Type II Shifted Multivariate Asymmetric Laplace Distribution based on Mixture of Normal Distribution

Yajie Duan, Guoliang Tian

Department of Mathematics,
Southern University of Science and Technology, Shenzhen, Guangdong Province, P. R. China
July 2019

Abstract. In this paper, a new shifted multivariate asymmetric Laplace distribution from normal variance mixture model is proposed and called Type II shifted multivariate asymmetric Laplace distribution. Kotz et al. (2001) proposed the original multivariate asymmetric Laplace distribution and the generalized shifted multivariate asymmetric Laplace distribution was introduced by Franczak et al. (2012). Unlike these two distributions, the random components in the new distribution are correlated only through the dependence structure among the normal random vector. Thus, it contains the multiplication of i.i.d. univariate asymmetric Laplace distributions as a special case if the normal variance matrix is diagonal. A tractable stochastic representation is useful in deriving the probability density function and other statistical properties. Various methods of parameter estimation including moments estimation, MLE and Bayesian estimation are discussed. In addition, hypothesis testing of independence among components of this new multivariate distribution are considered. Simulation studies are conducted to evaluate the performance of the proposed methods.

Keywords: Asymmetric Laplace Distribution; Stochastic Representation; MLE

1. Introduction

Laplace distribution is appropriate to model heavy-tailed data. Andrews and Mallows (1974) proposed a scale mixture of Gaussian or normal variance mixture models for univariate distribution. Because of the wide range of applications, multivariate Laplace distribution has been discussed by McGraw and Wagner (1968), Pillai (1985), Anderson (1992) and Kotz *et al.*(2000). In order to account for both the asymmetry and heavy tail of data, the univariate asymmetric Laplace (AL) distribution was introduced in Hinkley and Revankar (1977) and studied in Kozubowski and Podgrski (2000). Moreover, an extension of both the symmetric multivariate Laplace distributions and the univariate AL distributions, which is termed (centralized) multivariate asymmetric Laplace distributions, is proposed and studied in Kotz *et al.*(2001). Centralized multivariate asymmetric Laplace distributions are flexible in fitting financial data which has the features of asymmetry, steep peak at the origin, and heavier than normal tails.(See Kotz *et al.*(2001)) Franczak *et al.*(2012) introduced a shifted (non-centralized) multivariate asymmetric Laplace (SAL) distribution for the purpose of clustering and classification of data. In this paper, the density of $Y \sim \text{Gamma}(\alpha, \beta)$ is defined by $[\beta^\alpha/\Gamma(\alpha)]\, y^{\alpha-1} \exp(-\beta y)$ for $y > 0$, and $\text{Gamma}(1, \beta)$ is the exponential distribution, denoted by $\text{Exp}(\beta)$.

According to Kozubowski and Podgrski (2000), a random variable Y is said to have an asymmetric Laplace (AL) distribution with parameters $\mu \in \mathbb{R}, m \in \mathbb{R}$ and $\sigma \geq 0$ and *probability density function* (pdf)

$$f_{\mu,\kappa,\sigma}(y) = \frac{\sqrt{2}}{\sigma} \frac{\kappa}{1+\kappa^2} \begin{cases} \exp\left(-\frac{\sqrt{2}\kappa}{\sigma}|y-\mu|\right), & \text{if } y \geq \mu \\ \exp\left(-\frac{\sqrt{2}}{\sigma\kappa}|y-\mu|\right), & \text{if } y < \mu \end{cases}$$

where

$$\kappa = \frac{\sqrt{2}\sigma}{m + \sqrt{2\sigma^2 + m^2}} = \frac{\sqrt{2\sigma^2 + m^2} - m}{\sqrt{2}\sigma},$$

denoted by $Y \sim \mathcal{AL}(\mu, m, \sigma)$. Kozubowski and Podgrski showed that univariate AL random variable has the following *stochastic representation* (SR) based on the mixture of normal distributions:

$$Y \stackrel{d}{=} \mu + mZ + \sqrt{Z}X \tag{1.1}$$

where $Z \sim \text{Exp}(1), X \sim N(0, \sigma^2)$ and Z is independent of X (denoted by $Z \perp\!\!\!\perp X$). From (1.1), it is easy to obtain $E(Y) = \mu + m$ and $\text{Var}(Y) = m^2 + \sigma^2$. As an extension of the symmetric multivariate Laplace distributions and of the univariate asymmetric Laplace distributions, centralized multivariate asymmetric Laplace distribution is proposed by Kotz *et al.*(2001) with the notation $\mathbf{y} \sim \mathcal{AL}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$, where $\boldsymbol{m} \in \mathbb{R}^d$ represents the skewness in each dimension and $\boldsymbol{\Sigma} = (\sigma_{ij})$ is the scale parameter matrix. Kotz *et al.*(2001) note that the random variable $\mathbf{y} \sim \mathcal{AL}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$ can be generated through the SR:

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = Z \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix} + \sqrt{Z} \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = Z\boldsymbol{m} + \sqrt{Z}\mathbf{x} \tag{1.2}$$

where $Z \sim \text{Exp}(1)$ and $\mathbf{x} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$. From (1.2), it is easy to obtain $E(\mathbf{y}) = \boldsymbol{m}$, $\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma} + \boldsymbol{m}\boldsymbol{m}^\top$ and $Y_i \sim \mathcal{AL}\left(0, m_i, \sigma_{ii}\right), i = 1, \ldots, d$. The joint pdf of $\mathbf{y} \sim \mathcal{AL}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$ is shown in Appendix A.

Because the centralized multivariate AL distribution would force each component density to be joined at the same origin, Franczak *et al.*(2012) extended it to the shifted (non-centralized) multivariate asymmetric Laplace (SMAL) distribution. They introduce a shift parameter $\boldsymbol{\mu} \in \mathbb{R}^d$ based on the centralized multivariate AL distribution. The random vector $\mathbf{y}$ is said to follow the original or Type I shifted multivariate asymmetric Laplace distribution with the following SR:

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + Z \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix} + \sqrt{Z} \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix} = \boldsymbol{\mu} + Z\boldsymbol{m} + \sqrt{Z}\mathbf{x}, \tag{1.3}$$

denoted by $\mathbf{y} \sim \mathcal{SAL}_d^{(\mathrm{I})}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, where $Z \sim \text{Exp}(1)$ and $\mathbf{x} \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$. $\boldsymbol{\mu} \in \mathbb{R}^d$ is the shift parameter, $\boldsymbol{m} \in \mathbb{R}^d$ represents the skewness in each dimension and $\boldsymbol{\Sigma} = (\sigma_{ij})$ is the scale parameter matrix. The joint pdf of $\mathbf{y} \sim \mathcal{SAL}_d^{(\mathrm{I})}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ is shown in Appendix C.

From the SR(1.3), it is easy to obtain

$$E(\mathbf{y}) = \boldsymbol{\mu} + \boldsymbol{m}, \text{Var}(\mathbf{y}) = \boldsymbol{\Sigma} + \boldsymbol{m}\boldsymbol{m}^\top \tag{1.4}$$

And we have $Y_i = \mu_i + Zm_i + \sqrt{Z}X_i$, i.e.

$$Y_i \sim \mathcal{AL}\left(\mu_i, m_i, \sigma_{ii}\right), i = 1, \ldots, d. \tag{1.5}$$

When we consider the Type I shifted multivariate asymmetric Laplace distribution, the representation (1.3) indicates that the association between any pair of random components Y_i and $Y_j (i \neq j)$ in $\mathbf{y}$ comes from both the common factor Z and correlation of the corresponding components in $\mathbf{x}$. In other words, even X_i and X_j are independent, Y_i and Y_j are not independent. Therefore, it does not include the multiplication of d independent univariate asymmetric Laplace distributions as a special case.

Our aim is to propose a new shifted multivariate asymmetric Laplace distribution by replacing the unique Z with i.i.d. standard exponential random variables. Thus, the dependency among components of $\mathbf{y}$ only relies on the correlation structure in $\mathbf{x}$ such that it will reduce to i.i.d. univariate asymmetric Laplace distributions under some condition.

The rest of this paper is organized as follows. In Section 2, the Type II shifted multivariate asymmetric Laplace (SMAL) distribution is proposed and the joint probability density function and other distributional properties are explored. In Section 3, various methods of parameter estimation including moments estimation, MLE and Bayesian estimation are derived. In Section 4, hypothesis testing of independence among components of multivariate data are considered. In Section 5, simulations are conducted to evaluate the performance of the proposed methods. Finally, a discussion is given in Section 6.

