alpha_1\nu_1\nu_2 M^2(\mathbf{0},\nu_2^{-1}\boldsymbol{\Sigma};\boldsymbol{\Omega}^*) + \alpha_2(1-\nu_1)M^2(\mathbf{0},\boldsymbol{\Sigma};\boldsymbol{\Omega}^*)}{\nu_1\alpha_1 + (1-\nu_1)\alpha_2}, \quad (14)$$

where $M^2(\boldsymbol{\mu},\boldsymbol{\Sigma};\mathbb{A})$ denotes the second moment of a truncated multivariate normal distribution on $\mathbb{A}$ with location parameter μ and covariance matrix $\boldsymbol{\Sigma}$.

Updating $\hat{\nu}_1^{(k)}$ and $\hat{\nu}_2^{(k)}$ require a bi-dimensional search for the root of ν_1 and ν_2. As pointed out by Liu and Rubin[21], finding the roots can be a slow process in some situations. To avoid this obstacle and also eliminate the need for calculating the conditional expectation $\hat{s}_i^{(k)}$, we may resort to maximizing the restricted actual log-likelihood function, yielding the following step instead:

CML-step: Update $\hat{\nu}_1^{(k)}$ and $\hat{\nu}_2^{(k)}$ by optimizing the following constrained log-likelihood function:

$$\left(\hat{\nu}_1^{(k+1)}, \hat{\nu}_2^{(k+1)}\right) = \arg\max_{(\nu_1,\nu_2)} \sum_{i=1}^{n} \log f_{BMCN}(\mathbf{y}_i; \widehat{\boldsymbol{\mu}}^{(k+1)}, \widehat{\boldsymbol{\Sigma}}^{(k+1)}, \nu_1, \nu_2).$$

The iterations of the above algorithm must be performed repeatedly until a predetermined convergence criterion is met. In this paper, this criterion is defined in terms of the relative change in the log-likelihood function, specifically given by the expression $||\ell(\widehat{\boldsymbol{\theta}}^{(k+1)})/\ell(\widehat{\boldsymbol{\theta}}^{(k)}) - 1||$, where $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_{BMCN}(y_i; \boldsymbol{\theta})$.

The goal is for this value to become sufficiently small, indicating that the algorithm is converging toward a stable solution. For practical purposes, we can set this threshold to be 10^{-6}, meaning that we will continue to iterate until the relative change in the log-likelihood falls below this level.

Finite mixtures of bounded multivariate contaminated normal distributions (FM-BMCN)

Let $\mathbf{Y}_1,\ldots,\mathbf{Y}_n$ be a random sample from a G-component mixture of BMCN distributions. The pdf of this mixture model is

$$f(y_i; \boldsymbol{\Theta}) = \sum_{g=1}^{G} \pi_g f_{BMCN}(y_i; \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_{1g}, \nu_{2g}),$$

where π_g represents the mixing probability with $0 \le \pi_g \le 1$, $\sum_{g=1}^{G} \pi_g = 1$ and $\Theta = (\pi_1,\ldots,\pi_{G-1},\boldsymbol{\theta}_1^{\top},\ldots,\boldsymbol{\theta}_G^{\top})^{\top}$ denotes the vector of parameters, with $\boldsymbol{\theta}_g = (\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g,\nu_{1g},\nu_{2g})^{\top}$ denoting the parameters of component g.

