Statistics and Its Interface Volume 15 (2022) 475–485

Multivariate skew Laplace normal distribution for modeling skewness and heavy-tailedness in multivariate data sets

Fatma Zehra Doğru* and Olcay Arslan

Modeling both skewness and heavy-tailedness in multivariate data sets is a challenging problem. The main goal of this paper is to introduce a multivariate skew Laplace normal (MSLN) distribution to deal with the issue by providing a flexible model for modeling skewness and heavy-tailedness simultaneously. This distribution will be an alternative to some multivariate skew distributions including the multivariate skew-t-normal (MSTN) distribution introduced by [28]. This is due to the fact that the MSLN distribution has fewer parameters than most of these distributions, which causes computationally advantageous for the MSLN distribution over these distributions. The definition, some distributional properties of this distribution are studied. The maximum likelihood (ML) estimators for the parameters of the MSLN distribution are obtained via the expectation-maximization (EM) algorithm. A simulation study and a real data example are also provided to illustrate the capability of the MSLN distribution for modeling data sets in multivariate settings.

AMS 2000 subject classifications: Primary 60E05, 62H10; secondary 62H12.
Keywords and phrases: EM algorithm, ML estimation, MSGLN, MSLN, MSTN.

1. INTRODUCTION

In general, the distribution of a model is assumed to be normal thanks to its tractability and huge applicability. In practice, however, the data sets may be asymmetric and/or heavy-tailed. For example, there are some studies based on multivariate modeling using asymmetric and/or heavy-tailed distributions. The multivariate t distribution was proposed by [11] and [17] as a heavy-tailed alternative to the multivariate normal distribution and studied by [22], [33], [34], [23], and so on. The multivariate skew normal (MSN) distribution ([8], [7] and [6]) has recently been very popular both in theoretical and applied studies for modelling skew data sets.

*Corresponding author.

The MSN distribution can be defined as follows: Let $\boldsymbol{Z}$ be a p-dimensional random vector. It is said to have an MSN distribution ($\boldsymbol{Z} \sim MSN_p(\Sigma_{\boldsymbol{Z}}, \boldsymbol{\lambda})$) if it has the following pdf

$$f_Z(\boldsymbol{z}; \Sigma_{\boldsymbol{Z}}, \boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{z}; \Sigma_{\boldsymbol{Z}})\,\Phi\left(\boldsymbol{\lambda}^T \boldsymbol{z}\right), \quad \boldsymbol{z} \in R^p, \tag{1}$$

where $\phi_p(\boldsymbol{z}; \Sigma_{\boldsymbol{Z}})$ is the p-dimensional normal pdf with zero mean and $\Sigma_{\boldsymbol{Z}}$ correlation matrix, and $\boldsymbol{\lambda}$ is the skewness parameter. Here, $\Phi(\cdot)$ shows the cumulative density function (cdf) of $N(0,1)$. We can also write the following transformation to comprise location and scale parameters

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{w}\boldsymbol{Z}, \tag{2}$$

where $\boldsymbol{\mu}$ is the $p \times 1$ location vector, $\boldsymbol{w} = diag(w_1, \ldots, w_p)$, $w_i = \sqrt{w_{ii}}$, and $\Sigma = (w_{ij})$ is a full rank $p \times p$ covariance matrix. If the p-dimensional random vector $\boldsymbol{X} \in R^p$ has the following probability density function (pdf) it is said that it has an MSN distribution ($\boldsymbol{X} \sim MSN_p(\boldsymbol{\mu},\ \Sigma, \boldsymbol{\lambda})$):

$$f_{MSN}(\boldsymbol{x}; \boldsymbol{\mu},\ \Sigma, \boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{x}; \boldsymbol{\mu},\ \Sigma)\,\Phi\left(\boldsymbol{\lambda}^T \boldsymbol{w}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{3}$$

where $\Sigma = \boldsymbol{w}\Sigma_{\boldsymbol{Z}}\boldsymbol{w}$. Alternatively, it can be written as follows:

$$f_{MSN}(\boldsymbol{x}; \boldsymbol{\mu},\ \Sigma, \boldsymbol{\lambda}) = 2\phi_p(\boldsymbol{x}; \boldsymbol{\mu},\ \Sigma)\,\Phi\left(\boldsymbol{\lambda}^T \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})\right). \tag{4}$$

The random vector $\boldsymbol{X}$ has the following stochastic representation ([2])

$$\boldsymbol{X} = \boldsymbol{\mu} + \Sigma^{1/2}\boldsymbol{\delta}\,|U_1| + \Sigma^{1/2}\left(I_p - \boldsymbol{\delta}\boldsymbol{\delta}^T\right)^{1/2}\boldsymbol{U}_2, \quad U_1 \perp \boldsymbol{U}_2, \tag{5}$$

where $\boldsymbol{\delta} = \boldsymbol{\lambda}/\sqrt{1 + \boldsymbol{\lambda}^T\boldsymbol{\lambda}}$, $U_1 \sim N(0,1)$, $\boldsymbol{U}_2 \sim N_p(0, \boldsymbol{I}_p)$ and the symbol '$\perp$' denotes independence.

Proposition 1.1. *Let $\boldsymbol{X} \sim MSN_p(\boldsymbol{\mu},\ \Sigma, \boldsymbol{\lambda})$. Then the characteristic function of $\boldsymbol{X}$ is*

$$\begin{aligned} \Psi_{\boldsymbol{X}}(t) &= 2\exp\left(it^T\boldsymbol{\mu} - t^T\Sigma t/2\right)\,\Phi\left(i\boldsymbol{\eta}^T wt\right) \\ &= \exp\left(it^T\boldsymbol{\mu} - t^T\Sigma t/2\right)\left\{1 + i\tau\left(\boldsymbol{\eta}^T wt\right)\right\}, \quad t \in R^p \end{aligned} \tag{6}$$

where $\boldsymbol{\eta} = \Sigma_{\boldsymbol{Z}}\boldsymbol{\lambda}/\left(1 + \boldsymbol{\lambda}^T\Sigma_{\boldsymbol{Z}}\boldsymbol{\lambda}\right)^{1/2}$ and

$$\tau(x) = \int_0^x \sqrt{\frac{2}{\pi}}\exp\left(\frac{u^2}{2}\right) du, \quad x > 0, \quad \tau(-x) = -\tau(x). \tag{7}$$

See [24] for the detail of the characteristic function of the MSN distribution.

Using the stochastic representation given in Eq. (5) the mean vector and the covariance matrix of $\boldsymbol{X}$ are obtained as:

$$\begin{aligned} E(\boldsymbol{X}) &= \boldsymbol{\mu} + \sqrt{2/\pi}\Sigma^{1/2}\boldsymbol{\delta}, \\ cov(\boldsymbol{X}) &= \Sigma - \frac{2}{\pi}\Sigma^{1/2}\boldsymbol{\delta}\boldsymbol{\delta}^T\Sigma^{1/2} . \end{aligned} \tag{8}$$

Additionally, there are some multivariate distributions based on modifications of MSN distribution. For instance, [36] defines a new class of multivariate distributions including MSN distribution with applications to Bayesian regression models, [2] introduced a new class of MSN distributions, fundamental skew-normal distributions, and their canonical version, [5] examined many extensions of MSN distribution and some forms of semi-parametric formulations, [3] proposed the unification of skew-normal families, restricted MSN distribution was adopted by [26], unrestricted SN generalized-hyperbolic distribution was introduced by [29] applied for robust finite mixture modeling; and also, this distribution was used by [30] for Bayesian linear mixed models.

This paper focuses on both modeling skewness and heavy-tailedness in multivariate data sets. In this paper, we consider the multivariate skew Laplace normal (MSLN) distribution as a special case of the multivariate skew generalized Laplace normal (MSGLN) distribution proposed by [38]. The MSLN distribution is also more applicable than the MSN distribution thanks to its wider range of skewness and heavy-tailedness for modeling multivariate data sets. In the literature, multivariate skew and heavy-tailed data sets are usually modeled by using multivariate skew t (MST) ([35] and [27]) distribution or recently proposed multivariate skew-t-normal (MSTN) distribution ([28]). Additionally, MSGLN distribution was also proposed for modeling skew and heavy-tailed multivariate data settings. Although all of these distributions are very useful for modeling skew and heavy-tailed multivariate data sets, the MSTN distribution has an extra degrees of freedom parameter and the MSGLN distribution includes an extra shape parameter, which represents a very broad form of these families with several parameters. These extra parameters make computational intensity. On the other hand, for the MSLN case, we have only three parameters to deal with and this makes easy computation. Based on this important advantage, we propose the MSLN distribution as an alternative to the MSTN and MSGLN distributions to model both skewness and heavy-tailedness in the multivariate data settings.