2. Type II shifted multivariate asymmetric Laplace distribution

Let $\{Z_i\}_{i=1}^d \overset{\text{iid}}{\sim} \text{Exp}(1), \mathbf{x} = (X_1, \ldots, X_d)^\top \sim N_d(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{z} = (Z_1, \ldots, Z_d)^\top \perp\!\!\!\perp \mathbf{x}$. Define

$$\begin{aligned}\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_d \end{pmatrix} &= \boldsymbol{\mu} + \mathbf{Z}\boldsymbol{m} + \mathbf{Z}^{1/2}\mathbf{x} \\ &= \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_d \end{pmatrix} + \begin{pmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_d \end{pmatrix} \begin{pmatrix} m_1 \\ \vdots \\ m_d \end{pmatrix} + \begin{pmatrix} \sqrt{Z_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{Z_d} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}\end{aligned} \tag{2.1}$$

with $\mathbf{Z} = \text{diag}\,(Z_1, \ldots, Z_d)$, $\mathbf{Z}^{1/2} = \text{diag}\left(\sqrt{Z_1}, \ldots, \sqrt{Z_d}\right)$, then $\mathbf{y}$ is said to follow the d-dimensional Type II shifted multivariate asymmetric Laplace distribution (SMAL), denoted by $\mathbf{y} \sim \mathcal{SAL}_d^{(\text{II})}(\boldsymbol{\mu}, \boldsymbol{m}, \mathbf{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^d$ is the shift parameter, $\boldsymbol{m} \in \mathbb{R}^d$ represents the skewness in each dimension and $\mathbf{\Sigma} = (\sigma_{ij})$ is the scale parameter matrix. When $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{m} = \mathbf{0}$, $\mathbf{y}$ follows the Type II symmetric Laplace distribution with scale parameter matrix $\mathbf{\Sigma}$.

Comparing with the Type I SMAL distribution, the Type II SMAL random vector y could contain components which are independent. In fact, when X_i and $X_{i'}$ ($i, i' = 1, \ldots, d$ but $i \neq i'$) are independent (i.e., $\sigma_{ii'} = 0$), Y_i and $Y_{i'}$ are also independent because in the SR (2.1), Z_i and $Z_{i'}$ are independent. While in Type I SMAL distribution with SR (1.3), Y_i and $Y_{i'}$ are always dependent since they share a common factor Z, no matter whether X_i and $X_{i'}$ are independent or not.

2.1 Joint density function of Type II SMAL distribution

From the SR (2.1), we can obtain an equivalent mixture representation

$$Z_1, \ldots, Z_d \overset{\text{iid}}{\sim} \operatorname{Exp}(1) \quad \text{and} \quad \mathbf{y}|(\mathbf{z}=\boldsymbol{z}) \sim N_d\left(\boldsymbol{\mu}+\boldsymbol{Z}\boldsymbol{m}, \boldsymbol{Z}^{1/2}\boldsymbol{\Sigma}\boldsymbol{Z}^{1/2}\right) \tag{2.2}$$

where $\boldsymbol{z} = (z_1, \ldots, z_d)^\top$ is the realization of $\mathbf{z} = (Z_1, \ldots, Z_d)^\top$,$\boldsymbol{Z} = \operatorname{diag}(z_1, \ldots, z_d) = \operatorname{diag}(\boldsymbol{z})$ is the realization of $\mathbf{Z} = \operatorname{diag}(\mathbf{z})$, $\boldsymbol{Z}^{1/2} = \operatorname{diag}\left(\sqrt{z_1}, \ldots, \sqrt{z_d}\right) = \operatorname{diag}\left(\boldsymbol{z}^{1/2}\right)$ is the realization of $\mathbf{Z}^{1/2} = \operatorname{diag}\left(\mathbf{z}^{1/2}\right)$. The joint pdf of $\mathbf{y}$ is obtained as

$$\begin{aligned}
f_{\mathbf{y}}(\boldsymbol{y}) &= \int_{\mathbb{R}_+^d} f_{\mathbf{y}|\mathbf{z}}(\boldsymbol{y}|\boldsymbol{z}) f_{\mathbf{z}}(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \\
&= (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \int_{\mathbb{R}_+^d} z_\bullet^{-\frac{1}{2}} \exp\left[-\frac{(\boldsymbol{y}^*-\boldsymbol{Z}\boldsymbol{m})^\top \left(\boldsymbol{Z}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}^{-\frac{1}{2}}\right)(\boldsymbol{y}^*-\boldsymbol{Z}\boldsymbol{m})}{2} - z_+\right] \mathrm{d}\boldsymbol{z} \\
&= (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \int_{\mathbb{R}_+^d} z_\bullet^{-\frac{1}{2}} \exp\left\{-\frac{\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top \boldsymbol{Y}^* - \left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top \boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}} - \boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)}{2} - z_+\right\} \mathrm{d}\boldsymbol{z}
\end{aligned} \tag{2.3}$$

where $z_\bullet = \prod_{i=1}^d z_i$, $\boldsymbol{y}^* = \boldsymbol{y} - \boldsymbol{\mu} = (y_1-\mu_1, \ldots, y_d-\mu_d)^\top$, $\boldsymbol{Y}^* = \operatorname{diag}(\boldsymbol{y}-\boldsymbol{\mu}) = \operatorname{diag}(y_1-\mu_1, \ldots, y_d-\mu_d)$, $\boldsymbol{M} = \operatorname{diag}(\boldsymbol{m}) = \operatorname{diag}(m_1, \ldots, m_d)$ and $z_+ = \sum_{i=1}^d z_i$. It is clear from (2.3) that this distribution is not of the elliptical form.

2.2 Marginal distributions and moments

From the SR (2.1), we can easily have $Y_i = \mu_i + m_i Z_i + \sqrt{Z_i} X_i$, indicating that the marginal distributions still belong to the family of the univariate asymmetric Laplace distributions (AL), i.e.

$$Y_i \sim \mathcal{AL}(\mu_i, m_i, \sigma_{ii}), i = 1, \ldots, d. \tag{2.4}$$

More generally, for any positive integers $k_1, \ldots, k_r$ satisfying $1 \leqslant k_1 < \cdots < k_r \leqslant d$, we have

$$(Y_{k_1}, \ldots, Y_{k_r})^\top \sim \mathcal{SAL}_r^{(\mathrm{II})}(\boldsymbol{\mu}^*, \boldsymbol{m}^*, \boldsymbol{\Sigma}^*) \tag{2.5}$$

where $\boldsymbol{\mu}^* = (\mu_{k_1}, \ldots, \mu_{k_r})^\top$, $\boldsymbol{m}^* = (m_{k_1}, \ldots, m_{k_r})^\top$ and $\boldsymbol{\Sigma}^* = \left(\sigma_{ij}^*\right)$ is an $r \times r$ sub-matrix of $\boldsymbol{\Sigma}$ with $\sigma_{ij}^* = \sigma_{ij}$ for $i, j = k_1, \ldots, k_r$. For the moments, it is easy to derive that

$$E(\mathbf{y}) = \boldsymbol{\mu} + \boldsymbol{m} \quad \text{and} \quad \operatorname{Var}(\mathbf{y}) = \frac{\pi}{4}\boldsymbol{\Sigma} + \left(1 - \frac{\pi}{4}\right)\operatorname{diag}(\boldsymbol{\Sigma}) \tag{2.6}$$

Obviously, the variance-covariance matrix of $\mathbf{y}$ is different from that of the Type I shifted multivariate asymmetric Laplace distribution given by (1.4). The correlation coefficient between any two components Y_i and Y_j with $i \neq j$ is given by

$$\operatorname{Corr}(Y_i, Y_j) = \frac{\pi \sigma_{ij}}{4\sqrt{\sigma_{ii}\sigma_{jj}}} = \frac{\pi}{4}\operatorname{Corr}(X_i, X_j) \tag{2.7}$$

which is either positive or negative depending on σ_{ij}. In the form of matrix, we have

$$\operatorname{Corr}(\mathbf{y}) = \frac{\pi}{4}\boldsymbol{R} \tag{2.8}$$

where $\boldsymbol{R}$ is the corresponding correlation matrix associated with $\boldsymbol{\Sigma}$.

From the SR (2.1), we can find that because of the independence of Z_i and Z_j, the dependency of Y_i and Y_j only comes from the dependency of X_i and X_j. Moreover, if $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_{11}, \ldots, \sigma_{dd})$, the Type II shifted multivariate asymmetric Laplace distribution becomes the product of d independent univariate asymmetric Laplace distributions, where $Y_i \sim \mathcal{AL}(\mu_i, m_i, \sigma_{ii})$ for $i = 1, \ldots, d$.

