REVSTAT – Statistical Journal
Volume 19, Number 1, January 2021, 35–46

FINITE MIXTURES OF MULTIVARIATE SKEW LAPLACE DISTRIBUTIONS

Authors: Fatma Zehra Doğru
– Department of Statistics, Faculty of Arts and Science, Giresun University, Giresun, Turkey
fatma.dogru@giresun.edu.tr

Y. Murat Bulut
– Department of Statistics, Faculty of Science and Letters, Eskisehir Osmangazi University, Eskişehir, Turkey
ymbulut@ogu.edu.tr

Olcay Arslan
– Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey
oarslan@ankara.edu.tr

Received: February 2017 Revised: September 2018 Accepted: September 2018

Abstract:

- This paper proposes finite mixtures of multivariate skew Laplace distributions in order to model both skewness and heavy-tailedness in heterogeneous data sets. Maximum likelihood estimators for the parameters of interest are obtained using the EM algorithm. The paper offers a small simulation study and a real data example to illustrate the performance of the proposed mixture model.

Keywords:

- *EM algorithm; ML estimation; multivariate mixture model; MSL.*

AMS Subject Classification:

- 62H12, 65C60.

1. INTRODUCTION

Finite mixture models are used to model heterogeneous data sets thanks to their flexibility. These models are commonly applied in fields such as classification, cluster and latent class analysis, density estimation, data mining, image analysis, genetics, medicine, pattern recognition and suchlike; for more detail see [7, 12, 20, 21, 27].

In general, the distribution of mixture model components is assumed to be normal because of its tractability and wide applicability. In practice, however, the data sets may be asymmetric and/or heavy-tailed. For instance, there have been a number of studies focusing on multivariate mixture modeling using asymmetric and/or heavy-tailed distributions: [21] propose finite mixtures of multivariate t distributions as a robust extension of the multivariate normal mixture model ([20]); [16] introduces multivariate skew normal mixture models; [24] and [17] examine finite mixtures of restricted and unrestricted variants of the multivariate skew t distributions of [25]; [8] explore multivariate mixture modeling based on skew-normal independent distributions; and [18] introduce flexible mixture modeling based on skew-t-normal distribution.

In multivariate analysis, the multivariate skew normal (MSN) distribution, [5], [14] and [2], is proposed as an alternative to the multivariate normal (MN) distribution in order to deal with skewness in the data. However, certain alternative heavy-tailed skew distributions are required to model skewness and heavy-tailedness because MSN distribution is not heavy-tailed. One such example of heavy-tailed skew distribution is the multivariate skew t (MST) distribution, which is defined by [4] and [13]. [3] also proposes another heavy-tailed skew distribution called the multivariate skew Laplace (MSL) distribution, using a variance-mean mixture of the normal distribution. One advantage of the MSL distribution is that it has a smaller number of parameters than the MST distribution and has the same number of parameters as the MSN distribution. Regarding finite mixtures of the multivariate skew distributions, finite mixtures of MSN distributions were proposed by [16] to model heterogeneous data sets as they may not be able to modeled by mixtures of MN distributions due to the skew feature of data. On the other hand, data sets may not only have a skewness problem, but may also have a heavy-tailedness problem to be dealt with. For this reason, in this study, finite mixtures of MSL distributions as an alternative to finite mixtures of MSN distributions are explored in order to deal with both skewness and heavy-tailedness in heterogeneous data sets.

The rest of the paper is organized as follows: Section 2 summarizes certain properties of the MSL distribution; see [3] for further details of the MSL distribution. Section 3 presents mixtures of MSL distributions and gives the Expectation-Maximization (EM) algorithm to obtain maximum likelihood (ML) estimators for the parameters of the proposed mixture model. Section 4 offers the empirical information matrix of MSL distribution to compute standard errors of proposed estimators. Section 5 provides a small simulation study and a real data example to illustrate the performance of the proposed mixture model. Finally, Section 6 is devoted to conclusions.

2. MULTIVARIATE SKEW LAPLACE DISTRIBUTION

Let $\boldsymbol{Y} \in R^p$ be a p-dimensional random vector which has the MSL distribution ($\boldsymbol{Y} \sim \mathrm{MSL}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$) proposed by [3]. The probability density function (pdf) of this distribution is given below:

$$
\begin{aligned}
f_{\mathrm{MSL}}(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) \;&=\; \frac{|\,\Sigma\,|^{-\frac{1}{2}}}{2^p \pi^{\frac{p-1}{2}} \alpha \Gamma\left(\frac{p+1}{2}\right)} \\
&\quad \times \;\exp\left\{-\alpha\sqrt{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}} \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{\mu})} + (\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}} \Sigma^{-1} \boldsymbol{\gamma}\right\},
\end{aligned} \tag{2.1}
$$

where $\alpha = \sqrt{1+\boldsymbol{\gamma}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma}}$, $\boldsymbol{\mu} \in R^p$ is the location parameter, $\boldsymbol{\gamma} \in R^p$ is the skewness parameter, Σ is the positive definite scatter matrix and $\Gamma(\cdot)$ represents the complete gamma function.