We explore the definition of this distribution in detail by providing the stochastic representation which will be useful in the maximum likelihood (ML) estimation. We study the expectation-maximization (EM) algorithm to compute the ML estimates and provide a simulation study as well as a real data example to demonstrate the modeling performance of the proposed distribution. Results of our numerical studies confirm that the proposed EM algorithm works efficiently to find the ML estimates and the results also exhibit the superiority of MSLN distribution over the MSTN and MSGLN distributions for modeling skewness and heavy-tailedness in the real data example. Moreover, our results confirm that MSLN distribution has less computational time in seconds than the MSTN and MSGLN distributions are required to carry on parameter estimation.

The paper is organized as follows. Section 2 sketches the MSLN distribution with some properties. Section 3 proposes the ML estimators for the parameters of the MSLN distribution via the EM algorithm. Section 4 provides a simulation study and a real data example to illustrate the performance of the MSLN distribution. Section 5 is devoted to the conclusions.

2. THE MULTIVARIATE SKEW LAPLACE NORMAL DISTRIBUTION: DEFINITION AND PROPERTIES

Definition 2.1. A random vector $\boldsymbol{Y} \in R^p, p \geq 1$, is said to have a p-dimensional MSLN distribution ($\boldsymbol{Y} \sim MSLN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$) with the $p \times 1$ location vector $\boldsymbol{\mu}$, $p \times p$ dispersion matrix Σ, and $p \times 1$ skewness parameter vector $\boldsymbol{\lambda}$, if its pdf is given by

$$\psi(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}) = 2 f_{MLap}(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)\, \Phi\left(\boldsymbol{\lambda}^T \Sigma^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})\right), \tag{9}$$

where $f_{MLap}(\cdot\,; \boldsymbol{\mu}, \Sigma)$ is the pdf of multivariate Laplace (MLap) distribution

$$f_{MLap}(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma) = \frac{|\Sigma|^{-1/2}}{2^p \pi^{\frac{p-1}{2}} \Gamma\left(\frac{p+1}{2}\right)} e^{-\sqrt{(\boldsymbol{y}-\boldsymbol{\mu})^T \Sigma^{-1}(\boldsymbol{y}-\boldsymbol{\mu})}}, \quad y \in R^p,\ p \geq 1.$$

For the details about the MLap distribution see [19], [25], [20], [18] and [32].

Proposition 2.1. *Let* $U_1 \sim N(0,1)$, $\boldsymbol{U}_2 \sim N_p(0, \boldsymbol{I}_p)$, *and* V *has the inverse gamma distribution with the pdf*

$$g(v) = \frac{1}{\Gamma\left(\frac{p+1}{2}\right) 2^{\frac{p+1}{2}}} v^{-\left(\frac{p+1}{2}+1\right)} e^{-\frac{1}{2v}}, \quad v > 0, \tag{10}$$

and suppose that U_1, $\boldsymbol{U}_2$ *and* V *are mutually independent. Then, the random variable*

$$\boldsymbol{Y} = \boldsymbol{\mu} + \Sigma^{1/2}\left[\frac{\boldsymbol{\lambda}|U_1|}{\sqrt{V(V + \boldsymbol{\lambda}^T\boldsymbol{\lambda})}} + \left(V I_p + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)^{-1/2}\boldsymbol{U}_2\right] \tag{11}$$

has an MSLN distribution with the pdf given in Eq. (9).

Proof. Let $\gamma = \sqrt{V^{-1}\left(V + \boldsymbol{\lambda}^T\boldsymbol{\lambda}\right)}\,|U_1|$. Then, the hierarchical representation can be obtained as follows:

$$\boldsymbol{Y}|\left(\gamma, v\right) \sim N_p\left(\boldsymbol{\mu} + \frac{\Sigma^{1/2}\boldsymbol{\lambda}\gamma}{v + \boldsymbol{\lambda}^T\boldsymbol{\lambda}}, \Sigma^{1/2}\left(vI_p + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)^{-1}\Sigma^{1/2}\right),$$
$$\gamma|v \sim TN\left(0, \frac{v + \boldsymbol{\lambda}^T\boldsymbol{\lambda}}{v}; (0, \infty)\right),$$
$$v \sim g\left(v\right), \tag{12}$$

where $TN\left(\cdot\right)$ denotes the truncated normal distribution. Using the hierarchical representation given in (12), the joint pdf of $\boldsymbol{Y}$, γ, and V can be written as:

$$\begin{aligned} f\left(\boldsymbol{y}, \gamma, v\right) &= \frac{1}{2^p\pi^{\frac{p+1}{2}}|\Sigma|^{1/2}}\frac{v^{-3/2}e^{-1/2v}}{\Gamma\left(\frac{p+1}{2}\right)} \\ &\times\exp\left\{-\frac{1}{2}\left(v\boldsymbol{u}^T\boldsymbol{u} + \left(\gamma - \boldsymbol{\lambda}^T\boldsymbol{u}\right)^2\right)\right\}, \end{aligned} \tag{13}$$

where $\boldsymbol{u} = \Sigma^{-1/2}\left(\boldsymbol{y} - \boldsymbol{\mu}\right)$. After taking the integral of (13) over γ, we have

$$\begin{aligned} f\left(\boldsymbol{y}, v\right) &= \frac{1}{2^{p-\frac{1}{2}}\pi^{\frac{p}{2}}|\Sigma|^{1/2}}\frac{v^{-3/2}e^{-1/2v}}{\Gamma\left(\frac{p+1}{2}\right)} \\ &\times\exp\left\{-\frac{v\boldsymbol{u}^T\boldsymbol{u}}{2}\right\}\Phi\left(\boldsymbol{\lambda}^T\boldsymbol{u}\right). \end{aligned} \tag{14}$$

Then, taking the integral of Eq. (14) over v, the pdf given in Eq. (9) is obtained. □

Note that as in the Introduction we have already mentioned that this distribution becomes a special case of the MSGLN distribution ([38]) when we take the parameter α in MSGLN distribution as $(p+1)/2$ we get MSLN distribution. However, the advantage of the MSLN distribution is that it has a fewer number of parameters to deal with, which provides computational tractability. Finally, we can also note that the distribution of $\boldsymbol{Y}$ will be the MLap distribution with pdf $f_{MLap}\left(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma\right)$ when $\boldsymbol{\lambda} = \boldsymbol{0}$.

When $p = 1$ the density function of the MSLN distribution will be reduced to skew Laplace normal distribution with the following pdf:

$$f_{SLN}\left(y\right) = 2f_L\left(y; \mu, \sigma\right)\Phi\left(\lambda\left(\frac{\lambda\left(y - \mu\right)}{\sigma}\right)\right), \tag{15}$$

where $\mu \in R$ is the location parameter, $\sigma^2 \in (0, \infty)$ is the scale parameter, $\lambda \in R$ is the skewness parameter, and $f_L\left(y; \mu, \sigma\right)$ represents the pdf of Laplace distribution with

$$f_L\left(y; \mu, \sigma\right) = \frac{1}{2\sigma}e^{-\frac{|y-\mu|}{\sigma}}.$$

For the applicability of the univariate skew Laplace normal distribution ([21]), one can see papers by [14], [15], and [16].

To illustrate the shape of the MSLN distribution, we display examples of the MSLN densities and the contour plots for $\boldsymbol{\lambda} = (1, 1)^T$ and $\boldsymbol{\lambda} = (1, -0.5)^T$ in Figure 1. These plots show the peakedness, heavy-tailedness, and skewness of this distribution.