2.3 Comparison of Type I and II SMAL densities

Given the same $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, from (1.5) and (2.4), the marginal distributions of each component from both Type I and Type II shifted multivariate asymmetric Laplace distributions are identical. However, from the joint density functions given in (C.1) and (2.3), the expressions and shapes of their joint pdfs appear to be totally different. Various shapes of bivariate densities under different settings of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ are shown in Figures 1−2. Table 1 summarizes the maximum and minimum of each scenario for both densities.

Table 1 Maximum and minimum of Type I and Type II SMAL densities

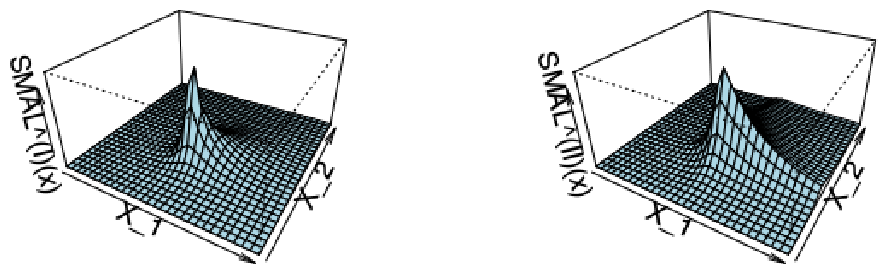

Figure 1: Comparison of type I and type II SMAL densities under $\boldsymbol{\mu} = (0, -1)^\top$, $\boldsymbol{m} = (1, 1)^\top$, $\boldsymbol{\Sigma} = diag(\mathbf{1}_2)$

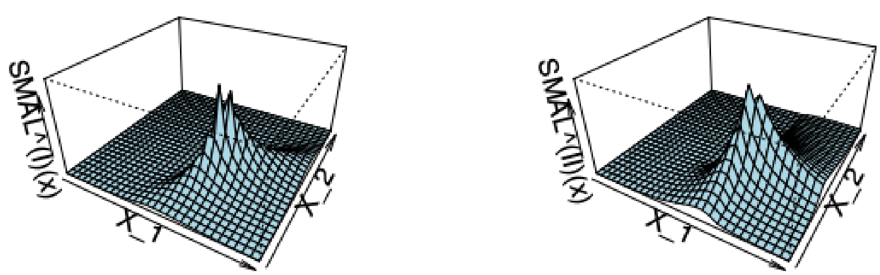

Figure 2: Comparison of type I and type II SMAL densities under $\boldsymbol{\mu} = (1, -1)^\top$, $\boldsymbol{m} = (1, 0.5)^\top$, $\boldsymbol{\Sigma} = [2, 2; 2, 3]$

	$\boldsymbol{\mu} = (0, -1)^\top$, $\boldsymbol{m} = (1, 1)^\top$ $\boldsymbol{\Sigma} = diag(\mathbf{1}_2)$		$\boldsymbol{\mu} = (1, -1)^\top$, $\boldsymbol{m} = (1, 0.5)^\top$ $\boldsymbol{\Sigma} = [2, 2; 2, 3]$	
	$f^{(\mathrm{I})}$	$f^{(\mathrm{II})}$	$f^{(\mathrm{I})}$	$f^{(\mathrm{II})}$
Maximum	0.5849137	0.2938082	0.4719195	0.2448336
Minimum	0.0000007	0.0000003	1e-10	3e-11

$f^{(\mathrm{I})}$: Density function of Type I SMAL distribution given by (C.1)

$f^{(\mathrm{II})}$: Density function of Type II SMAL distribution given by (2.3)

3 Estimation of parameters in Type II SMAL distribution

Let $\mathbf{y}_1, \ldots, \mathbf{y}_n \overset{\text{iid}}{\sim} \mathcal{SAL}_d^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ and $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{dj})^\top$ denote the realization of $\mathbf{y}_j = (Y_{1j}, \ldots, Y_{dj})^\top$ for $j = 1, \ldots, n$, then the observed data are denoted by $Y_{\text{obs}} = \{\boldsymbol{y}_j\}_{j=1}^n$. There are some different methods to make parameter estimation for Type II SMAL distribution.

3.1 Moment estimation

Although we could not separately estimate $\boldsymbol{\mu}$ and $\boldsymbol{m}$ from the expectation derived in (2.6), we can estimate $\boldsymbol{\Sigma}$ by the method of moments based on the variance−covariance specified by (2.6), resulting in the following moment estimate (ME):

$$\hat{\boldsymbol{\Sigma}}^{\mathrm{M}} = \left(\hat{\sigma}_{ii'}^{\mathrm{M}}\right), \tag{3.1}$$

where

$$\hat{\sigma}_{ii}^{\mathrm{M}} = \frac{\sum_{j=1}^n (y_{ij} - \overline{y}_i)^2}{(n-1)}, \quad i = 1, \ldots, d$$
$$\hat{\sigma}_{ii'}^{\mathrm{M}} = \frac{4\sum_{j=1}^n (y_{ij} - \overline{y}_i)(y_{i'j} - \overline{y}_{i'})}{(n-1)\pi}, \quad i, i' = 1, \ldots, d, \quad i \neq i',$$

and $\overline{y}_i = \sum_{j=1}^n y_{ij}/n$.

3.2 Maximum Likelihood Estimation of parameters

3.2.1 MLE via an ECM algorithm

Based on the mixture representation (2.2), for a fixed $j (j = 1, \ldots, n)$, we introduce latent variables $Z_{1j}, \ldots, Z_{dj} \overset{\text{iid}}{\sim} \text{Exp}(1)$ such that $\mathbf{y}_j | (\mathbf{z}_j = \boldsymbol{z}_j) \sim N_d\left(\boldsymbol{\mu} + \boldsymbol{Z}_j\boldsymbol{m}, \boldsymbol{Z}_j^{1/2}\boldsymbol{\Sigma}\boldsymbol{Z}_j^{1/2}\right)$, where $\boldsymbol{z_j} = (z_{1j}, \ldots, z_{dj})^\top$ is the realization of $\mathbf{z_j} = (Z_{1j}, \ldots, Z_{dj})^\top$, $\boldsymbol{Z}_j = \mathrm{diag}\,(z_{1j}, \ldots, z_{dj}) = \mathrm{diag}\,(\boldsymbol{z}_j)$ is the realization of $\mathbf{Z_j} = \mathrm{diag}\,(\mathbf{z}_j)$, $\boldsymbol{Z}_j^{1/2} = \mathrm{diag}\left(\sqrt{z_{1j}}, \ldots, \sqrt{z_{dj}}\right) = \mathrm{diag}\left(\boldsymbol{z}_j^{1/2}\right)$ is the realization of $\mathbf{Z}_j^{1/2} = \mathrm{diag}\left(\mathbf{z}_j^{1/2}\right)$. The missing data are denoted by $Y_{\text{mis}} = \{\boldsymbol{z}_j\}_{j=1}^n$ and the complete data are $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\}$. Then the complete-data likelihood function is

$$\begin{aligned}
& L\left(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma} | Y_{\text{com}}\right) \\
= \quad & \prod_{j=1}^n \left\{ (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}} \exp\left[-\frac{\left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{-\frac{1}{2}} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right)}{2} - \sum_{i=1}^d z_{ij} \right] \right\} \\
\propto \quad & |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} \sum_{j=1}^n \left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{-\frac{1}{2}} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right) \right]
\end{aligned} \tag{3.2}$$

so that the log-likelihood function is

$$\ell\left(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma} | Y_{\text{com}}\right) = c - \frac{n}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} \operatorname{tr}\left\{ \boldsymbol{\Sigma}^{-1} \sum_{j=1}^n \left[\boldsymbol{Z}_j^{-\frac{1}{2}} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right)\right] \left[\boldsymbol{Z}_j^{-\frac{1}{2}} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j \boldsymbol{m}\right)\right]^\top \right\} \tag{3.3}$$

We employ the expectation/conditional maximization (ECM) algorithm of Meng & Rubin (1993). The three CM-steps are to obtain the following conditional maximizers

$$\hat{\boldsymbol{\mu}} = \left(\sum_{j=1}^n \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{-\frac{1}{2}}\right)^{-1} \left(\sum_{j=1}^n \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{y}_j - \sum_{j=1}^n \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{\frac{1}{2}} \boldsymbol{m}\right), \tag{3.4}$$

$$\hat{\boldsymbol{m}} = \left(\sum_{j=1}^n \boldsymbol{Z}_j^{\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{\frac{1}{2}}\right)^{-1} \left(\sum_{j=1}^n \boldsymbol{Z}_j^{\frac{1}{2}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{y}_j^*\right), \tag{3.5}$$

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}} = \quad & \frac{1}{n} \sum_{j=1}^n \boldsymbol{Z}_j^{-\frac{1}{2}} \left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j \boldsymbol{m}\right) \left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j \boldsymbol{m}\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}} \\
= \quad & \frac{1}{n} \sum_{j=1}^n \left[\boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{y}_j^* \left(\boldsymbol{y}_j^*\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}} - \boldsymbol{Z}_j^{-\frac{1}{2}} \boldsymbol{y}_j^* \boldsymbol{m}^\top \boldsymbol{Z}_j^{\frac{1}{2}} - \boldsymbol{Z}_j^{\frac{1}{2}} \boldsymbol{m} \left(\boldsymbol{y}_j^*\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}} + \boldsymbol{Z}_j^{\frac{1}{2}} \boldsymbol{m} \boldsymbol{m}^\top \boldsymbol{Z}_j^{\frac{1}{2}} \right],
\end{aligned} \tag{3.6}$$

where $\boldsymbol{y}_j^* = \boldsymbol{y}_j - \boldsymbol{\mu} = (y_{1j} - \mu_1, \ldots, y_{dj} - \mu_d)^\top$.