Proposition 2.1. *The characteristic function of* $\mathrm{MSL}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ *is*

$$
\Phi_{\boldsymbol{Y}}(t) = e^{it^{\mathsf{T}}\boldsymbol{\mu}} \left[1 + t^{\mathsf{T}}\Sigma t - 2it^{\mathsf{T}}\boldsymbol{\gamma}\right]^{-\frac{p+1}{2}}, \quad t \in R^p.
$$

See [3] for proof of this proposition.

If $\boldsymbol{Y} \sim \mathrm{MSL}_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ then the expectation and variance of $\boldsymbol{Y}$ are:

$$
\begin{aligned}
E(\boldsymbol{Y}) \;&=\; \boldsymbol{\mu} + (p+1)\,\boldsymbol{\gamma}, \\
\mathrm{Var}(\boldsymbol{Y}) \;&=\; (p+1)\left(\Sigma + 2\boldsymbol{\gamma}\boldsymbol{\gamma}^{\mathsf{T}}\right).
\end{aligned}
$$

The MSL distribution can be obtained as a variance-mean mixture of MN distribution and inverse gamma (IG) distribution. The variance-mean mixture representation is given as follows:

$$
\boldsymbol{Y} = \boldsymbol{\mu} + V^{-1}\boldsymbol{\gamma} + \sqrt{V^{-1}}\Sigma^{\frac{1}{2}}\boldsymbol{X}, \tag{2.2}
$$

where $\boldsymbol{X} \sim N_p(\mathbf{0}, I_p)$ and $V \sim IG\left(\frac{p+1}{2}, \frac{1}{2}\right)$. Note that if $\boldsymbol{\gamma} = \mathbf{0}$, the density function of $\boldsymbol{Y}$ reduces to the density function of symmetric multivariate Laplace distribution given by [22]. In addition, the conditional distribution of $\boldsymbol{Y}$ given $V = v$ will be:

$$
\boldsymbol{Y}|v \sim N_p\left(\boldsymbol{\mu} + v^{-1}\boldsymbol{\gamma}, v^{-1}\Sigma\right).
$$

The joint density function of $\boldsymbol{Y}$ and V is:

$$
\begin{aligned}
f(\boldsymbol{y}, v) \;&=\; \frac{|\,\Sigma\,|^{-\frac{1}{2}}\, e^{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma}}}{2^p \pi^{\frac{p-1}{2}} \alpha \Gamma\left(\frac{p+1}{2}\right)} \\
&\quad \times \left\{v^{-\frac{3}{2}} e^{-\frac{1}{2}\left\{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})v + \left(1+\boldsymbol{\gamma}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma}\right)v^{-1}\right\}}\right\}.
\end{aligned}
$$

Then, we have the following conditional density function of V given $\boldsymbol{Y}$:

$$
\begin{aligned}
f(v|\boldsymbol{y}) \;&=\; \frac{\alpha}{\sqrt{2\pi}} e^{\alpha\sqrt{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}} \\
&\quad \times\; v^{-\frac{3}{2}} e^{-\frac{1}{2}\left\{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})v + \alpha^2 v^{-1}\right\}}, \quad v > 0.
\end{aligned} \tag{2.3}
$$

Using the conditional density function given in (2.3), the conditional expectations can be obtained as follows:

$$E(V|\boldsymbol{y}) = \frac{\sqrt{1+\boldsymbol{\gamma}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma}}}{\sqrt{(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}}, \tag{2.4}$$

$$E\left(V^{-1}|\boldsymbol{y}\right) = \frac{1+\sqrt{(1+\boldsymbol{\gamma}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma})(\boldsymbol{y}-\boldsymbol{\mu})^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}}{1+\boldsymbol{\gamma}^{\mathsf{T}}\Sigma^{-1}\boldsymbol{\gamma}}. \tag{2.5}$$

Note that these conditional expectations will be used in the EM algorithm given in subsection 3.1; see [3] for further details of the MSL distribution.

3. FINITE MIXTURES OF MSL DISTRIBUTIONS

Let $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n$ be p-dimensional random sample which comes from a g-component mixtures of MSL distributions. The pdf of a g-component finite mixtures of MSL distributions is given by:

$$f(\boldsymbol{y}|\boldsymbol{\Theta}) = \sum_{i=1}^{g} \pi_i f(\boldsymbol{y}; \boldsymbol{\mu}_i, \Sigma_i, \boldsymbol{\gamma}_i), \tag{3.1}$$

where π_i denotes the mixing probability with $\sum_{i=1}^{g} \pi_i = 1$, $0 \leq \pi_i \leq 1$, $f(\boldsymbol{y}; \boldsymbol{\mu}_i, \Sigma_i, \boldsymbol{\gamma}_i)$ represents the pdf of the i-th component (pdf of the MSL distribution) given in (2.1) and $\boldsymbol{\Theta} = (\pi_1, ..., \pi_g, \boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_g, \Sigma_1, ..., \Sigma_g, \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_g)^{\mathsf{T}}$ is the unknown parameter vector.