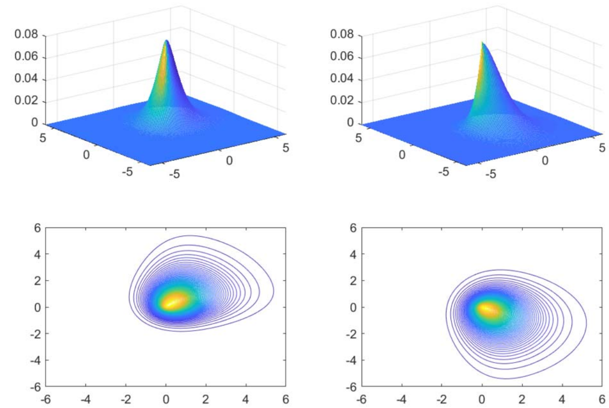

Figure 1. Two examples of MSLN densities for $\boldsymbol{\mu} = (0, 0)^T, \Sigma = I, \boldsymbol{\lambda} = (1, 1)^T$ (left side) and $\boldsymbol{\mu} = (0, 0)^T, \Sigma = I, \boldsymbol{\lambda} = (1, -0.5)^T$ (right side).

Proposition 2.2. *For $t \in R^p$, the characteristic function of $\boldsymbol{Y} \sim MSLN_p\left(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}\right)$ is*

$$\Psi_{\boldsymbol{Y}}\left(t\right) = e^{it^T\boldsymbol{\mu}}E_V\left(\exp(-V^{-1}t^T\Sigma t/2)\left(1 + i\tau(\boldsymbol{\kappa}^Twt)\right)\right), \tag{16}$$

where $\boldsymbol{\kappa} = \frac{V^{-3/2}\Sigma_{\boldsymbol{Z}}\boldsymbol{\lambda}}{\left(1+V^{-2}\boldsymbol{\lambda}^T\Sigma_{\boldsymbol{Z}}\boldsymbol{\lambda}\right)^{1/2}}$ and the function $\tau\left(\cdot\right)$ is defined in Eq. (7). We note that the expectation given in Eq. (16) can be calculated by numerical methods.

The following is another scale mixture representation of the MSLN distribution

$$\boldsymbol{Y}|v \sim MSN_p\left(\boldsymbol{\mu}, v^{-1}\Sigma, v^{-1/2}\boldsymbol{\lambda}\right),$$
$$v \sim g\left(v\right). \tag{17}$$

By using this representation and the equation given in (8), we get the expectation and covariance matrix of MSLN distribution as

$$\begin{aligned} E\left(\boldsymbol{Y}\right) &= \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\Sigma^{1/2}\boldsymbol{\lambda}\,\zeta, \\ Cov\left(\boldsymbol{Y}\right) &= (p+1)\,\Sigma - \frac{2}{\pi}\Sigma^{1/2}\boldsymbol{\lambda}\boldsymbol{\lambda}^T\Sigma^{1/2}\zeta^2, \end{aligned}$$

where $\zeta = E\left(\frac{v^{-1/2}}{\sqrt{v+\boldsymbol{\lambda}^T\boldsymbol{\lambda}}}\right)$, which can be computed by using numerical methods.

In the following propositions, the distribution of the linear transformation, marginal distributions and conditional distributions of a random vector with MSLN distribution are given. Since these propositions and their proofs are very similar to those given by [4] for the multivariate skew Laplace distribution of different type, we will briefly outline the proofs of these propositions for our proposed distribution.

Proposition 2.3. *If* $\boldsymbol{Y} \sim MSLN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$ *and* $\boldsymbol{T} = A\boldsymbol{Y} + \boldsymbol{b}$ *for* $A \in R^{q\times p}$ *and* $\boldsymbol{b} \in R^{q\times 1}$ *then* $\boldsymbol{T} \sim MSLN_q\left(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T, A\boldsymbol{\lambda}\right)$.

Proof. For $t \in R^q$, the characteristic function of $\boldsymbol{T}$ is

$$\Psi_{\boldsymbol{T}}(t) = E\left(e^{it^T(A\boldsymbol{Y}+\boldsymbol{b})}\right) = e^{it^T\boldsymbol{b}}\Psi_{\boldsymbol{Y}}\left(A^T t\right)$$
$$= \mathrm{e}^{it^T(A\boldsymbol{\mu}+\boldsymbol{b})}E_V(\exp(-V^{-1}t^T \mathrm{A}\Sigma\mathrm{A}^{\mathrm{T}} t/2)(1 + i\tau(\boldsymbol{\eta}^T w A^T t))).$$

Then, $\boldsymbol{T} \sim MSLN_q\left(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T, A\boldsymbol{\lambda}\right)$. □

From Proposition 2.3, we can conclude that if $\boldsymbol{Y} \sim MSLN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$, for any $\boldsymbol{b} \in R^p$, $\boldsymbol{b}^T\boldsymbol{Y} \sim MSLN_1\left(\boldsymbol{b}^T\boldsymbol{\mu}, \boldsymbol{b}^T\Sigma\boldsymbol{b}, \boldsymbol{b}^T\boldsymbol{\lambda}\right)$.

Proposition 2.4. *Let* $\boldsymbol{Y} \sim MSLN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$ *and* $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$ *be the partition of* $\boldsymbol{Y}$ *whose dimensions are* k *and* s, $k + s = p$. *Assume that* $\boldsymbol{\mu}$, Σ, *and* $\boldsymbol{\lambda}$ *are suitably partitioned as*

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{bmatrix} \ \text{and}$$
$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\boldsymbol{Y}_1, \boldsymbol{\mu}_1$ *and* $\boldsymbol{\lambda}_1$ *are* $k \times 1$ *vectors and* Σ_{11} *is* $k \times k$ *matrix. Then, we can have the following results:*

i) $\boldsymbol{Y}_1 \sim MSLN_k(\boldsymbol{\mu}_1, \Sigma_{11}, \boldsymbol{\lambda}_1)$.
ii) The conditional distribution of $\boldsymbol{Y}_2$ *given* $\boldsymbol{Y}_1 = y_1$ *is a normal variance-mean mixture distribution with the parameters* $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{y}_1 - \boldsymbol{\mu}_1)$, $\boldsymbol{\lambda}_{2.1} = \boldsymbol{\lambda}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\lambda}_1$ *and* $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Proof. We can express proofs of i) and ii) as follows:
i) Let define $A\boldsymbol{Y} = \boldsymbol{Y}_1$. Here, $A = \begin{bmatrix} I_k & \boldsymbol{0}_{k\times(p-k)} \end{bmatrix}$ is a $k \times p$ matrix, I_k shows the identity matrix, and $\boldsymbol{0}_{k\times(p-k)}$ represents a null matrix. The distribution of $A\boldsymbol{Y} = \boldsymbol{Y}_1$ will be an MSLN distribution with the parameters $A\boldsymbol{\mu} = \boldsymbol{\mu}_1$, $A\Sigma A^T = \Sigma_{11}$, and $A\boldsymbol{\lambda} = \lambda$ according to Proposition 2.3.
ii) Let $\boldsymbol{T}$ be a linear transformation of $\boldsymbol{Y}$ defined as follows:

$$\begin{aligned}\boldsymbol{T} &= \begin{bmatrix} I_k & \boldsymbol{0}_{k\times(p-k)} \\ -\Sigma_{21}\Sigma_{11}^{-1} & I_{p-k} \end{bmatrix}\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \end{bmatrix} \\ &+ \begin{bmatrix} \boldsymbol{0}_k \\ \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\mu}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{U} \end{bmatrix},\end{aligned}$$

where $\boldsymbol{U} = \boldsymbol{Y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1)$. Then, using Proposition 2.3, the distribution of $\boldsymbol{T}$ will be an MSLN distribution with the following parameters

$$\boldsymbol{\mu_T} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{\lambda_T} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{\lambda}_1 \end{bmatrix} \ \text{and}$$
$$\Sigma = \begin{bmatrix} \Sigma_{11} & \boldsymbol{0}_{k\times(p-k)} \\ \boldsymbol{0}_{(p-k)\times k} & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix}.$$