The E-step is to replace $z_{ij}^{-\frac{1}{2}} z_{kj}^{-\frac{1}{2}}$, $z_{ij}^{-\frac{1}{2}} z_{kj}^{\frac{1}{2}}$, $z_{ij}^{\frac{1}{2}} z_{kj}^{\frac{1}{2}}$ in (3.4), (3.5) and (3.6) by their conditional expectations $E\left(Z_{ij}^{-\frac{1}{2}} Z_{kj}^{-\frac{1}{2}} | Y_{\text{obs}}, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$, $E\left(Z_{ij}^{-\frac{1}{2}} Z_{kj}^{\frac{1}{2}} | Y_{\text{obs}}, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$, $E\left(Z_{ij}^{\frac{1}{2}} Z_{kj}^{\frac{1}{2}} | Y_{\text{obs}}, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$ for $i, k = 1, \ldots, d$ and $j = 1, \ldots, n$.

From (2.3), we can derive the conditional distribution of $\mathbf{z}_j|\mathbf{y}_j$ given by

$$
\begin{aligned}
f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{z}_j|\boldsymbol{y}_j\right) = &\ \frac{1}{c_j}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right)^\top \left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right) - \sum_{i=1}^{d} z_{ij}\right] \\
= &\ \frac{1}{c_j}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top \boldsymbol{Y}^* - \left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top \boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}} - \boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)\right. \\
&\ \left. - \sum_{i=1}^{d} z_{ij}\right\}
\end{aligned}
\tag{3.7}
$$

where $\boldsymbol{y}^* = \boldsymbol{y} - \boldsymbol{\mu}$, $\boldsymbol{Y}^* = \mathrm{diag}(\boldsymbol{y} - \boldsymbol{\mu})$, $\boldsymbol{M} = \mathrm{diag}(\boldsymbol{m})$ and c_j is the normalizing constant. The values of the c_j's are calculated via

$$
\begin{aligned}
c_j = &\ \int_{\mathbb{R}_+^d}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right)^\top \left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right) - \sum_{i=1}^{d} z_{ij}\right] \mathrm{d}\boldsymbol{z}_j \\
= &\ \int_{\mathbb{R}_+^d}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top \boldsymbol{Y}^* - \left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top \boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}} - \boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)\right. \\
&\ \left. - \sum_{i=1}^{d} z_{ij}\right\} \mathrm{d}\boldsymbol{z}_j
\end{aligned}
\tag{3.8}
$$

for $j = 1, \ldots, n$, and the corresponding conditional expectations are given by

$$
\begin{aligned}
&E\left(Z_{ij}^{-\frac{1}{2}} Z_{kj}^{-\frac{1}{2}}|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) \\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{-\frac{1}{2}} z_{kj}^{-\frac{1}{2}}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right)^\top \left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right) - \sum_{i=1}^{d} z_{ij}\right] \mathrm{d}\boldsymbol{z}_j \\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{-\frac{1}{2}} z_{kj}^{-\frac{1}{2}}\left(\prod_{i=1}^{d} z_{ij}\right)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top \boldsymbol{Y}^* - \left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top \boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}} - \boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)\right. \\
&\left. - \sum_{i=1}^{d} z_{ij}\right\} \mathrm{d}\boldsymbol{z}_j
\end{aligned}
\tag{3.9}
$$

$$
\begin{aligned}
&E\left(Z_{ij}^{-\frac{1}{2}}Z_{kj}^{\frac{1}{2}}|\boldsymbol{y}_j,\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma}\right)\\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{-\frac{1}{2}}z_{kj}^{\frac{1}{2}}\left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)^\top\left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)-\sum_{i=1}^d z_{ij}\right]\mathrm{d}\boldsymbol{z}_j\\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{-\frac{1}{2}}z_{kj}^{\frac{1}{2}}\left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top\boldsymbol{Y}^*-\left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top\boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}}-\boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)\right.\\
&\left.-\sum_{i=1}^d z_{ij}\right\}\mathrm{d}\boldsymbol{z}_j
\end{aligned}
\tag{3.10}
$$

$$
\begin{aligned}
&E\left(Z_{ij}^{\frac{1}{2}}Z_{kj}^{\frac{1}{2}}|\boldsymbol{y}_j,\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma}\right)\\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{\frac{1}{2}}z_{kj}^{\frac{1}{2}}\left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)^\top\left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)-\sum_{i=1}^d z_{ij}\right]\mathrm{d}\boldsymbol{z}_j\\
=&\frac{1}{c_j}\int_{\mathbb{R}_+^d} z_{ij}^{\frac{1}{2}}z_{kj}^{\frac{1}{2}}\left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left[\left(\boldsymbol{z}^{-\frac{1}{2}}\right)^\top\boldsymbol{Y}^*-\left(\boldsymbol{z}^{\frac{1}{2}}\right)^\top\boldsymbol{M}\right]\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{Y}^*\boldsymbol{z}^{-\frac{1}{2}}-\boldsymbol{M}\boldsymbol{z}^{\frac{1}{2}}\right)\right.\\
&\left.-\sum_{i=1}^d z_{ij}\right\}\mathrm{d}\boldsymbol{z}_j
\end{aligned}
\tag{3.11}
$$

for $i,k=1,\ldots,d$ and $j=1,\ldots,n$.

From (3.7), we can find that the closed-form expression of the conditional density of $\mathbf{z}_j|\mathbf{y}_j$ is not available. Thus, we can numerically calculate the values of the c_j's in (3.8) and the conditional expectations in (3.9), (3.10) and (3.11) by using the build-in function **adaptIntegrate** in the **R** package **cubature**, which evaluates multivariate integration over hypercubes (Johnson & Narasimhan, 2013).

Thus, the ECM algorithm for calculating the *maximum likelihood estimates* (MLEs) of $(\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma})$ is to combine the three CM-steps (3.4), (3.5) and (3.6) with the E-step (3.9), (3.10) and (3.11). Let $\boldsymbol{\theta}^{(t+1)}$ be the $(t+1)$ -th approximation of $\hat{\boldsymbol{\theta}}$, then the stopping rule of the ECM algorithm is set as $\left\|\boldsymbol{\theta}^{(t+1)}-\boldsymbol{\theta}^{(t)}\right\|\leqslant\delta$, where δ is a predetermined precision.