3.1. ML estimation

The ML estimator of $\boldsymbol{\Theta}$ can be found by maximizing the following log-likelihood function:

$$\ell(\boldsymbol{\Theta}) = \sum_{j=1}^{n} \log\left(\sum_{i=1}^{g} \pi_i f(\boldsymbol{y}_j; \boldsymbol{\mu}_i, \Sigma_i, \boldsymbol{\gamma}_i)\right). \tag{3.2}$$

However, there is not an explicit maximizer of (3.2). Therefore, in general, the EM algorithm ([9]) is used to obtain the ML estimator of $\boldsymbol{\Theta}$. Here, we will use the following EM algorithm:

Let $\boldsymbol{Z}_j = (Z_{1j}, ..., Z_{gj})^{\mathsf{T}}$ be the latent variables with

$$Z_{ij} = \begin{cases} 1, & \text{if } j^{\text{th}} \text{ observation belongs to } i^{\text{th}} \text{ component,} \\ 0, & \text{otherwise,} \end{cases} \tag{3.3}$$

where $j = 1, ..., n$ and $i = 1, ..., g$. To implement the steps of the EM algorithm, we will use the stochastic representation of the MSL distribution given in (2.2). If we do so, the hierarchical representation for the mixtures of MSL distributions will be:

$$\begin{aligned} \mathbf{Y}_j | v_j, z_{ij} = 1 &\sim N\left(\boldsymbol{\mu} + v_j^{-1}\boldsymbol{\gamma}, v_j^{-1}\Sigma\right), \\ V_j | z_{ij} = 1 &\sim IG\left(\frac{p+1}{2}, \frac{1}{2}\right). \end{aligned} \tag{3.4}$$

Let $(\boldsymbol{y},\boldsymbol{v},\boldsymbol{z})$ be the complete data, where $\boldsymbol{y}=\left(\boldsymbol{y}_1^\top,...,\boldsymbol{y}_n^\top\right)^\top$, $\boldsymbol{v}=(v_1,...,v_n)$ and $\boldsymbol{z}=(z_1,...,z_n)^\top$. Using the hierarchical representation given above and ignoring the constants, the complete data log-likelihood function can be written as:

$$\begin{aligned}\ell_c(\boldsymbol{\Theta};\boldsymbol{y},\boldsymbol{v},\boldsymbol{z}) &= \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\left\{\log\pi_i-\frac{1}{2}\log|\Sigma_i|+(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}\gamma_i\right.\\ &\quad\left.-\frac{1}{2}v_j(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}(\boldsymbol{y}_j-\boldsymbol{\mu}_i)-\frac{1}{2}\gamma_i^\top\Sigma_i^{-1}\gamma_i v_j^{-1}-\frac{1}{2}\left(3\log v_j+v_j^{-1}\right)\right\}. \qquad (3.5)\end{aligned}$$

To overcome the latency of the latent variables given in (3.5), we have to take the conditional expectation of the complete data log-likelihood function given the observed data $\boldsymbol{y}_j$

$$\begin{aligned}E(\ell_c(\boldsymbol{\Theta};\boldsymbol{y},\boldsymbol{v},\boldsymbol{z}|\boldsymbol{y}_j)) &= \sum_{j=1}^{n}\sum_{i=1}^{g} E(Z_{ij}|\boldsymbol{y}_j)\left\{\log\pi_i-\frac{1}{2}\log|\Sigma_i|-(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}\gamma_i\right.\\ &\quad\left.-\frac{1}{2}E(V_j|\boldsymbol{y}_j)(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}(\boldsymbol{y}_j-\boldsymbol{\mu}_i)-\frac{1}{2}\gamma_i^\top\Sigma_i^{-1}\gamma_i E\left(V_j^{-1}|\boldsymbol{y}_j\right)\right\}. \qquad (3.6)\end{aligned}$$

Since the last part of the complete data log-likelihood function does not include the parameters of interest it is omitted and the conditional expectation of the other terms are taken. The conditional expectations $E(V_j|\boldsymbol{y}_j)$ and $E\left(V_j^{-1}|\boldsymbol{y}_j\right)$ can be calculated using the conditional expectations given in (2.4) and (2.5), and the conditional expectation $E(Z_{ij}|\boldsymbol{y}_j)$ can be computed using the classical theory of mixture modeling. Next, the steps of the EM algorithm can be formed as follows:

EM algorithm:

1. Set initial parameter estimate $\boldsymbol{\Theta}^{(0)}$ and a stopping rule Δ.

2. E-Step: Compute the following conditional expectations for $k=0,1,2,...$ iteration

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij}|\boldsymbol{y}_j,\hat{\boldsymbol{\Theta}}^{(k)}\right)=\frac{\hat{\pi}_i^{(k)}f\left(\boldsymbol{y}_j;\hat{\boldsymbol{\mu}}_i^{(k)},\hat{\Sigma}_i^{(k)},\hat{\boldsymbol{\gamma}}_i^{(k)}\right)}{f\left(\boldsymbol{y}_j;\hat{\boldsymbol{\Theta}}^{(k)}\right)}, \qquad (3.7)$$