Then, the ratio of the density function of $\boldsymbol{T}$ and $\boldsymbol{Y}_1$ equals to the conditional density function of $\boldsymbol{U} = \boldsymbol{Y}_2 - \boldsymbol{\mu}_2 - \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{Y}_1 - \boldsymbol{\mu}_1)$ given $\boldsymbol{Y}_1 = \boldsymbol{y}_1$ with the parameters $\boldsymbol{\mu}_{2.1}$, $\boldsymbol{\lambda}_{2.1}$, and $\Sigma_{22.1}$. We can reveal that the conditional distribution of $\boldsymbol{Y}_2$ given $\boldsymbol{Y}_1$ is not an MSLN distribution, rather it is in the class of normal variance–mean mixture distributions. □

Further, we can also obtain the following conditional distribution. First, when $\boldsymbol{y}$ and v are given, the pdf of the conditional distribution of γ is given as follows:

$$f(\gamma|\boldsymbol{y}, v) = \frac{1}{\sqrt{2\pi}\Phi\left(\boldsymbol{\lambda}^T\boldsymbol{u}\right)}\exp\left\{-\frac{1}{2}\left(\gamma - \boldsymbol{\lambda}^T\boldsymbol{u}\right)^2\right\}. \tag{18}$$

This can be easily found by dividing Eq. (13) by Eq. (14). From the density function given in Eq. (18), it is obvious that γ and V are conditionally independent; therefore, the distribution of $\gamma|\boldsymbol{y}$ is:

$$\gamma|\boldsymbol{y} \sim TN\left(\boldsymbol{\lambda}^T\boldsymbol{u}, 1; (0, \infty)\right). \tag{19}$$

Finally, after dividing (14) by (9) we have the following conditional density function:

$$f(v \mid \boldsymbol{y}) = \frac{1}{\sqrt{2\pi}}v^{-3/2}\exp\left\{\frac{1}{2}\boldsymbol{u}^T\boldsymbol{u} - \frac{1}{2}\left(v\boldsymbol{u}^T\boldsymbol{u} + v^{-1}\right)\right\}. \tag{20}$$

Proposition 2.5. *Using the hierarchical representation given in (12) and the conditional distribution (19) and (20), we get the following conditional expectations:*

$$\begin{aligned} E(V|\boldsymbol{y}) &= \left((\boldsymbol{y} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right)^{-1/2}, \\ E(\gamma|\boldsymbol{y}) &= \boldsymbol{\lambda}^T\boldsymbol{u} + \frac{\phi\left(\boldsymbol{\lambda}^T\boldsymbol{u}\right)}{\Phi\left(\boldsymbol{\lambda}^T\boldsymbol{u}\right)}. \end{aligned} \tag{21}$$

We note that these conditional expectations will be used in the EM algorithm given in the next section.

3. THE ML ESTIMATION

Let $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ be a p-dimensional random sample from MSLN distribution with the unknown parameters $\boldsymbol{\Theta} = (\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$. The ML estimator for $\boldsymbol{\Theta}$ can be obtained by maximizing the following log-likelihood function:

$$\ell(\boldsymbol{\Theta}) = \sum_{i=1}^{n}\log\left(\psi\left(\boldsymbol{y}_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}\right)\right). \tag{22}$$

Since there is no explicit solution for the log-likelihood function, we need a numerical method. Mostly, the EM algorithm ([12]) is used to get the ML estimator of $\boldsymbol{\Theta}$.

Let V and γ be the latent variables and $(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{v})$ be the complete data, where $\boldsymbol{y}=\left(\boldsymbol{y}_1^T,\ldots,\boldsymbol{y}_n^T\right)^T$, $\boldsymbol{\gamma}=(\gamma_1,\ldots,\gamma_n)^T$ and $\boldsymbol{v}=(v_1,\ldots,v_n)$. Using the hierarchical representation given in (12) and ignoring the constants, the complete data log-likelihood function can be written by:

$$\begin{aligned}(23)\quad \ell_c(\boldsymbol{\Theta};\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{v}) &= -\frac{n}{2}\log|\Sigma| \;-\frac{1}{2}\sum_{i=1}^n 3logv_i+v_i^{-1}\\ &-\frac{1}{2}\sum_{i=1}^n v_i(\boldsymbol{y}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})-\frac{1}{2}\sum_{i=1}^n\gamma_i^2\\ &-\frac{1}{2}\sum_{i=1}^n\left[-2\gamma_i\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu})+\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\beta}\right],\end{aligned}$$

where $\boldsymbol{\beta}=\Sigma^{-1/2}\boldsymbol{\lambda}$ is a reparameterized parameter vector. To get rid of the latency problem in the complete data log-likelihood function, we have to take the conditional expectation of the complete data log-likelihood function given the observed data $\boldsymbol{y}_i$

(24)

$$\begin{aligned}E\left(\ell_c(\boldsymbol{\Theta};\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{v})\,|\boldsymbol{y}_i\right) &= -\frac{n}{2}\log|\Sigma| \;-\frac{1}{2}\sum_{i=1}^n 3E\left(logV_i|\boldsymbol{y}_i\right)\\ &-\frac{1}{2}\sum_{i=1}^n E\left(V_i|\boldsymbol{y}_i\right)(\boldsymbol{y}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})+E\left(V_i^{-1}\mid\boldsymbol{y}_i\right)\\ &-\frac{1}{2}\sum_{i=1}^n(E(\gamma_i^2|\boldsymbol{y}_i)-2E(\gamma_i\mid\boldsymbol{y}_i)\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu}))\\ &-\frac{1}{2}\sum_{i=1}^n\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\beta}.\end{aligned}$$

Since some conditional expectations are not related to the parameters, we only compute the conditional expectations $E(V_i|\boldsymbol{y}_i)$ and $E(\gamma_i\mid\boldsymbol{y}_i)$ using Proposition 2.5. Now, we are ready to implement the EM algorithm by using the following steps:

EM algorithm:
1. Take initial parameter estimate $\boldsymbol{\Theta}^{(0)}$ and a stopping rule Δ.
2. E-Step: Calculate the following conditional expectations for $k=0,1,2,\ldots$

$$\begin{aligned}\hat{v}_i^{(k)} &= E(V_i|\boldsymbol{y}_i)\\ (25)\qquad &= \left(\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)^T\widehat{\Sigma}^{(k)-1}\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)\right)^{-1/2},\end{aligned}$$

$$(26)\qquad \widehat{\gamma}_i^{(k)}=E(\gamma_i|\boldsymbol{y}_i)=\widehat{\boldsymbol{\lambda}}_i^{(k)T}\boldsymbol{u}_i+\frac{\phi\left(\widehat{\boldsymbol{\lambda}}_i^{(k)T}\boldsymbol{u}_i\right)}{\Phi\left(\widehat{\boldsymbol{\lambda}}_i^{(k)T}\boldsymbol{u}_i\right)}.$$

Using these conditional expectations, we obtain the following objective function which will be maximized with respect to the unknown parameters:

$$\begin{aligned}Q(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}^{(k)}) &= -\frac{n}{2}\log|\Sigma| \;+\sum_{i=1}^n\widehat{\gamma}_i^{(k)}\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu})\\ &-\frac{1}{2}\sum_{i=1}^n\hat{v}_i^{(k)}(\boldsymbol{y}_i-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})\\ (27)\qquad &-\frac{1}{2}\sum_{i=1}^n\boldsymbol{\beta}^T(\boldsymbol{y}_i-\boldsymbol{\mu})(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\beta}.\end{aligned}$$

3. M-Step: Maximizing $Q(\boldsymbol{\Theta};\widehat{\boldsymbol{\Theta}}^{(k)})$ with respect to $\boldsymbol{\Theta}$ yields the following updating equations to compute the $(k+1)\,th$ parameter estimates:

$$\begin{aligned}(28)\quad \widehat{\boldsymbol{\mu}}^{(k+1)} &= \left(\sum_{i=1}^n\hat{v}_i^{(k)}\Sigma^{-1}+n\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)T}\right)^{-1}\\ &\times\left(\sum_{i=1}^n\hat{v}_i^{(k)}\Sigma^{-1}\boldsymbol{y}_i-\sum_{i=1}^n\widehat{\gamma}_i^{(k)}\widehat{\boldsymbol{\beta}}^{(k)}+\widehat{\boldsymbol{\beta}}^{(k)}\widehat{\boldsymbol{\beta}}^{(k)T}\sum_{i=1}^n\boldsymbol{y}_i\right),\end{aligned}$$

$$(29)\quad \widehat{\Sigma}^{(k+1)}=\frac{1}{n}\sum_{i=1}^n\hat{v}_i^{(k)}\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)^T,$$

$$\begin{aligned}(30)\quad \widehat{\boldsymbol{\beta}}^{(k+1)} &= \left(\sum_{i=1}^n\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)^T\right)^{-1}\\ &\times\left(\sum_{i=1}^n\widehat{\gamma}_i^{(k)}\left(\boldsymbol{y}_i-\widehat{\boldsymbol{\mu}}^{(k)}\right)\right),\end{aligned}$$

$$(31)\quad \widehat{\boldsymbol{\lambda}}^{(k+1)}=\left(\widehat{\Sigma}^{(k+1)}\right)^{1/2}\widehat{\boldsymbol{\beta}}^{(k+1)}.$$

4. Repeat E and M steps until the convergence rule $\left\|\widehat{\boldsymbol{\Theta}}^{(k+1)}-\widehat{\boldsymbol{\Theta}}^{(k)}\right\|<\Delta$ is satisfied.