3.2.2 Bootstrap confidence intervals of parameters based on MLE

Based on the MLE of parameters, we can calculate the bootstrap *confidence intervals* (CIs) of parameters for small to moderate sample sizes by the bootstrap approach. For an arbitrary function of h, say $\vartheta = h(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, let $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}})$ denote the MLEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ calculated by the ECM algorithm specified in (3.4)-(3.6) and (3.9)-(3.11), then $\hat{\vartheta} = h(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}})$ is the MLE of ϑ. Based on the obtained MLEs $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}})$, we can generate $\mathbf{y}_1^*, \ldots, \mathbf{y}_n^* \overset{\text{iid}}{\sim} \mathrm{SAL}_d^{(\mathrm{II})}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}})$ via the SR (2.1). Based on the bootstrap sample $Y_{\text{obs}}^* = \{\boldsymbol{y}_j^*\}_{j=1}^n$, we first compute the MLEs $\left(\hat{\boldsymbol{\mu}}^*, \hat{\boldsymbol{m}}^*, \hat{\boldsymbol{\Sigma}}^*\right)$ and then obtain a bootstrap replication $\hat{\vartheta}^* = h\left(\hat{\boldsymbol{\mu}}^*, \hat{\boldsymbol{m}}^*, \hat{\boldsymbol{\Sigma}}^*\right)$. Independently repeating the above process G times, we obtain G bootstrap samples $\{Y_{\text{obs}}^*(g)\}_{g=1}^G$ and G bootstrap replications $\left\{\hat{\vartheta}_g^*\right\}_{g=1}^G$. The standard error of $\hat{\vartheta}, \mathrm{se}(\hat{\vartheta})$, can be estimated by the sample standard deviation of the G replications, i.e.,

$$\widehat{\mathrm{se}}(\hat{\vartheta}) = \left\{\frac{1}{G-1}\sum_{g=1}^{G}\left[\hat{\vartheta}_g^* - \left(\hat{\vartheta}_1^* + \cdots + \hat{\vartheta}_G^*\right)/G\right]^2\right\}^{1/2}. \tag{3.12}$$

If $\left\{\hat{\vartheta}_g^*\right\}_{g=1}^G$ is approximately normally distributed, the $(1-\alpha)100\%$ simple bootstrap CI for ϑ is

$$\left[\hat{\vartheta} - z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat{\vartheta}), \hat{\vartheta} + z_{\alpha/2}\cdot\widehat{\mathrm{se}}(\hat{\vartheta})\right]. \tag{3.13}$$

Alternatively, if $\left\{\hat{\vartheta}_g^*\right\}_{g=1}^G$ is non-normally distributed, the $(1-\alpha)100\%$ bootstrap percentile CI of ϑ is given by

$$\left[\hat{\vartheta}_{\mathrm{L}}, \hat{\vartheta}_{\mathrm{U}}\right], \tag{3.14}$$

where $\hat{\vartheta}_{\mathrm{L}}$ and $\hat{\vartheta}_{\mathrm{U}}$ are the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of $\left\{\hat{\vartheta}_g^*\right\}_{g=1}^G$, respectively.

3.3 Bayesian maximum a posteriori (MAP) estimation of parameters

Based on the Bayesian methods, we can estimate parameters in the Type II SMAL distribution by the *Maximum a posteriori estimation* (MAPs), specifically by the posterior modes which are obtained via an ECM algorithm. Two kinds of prior distributions are considered including the non-informative prior and the conjugate prior.

3.3.1 The non-informative prior

If little information is known about the parameters before experiments, then we could consider the following non-informative (or diffuse) prior distribution, assuming the independence of parameter,

$$\pi(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{d+1}{2}} \tag{3.15}$$

Especially, when $d = 1$, (3.15) becomes $1/\sigma^2$. By introducing the same latent variables Z_{ij} 's as in the MLE of parameters, we combine (3.2) with (3.15) and obtain the following complete-data posterior distribution

$$p\left(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma} | Y_{\mathrm{com}}\right) \propto |\boldsymbol{\Sigma}|^{-\frac{n+d+1}{2}} \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j\boldsymbol{m}\right)^{\top} \boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j\boldsymbol{m}\right)\right] \tag{3.16}$$

The three CM-steps are to calculate the complete-data posterior modes as

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{A}_1^{-1}\boldsymbol{b}_1, \quad \tilde{\boldsymbol{m}} = \boldsymbol{P}_1^{-1}\boldsymbol{q}_1, \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} = \frac{\boldsymbol{C}_1}{n+d+1} \tag{3.17}$$

where

$$\begin{aligned}
\boldsymbol{A}_1 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}, \quad \boldsymbol{b}_1 = \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{y}_j - \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{m} \\
\boldsymbol{P}_1 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{\frac{1}{2}}, \quad \boldsymbol{q}_1 = \sum_{j=1}^{n}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}(\boldsymbol{y}_j - \boldsymbol{\mu}) \\
\boldsymbol{C}_1 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j\boldsymbol{m}\right)\left(\boldsymbol{y}_j - \boldsymbol{\mu} - \boldsymbol{Z}_j\boldsymbol{m}\right)^{\top}\boldsymbol{Z}_j^{-\frac{1}{2}}.
\end{aligned}$$

The E-step is to replace $z_{ij}^{-\frac{1}{2}}z_{kj}^{-\frac{1}{2}}$, $z_{ij}^{-\frac{1}{2}}z_{kj}^{\frac{1}{2}}$, $z_{ij}^{\frac{1}{2}}z_{kj}^{\frac{1}{2}}$ in (3.17) by their conditional expectations in (3.9), (3.10) and (3.11) for $i, k = 1, \ldots, d$ and $j = 1, \ldots, n$.

To implement the DA algorithm to be introduced later, we need to derive the conditional posterior distributions of $\boldsymbol{\mu}|\left(Y_{\mathrm{com}}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$, $\boldsymbol{m}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ and $\boldsymbol{\Sigma}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{m}\right)$ when choosing the non-informative prior, which are given by

$$\begin{aligned}
\boldsymbol{m}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) &\sim N_d\left(\boldsymbol{P}_1^{-1}\boldsymbol{q}_1, \boldsymbol{P}_1^{-1}\right), \\
\boldsymbol{\mu}|\left(Y_{\mathrm{com}}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) &\sim N_d\left(\boldsymbol{A}_1^{-1}\boldsymbol{b}_1, \boldsymbol{A}_1^{-1}\right), \\
\text{and} \quad \boldsymbol{\Sigma}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{m}\right) &\sim \mathrm{IWishart}_d\left(\boldsymbol{C}_1, n\right),
\end{aligned} \tag{3.18}$$

respectively.

3.3.2 The conjugate prior

We consider the following conjugate prior

$$\boldsymbol{m}|\boldsymbol{\mu},\boldsymbol{\Sigma} \sim N_d\left(\boldsymbol{m}_0+\boldsymbol{\mu},\boldsymbol{\Sigma}/\xi_0\right), \quad \boldsymbol{\mu}|\boldsymbol{\Sigma} \sim N_d\left(\boldsymbol{\mu}_0,\boldsymbol{\Sigma}/\kappa_0\right) \quad \text{and} \quad \boldsymbol{\Sigma} \sim \text{ IWishart }_d\left(\boldsymbol{\Lambda}_0,\nu_0\right) \tag{3.19}$$

where $\nu_0 > d-1$. Then the complete-data posterior distribution is given by

$$\begin{aligned}
p\left(\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma}|Y_{\text{com}}\right) &\propto L\left(\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma}|Y_{\text{com}}\right)\cdot\pi(\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma})\\
&\propto |\boldsymbol{\Sigma}|^{-\frac{n+\nu_0+d+3}{2}}\exp\left[-\frac{1}{2}\sum_{j=1}^{n}\left(\boldsymbol{y}_j-\boldsymbol{\mu}-\boldsymbol{Z}_j\boldsymbol{m}\right)^\top \boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\left(\boldsymbol{y}_j-\boldsymbol{\mu}-\boldsymbol{Z}_j\boldsymbol{m}\right)\right.\\
&\quad -\frac{1}{2}\xi_0\left(\boldsymbol{m}-\boldsymbol{m}_0-\boldsymbol{\mu}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{m}-\boldsymbol{m}_0-\boldsymbol{\mu}\right)-\frac{1}{2}\kappa_0\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)\\
&\quad \left.-\frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Lambda}_0\boldsymbol{\Sigma}^{-1}\right)\right]
\end{aligned} \tag{3.20}$$

The three CM-steps are to calculate the complete-data posterior modes as

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{A}_2^{-1}\boldsymbol{b}_2, \quad \tilde{\boldsymbol{m}} = \boldsymbol{P}_2^{-1}\boldsymbol{q}_2, \quad \text{and} \quad \tilde{\boldsymbol{\Sigma}} = \frac{\boldsymbol{C}_2}{n+\nu_0+d+3} \tag{3.21}$$

where

$$\begin{aligned}
\boldsymbol{A}_2 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}+\kappa_0\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}^{-1},\\
\boldsymbol{b}_2 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{y}_j-\sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{m}+\kappa_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu_0}+\boldsymbol{\Sigma}^{-1}(\boldsymbol{m}-\boldsymbol{m_0})\\
\boldsymbol{P}_2 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{\frac{1}{2}}+\xi_0\boldsymbol{\Sigma}^{-1},\\
\boldsymbol{q}_2 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}(\boldsymbol{y}_j-\boldsymbol{\mu})+\xi_0\boldsymbol{\Sigma}^{-1}(\boldsymbol{m_0}+\boldsymbol{\mu})\\
\boldsymbol{C}_2 &= \sum_{j=1}^{n}\boldsymbol{Z}_j^{-\frac{1}{2}}\left(\boldsymbol{y}_j-\boldsymbol{\mu}-\boldsymbol{Z}_j\boldsymbol{m}\right)\left(\boldsymbol{y}_j-\boldsymbol{\mu}-\boldsymbol{Z}_j\boldsymbol{m}\right)^\top\boldsymbol{Z}_j^{-\frac{1}{2}}+\xi_0\left(\boldsymbol{m}-\boldsymbol{m}_0-\boldsymbol{\mu}\right)\left(\boldsymbol{m}-\boldsymbol{m}_0-\boldsymbol{\mu}\right)^\top\\
&\quad +\kappa_0\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0\right)^\top+\boldsymbol{\Lambda}_0
\end{aligned}$$

The E-step is to replace $z_{ij}^{-\frac{1}{2}}z_{kj}^{-\frac{1}{2}}$, $z_{ij}^{-\frac{1}{2}}z_{kj}^{\frac{1}{2}}$, $z_{ij}^{\frac{1}{2}}z_{kj}^{\frac{1}{2}}$ in (3.21) by their conditional expectations in (3.9), (3.10) and (3.11) for $i,k=1,\ldots,d$ and $j=1,\ldots,n$.