$$\hat{v}_{1ij}^{(k)} = E\left(V_j|\boldsymbol{y}_j,\hat{\boldsymbol{\Theta}}^{(k)}\right)=\frac{\sqrt{1+\hat{\boldsymbol{\gamma}}_i^{(k)\top}\hat{\Sigma}_i^{(k)^{-1}}\hat{\boldsymbol{\gamma}}_i^{(k)}}}{\sqrt{\left(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}}_i^{(k)}\right)^\top\hat{\Sigma}_i^{(k)^{-1}}\left(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}}_i^{(k)}\right)}}, \qquad (3.8)$$

$$\begin{aligned}\hat{v}_{2ij}^{(k)} &= E\left(V_j^{-1}|\boldsymbol{y}_j,\hat{\boldsymbol{\Theta}}^{(k)}\right)\\ &= \frac{1+\sqrt{\left(1+\hat{\boldsymbol{\gamma}}_i^{(k)\top}\hat{\Sigma}_i^{(k)^{-1}}\hat{\boldsymbol{\gamma}}_i^{(k)}\right)\left(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}}_i^{(k)}\right)^\top\hat{\Sigma}_i^{(k)^{-1}}\left(\boldsymbol{y}_j-\hat{\boldsymbol{\mu}}_i^{(k)}\right)}}{1+\hat{\boldsymbol{\gamma}}_i^{(k)\top}\hat{\Sigma}_i^{(k)^{-1}}\hat{\boldsymbol{\gamma}}_i^{(k)}}. \qquad (3.9)\end{aligned}$$

Using these conditional expectations, we form the following objective function:

$$\begin{aligned}Q\left(\boldsymbol{\Theta};\hat{\boldsymbol{\Theta}}^{(k)}\right) &= \sum_{j=1}^{n}\sum_{i=1}^{g}\hat{z}_{ij}^{(k)}\left\{\log\pi_i-\frac{1}{2}\log|\Sigma_i|-(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}\gamma_i\right.\\ &\quad\left.-\frac{1}{2}\hat{v}_{1ij}^{(k)}(\boldsymbol{y}_j-\boldsymbol{\mu}_i)^\top\Sigma_i^{-1}(\boldsymbol{y}_j-\boldsymbol{\mu}_i)-\frac{1}{2}\hat{v}_{2ij}^{(k)}\gamma_i^\top\Sigma_i^{-1}\gamma_i\right\}. \qquad (3.10)\end{aligned}$$

3. M-Step: Maximize the $Q\big(\boldsymbol{\Theta}; \widehat{\boldsymbol{\Theta}}^{(k)}\big)$ with respect to $\boldsymbol{\Theta}$ to get the $(k+1)$-th parameter estimates for the parameters. This maximization yields the following updating equations:

$$\widehat{\pi}_i^{(k+1)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(k)}}{n}, \tag{3.11}$$

$$\widehat{\boldsymbol{\mu}}_i^{(k+1)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)} \boldsymbol{y}_j - \sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{\boldsymbol{\gamma}}_i^{(k)}}{\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)}}, \tag{3.12}$$

$$\widehat{\boldsymbol{\gamma}}_i^{(k+1)} = \frac{\Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)}\Big)\Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \boldsymbol{y}_j\Big) - \Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)}\Big)\Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)} \boldsymbol{y}_j\Big)}{\Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)}\Big)\Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{2ij}^{(k)}\Big) - \Big(\sum_{j=1}^n \widehat{z}_{ij}^{(k)}\Big)^2}, \tag{3.13}$$

$$\widehat{\Sigma}_i^{(k+1)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{1ij}^{(k)} \Big(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i^{(k)}\Big)\Big(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i^{(k)}\Big)^{\mathsf{T}} - \widehat{\boldsymbol{\gamma}}_i^{(k)} \widehat{\boldsymbol{\gamma}}_i^{(k)\mathsf{T}} \sum_{j=1}^n \widehat{z}_{ij}^{(k)} \widehat{v}_{2ij}^{(k)}}{\sum_{j=1}^n \widehat{z}_{ij}^{(k)}}. \tag{3.14}$$

4. Repeat E and M steps until the convergence rule $\|\, \widehat{\boldsymbol{\Theta}}^{(k+1)} - \widehat{\boldsymbol{\Theta}}^{(k)} \,\| < \Delta$ is obtained. Alternatively, the absolute difference of the actual log-likelihood $\big\|\ell\big(\widehat{\boldsymbol{\Theta}}^{(k+1)}\big) - \ell\big(\widehat{\boldsymbol{\Theta}}^{(k)}\big)\big\| < \Delta$ or $\big\|\ell\big(\widehat{\boldsymbol{\Theta}}^{(k+1)}\big)/\ell\big(\widehat{\boldsymbol{\Theta}}^{(k)}\big) - 1\big\| < \Delta$ can be used as a stopping rule ([10]).