3.1 Estimation of standard errors

We will calculate the standard errors of the ML estimators by using the method proposed by [31]. This method is based on an approximation to the observed information matrix by the score vector which can be used to estimate the standard errors of parameters. This observed information matrix can be transformed as:

$$(32)\qquad \hat{I}_e=\sum_{i=1}^n\widehat{\boldsymbol{s}}_i\widehat{\boldsymbol{s}}_i^T\;,$$

where $\widehat{\boldsymbol{s}}_i=E_{\widehat{\boldsymbol{\Theta}}}\left(\frac{\partial\ell_{ci}(\boldsymbol{\Theta};\boldsymbol{y}_i,\gamma_i,v_i)}{\partial\boldsymbol{\Theta}}\middle|\boldsymbol{y}_i\right)$, $i=1,\ldots,n$ are the individual scores and $\ell_{ci}(\boldsymbol{\Theta};\boldsymbol{y}_i,\gamma_i,v_i)$ is the complete data log-likelihood function for the ith observation. Let $S=\Sigma^{1/2}$

and $\mathbf{\Theta} = (\boldsymbol{\mu}, S, \boldsymbol{\lambda})$ be the parameter vector. Then, we get

$$\begin{aligned}\ell_{ci}(\mathbf{\Theta};\boldsymbol{y}_i,\gamma_i,v_i) &= -\log|S| - \frac{1}{2}\left(3logv_i + v_i^{-1}\right)\\ &- \frac{1}{2}\left(v_i\boldsymbol{u}_i^T\boldsymbol{u}_i + \left(\gamma_i - \boldsymbol{\lambda}^T\boldsymbol{u}_i\right)^2\right),\end{aligned} \tag{33}$$

where $\boldsymbol{u}_i = S^{-1}((\boldsymbol{y}_i - \boldsymbol{\mu}))$. The score vector $\widehat{\boldsymbol{s}}_i$ is $(\widehat{\boldsymbol{s}}_{i,\boldsymbol{\mu}}, \widehat{\boldsymbol{s}}_{i,s}, \widehat{\boldsymbol{s}}_{j,\boldsymbol{\lambda}})$, where $\boldsymbol{s} = vech(S)$. After taking the derivative of ℓ_{ci} with respect to the parameters, the elements of $\widehat{\boldsymbol{s}}_i$ are obtained as follows:

$$\begin{aligned}&\widehat{\boldsymbol{s}}_{i,\boldsymbol{\mu}} = \hat{v}_i\hat{S}^{-1}\,\widehat{\boldsymbol{u}}_i - \left(\widehat{\gamma}_i - \widehat{\boldsymbol{\lambda}}^T\widehat{\boldsymbol{u}}_i\right)\hat{S}^{-1}\widehat{\boldsymbol{\lambda}},\\ &\widehat{\boldsymbol{s}}_{i,s} = vech\left\{-\left(2\hat{S}^{-1} - Diag\left(\hat{S}^{-1}\right)\right)\right.\\ &+\hat{v}_i\left(\hat{B}_i + \hat{B}_i^T - Diag\left(\hat{B}_i\right)\right)\\ &\left.-\left(\widehat{\gamma}_i - \widehat{\boldsymbol{\lambda}}^T\widehat{\boldsymbol{u}}_i\right)\left(\hat{C}_i + \hat{C}_i^T - Diag\left(\hat{C}_i\right)\right)\right\},\\ &\widehat{\boldsymbol{s}}_{j,\boldsymbol{\lambda}} = \left(\widehat{\gamma}_i - \widehat{\boldsymbol{\lambda}}^T\widehat{\boldsymbol{u}}_i\right)\widehat{\boldsymbol{u}}_i,\end{aligned}$$

where $\hat{S} = \widehat{\Sigma}^{1/2}$, $\widehat{\boldsymbol{u}}_i = \hat{S}^{-1}((\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}}))$, $\hat{B}_i = \hat{S}^{-1}\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i^T$, $\hat{C}_i = \widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{\lambda}}^T\hat{S}^{-1}$, and $\hat{v}_i$ and $\widehat{\gamma}_i$ can be computed by using the equations given in (25) and (26) evaluated at $\widehat{\mathbf{\Theta}}$. Consequently, the standard errors of the estimates can be obtained by using the square roots of the diagonal elements of $\hat{I}_e^{-1}$.

4. APPLICATIONS

This section will provide a simulation study and a real data example in order to demonstrate the applicability of the MSLN distribution. We summarize the following computational details for the simulation study and the real data example.

Details of the computation:

i) All simulation studies and the real data example are carried out by using MATLAB R2017b software.

ii) We take the stopping rule Δ as 10^{-6} for all numerical computations.

iii) We set the true parameter values as initial values for the EM algorithm in the simulation study. In the real data example, we use the mean vector, covariance matrix, and the skewness vector of the data set as initial values for the location vector, dispersion matrix, and the skewness vector of the MSLN distribution.

iv) For the simulation study, the number of replications is set as $N = 500$ and the sample sizes (n) are taken as 200, 400, and 600 respectively.

v) For the simulation study, the data set is generated by using the following procedure:

– Sample $U_1 \sim N(0,1)$, $\boldsymbol{U}_2 \sim N_p(0, \boldsymbol{I}_p)$, and V from the inverse gamma distribution given in (10) independently. We note that we can easily sample data from the inverse gamma distribution by applying the relation $1/V \sim Gamma((p+1)/2, 2)$.

– Then, $\boldsymbol{Y} = \boldsymbol{\mu} + \Sigma^{1/2}\left[\frac{\boldsymbol{\lambda}|U_1|}{\sqrt{V(V+\boldsymbol{\lambda}^T\boldsymbol{\lambda})}} + \left(VI_p + \boldsymbol{\lambda}\boldsymbol{\lambda}^T\right)^{-1/2}\boldsymbol{U}_2\right]$ generates the sample from the $MSLN_p(\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda})$ distribution.

4.1 Simulation study

In the simulation study, the data set are generated from the MSLN distribution by using the procedure given in item (v) of details of the computation. The performance of the proposed distribution is assessed by using the bias and the mean Euclidean distance values of estimates. The formula for the bias is given by:

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

where θ is the true parameter value, $\widehat{\theta}_i$ is the estimate of θ for the ith simulated data, and $\overline{\theta} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_i$. For instance, the formula of mean Euclidean distance of $\widehat{\boldsymbol{\mu}}$ is given below:

$$\|\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\| = \frac{1}{N}\left(\sum_{i=1}^{N}(\widehat{\mu}_i - \mu_i)^2\right)^{\frac{1}{2}}.$$

This simulation study focuses on modeling skewness and heavy-tailedness in different configurations. First, the data set is generated from the following bivariate MSLN distribution with given parameters:

$$\begin{aligned}&\psi(\boldsymbol{y}_i;\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}), \quad \boldsymbol{\mu} = (\mu_1, \mu_2)^T,\\ &\Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}, \quad \boldsymbol{\lambda} = (\lambda_1, \lambda_2)^T,\ i = 1,\ldots,n.\end{aligned} \tag{34}$$

Next, the data set are generated from the following three-variate MSLN distribution with given parameters:

$$\begin{aligned}&\psi(\boldsymbol{y}_i;\boldsymbol{\mu}, \Sigma, \boldsymbol{\lambda}), \quad \boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^T,\\ &\Sigma = \begin{bmatrix}\sigma_{11} & \sigma_{12} & \sigma_{13}\\ \sigma_{21} & \sigma_{22} & \sigma_{23}\\ \sigma_{31} & \sigma_{32} & \sigma_{33}\end{bmatrix},\\ &\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)^T,\ i = 1,\ldots,n.\end{aligned} \tag{35}$$

True parameters are represented in Table 1 and also simulated data examples are displayed with Figures 2 and 3. We can observe the peakedness, skewness and heavy-tailedness of the simulated data sets from these figures.