We introduce a set of membership component indicators $\boldsymbol{Z}_i = (Z_{i1},\ldots,Z_{iG})^{\top}$, which are all binary variables. In other words, if this element is in the jth position of vector $\boldsymbol{Z}_i$, then $\boldsymbol{Y}_i$ is allocated to component G. The distribution of $\boldsymbol{Z}_i$ is multinomial distribution with 1 trial and cell probabilities $\pi_1,\ldots,\pi_G$ and it is denoted denoted by

$$\boldsymbol{Z}_i \sim Mult(1;\pi_1,\ldots,\pi_G),$$

$$\boldsymbol{Y}_i | Z_{ig}=1 \sim \text{BMCN}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g, \nu_{1g}, \nu_{2g}),$$

$$f_{\gamma_i|\mathbf{Y}_i=\mathbf{y}_i,Z_{ig}=1}(\gamma_i) = \frac{\phi_p(\mathbf{y}_i;\boldsymbol{\mu}_g,\gamma_i^{-1}\boldsymbol{\Sigma}_g) f_\tau(\gamma_i;\nu_{1g},\nu_{2g})}{\nu_{1g}\phi_p(\mathbf{y}_i;\boldsymbol{\mu}_g,\nu_{2g}^{-1}\boldsymbol{\Sigma}_g)+(1-\nu_{1g})\phi_p(\mathbf{y}_i;\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)}. \quad (15)$$

Maximum likelihood estimates of FM-BMCN parameters

From (15), the log-likelihood function of Θ, ignoring additive constants, based on the complete data including the observed data $\mathbf{y}=(y_1,\ldots,y_n)^\top$ and latent variables $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_n)^\top$ and $\mathbf{Z}=(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)^\top$, is

$$\begin{aligned}\ell_c(\boldsymbol{\Theta}|\mathbf{y}_c) = \sum_{i=1}^n \sum_{g=1}^G Z_{ig}\{&\log\pi_g - \frac{1}{2}\log \\ &-\frac{1}{2}\log|\boldsymbol{\Sigma}_g| - \frac{\gamma_i}{2}(\mathbf{y}_i-\boldsymbol{\mu}_g)^\top\boldsymbol{\Sigma}_g^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_g) \\ &+\log f_\tau(\gamma_i;\nu_{1g},\nu_{g2}) - \\ &\log(\nu_{1g}\int_{\boldsymbol{\Omega}_g}\phi_p(\mathbf{y};\boldsymbol{\mu}_g,\nu_{2g}^{-1}\boldsymbol{\Sigma}_g)d\mathbf{y} + \\ &+(1-\nu_{1g})\int_{\boldsymbol{\Omega}_g}\phi_p(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)d\mathbf{y})\}, \end{aligned} \quad (16)$$

where $\boldsymbol{\Omega}_g=(\mathbf{a}_g,\mathbf{b}_g)$ is the support of the gth component.

The conditional expectations involved in the Q-function, given current parameter $\boldsymbol{\Theta}^{(k)}$, that are required are

$$\hat{z}_{ig}^{(k)} = E(Z_{ig}|\mathbf{y}_i,\widehat{\boldsymbol{\Theta}}^{(k)}) = \frac{\hat{\pi}_j^{(k)} f_{BMCN}(\mathbf{y}_i;\widehat{\boldsymbol{\theta}}_g^{(k)})}{f(y_i;\widehat{\boldsymbol{\Theta}}^{(k)})}, \quad (17)$$

$$\hat{\gamma}_{ig}^{(k)} = E(\gamma_i|\mathbf{y}_i,Z_{ig}=1,\widehat{\boldsymbol{\Theta}}^{(k)}) = E(\gamma_i|y_i,\widehat{\boldsymbol{\theta}}_g^{(k)}) \quad (18)$$

$$\hat{s}_{ig}^{(k)} = E(\log h_\tau(\gamma_i;\nu_{1g},\nu_{2g})|\mathbf{y}_i,Z_{ig} = 1,\widehat{\boldsymbol{\Theta}}^{(k)}) = E(\log\gamma_i|\mathbf{y}_i,\widehat{\boldsymbol{\theta}}_g^{(k)}). \quad (19)$$