3.2. Initial values

In order to determine the initial values for the EM algorithm, the following procedure given by [16] will be used:

- **i)** Perform the K-means clustering algorithm ([15]).
- **ii)** Initialize the component labels $\widehat{\boldsymbol{z}}_j^{(0)} = \{z_{ij}\}_{i=1}^g$ according to the K-means clustering results.
- **iii)** The initial values of mixing probabilities, component locations and component scale variances can be set as:

$$\widehat{\pi}_i^{(0)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(0)}}{n},$$

$$\widehat{\boldsymbol{\mu}}_i^{(0)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(0)} \boldsymbol{y}_j}{\sum_{j=1}^n \widehat{z}_{ij}^{(0)}},$$

$$\widehat{\Sigma}_i^{(0)} = \frac{\sum_{j=1}^n \widehat{z}_{ij}^{(0)} \Big(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i^{(0)}\Big)\Big(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i^{(0)}\Big)^{\mathsf{T}}}{\sum_{j=1}^n \widehat{z}_{ij}^{(0)}}.$$

- **iv)** For the skewness parameters, use the skewness coefficient vector of each cluster.

4. THE EMPIRICAL INFORMATION MATRIX

We will compute the standard errors of ML estimators using the information based method given by [6]. At this point, we will use the inverse of the empirical information matrix in order to have an approximation to the asymptotic covariance matrix of the estimators.

This information matrix can be obtained as:

$$\widehat{I}_e = \sum_{j=1}^{n} \widehat{s}_j \widehat{s}_j^{\mathsf{T}}, \tag{4.1}$$

where $\widehat{s}_j = E_{\widehat{\boldsymbol{\Theta}}}\left(\frac{\partial \ell_{cj}(\boldsymbol{\Theta};\boldsymbol{y}_j,\boldsymbol{v}_j,\boldsymbol{z}_j)}{\partial \boldsymbol{\Theta}} | \boldsymbol{y}_j\right)$, $j = 1, 2, ..., n$ are the individual scores and $\ell_{cj}(\boldsymbol{\Theta}; \boldsymbol{y}_j, \boldsymbol{v}_j, \boldsymbol{z}_j)$ is the complete data log-likelihood function for the j-th observation. The components of the score vector $\widehat{\boldsymbol{s}}_j$ are $\left(\widehat{\boldsymbol{s}}_{j,\pi_1}, ..., \widehat{\boldsymbol{s}}_{j,\pi_{g-1}}, \widehat{\boldsymbol{s}}_{j,\boldsymbol{\mu}_1}, ..., \widehat{\boldsymbol{s}}_{j,\boldsymbol{\mu}_g}, \widehat{\boldsymbol{s}}_{j,\sigma_1}, ..., \widehat{\boldsymbol{s}}_{j,\sigma_g}, \widehat{\boldsymbol{s}}_{j,\gamma_1}, ..., \widehat{\boldsymbol{s}}_{j,\gamma_g}\right)$. After straightforward algebra, we obtain these components as follows:

$$\widehat{\boldsymbol{s}}_{j,\pi_r} = \frac{\widehat{z}_{rj}}{\widehat{\pi}_r} - \frac{\widehat{z}_{gj}}{\widehat{\pi}_g}, \quad r = 1, ..., g-1, \tag{4.2}$$

$$\widehat{\boldsymbol{s}}_{j,\boldsymbol{\mu}_i} = \widehat{z}_{ij} \widehat{\Sigma}_i^{-1} \left(\widehat{v}_{1ij} \left(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i\right) - \widehat{\boldsymbol{\gamma}}_i\right), \tag{4.3}$$

$$\begin{aligned} \widehat{\boldsymbol{s}}_{j,\sigma_i} = {} & \operatorname{vech}\Big(\widehat{z}_{ij}\Big\{-\Big(\widehat{\Sigma}_i^{-1} - \widehat{v}_{1ij}\widehat{\Sigma}_i^{-1}\left(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i\right)\left(\boldsymbol{y}_j - \widehat{\boldsymbol{\mu}}_i\right)^{\mathsf{T}}\widehat{\Sigma}_i^{-1} - \widehat{v}_{2ij}\widehat{\Sigma}_i^{-1}\widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i^{\mathsf{T}}\widehat{\Sigma}_i^{-1}\Big) \\ & + \frac{1}{2}\operatorname{diag}\Big(\widehat{\Sigma}_i^{-1} - \widehat{v}_{1ij}\widehat{\Sigma}_i^{-1}\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i\right)\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i\right)^{\mathsf{T}}\widehat{\Sigma}_i^{-1} - \widehat{v}_{2ij}\widehat{\Sigma}_i^{-1}\widehat{\boldsymbol{\gamma}}_i\widehat{\boldsymbol{\gamma}}_i^{\mathsf{T}}\Big)\Big\}\Big), \end{aligned} \tag{4.4}$$

$$\boldsymbol{s}_{j,\boldsymbol{\gamma}_i} = \widehat{z}_{ij} \widehat{\Sigma}_i^{-1} \left(\left(\boldsymbol{y}_j - \boldsymbol{\mu}_i\right) - \widehat{v}_{2ij} \boldsymbol{\gamma}_i\right). \tag{4.5}$$

Therefore, using these equations we can form the information matrix I_e given in (4.1). After this, the standard errors of the ML estimators $\widehat{\boldsymbol{\Theta}}$ will be found using the square root of the matrix $\widehat{I}_e^{-1}$.