Tables 2 and 3 summarize the simulation results of Cases I–IV implemented for the sample sizes 200, 400, and 600. These tables are simulation examples of bivariate MSLN distribution, which consist of the true parameter values, bias and the mean Euclidian distance values of estimates. It can be observed from the tables that the proposed algorithm works accurately and estimate all parameters. Furthermore, the mean Euclidian distances for all parameter estimates are getting smaller when the sample size (n) increases and according to the bias values of estimates, the parameter estimates are very close to the true

Table 1. True parameter values for the MSLN distribution

Case	μ_1	μ_2	μ_3	σ_{11}	σ_{12}	σ_{13}	σ_{22}	σ_{23}	σ_{33}	λ_1	λ_2	λ_3
I	4	4	–	1.5	0	–	1.5	–	–	5	5	–
II	4	4	–	1.5	0	–	1.5	–	–	5	−5	–
III	4	4	–	5	0	–	0.5	–	–	1	1	–
IV	4	4	–	1	0	–	1	–	–	5	−0.1	–
V	4	4	1	1.5	0	0	1.5	0	1.5	−2	−2	2
VI	4	4	1	2	0	0	2	0	2	5	5	5

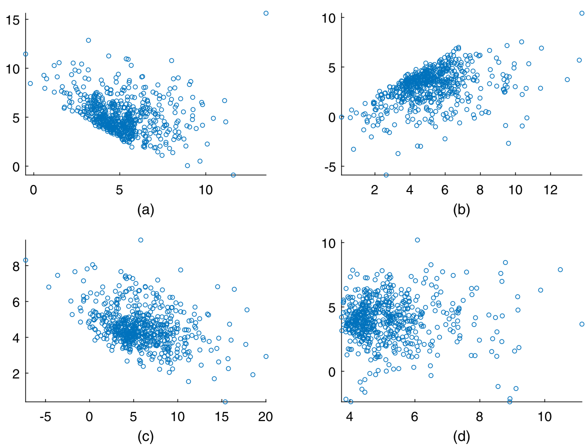

Figure 2. Simulated data set of $n = 600$ from bivariate MSLN distribution for Cases I–IV. For the sake of simplicity, it is taken as $\boldsymbol{\mu} = (4,4)^T$. a) $\boldsymbol{\lambda} = (5,5)^T,\ \sigma_{11} = 1.5,\ \sigma_{12} = 0$ and $\sigma_{22} = 1.5$. b) $\boldsymbol{\lambda} = (5,-5)^T,\ \sigma_{11} = 1.5,\ \sigma_{12} = 0$ and $\sigma_{22} = 1.5$. c) $\boldsymbol{\lambda} = (1,1)^T,\ \sigma_{11} = 5,\ \sigma_{12} = 0$ and $\sigma_{22} = 0.5$. d) $\boldsymbol{\lambda} = (5,-0.1)^T,\ \sigma_{11} = 1,\ \sigma_{12} = 0$ and $\sigma_{22} = 1$.

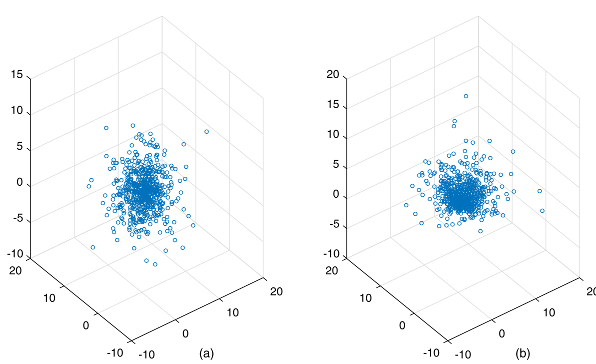

Figure 3. Simulated data set of $n = 600$ from three-variate MSLN distribution for Cases V–VI. For simplicity, it is set as $\boldsymbol{\mu} = (4,4,1)^T$. a) $\boldsymbol{\lambda} = (-2,-2,2)^T,\ \sigma_{11} = 1.5,\ \sigma_{12} = 0, \sigma_{13} = 0,\ \sigma_{22} = 1.5,\ \sigma_{23} = 0$ and $\sigma_{33} = 1.5$. b) $\boldsymbol{\lambda} = (5,5,5)^T,\ \sigma_{11} = 2,\ \sigma_{12} = 0, \sigma_{13} = 0,\ \sigma_{22} = 2,\ \sigma_{23} = 0$ and $\sigma_{33} = 2$.

parameter values. In addition to that, the center, scale, and skewness of the simulated data generated from (34) can be captured superbly.

Similarly, Table 4 gives the simulation results for Cases V and VI. These tables are simulation examples of three-variate MSLN distribution conducted for the sample sizes 200, 400, and 600. Also, this table includes the true parameter values, bias, and the mean Euclidian distance values of estimates. The observed algorithm gives exact results for the three-variate case. We observe the same results with the bivariate case as when the sample size (n) increases, the mean Euclidian distances for all parameter estimates decrease for all parameters. Moreover, the center, scale, and skewness of the simulated data generated from (35) can be predicted perfectly.

4.2 Real data example: Mastitis in dairy cattle

We will analyze the Mastitis data set given by [13] in order to illustrate the applicability of the proposed distribution. The data set was obtained from the appearance of the infectious disease called mastitis in dairy cows and was used to compare the average of the milk yield in two years, consists of the milk yields (in thousands of liters) of 107 dairy cows from a single herd in two consecutive years. For this data set, although all animals were healthy in the first year, in the second year 27 cows became missing data as they were infected. For further information see [13]. This data set was used by [10] for a multivariate generalization of the Heckman model to see the dropouts in repeated continuous responses. Also, [9] analyzed this data set for the ECM estimation in multivariate skew-normal data with dropout.

Since this data set includes extreme points, we consider that it may be an adequate example of modeling skewness and heavy-tailedness. Missing observations are ignored in the case of this example and analyses were made for the rest of the data set. Thus, we will use this data set to compare the performance of the MSLN distribution with the performance of MSTN and MSGLN distributions. We will use the following criteria for the comparison of distributions:

$$-2\ell\left(\widehat{\boldsymbol{\Theta}}\right) + mc_n$$

where $\ell(\cdot)$ represents the maximized log-likelihood, m is the number of free parameters to be estimated in the model and c_n is the penalty term. Here, we take $c_n = 2$ for the Akaike

Table 2. Bias and mean Euclidean distance values of estimates for the sample size 200, 400 *and* 600 *for Cases I and II*

	Case I				Case II			
n	Parameter	True	Bias	Distance	Parameter	True	Bias	Distance
200	μ_1	4	0.0075	0.1645	μ_1	4	0.0107	0.1614
	μ_2	4	0.0063		μ_2	4	-0.0048	
	σ_{11}	1.5	0.0076	0.2000	σ_{11}	1.5	0.0080	
	σ_{12}	0	-0.0119		σ_{12}	0	0.0192	0.2037
	σ_{22}	1.5	0.0112		σ_{22}	1.5	0.0119	
	λ_1	5	0.0491	0.6228	λ_1	5	0.0347	0.6356
	λ_2	5	0.0647		λ_2	-5	-0.0374	
400	μ_1	4	0.0005	0.1167	μ_1	4	0.0067	0.1135
	μ_2	4	0.0052		μ_2	4	0.0014	
	σ_{11}	1.5	0.0081	0.1443	σ_{11}	1.5	-0.0026	0.1394
	σ_{12}	0	-0.0079		σ_{12}	0	0.0088	
	σ_{22}	1.5	0.0022		σ_{22}	1.5	0.0122	
	λ_1	5	0.0217	0.4659	λ_1	5	0.0043	0.4389
	λ_2	5	0.0099		λ_2	-5	-0.0251	
600	μ_1	4	0.0027	0.0947	μ_1	4	0.0019	0.0924
	μ_2	4	-0.0016		μ_2	4	-0.0020	
	σ_{11}	1.5	0.0050	0.1168	σ_{11}	1.5	0.0057	0.1164
	σ_{12}	0	0.0038		σ_{12}	0	-0.0025	
	σ_{22}	1.5	0.0021		σ_{22}	1.5	0.0010	
	λ_1	5	0.0429	0.3686	λ_1	5	0.0245	0.3652
	λ_2	5	0.0259		λ_2	-5	-0.0213	