The resulting Q function is given by

$$\begin{aligned}Q(\boldsymbol{\Theta}|\widehat{\boldsymbol{\Theta}}^{(k)}) = \sum_{i=1}^n \sum_{g=1}^G &\hat{z}_{ig}^{(k)}\{\log\pi_g - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| \\ &-\frac{\gamma_{ig}}{2}(\mathbf{y}_i-\boldsymbol{\mu}_g)^\top \\ &\boldsymbol{\Sigma}_g^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_g)+\hat{s}_{ig}^{(k)} - \\ &\log(\nu_{1g}\int_{\boldsymbol{\Omega}_g}\phi_p(\mathbf{y};\boldsymbol{\mu}_g,\nu_{2g}^{-1}\boldsymbol{\Sigma}_g)d\mathbf{y} + \\ &(1-\nu_{1g})\int_{\boldsymbol{\Omega}_g}\phi_p(\mathbf{y};\boldsymbol{\mu}_g,\boldsymbol{\Sigma}_g)d\mathbf{y})\}. \end{aligned} \quad (20)$$

With all these, the implementation of the ECME algorithm proceeds as follows:

E-step: Given $\boldsymbol{\Theta}=\hat{\Theta}^{(k)}$, compute $\hat{z}_{ij}^{(k)}$, $\hat{\gamma}_{ig}^{(k)}$ and $\hat{s}_{ig}^{(k)}$ for $i=1,\ldots,n$ and $g=1,\ldots,G$ using (17)-(19);

CM-steps: Update $\hat{\Theta}^{(k)}$ by maximizing $Q(\boldsymbol{\Theta}|\hat{\Theta}^{(k)})$ over $\boldsymbol{\Theta}$ which leads to

$$\hat{\pi}_g^{(k+1)} = \frac{1}{n}\sum_{i=1}^n \hat{z}_{ig}^{(k)},$$

$$\hat{\boldsymbol{\mu}}_g^{(k+1)} = \frac{1}{\sum_{i=1}^n \hat{z}_{ij}^{(k)}\hat{\gamma}_{ig}^{(k)}}\Big(\sum_{i=1}^n \hat{z}_{ij}^{(k)}\hat{\gamma}_{ig}^{(k)}\mathbf{y}_i - \kappa_1(\hat{\theta}_g^{(k)};\boldsymbol{\Omega}_g^*)\sum_{i=1}^n \hat{z}_{ij}^{(k)}\Big),$$

$$\hat{\boldsymbol{\Sigma}}_g^{(k+1)} = \frac{1}{\sum_{i=1}^n \hat{z}_{ij}^{(k)}}\sum_{i=1}^n \hat{z}_{ij}^{(k)}\hat{\gamma}_{ig}^{(k)}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_g^{(k+1)})(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_g^{(k+1)})^\top + \kappa_2(\hat{\theta}_g^{(k)},\boldsymbol{\Omega}_g^*),$$

Where

$$\boldsymbol{\Omega}_g^* = (\mathbf{a}_g - \boldsymbol{\mu}_g, \mathbf{b}_g - \boldsymbol{\mu}_g)$$

CML-step: Update $\hat{\nu}_{1g}^{(k)}$ and $\hat{\nu}_{2g}^{(k)}$ by optimizing constrained log-likelihood function, given by

$$(\hat{\nu}_{1g}^{(k+1)}, \hat{\nu}_{2g}^{(k+1)}) = \arg\max_{(\nu_{1g},\nu_{2g})} \sum_{i=1}^{n} \log \Big(\sum_{g=1}^{G} \hat{\pi}_g^{(k+1)} f_{BMCN}(y_i; \hat{\boldsymbol{\mu}}_g^{(k+1)}, \hat{\boldsymbol{\Sigma}}_g^{(k+1)}, \nu_{1g}, \nu_{2g}) \Big).$$