To implement the DA algorithm to be introduced later, we need to derive the conditional posterior distributions of $\boldsymbol{\mu}|\left(Y_{\mathrm{com}}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$, $\boldsymbol{m}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ and $\boldsymbol{\Sigma}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{m}\right)$ when choosing the conjugate prior, which are given by

$$\begin{aligned}
& \boldsymbol{m}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \sim N_d\left(\boldsymbol{P}_2^{-1}\boldsymbol{q}_2, \boldsymbol{P}_2^{-1}\right), \\
& \boldsymbol{\mu}|\left(Y_{\mathrm{com}}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) \sim N_d\left(\boldsymbol{A}_2^{-1}\boldsymbol{b}_2, \boldsymbol{A}_2^{-1}\right), \\
\text{and} \quad & \boldsymbol{\Sigma}|\left(Y_{\mathrm{com}}, \boldsymbol{\mu}, \boldsymbol{m}\right) \sim \mathrm{IWishart}_d\left(\boldsymbol{C}_2, n+\nu_0+2\right),
\end{aligned} \tag{3.22}$$

respectively.

3.4 Bayesian expected a posteriori (EAP) estimation of parameters

In addition to estimating parameters by posterior modes, we can also obtain posterior means and their Bayesian credible intervals as parameter estimates by generation of posterior samples. Here we use the data augmentation (DA) algorithm (Tanner and Wong, 1987) to generate the posterior samples form the observed posterior distribution $p\left(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}|Y_{\mathrm{com}}\right)$.

3.4.1 Generation of posterior samples via a DA algorithm embedded with an AR algorithm

The DA algorithm includes I-step and P-step. The I-step is to draw the missing values from the conditional distribution and the P-step is to draw the parameters given complete-data and generate posterior samples.

For the I-step, we need draw the missing values $\{\boldsymbol{z}_j\}_{j=1}^n$ of $\{\mathbf{z}_j\}_{j=1}^n$ from the conditional distribution $f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{z}_j|\boldsymbol{y}_j\right)$ in (3.7), which is given by

$$f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{z}_j|\boldsymbol{y}_j\right) \propto \left(\prod_{i=1}^d z_{ij}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right)^\top \left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^* - \boldsymbol{Z}_j\boldsymbol{m}\right) - \sum_{i=1}^d z_{ij}\right] \tag{3.23}$$

Because no explicit form is available, we adopt the acceptance–rejection (AR) algorithm (von Neumann, 1951) to implement this I-step. For each un-normalizing target density function

$f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{z}_j|\boldsymbol{y}_j\right)$, we choose the following envelope density

$$g_{\mathbf{z}_j}\left(\boldsymbol{z}_j\right)=\prod_{i=1}^{d}\operatorname{Gamma}\left(z_{ij}|0.5,1\right)$$

where $\operatorname{Gamma}(z|\alpha,\beta)$ denotes the density of the Gamma (α,β) distribution. Since $\boldsymbol{\Sigma}$ is a nonnegative definite matrix as the covariance matrix of the multivariate normal distribution, then $\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}$ is also nonnegative definite. Thus,
$\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)^\top\left(\boldsymbol{Z}_j^{-\frac{1}{2}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_j^{-\frac{1}{2}}\right)\left(\boldsymbol{y}_j^*-\boldsymbol{Z}_j\boldsymbol{m}\right)\geqslant 0$ holds. Then we have

$$f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{z}_j|\boldsymbol{y}_j\right)\leqslant\prod_{i=1}^{d}z_{ij}^{-\frac{1}{2}}\mathrm{e}^{-z_{ij}}=(\sqrt{\pi})^d\cdot g\left(\boldsymbol{z}_j\right)$$

The envelope constant $c=(\sqrt{\pi})^d>1$. The procedure for generating $\{\mathbf{z}_j=\boldsymbol{z}_j\}_{j=1}^n$ via the AR algorithm is as follows:

Step 1: Draw $V_j=v_j\sim U(0,1)$ and independently generate $U_{ij}=u_{ij}\overset{\text{iid}}{\sim}\operatorname{Gamma}(0.5,1)$ for $i=1,\ldots,d$. Let $\mathbf{u}_j=\left(U_{1j},\ldots,U_{dj}\right)^\top$ and $\boldsymbol{u}_j=\left(u_{1j},\ldots,u_{dj}\right)^\top$.

Step 2: If $v_j\leqslant f_{\mathbf{z}_j|\mathbf{y}_j}\left(\boldsymbol{u}_j|\boldsymbol{y}_j\right)/\left[(\sqrt{\pi})^d\prod_{i=1}^d\operatorname{Gamma}\left(u_{ij}|0.5,1\right)\right]$, set $\mathbf{z}_j=\boldsymbol{u}_j$; otherwise, go back to Step 1.

The above procedure is conducted from $j=1$ to n and then we can obtain $\{\mathbf{z}_j=\boldsymbol{z}_j\}_{j=1}^n$. Next, for the P-step, given $(Y_{\text{obs}},\boldsymbol{z}_1,\ldots,\boldsymbol{z}_n)$, we can draw $(\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma})$ via the Gibbs sampling method. We can draw $\boldsymbol{\Sigma}$ for given $(Y_{\text{obs}},\boldsymbol{\mu},\boldsymbol{m})$, update $\boldsymbol{\mu}$ for given $(Y_{\text{obs}},\boldsymbol{m},\boldsymbol{\Sigma})$ and then update $\boldsymbol{m}$ for given $(Y_{\text{obs}},\boldsymbol{\mu},\boldsymbol{\Sigma})$ from their conditional posterior distributions, $\boldsymbol{\mu}|\left(Y_{\text{com}},\boldsymbol{m},\boldsymbol{\Sigma}\right)$, $\boldsymbol{m}|\left(Y_{\text{com}},\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ and $\boldsymbol{\Sigma}|\left(Y_{\text{com}},\boldsymbol{\mu},\boldsymbol{m}\right)$, shown in (3.18) when choosing the non-informative prior or (3.22) when choosing the conjugate prior distributions. Repeat L times in the procedure of Gibbs sampling, we can obtain L posterior samples for $\boldsymbol{\mu},\boldsymbol{m},\boldsymbol{\Sigma}$ respectively, denoted by $\{\boldsymbol{\mu}_l^*\}_{l=1}^L$, $\{\boldsymbol{m}_l^*\}_{l=1}^L$ and $\{\boldsymbol{\Sigma}_l^*\}_{l=1}^L$.

3.4.2 Posterior means and Bayesian credible intervals

After discarding the first half of the posterior samples generated by DA algorithm, we can calculate the posterior means as Bayesian expected a posteriori (EAP) estimates of parameters, the posterior standard deviations and the Bayesian credible intervals of the parameters.

The Bayesian expected a posteriori (EAP) estimates of parameters are

$$\widetilde{\boldsymbol{\mu}} = \frac{\sum_{l=L/2+1}^{L} \boldsymbol{\mu}_l^*}{L/2}, \quad \widetilde{\boldsymbol{m}} = \frac{\sum_{l=L/2+1}^{L} \boldsymbol{m}_l^*}{L/2}, \quad \text{and} \quad \widetilde{\boldsymbol{\Sigma}} = \frac{\sum_{l=L/2+1}^{L} \boldsymbol{\Sigma}_l^*}{L/2}. \tag{3.24}$$

For an arbitaray function of $\boldsymbol{\mu}, \boldsymbol{m}$ and $\boldsymbol{\Sigma}$, say $\vartheta = h(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, let $(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\Sigma}})$ denote the EAP estimates of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ calculated by (3.24), then $\widetilde{\vartheta} = h(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{m}}, \widetilde{\boldsymbol{\Sigma}})$ is the Bayesian expected a posteriori (EAP) estimates of ϑ. And $\{\vartheta_l^*\}_{l=L/2+1}^{L} = \{h(\boldsymbol{\mu}_l^*, \boldsymbol{m}_l^*, \boldsymbol{\Sigma}_l^*)\}_{l=L/2+1}^{L}$ is the posterior samples, from which we can obtain the $(1-\alpha)100\%$ Bayesian credible interval of ϑ is given by

$$\left[\widetilde{\vartheta}_{\mathrm{L}}, \widetilde{\vartheta}_{\mathrm{U}}\right], \tag{3.25}$$

where $\widetilde{\vartheta}_{\mathrm{L}}$ and $\widetilde{\vartheta}_{\mathrm{U}}$ are the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of $\{\vartheta_l^*\}_{l=L/2+1}^{L}$, respectively.