Table 3. Bias and mean Euclidean distance values of estimates for the sample size 200, 400 *and* 600 *for Cases III and IV*

	Case III				Case IV			
n	Parameter	True	Bias	Distance	Parameter	True	Bias	Distance
200	μ_1	4	-0.0165	0.3334	μ_1	4	0.0039	0.1451
	μ_2	4	0.0064		μ_2	4	-0.0198	
	σ_{11}	5	0.0704	0.5848	σ_{11}	1	-0.0033	0.1455
	σ_{12}	0	-0.0111		σ_{12}	0	0.0154	
	σ_{22}	0.5	0.0030		σ_{22}	1	0.0186	
	λ_1	1	0.0982	0.3897	λ_1	5	0.0941	0.6195
	λ_2	1	0.0743		λ_2	-0.1	0.0424	
400	μ_1	4	-0.0097	0.2132	μ_1	4	0.0025	0.0954
	μ_2	4	0.0010		μ_2	4	0.0014	
	σ_{11}	5	0.0207	0.3744	σ_{11}	1	0.0011	0.0943
	σ_{12}	0	0.0023		σ_{12}	0	-0.0037	
	σ_{22}	0.5	0.0022		σ_{22}	1	0.0053	
	λ_1	1	0.0473	0.2555	λ_1	5	0.0495	0.4345
	λ_2	1	0.0411		λ_2	-0.1	-0.0092	
600	μ_1	4	0.0007	0.1716	μ_1	4	0.0012	0.0795
	μ_2	4	-0.0014		μ_2	4	-0.0004	
	σ_{11}	5	0.0001	0.3041	σ_{11}	1	-0.0010	0.0768
	σ_{12}	0	0.0060		σ_{12}	0	0.0008	
	σ_{22}	0.5	0.0027		σ_{22}	1	0.0055	
	λ_1	1	0.0229	0.1949	λ_1	5	0.0405	0.3583
	λ_2	1	0.0237		λ_2	-0.1	-0.0039	

information criteria (AIC) ([1]) and $c_n = \log(n)$ for the Bayesian information criteria (BIC) ([37]).

We summarize the estimation results in Table 5. It includes ML estimates, standard errors (SE) of the estimates, information criterion values of estimates and CPU times (CT) in seconds, for MSLN, MSTN and MSGLN distributions. Here, the SEs of estimators for the parameters of the MSLN distribution are computed by using the method given by [31]. Details about the computation of the SEs can be found in sub-section 3.1. Based on the values of informa-

Table 4. Bias and mean Euclidean distance values of estimates for the sample size 200, 400 and 600 for Cases V and VI

	Case V				Case VI			
n	Parameter	True	Bias	Distance	Parameter	True	Bias	Distance
200	μ_1	4	-0.0042	0.3522	μ_1	4	0.0122	0.3461
	μ_2	4	-0.0078		μ_2	4	-0.0005	
	μ_3	1	0.0058		μ_3	1	0.0098	
	σ_{11}	1.5	0.0121	0.3586	σ_{11}	2	0.0293	0.4612
	σ_{12}	0	-0.0074		σ_{12}	0	-0.0025	
	σ_{13}	0	0.0159		σ_{13}	0	-0.0190	
	σ_{22}	1.5	0.0187		σ_{22}	2	0.0176	
	σ_{23}	0	0.0169		σ_{23}	0	-0.0169	
	σ_{33}	1.5	0.0144		σ_{33}	2	0.0053	
	λ_1	-2	-0.0830	0.6557	λ_1	5	0.0390	0.9263
	λ_2	-2	-0.0795		λ_2	5	0.0130	
	λ_3	2	0.0640		λ_3	5	-0.0168	
400	μ_1	4	-0.0105	0.2404	μ_1	4	0.0073	0.2451
	μ_2	4	0.0064		μ_2	4	0.0018	
	μ_3	1	0.0094		μ_3	1	0.0095	
	σ_{11}	1.5	-0.0010	0.2534	σ_{11}	2	0.0144	0.3264
	σ_{12}	0	-0.0065		σ_{12}	0	-0.0010	
	σ_{13}	0	0.0062		σ_{13}	0	-0.0156	
	σ_{22}	1.5	0.0253		σ_{22}	2	0.0120	
	σ_{23}	0	0.0033		σ_{23}	0	-0.0035	
	σ_{33}	1.5	-0.0016		σ_{33}	2	-0.0001	
	λ_1	-2	-0.0414	0.4623	λ_1	5	0.0181	0.6267
	λ_2	-2	-0.0616		λ_2	5	0.0230	
	λ_3	2	0.0433		λ_3	5	-0.0279	
600	μ_1	4	-0.0076	0.1924	μ_1	4	0.0062	0.2085
	μ_2	4	0.0028		μ_2	4	0.0055	
	μ_3	1	0.0006		μ_3	1	-0.0017	
	σ_{11}	1.5	-0.0034	0.2080	σ_{11}	2	0.0060	0.2637
	σ_{12}	0	-0.0052		σ_{12}	0	-0.0085	
	σ_{13}	0	0.0018		σ_{13}	0	0.0005	
	σ_{22}	1.5	0.0108		σ_{22}	2	0.0053	
	σ_{23}	0	-0.0026		σ_{23}	0	-0.0001	
	σ_{33}	1.5	0.0077		σ_{33}	2	0.0064	
	λ_1	-2	-0.0216	0.3705	λ_1	5	0.0023	0.5360
	λ_2	-2	-0.0354		λ_2	5	0.0022	
	λ_3	2	0.0422		λ_3	5	0.0149	

tion criteria given in Table 5, the best fit for this data set is obtained from MSLN distribution. It can also be observed that the computation time of the MSLN distribution is superior to the computation time of the MSTN and MSGLN distributions. The reason is that the MSLN distribution has a fewer number of parameters than the MSTN and MSGLN distributions. We note that the MSTN distribution includes a degrees of freedom parameter (ν) and the MSGLN distribution has a shape parameter (α). Further, we display the scatter plot of the data set with contour plots of the fitted densities obtained from MSLN, MSTN and MSGLN distributions in Figure 4 (a), (b) and (c). This figure indicates that the MSLN distribution can provide a more accurate fit to the data than the MSTN and MSGLN distributions.