Results

A simulation study: recovery of the true underlying parameters

In order to validate the efficacy of the proposed algorithm, we have carried out a comprehensive simulation study using various sample sizes. The primary objective here is to assess the accuracy of the estimates of the true parameter values. For this propose, we used 500 Monte Carlo (MC) samples of sizes n =100, 250, 500, and 1000 from a two-component FM-BMCN distribution under the following scenario: data are simulated from FM-BMCN model with component parameters π_1=0.3, $\boldsymbol{\mu}_1$=(1,-1), $\boldsymbol{\sigma}_1$=vech($\boldsymbol{\Sigma}_1$)=(1,0.5,4), $\boldsymbol{\nu}_1$=(ν_{11},ν_{21})=(0.2,0.7) and $\boldsymbol{\mu}_2$=(3,-2), $\boldsymbol{\sigma}_2$=vech($\boldsymbol{\Sigma}_2$)=(1,0,1), $\boldsymbol{\nu}_2$=(ν_{12}, ν_{22})=(0.5,0.5), were vech(.) is the half-vectorization operator that stacks the lower triangular elements of a $p \times p$ symmetric matrix into a single $p(p+1)/2$ vector. We use the mean absolute bias (MAB) and square root of mean squared error (RMSE) as efficiency criteria. For a vector of parameters $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)$, these measures are, respectively, defined as

$$\text{MAB} = \frac{1}{pR} \sum_{k=1}^{p} \sum_{r=1}^{R} \left| \hat{\theta}_{kr} - \theta_k^A \right|,$$

and

$$\text{RMSE} = \sqrt{\frac{1}{pR} \sum_{k=1}^{p} \sum_{r=1}^{R} \left(\hat{\theta}_{kr} - \theta_k^A \right)^2}.$$

where $\hat{\theta}_{kr}$ denotes the ML estimate of the k-th parameter at the r-th replication and θ_k^A represents the actual value of θ_k.

The experimental results are summarized in Table 1, which indicate that the MAB and RMSE values tend to decrease with increasing sample size. This finding suggests that the proposed ECME algorithm can provide reliable estimates for the FM-BMCN model. Moreover, estimating the $\boldsymbol{\nu}$=(ν_1,ν_2) parameter is complicated due to the flatness of the likelihood with respect to it, and this results in greater variation.

An application in skin cancer detection

One of the important steps in image analysis is medical image segmentation. The objective of the skin cancer detection project is to develop a framework to analyze and assess the risk of melanoma using dermatological photographs taken with a standard consumer-grade camera. Segmentation of the lesion is a crucial step to develop a skin cancer detection framework. The objective is to find the border of the skin lesion. It is important that this step is performed accurately because many features

used to assess the risk of melanoma are derived based on the lesion border. The set of images includes images extracted from the public databases DermIS and DermQuest, along with manual segmentations (ground truth) of the lesions is available at
https://uwaterloo.ca/vision-image-processing-lab/research-demos/skin-cancer-detection.
Two real skin images are displayed in Figures 1 (a) and 2 (a). The objective now is to segment these images in two labels. Each pixel will have a three RGB intensities between 0 and 255 which can be transformed into $\boldsymbol{\Omega}$=($\mathbf{a}$=(0,0,0),$\mathbf{b}$=(1,1,1)). The darker the color, the lower the intensity, and vice versa. Each pixel will then be grouped into G=2 clusters, where every cluster
will be assumed to have a different distribution. Here, we supposes the pixels are independent and does not take into account the spatial relationship among the pixels.
For comparative purpose, we used the

.Table 1. Simulation results based on 500 replications with different sample sizes

Parameter	n=100		n=250		n=500		n=1000	
	MAB	RMSE	MAB	RMSE	MAB	RMSE	MAB	RMSE
π_1	0.108	0.116	0.084	0.103	0.065	0.088	0.054	0.072
$\boldsymbol{\mu}_1$	0.092	0.125	0.057	0.077	0.037	0.048	0.028	0.033
$\boldsymbol{\mu}_2$	0.115	0.157	0.068	0.092	0.043	0.055	0.033	0.045
$\boldsymbol{\sigma}_1$	0.429	0.642	0.306	0.489	0.191	0.314	0.127	0.199
$\boldsymbol{\sigma}_2$	0.259	0.382	0.143	0.209	0.115	0.168	0.081	0.125
$\boldsymbol{\nu}_1$	0.652	0.810	0.430	0.621	0.266	0.366	0.206	0.265
$\boldsymbol{\nu}_2$	0.532	0.733	0.424	0.573	0.259	0.354	0.201	0.259