4 Testing hypothesis of independence among components

In Type I SMAL distribution with SR (1.3), Y_i and $Y_{i'}$ are always dependent since they share a common factor Z. While the Type II SMAL random vector y could contain components which are independent. From the SR (2.1) we know that $\{Z_i\}_{i=1}^{d}$ are mutually independent. If all $\sigma_{ij} = 0$ for $i, j = 1, \ldots, d$ and $i \neq j$, then all components $\{Y_i\}_{i=1}^{d}$ are independent. Suppose that we want to test the independence among components, i.e. the following null hypothesis

$$H_0 : \sigma_{ij} = 0 \text{ for } i, j = 1, \ldots, d, i \neq j \quad \text{against} \quad H_1 : H_0 \tag{4.1}$$

Under H_0, the *likelihood ratio test* (LRT) statistic is given by

$$T = -2\left\{\ell\left(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{m}}_0, \hat{\boldsymbol{\Sigma}}_0 | Y_{\mathrm{obs}}\right) - \ell\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}} | Y_{\mathrm{obs}}\right)\right\} \sim \chi^2(d(d-1)/2) \tag{4.2}$$

where $\left(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{m}}_0, \hat{\boldsymbol{\Sigma}}_0\right)$ are the constrained MLEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ under H_0, and $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{\Sigma}}\right)$ are the unconstrained MLEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$. The constrained MLEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ are the combined MLEs of $(\mu_i, m_i, \sigma_{ii})$ for each marginal component, which follows the univariate asymmetric Laplace distributions (AL), i.e. $Y_i \sim \mathrm{AL}(\mu_i, m_i, \sigma_{ii}), i = 1, \ldots, d$.

Under H_0, the test statistic T approximately follows a chi-squared distribution with $d(d-1)/2$ degrees of freedom, and the corresponding p -value is given by

$$p = \begin{cases} \Pr\left(T > t|H_0\right), & \text{if } d \leqslant 2 \\ 2\min\left\{\Pr\left(T > t|H_0\right), \Pr\left(T \leqslant t|H_0\right)\right\}, & \text{if } d \geqslant 3 \end{cases} \tag{4.3}$$

For a given significant level α, the null hypothesis should be rejected if $p \leqslant \alpha$

5 Simulation Studies

We conduct some simulation studies to evaluate the performance of the proposed parameter estimating methods including the methods of moment estimation (ME), maximum likelihood estimation (MLE) and Bayesian maximum a posteriori estimation (MAPE) adopting the conjugate prior distributions of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$. We set the sample size $n = 20, 50$ for $d = 2$ and $n = 20, 30$ for $d = 3$. We investigate the accuracy of point estimators of the parameters. The parameter configurations are summarized in Table 2.

For a given combination of $(n, d, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, we generate $\left\{\mathbf{y}_j = \boldsymbol{y}_j\right\}_{j=1}^n \overset{\text{iid}}{\sim} \mathcal{SAL}_d^{(\mathrm{II})}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ via the SR (2.1). Based on the generated samples $\left\{\boldsymbol{y}_j\right\}_{j=1}^n$, we can obtain MEs, MLEs and MAPEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ by (2.6), the ECM algorithm for MLE (3.4)-(3.6) and (3.9)-(3.11), and the ECM algorithm for Bayesian MAP estimation (3.21) and (3.9)-(3.11) , respectively. By repeating the process for 50 or 100 times, we obtain the average MEs, MLEs, MAPEs and the corresponding mean squared errors (MSEs) of parameters. Tables 3-4 summarize the simulation results for d = 2,3, respectively.

It is seen that the MSEs obtained by the method of moment estimation is always larger than those obtained by the method of maximum likelihood estimation and Bayesian MAP estimation. Moreover, the MSEs and the number of iteration of algorithm become smaller as the sample size increases. And overall, the method of MLE performs better than Bayesian MAP estimation and moment method. And those methods perform better as the sample size increases.

Table 2 Parameter configurations

Parameter	d=2	d=3
$\boldsymbol{\mu}$	$(0,0)^\top$	$(-1,0,1)^\top$
$\boldsymbol{m}$	$(0.1,0.25)^\top$	$(0.25,0.25,0.5)^\top$
$\boldsymbol{\Sigma}$	$\begin{pmatrix} 0.2 & 0.2 \\ 0.2 & 0.3 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 & -0.5 \\ 0.5 & 1 & -0.5 \\ -0.5 & -0.5 & 1 \end{pmatrix}$

Table 3 Comparisons between MLEs, Bayesian MAPEs and MEs of parameters in Type II bivariate SAL distribution

Sample Size	Method	Parameter							Iteration No.
		μ_1	μ_2	m_1	m_1	σ_{11}	σ_{12}	σ_{22}	
		0	0	0.1	0.25	0.2	0.2	0.3	
$n=20$	MLE	-0.0021	0.0017	0.0903	0.2357	0.1858	0.1889	0.2841	7.14
		(0.0012)	(0.0021)	(0.0058)	(0.0110)	(0.0064)	(0.0075)	(0.0150)	
	MAPE	-0.0016	0.0157	0.1153	0.2309	0.1770	0.1146	0.2313	6.50
		(0.0016)	(0.0033)	(0.0093)	(0.0113)	(0.0036)	(0.0099)	(0.0107)	
	ME					0.1954	0.2014	0.3594	
						(0.0084)	(0.0093)	(0.0406)	
$n=50$	MLE	0.0001	0.0006	0.0923	0.2399	0.1980	0.1992	0.2983	5.05
		(0.0002)	(0.0003)	(0.0028)	(0.0054)	(0.0027)	(0.0027)	(0.0052)	
	MAPE	0.0012	0.0116	0.1009	0.2387	0.1946	0.1622	0.2729	5.15
		(0.0002)	(0.0007)	(0.0027)	(0.0046)	(0.0021)	(0.0035)	(0.0047)	
	ME					0.1988	0.1840	0.3474	
						(0.0034)	(0.0040)	(0.0176)	

* MLE is the average of 100 MLEs of the parameters via the ECM algorithm with precision 0.005; MAPE is the average of 100 MAPEs of the parameters by the Bayesian posterior modes via the ECM algorithm with precision 0.005; ME is the average of 100 MEs of the parameters; MSEs are given in the parentheses; Iteration No. is the average iterative times for the ECM algorithm.

Table 4 Comparisons between MLEs, Bayesian MAPEs and MEs of parameters in Type II trivariate SAL distribution

Sample Size	Method	Parameter						Iteration No.
		μ_1	μ_2	μ_3	m_1	m_1	m_3	
		-1	0	1	0.25	0.25	0.5	
	MLE	-1.0166	-0.0412	0.9971	0.2517	0.3280	0.4656	6.21
		(0.0060)	(0.0080)	(0.0047)	(0.0638)	(0.0659)	(0.0527)	
	MAPE	-0.9712	-0.0510	1.1031	0.2368	0.2012	0.5694	6.03
		(0.0070)	(0.0071)	(0.0045)	(0.0540)	(0.0701)	(0.0624)	
	ME							
		σ_{11}	σ_{12}	σ_{13}	σ_{22}	σ_{23}	σ_{33}	
$n = 20$		1	0.5	-0.5	1	-0.5	1	
	MLE	0.8497	0.3867	-0.3749	0.8726	-0.4110	0.8286	
		(0.1859)	(0.1003)	(0.0854)	(0.1659)	(0.0906)	(0.1610)	
	MAPE	1.2304	0.3534	-0.3501	0.8012	-0.3912	1.1890	
		(0.1612)	(0.1104)	(0.0812)	(0.1820)	(0.0900)	(0.1684)	
	ME	1.0549	0.5376	-0.5030	1.0850	-0.5251	1.1831	
		(0.2811)	(0.1428)	(0.1191)	(0.2556)	(0.1471)	(0.3695)	
		μ_1	μ_2	μ_3	m_1	m_1	m_3	
		-1	0	1	0.25	0.25	0.5	
	MLE	-0.9956	-0.0209	1.0430	0.2359	0.2457	0.5147	5.98
		(0.0039)	(0.0062)	(0.0030)	(0.0225)	(0.0365)	(0.0197)	
	MAPE	-1.0143	0.0201	1.0510	0.2346	0.2399	0.4810	5.91
		(0.0031)	(0.0051)	(0.0010)	(0.0190)	(0.0398)	(0.0203)	
	ME							
		σ_{11}	σ_{12}	σ_{13}	σ_{22}	σ_{23}	σ_{33}	
$n = 30$		1	0.5	-0.5	1	-0.5	1	
	MLE	0.8821	0.4543	-0.4296	0.9047	-0.4639	0.8882	
		(0.1119)	(0.0744)	(0.1011)	(0.0923)	(0.0826)	(0.1515)	
	MAPE	0.8103	0.4310	-0.5949	0.8701	-0.4103	1.1730	
		(0.1023)	(0.0802)	(0.1043)	(0.0893)	(0.0921)	(0.1639)	
	ME	1.0330	0.5041	-0.4978	1.1207	-0.5128	1.1432	
		(0.1211)	(0.0520)	(0.0372)	(0.1691)	(0.0569)	(0.3050)	