5. APPLICATIONS

This section will illustrate the performance of the proposed mixture model based on a small simulation study and a real data example. All computations for the simulation study and real data example are conducted using an MATLAB R2013a. For all computations, the stopping rule Δ is taken as 10^{-6}. The codes are available upon request.

5.1. Simulation study

In the simulation study, the data set is generated from the following two-component mixtures of MSL distributions:

$$f\left(\boldsymbol{y}_i|\boldsymbol{\Theta}\right) = \pi_1 f_p\left(\boldsymbol{y}_j; \boldsymbol{\mu}_1, \Sigma_1, \boldsymbol{\gamma}_1\right) + \left(1-\pi_1\right) f_p\left(\boldsymbol{y}_j; \boldsymbol{\mu}_2, \Sigma_2, \boldsymbol{\gamma}_2\right),$$

where

$$\boldsymbol{\mu}_i = \left(\mu_{i1}, \mu_{i2}\right)^{\mathsf{T}}, \quad \Sigma_i = \begin{bmatrix} \sigma_{i,11} & \sigma_{i,12} \\ \sigma_{i,21} & \sigma_{i,22} \end{bmatrix}, \quad \boldsymbol{\gamma}_i = \left(\gamma_{i1}, \gamma_{i2}\right)^{\mathsf{T}}, \quad i = 1, 2,$$

with the parameter values

$$\boldsymbol{\mu}_1 = (2,2)^{\mathsf{T}}, \quad \boldsymbol{\mu}_2 = (-2,-2)^{\mathsf{T}}, \quad \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1.5 & 0 \\ 0 & 1.5 \end{bmatrix},$$

$$\boldsymbol{\gamma}_1 = (1,1)^{\mathsf{T}}, \quad \boldsymbol{\gamma}_2 = (-1,-1)^{\mathsf{T}}, \quad \pi_1 = 0.6.$$

The sample sizes are set as 500, 1000 and 2000 and the number of replicates (N) is taken as 500. The table contains the bias, standard errors (SEs) and the mean Euclidean distance values of the estimates. The formula of bias is given below:

$$\widehat{\text{bias}}\left(\widehat{\theta}\right) = \bar{\theta} - \theta,$$

where θ is the true parameter value, $\bar{\theta} = \frac{1}{N}\sum_{j=1}^{N} \widehat{\theta}_j$ and $\widehat{\theta}_j$ is the estimate of θ for the j-th simulated data. The mean Euclidean distances of the estimators are computed using the average of the Euclidian norm between the estimates and the true parameter values. For instance, for the mean Euclidean distance of $\widehat{\boldsymbol{\mu}}_i$ will be as follows:

$$\| \widehat{\boldsymbol{\mu}}_i - \boldsymbol{\mu}_i \| = \frac{1}{N}\left(\sum_{j=1}^{N} \left(\widehat{\mu}_{ij} - \mu_{ij}\right)^2\right)^{\frac{1}{2}}.$$

The other mean Euclidean distances of other estimates are also obtained in a similar way. The distance for $\widehat{\pi}_i$, on the other hand, will be the mean squared error (MSE). The formula of MSE is given as:

$$\widehat{\text{MSE}}\left(\widehat{\pi}\right) = \frac{1}{N}\sum_{j=1}^{N} \left(\widehat{\pi}_j - \pi\right)^2,$$

where π is the true parameter value, $\widehat{\pi}_j$ is the estimate of π for the j-th simulated data and $\bar{\pi} = \frac{1}{N}\sum_{j=1}^{N} \widehat{\pi}_j$. We calculate the SEs of estimates using the empirical information matrix of the finite mixture model based on the MSL distribution given in Section 4.

Table 1: Bias, SEs and mean Euclidean distance values of the estimates for $n = 500$, 1000 and 2000.