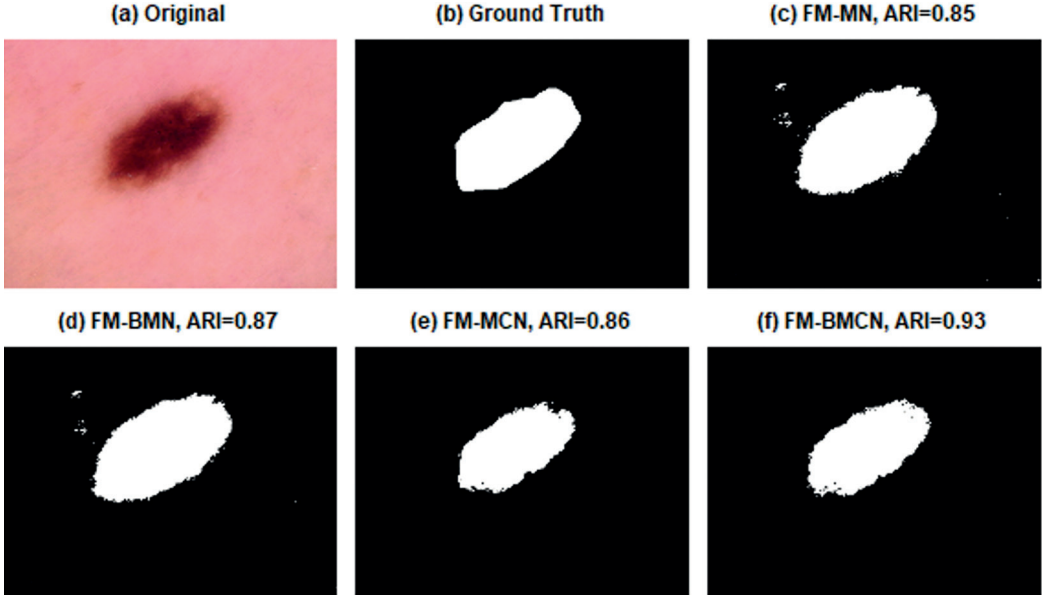

Figure 1. Segmentation of lesion (image no.ISIC_0024339): (a) Original, (b) ground truth, (c)-(f) segmented images obtained using different models.