5. CONCLUSIONS

In this paper, we have proposed the usage of MSLN distribution, which has fewer parameters than the MSTN distribution, as an alternative to the MSTN distribution to model skew and heavy-tailed multivariate data sets. We have given the EM algorithm to obtain the ML estimates for the parameters of the MSLN distribution. A simulation study has been provided to demonstrate the performance of the proposed distribution and it has confirmed that the proposed EM algorithm works accurately to estimate the parameters. Furthermore, we have given a real data example to show the applicability of the MSLN distribution over the MSTN and MSGLN distributions to model both skewness and heavy-tailedness in the data. We have observed that the MSLN distribution is superior to the MSTN and MSGLN distributions and can be used as an alternative distribution for modeling

Table 5. Estimation results obtained from MSLN, MSTN and MSGLN distributions

		MSTN		MSGLN		MSLN	
		Estimate	SE	Estimate	SE	Estimate	SE
Model	μ_{11}	5.0988	0.4692	5.2315	0.1783	4.7624	0.2829
	μ_{21}	5.4257	0.4694	5.6146	0.2321	5.1229	0.3270
	σ_{11}	1.1023	0.4868	0.9080	0.4127	1.3974	0.4917
	σ_{12}	1.1883	0.2692	0.9209	0.1813	1.5512	0.2644
	σ_{22}	1.9929	0.3571	1.5607	0.3328	2.2844	0.3388
	λ_{11}	0.6546	1.0051	0.4310	0.3801	1.2546	0.9977
	λ_{21}	1.2801	1.1900	0.9298	0.5014	1.7456	1.2947
	ν	6.4084	4.2291	–	–	–	–
	α	–	–	0.9963	0.3918	–	–
Information Criteria	$\ell(\widehat{\mathbf{\Theta}})$	-460.7897		-265.7059		**-255.5245**	
	BIC	956.6356		566.4681		**541.7232**	
	AIC	937.5794		547.4119		**525.0490**	
CPU times in seconds	CT	15.1563		21.3906		**7.8594**	

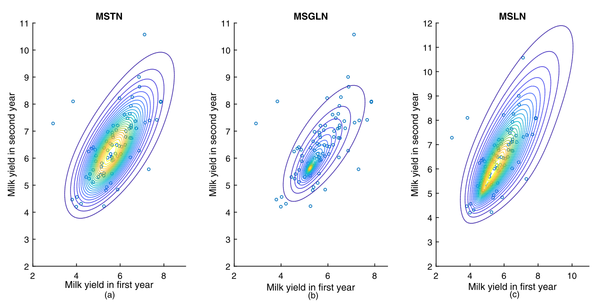

Figure 4. Scatter plot of the mastitis (milk yield in the first year, milk yield in the first year) data set along with the contour plots of the fitted MSTN (a), MSGLN (b), and MSLN (c) distributions.

this type of data set. Furthermore, for future study, similar definitions can be introduced using other forms of Azzalini type skew normal distribution mentioned in the Introduction.

ACKNOWLEDGEMENTS

We thank two anonymous referees and the Editor, whose comments and advice have greatly helped us to improve this paper.

Received 22 March 2021

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*, Proceeding of the Second International Symposium on Information Theory, B. N. Petrov and F. Caski, eds., 267–281, Akademiai Kiado, Budapest. MR0483125

[2] Arellano-Valle, R. B. and Genton, M. G. (2005). On fundamental skew distributions. *J. Multivar. Anal.* **96** 93–116. MR2202402

[3] Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scan. J. Stat.* **33**(3) 561–574. MR2298065

[4] Arslan, O. (2010). An alternative multivariate skew Laplace distribution: properties and estimation. *Stat. Pap.* **51**(4) 865–887. MR2749005

[5] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scan. J. Stat.* **32**(2) 159–188. MR2188669

[6] Azzalini, A. (2013). *The skew-normal and related families (Vol. 3)*. Cambridge University Press, New York. MR3468021

[7] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 579–602. MR1707862

[8] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika.* **83** 715–726. MR1440039

[9] Baghfalaki, T. and Ganjali, M. (2012). An ECM estimation approach for analyzing multivariate skew-normal data with dropout. *Commun.Stat. Simul. Comput.* **41**(10) 1970–1988. MR2944159

[10] Crouchley, R. and Ganjali, M. (2002). The common structure of several recent statistical models for dropout in repeated continuous responses. *Stat. Modelling.* **2** 39–62. MR1920758

[11] Cornish, E. A. (1954). The multivariate t-distribution associated with a set of normal sample deviates. *Aust. J. Phys.* **7** 531. MR0066598

[12] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B.* **39** 1–38. MR0501537

[13] Diggle, P. J. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *J. R. Stat. Soc. Ser. C Appl. Stat.* **43**(1) 49–93. MR2699422

[14] Doğru, F. Z. and Arslan, O. (2017). Parameter estimation for mixtures of skew Laplace normal distributions and application in mixture regression modeling. *Commun. Stat.-Theory and Methods.* **46**(21) 10879–10896. MR3740843

[15] Doğru, F. Z. and Arslan, O. (2019). Joint modelling of the location, scale and skewness parameters of the skew Laplace normal distribution. *Iran. J. Sci. Technol. Trans. A Sci.* **43**(3) 1249–1257. MR3977444

[16] Doğru, F. Z., Yu, K. and Arslan, O. (2019). Heteroscedastic and heavy-tailed regression with mixtures of skew Laplace normal distributions. *J. Stat. Comput. Simul.* **89**(17) 3213–3240. MR4008822

[17] Dunnett, C. W. and Sobel, M. (1954). A bivariate generalization of Student's t-distribution, with tables for certain special cases. *Biometrika.* **41**(1–2) 153–169. MR0061793

[18] Ernst, M. D. (1998). A multivariate generalized Laplace distribution. *Comput. Stat.* **13** 227–232. MR1649311

[19] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, 1st ed. Chapman and Hall, London. MR1071174

[20] Gómez, E., Gómez-Villegas, M. A. and Marín, J. M. (1998). A multivariate generalization of the power exponential family of Distributions. *Commun. Stat. Theory Methods.* **27** 589–600. MR1619030

[21] Gómez, H. W., Venegas, O. and Bolfarine, H. (2007). Skew-symmetric distributions generated by the distribution function of the normal distribution. *Environmetrics.* **18** 395–407. MR2370731

[22] Gupta, S. S. (1963). Probability integrals of multivariate normal and multivariate t. *Ann. Math. Stat.* 792–828. MR0152068

[23] Islam, M. Q. (2014). Estimation in multivariate nonnormal distributions with stochastic variance function. *J. Comput. Appl. Math.* **255** 698–714. MR3093453

[24] Kim, H. M. and Genton, M. G. (2011). Characteristic functions of scale mixtures of multivariate skew-normal distributions. *J. Multivar. Anal.* **102**(7) 1105–1117. MR2805651

[25] Lange, K. and Sinsheimer, J. S. (1993). Normal/Independent distributions and their applications in robust regression. *J. Comput. Graph. Stat.* **2** 175–198. MR1272391

[26] Lee, S. X. and McLachlan, G. J. (2013). On mixtures of skew normal and skew t-distributions. *Adv. Data Anal. Classif.* **7**(3) 241–266. MR3103965

[27] Lin, T. I. (2010). Robust mixture modeling using multivariate skew t distributions. *Stat. Comput.* **20** 343–356. MR2725392

[28] Lin, T. I., Ho, H. J. and Lee, C. R. (2014). Flexible mixture modelling using the multivariate skew-t-normal distribution. *Stat. and Comput.* **24**(4) 531–546. MR3223539

[29] Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2019). Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Stat. Comput.* **29**(3) 415–428. MR3969052

[30] Maleki, M., Wraith, D. and Arellano-Valle, R. B. (2019). A flexible class of parametric distributions for Bayesian linear mixed models. *Test* **28**(2) 543–564. MR3962066

[31] Meilijson, I. (1989). A fast improvement to the EM algorithm to its own terms. *J. R. Stat. Soc. Series B Stat. Methodol.* **51** 127–138. MR0984999

[32] Nadarajah, S. (2003). The Kotz-type distribution with application. *Stat.* **37** 341–358. MR1997185

[33] Nadarajah, S. and Kotz, S. (2005). Mathematical properties of the multivariate t distribution. *Acta Appl. Math.* **89**(1–3) 53–84. MR2220198

[34] Nadarajah, S. and Kotz, S. (2006). Percentage points of the multivariate t distribution. *Int. Stat. Rev.* **74**(1) 15–30. MR2236201

[35] Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T. I., Maier, L., BaecherAllan, 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.

[36] Sahu, S. K., Dey, D. K. and Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Can. J. Stat.* **31**(2) 129–150. MR2531833

[37] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **6**(2) 461–464. MR0468014

[38] Wang, W. L., Jamalizadeh, A. and Lin, T. I. (2020). Finite mixtures of multivariate scale-shape mixtures of skew-normal distributions. *Stat. Papers.* **61** 2643–2670. MR4167345

Fatma Zehra Doğru
Giresun University
Department of Statistics
Giresun
Turkey
E-mail address: fatma.dogru@giresun.edu.tr

Olcay Arslan
Ankara University
Department of Statistics
Ankara
Turkey
E-mail address: oarslan@ankara.edu.tr