n	Parameter	Component 1				Component 2			
		True	Bias	SE	Distance	True	Bias	SE	Distance
500	π_1	0.6	0.001489	0.122734	0.000533				
	μ_{i1}	2	−0.001200	0.174686	0.201026	−2	−0.014079	0.210878	0.273981
	μ_{i2}	2	0.010775	0.176092		−2	0.031688	0.208902	
	$\sigma_{i,11}$	1.5	−0.018232	0.182958	0.198493	1.5	−0.005309	0.232425	0.254685
	$\sigma_{i,12}$	0	0.002903	0.156863		0	−0.006157	0.170339	
	$\sigma_{i,22}$	1.5	−0.004315	0.188067		1.5	−0.033095	0.224071	
	γ_{i1}	1	−0.001810	0.256792	0.106922	−1	−0.000255	0.291138	0.139768
	γ_{i2}	1	−0.005965	0.255197		−1	−0.016220	0.290819	
1000	π_1	0.6	0.001075	0.082913	0.000236				
	μ_{i1}	2	−0.007428	0.122541	0.147008	−2	−0.001296	0.146090	0.180637
	μ_{i2}	2	−0.006677	0.122956		−2	−0.003134	0.146122	
	$\sigma_{i,11}$	1.5	−0.011271	0.127618	0.136552	1.5	−0.002545	0.155037	0.169090
	$\sigma_{i,12}$	0	−0.008758	0.106796		0	−0.000529	0.116473	
	$\sigma_{i,22}$	1.5	−0.000720	0.128693		1.5	−0.009886	0.155416	
	γ_{i1}	1	0.005704	0.174375	0.078061	−1	−0.001849	0.194632	0.093384
	γ_{i2}	1	0.006670	0.174589		−1	−0.002547	0.192690	
2000	π_1	0.6	0.000593	0.057421	0.000119				
	μ_{i1}	2	−0.002704	0.086564	0.103007	−2	0.004608	0.102633	0.126149
	μ_{i2}	2	0.000165	0.086560		−2	−0.003348	0.103218	
	$\sigma_{i,11}$	1.5	−0.006294	0.088880	0.098289	1.5	−0.008891	0.106630	0.116648
	$\sigma_{i,12}$	0	−0.003700	0.074673		0	−0.000537	0.080978	
	$\sigma_{i,22}$	1.5	0.001142	0.089714		1.5	0.000997	0.108406	
	γ_{i1}	1	0.002448	0.121117	0.056854	−1	−0.003021	0.133069	0.067364
	γ_{i2}	1	0.001445	0.120912		−1	0.000548	0.133948	

Table 1 shows the simulation results for the sample sizes 500, 1000 and 2000. We give the bias, SEs and mean Euclidean distance values of estimates and true parameter values. It can be seen from the table that the proposed model works accurately to obtain the estimates for all the parameters. Furthermore, the mean Euclidian distances get smaller when the sample sizes increase. We observe similar results for the SEs of the estimates. These values decrease as the sample sizes increase. All these findings confirm that the proposed finite mixture model will be an alternative finite mixture model for modelling heterogeneous data with skew and heavy-tail components.

5.2. Real data example

This real data example will investigate the bank data set, which was given in Tables 1.1 and 1.2 by [11] and examined by [19], to model through a skew-symmetric distribution. Concerning this data set, there are six measurements made on 100 genuine and 100 counterfeit old Swiss 1000 franc bills. This data set was also analyzed by [16] to model mixtures of MSN distributions. He used the variables X_1, the width of the right edge, and X_2, the length of the image diagonal, that reveal a bimodal distribution with asymmetric components. Following this, the current study uses Swiss bank data to illustrate the applicability of the finite mixtures of multivariate skew Laplace distributions (FM-MSL) and compares the results with the finite mixtures of multivariate skew normal distributions (FM-MSN). The estimation results are displayed in Table 2 for FM-MSN and FM-MSL. The table contains the ML estimates, standard errors of the estimates for all components, the log-likelihood, the values of the Akaike information criterion (AIC) ([1]) and the Bayesian information criterion (BIC) ([26]).

Table 2: ML estimation results of the Swiss bank data set for FM-MSN and FM-MSL.

	FM-MSN				FM-MSL			
	1		2		1		2	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
π_1	0.504	0.036	—	—	0.521	0.163	—	—
μ_{i1}	130.38	0.122	129.32	0.062	130.20	0.118	129.65	0.076
μ_{i2}	140.06	0.064	141.39	1.125	139.50	0.152	141.76	0.201
$\sigma_{i,11}$	0.068	0.023	0.037	0.016	0.067	0.054	0.104	0.030
$\sigma_{i,12}$	0.051	0.015	−0.012	0.015	0.001	0.037	−0.023	0.043
$\sigma_{i,22}$	0.056	0.027	0.154	0.032	0.371	0.100	0.194	0.218
γ_{i1}	−0.230	0.043	0.494	0.077	−0.017	0.108	0.034	0.060
γ_{i2}	−0.800	0.067	0.177	1.433	0.054	0.154	−0.148	0.198
$\ell(\widehat{\Theta})$			−310.07				**−152.30**	
AIC			650.14				**334.60**	
BIC			699.61				**384.08**	

Additionally, we give results and criterion values for FM-MSN which was computed by [16]. According to information criterion values, the FM-MSL has better fit than the FM-MSN.

Figure 1 displays a scatter plot of the data together with contour plots of the fitted two-component FM-MSL model. From this plot, it can be seen that the proposed mixture model of MSL distributions captures bimodality and asymmetry and provides a satisfactory fit to the data.