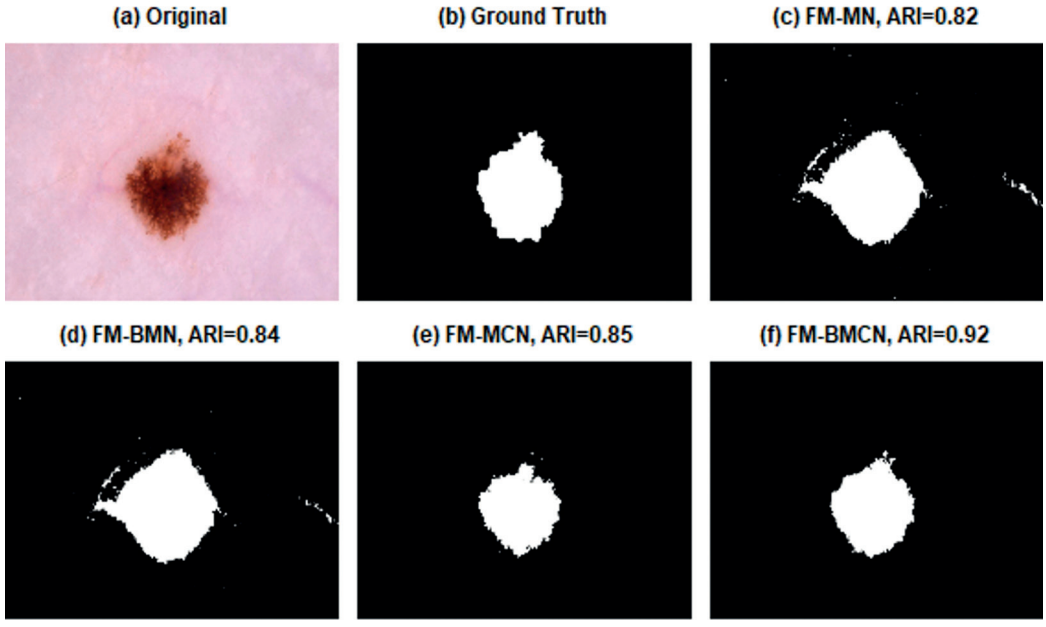

Figure 2. Segmentation of lesion (image no.ISIC_0024352): (a) Original, (b) ground truth, (c)-(f) segmented images obtained using different models.

mixsmsn R package[24,25] to fit finite mixtures of normal (FM-NM) distributions. Additionally, we employed the EM-based algorithms of Lee and Scott[18] to fit finite mixtures of bounded normal (FM-BNM) distribution. The finite mixtures of MCN (FM-MCN) distributions is also fitted as sub-model of the proposed FM-BMCN model. Figs 1 (c)-(f) and Figs 2 (c)-(f) reveal skin lesion can be caught when segmentation is conducted for 2 clusters. However, the segmentation accuracies of the normal mixture models including FM-NM and FM-BNM are quite poor. The proposed method yields a better segmentation result, with the highest Adjusted Rand Index[26] (ARI).

Discussion

Mixture models have been widely used in machine learning and pattern recognition for statistical modeling. These models have proven to be valuable in various applications, such as speech and image processing. However, one limitation of these models is that their distributions are unbounded, with a support range of $(-\infty;+\infty)$. In many real-life scenarios, observed data falls within bounded support regions. For instance, in image analysis, each pixel has a grayscale intensity between 0 and 255, which can be transformed into the range (0;1). Each pixel can then be grouped into clusters, with each cluster assumed to have a different distribution.

However real data, in addition to being characterized by underlying asymmetric clusters, are often "contaminated" by outliers or otherwise "bad" points. The "bad" points refers to points that have a deleterious effect on parameter estimation. Thus an important practical application is the development of methods capable of detecting bad points and performing robust parameter estimation when they are present.

This paper introduced mixtures of BMCN

distributions as a model-based clustering method for handling asymmetric clusters under the presence of outliers. Our research also explores the parameter estimation for the finite mixtures of BMCN distributions. To demonstrate the performance of the proposed algorithm, we have successfully applied it to cancer lesion detection. The proposed model has been able to accurately segment images, showcasing the potential of the developed method in image analysis and image processing. Overall, the proposed algorithm offers a reliable and efficient way to model many real-world problems, especially in the field of image analysis.

Conclusion

The robust model-based clustering was enhanced by incorporating finite mixtures of BMCN distributions. This approach is suitable for modeling complex data with outliers and bounded values. To estimate the model parameters, an ECME algorithm was devised, which includes a selection mechanism and closed-form expressions at the E-step. Empirical evaluations conducted on skin cancer detection images showed that the proposed method has successfully enhanced accuracy in segmenting skin lesions.

Conflict of Interests

Authors have no conflict of interests.

References

1. Tukey JW. A survey of sampling from contaminated distributions. Contributions to Probability and Statistics. 1960:448-85.

2. Aitkin M, Wilson GT. Mixture models, outliers, and the EM algorithm. Technometrics. 1980;22(3):325-31.

3. Hedelin P, Skoglund J. Vector quantization based on Gaussian mixture models. IEEE Transactions on Speech and Audio Processing. 2000;8(4):385-401.

4. Lindblom J, Samuelsson J. Bounded support Gaussian mixture modeling of speech spectra. IEEE Transactions on Speech and Audio Processing. 2003;11(1):88-99.

5. Flecher C, Allard D, Naveau P. Truncated skew-normal distributions: moments, estimation by weighted moments and application to climatic data. Metron. 2010;68:331-45.

6. Nguyen TM, Wu QJ. Bounded asymmetrical student's-t mixture model. IEEE Transactions on Cybernetics. 2013;44(6):857-69.

7. Xiong T, Yi Z, Zhang L. Grayscale image segmentation by spatially variant mixture model with student's t-distribution. Multimedia Tools and Applications. 2014;72:167-89.

8. Sun J, Ji Z. Bounded asymmetric Gaussian mixture model with spatial constraint for image segmentation. In 2016 International Conference on Progress in Informatics and Computing. 2016;369-73. IEEE.

9. Bi H, Tang H, Shu HZ, Dillenseger JL. Bounded Rayleigh mixture model for ultrasound image segmentation. In Eighth

International Conference on Graphic and Image Processing. 2017;10225:215-19. SPIE.

10. Punzo A, McNicholas PD. Parsimonious mixtures of multivariate contaminated normal distributions. Biometrical Journal. 2016;58(6):1506-37.

11. Mazza A, Punzo A. Mixtures of multivariate contaminated normal regression models. Statistical Papers. 2020;61(2):787-822.

12. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (methodological). 1977;39(1):1-22.

13. Atkinson SE. The performance of standard and hybrid EM algorithms for ML estimates of the normal mixture model with censoring. Journal of Statistical Computation and Simulation. 1992;44(1-2):105-15.

14. McLachlan GJ, Jones PN. Fitting mixture models to grouped and truncated data via the EM algorithm. Biometrics. 1988;1:571-8.

15. Nguyen TM, Wu QJ, Zhang H. Bounded generalized Gaussian mixture model. Pattern Recognition. 2014;47(9):3132-42.

16. Azam M, Bouguila N. Bounded generalized Gaussian mixture model with ICA. Neural Processing Letters. 2019;49:1299-320.

17. Azam M, Bouguila N. Multivariate bounded support Laplace mixture model. Soft Computing. 2020;24(17):13239-68.

18. Lee G, Scott C. EM algorithms for multivariate Gaussian mixture models with truncated and censored data. Computational Statistics & Data Analysis. 2012;56(9):2816-29.

19. Yu Q, Cao G, Shi H, Zhang Y, Fu P. EPLL image restoration with a bounded asymmetrical Student's t mixture model. Journal of Visual Communication and Image Representation. 2022;2088:103611.

20. Mahdavi A, Amirzadeh V, Jamalizadeh A, Lin TI. Maximum likelihood estimation for scale-shape mixtures of flexible generalized skew normal distributions via selection representation. Computational Statistics. 2021;36:2201-30.

21. Mahdavi A, Amirzadeh V, Jamalizadeh A, Lin TI. A Multivariate flexible skew-symmetric-normal distribution: Scale-shape mixtures and parameter estimation via selection representation. Symmetry. 2021;13(8):1343.

22. Meng XL, Rubin DB. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika. 1993;80(2):267-78.

23. Liu C, Rubin DB. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika. 1994;81(4):633-48.

24. Cabral CR, Lachos VH, Prates MO. Multivariate mixture modeling using skew-normal independent distributions.

Computational Statistics & Data Analysis. 2012;56(1):126-42.

25. Prates MO, Lachos VH, Cabral CR. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. Journal of Statistical Software. 2013;54:1-20.

26. Hubert L, Arabie P. Comparing partitions. Journal of Classification. 1985;2:193-218.