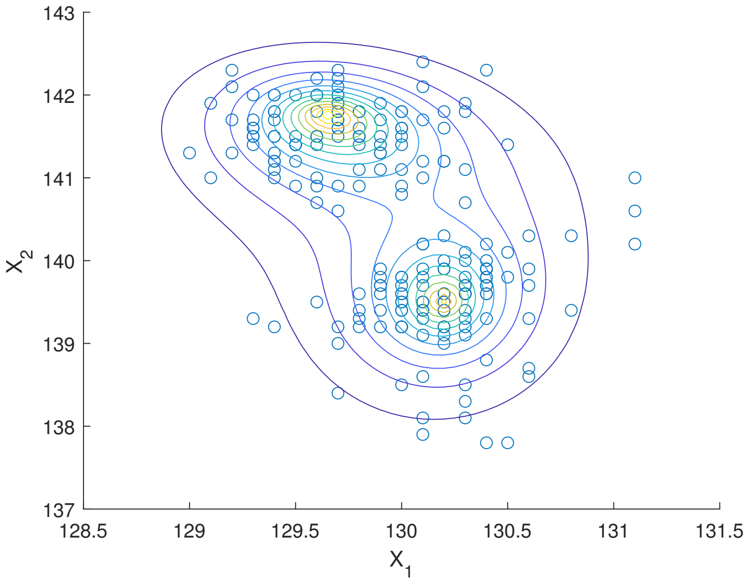

Figure 1: Scatter plot of the Swiss bank data set along with the contour plots of the fitted two-component FM-MSL model.

6. CONCLUSION

In this paper, we have proposed mixtures of MSL distributions and given the EM algorithm to obtain the estimates. A small simulation study has been provided to demonstrate the performance of the proposed mixture model. This shows that the proposed mixture model has accurately estimated the parameters. A real data example has also been offered to compare the mixtures of the MSL distributions with the mixtures of MSN distributions. This comparison proves that the proposed model has the best fit according to the information criterion values. This means that the proposed model can be used as an alternative mixture model to the mixtures of MSN distributions.

ACKNOWLEDGMENTS

We wish to thank the anonymous referee and the Associate Editor, whose comments and suggestions have helped us to improve this paper greatly.

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* In "Proceeding of the Second International Symposium on Information Theory" (B.N. Petrov and F. Caski, Eds.), Akademiai Kiado, Budapest, 267–281.

[2] Arellano-Valle, R.B. and Genton, M.G. (2005). On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**(1), 93–116.

[3] Arslan, O. (2010). An alternative multivariate skew Laplace distribution: properties and estimation, *Statistical Papers*, **51**(4), 865–887.

[4] Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **65**(2), 367–389.

[5] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**(4), 715–726.

[6] Basford, K.E.; Greenway, D.R.; McLachlan, G.J. and Peel, D. (1997). Standard errors of fitted means under normal mixture, *Computational Statistics*, **12**, 1–17.

[7] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*, Springer, Singapore.

[8] Cabral, C.R.B.; Lachos, V.H. and Prates, M.O. (2012). Multivariate mixture modeling using skew-normal independent distributions, *Computational Statistics & Data Analysis*, **56**(1), 126–142.

[9] Dempster, A.P.; Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

[10] Dias, J.G. and Wedel, M. (2004). An empirical comparison of EM, SEM and MCMC performance for problematic gaussian mixture likelihoods, *Statistics and Computing*, **14**, 323–332.

[11] Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics, a Practical Approach*, Cambridge University Press, Cambridge.

[12] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Springer, New York.

[13] Gupta, A.K. (2003). Multivariate skew t-distribution, *Statistics: A Journal of Theoretical and Applied Statistics*, **37**(4), 359–363.

[14] Gupta, A.K.; González-Farías, G. and Domínguez-Molina, J.A. (2004). A multivariate skew normal distribution, *Journal of Multivariate Analysis*, **89**(1), 181–190.

[15] Hartigan, J.A. and Wong, M.A. (1979). Algorithm AS 136: a k-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**(1), 100–108.

[16] Lin, T.I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models, *Journal of Multivariate Analysis*, **100**, 257–265.

[17] Lin, T.I. (2010). Robust mixture modeling using multivariate skew t distributions, *Statistics and Computing*, **20**(3), 343–356.

[18] Lin, T.I.; Ho, H.J. and Lee, C.R. (2014). Flexible mixture modelling using the multivariate skew-t-normal distribution, *Statistics and Computing*, **24**(4), 531–546.

[19] Ma, Y. and Genton, M.G. (2004). Flexible class of skew-symmetric distribtions, *Scandinavian Journal of Statististics*, **31**, 459–468.

[20] McLachlan, G.J. and Basford, K.E. (1988). *Mixture Models: Inference and Application to Clustering*, Marcel Dekker, New York.

[21] McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Models*, Wiley, New York.

[22] Naik, D.N. and Plungpongpun, K. (2006). *A Kotz-type distribution for multivariate statistical inference*. In "Advances in Distribution Theory, Order Statistics, and Inference" (pp. 111–124), Birkhäuser, Boston.

[23] Peel, D. and McLachlan, G.J. (2000). Robust mixture modelling using the t distribution, *Statistics and computing*, **10**(4), 339–348.

[24] Pyne, S.; Hu, X.; Wang, K.; Rossin, E.; Lin, T.I.; Maier, L.; Baecher-Allan, C.; McLachlan, G.J.; Tamayo, P.; Hafler, D.A.; De Jager, P.L. and Mesirov, J.P. (2009). Automated high-dimensional flow cytometric data analysis, *Proc. Natl. Acad. Sci. USA*, **106**, 8519–8524.

[25] Sahu, S.K.; Dey, D.K. and Branco, M.D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**(2), 129–150.

[26] Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.

[27] Titterington, D.M.; Smith, A.F.M. and Markov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, New York.