* MLE is the average of 50 MLEs of the parameters via the ECM algorithm with precision 0.02; MAPE is the average of 50 MAPEs of the parameters by the Bayesian posterior modes via the

ECM algorithm with precision 0.02; ME is the average of 50 MEs of the parameters; MSEs are given in the parentheses; Iteration No. is the average iterative times for the ECM algorithm.

6 Discussion

In this paper, we proposed a new shifted multivariate asymmetric Laplace distribution which is different from the previous versions proposed by Kotz (2001) and Franczak (2012). Although the density function for this Type II SMAL distribution is not of explicit form, a stochastic representation is provided for deriving many useful distributional properties and conducting both likelihood-based and Bayesian statistical inferences. Because of the extensive applications of asymmetric Laplace distribution in analyzing heavy-tailed asymmetric data, the further work is to apply the new distribution and proposed methods to real data analysis and consider Type II shifted multivariate asymmetric Laplace regression model by incorporating covariates.

References

Anderson, D.N. (1992). A multivariate Linnik distribution. *Statistics & Probability Letters* **14** (4), 333336.

Andrews, D.F. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B* **36**(1), 99102.

Hinkley, D. V., & Revankar, N. S. (1977). Estimation of the pareto law from underreported data : a further analysis. *Journal of Econometrics*, **5**(1), 1-11.

Kotz, S., Kozubowski, T.J. and Podgorski, K. (2001). Asymmetric multivariate Laplace distribution. *The Laplace Distribution and Generalizations.* Birkhauser Boston, 239 272.

Kotz, S. , Kozubowski, T. J. , & Podgrski, Krzysztof. (2002). Maximum entropy characterization of asymmetric laplace distribution. *Int.math.j*, **1**(1), 53-63.

Kotz, S. , Kozubowski, T. J. , & Krzysztof Podgrski. (2002). Maximum likelihood estimation of asymmetric laplace parameters. *Annals of the Institute of Statistical Mathematics*, **54**(4), 816-826.

Kozubowski, T.]. and Podgorski, K. (2000). Asymmetric Laplace distributions, *Math. Sci.* **25**, 37-46.

Kozubowski, T. J. . (2001). Asymmetric Laplace laws and modeling financial data.*Elsevier Science Publishers B. V.*

Franczak, B. C. , Browne, R. P. , and Mcnicholas, P. D. . (2014). Mixtures of shifted asymmetric Laplace distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**(6), 1149-1157.

Meng, X.L. and Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* **80**(2), 267278.

Zhang, C., Tian, G.-L. and Yuen, C.K.. (2018). A new robust regression model: Type II multivariate t distribution with applications, under review

Zhang, C., Tian, G.-L. and Tang, M.L.. (2019). A new multivariate Laplace distribution based on mixture of normal distribution, under review

Appendix A: Density function of centralized multivariate asymmetric Laplace distribution

Assuming a d-dimensional random vector $\mathbf{y}$ that follows a centralized asymmetric Laplace (CAL) distribution, i.e.$\mathbf{y} \sim \mathcal{AL}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$,, the density function of $\mathbf{y}$ is given by

$$f(\mathbf{y}|\boldsymbol{m}, \boldsymbol{\Sigma}) = \frac{2K_\nu(u)}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \left(\frac{\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y}}{2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}} \right)^{\nu/2} \exp\left\{ \mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m} \right\} \tag{.1}$$

where $\nu = (2-d)/2$, $u = \sqrt{(2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m})(\mathbf{y}^\top \boldsymbol{\Sigma}^{-1} \mathbf{y})}$, K_ν is the modified Bessel function of the third kind with index ν. $\boldsymbol{\Sigma}$ is a covariance matrix, and $\boldsymbol{m} \in \mathbb{R}^d$ represents the skewness in each dimension.

Appendix B: MLEs of parameters in centralized multivariate asymmetric Laplace distribution via EM algorithm

MLEs of parameters in centralized multivariate asymmetric Laplace distribution can be obtained via the EM algorithm. From the SR (1.2) of the centralized multivariate asymmetric Laplace random vector, we obtain an equivalent mixture representation: $Z \sim \mathrm{Exp}(1)$ and $\mathbf{y}|Z \sim N_d(Z\boldsymbol{m}, Z\boldsymbol{\Sigma})$. Assume $\mathbf{y}_j \overset{\text{iid}}{\sim} \mathcal{AL}_d(\boldsymbol{m}, \boldsymbol{\Sigma})$ for $j = 1, \ldots, n$ and $\boldsymbol{y}_j$ denote the corresponding realization of $\mathbf{y}_j$. For observations $Y_{\text{obs}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, we introduce latent variables $Z_1, ...Z_n \overset{\text{iid}}{\sim} Exp(1)$ and the missing data is $Y_{\text{mis}} = \{z_1, \ldots, z_n\}$. The complete data is $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\}$ and the complete-data likelihood function is given by

$$\begin{aligned} L(\boldsymbol{m}, \boldsymbol{\Sigma}|Y_{\text{com}}) &= \prod_{j=1}^n \left\{ (2\pi)^{-\frac{d}{2}} |z_j \boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{\left(\boldsymbol{y}_j - z_j \boldsymbol{m}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_j - z_j \boldsymbol{m}\right)}{2z_j} \right] \cdot \exp(-z_j) \right\} \\ &\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\sum_{j=1}^n \frac{\left(\boldsymbol{y}_j - z_j \boldsymbol{m}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_j - z_j \boldsymbol{m}\right)}{2z_j} \right] \end{aligned}.$$

Thus, for the M-step, the MLEs of $(\boldsymbol{m}, \boldsymbol{\Sigma})$ are given as

$$\begin{aligned}
\hat{\boldsymbol{m}} &= \frac{\sum_{j=1}^{n} \boldsymbol{y}_j}{\sum_{j=1}^{n} z_j} \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{j=1}^{n} \frac{\left(\boldsymbol{y}_j - z_j\hat{\boldsymbol{m}}\right)\left(\boldsymbol{y}_j - z_j\hat{\boldsymbol{m}}\right)^\top}{z_j} \\
&= \frac{1}{n}\sum_{j=1}^{n} \left(\boldsymbol{y}_j\boldsymbol{y}_j^\top/z_j - \boldsymbol{y}_j\hat{\boldsymbol{m}}^\top - \hat{\boldsymbol{m}}\boldsymbol{y}_j^\top + z_j\hat{\boldsymbol{m}}\hat{\boldsymbol{m}}^\top\right)
\end{aligned}$$

The E-step is to replace $\{z_j\}_{j=1}^n$ and $\{1/z_j\}_{j=1}^n$ by their conditional expectations, $E\left(Z_j|\boldsymbol{y}_j, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$ and $E\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$ for $j = 1, \ldots, n$. From the density function, we can derive the conditional distributions of $Z_j|\mathbf{y}_j$ and find that they follow a generalized inverse Gaussian (GIG) distribution, i.e., $Z_j|\mathbf{y}_j \sim GIG(a, b, \nu)$, where $\nu = (2-d)/2$, $a \equiv 2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}$ and $b \equiv \mathbf{y}_j^\top\boldsymbol{\Sigma}^{-1}\mathbf{y}_j$. Thus, the conditional expectations for $j = 1, \ldots, n$, are given by

$$E\left(Z_j|\boldsymbol{y}_j, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{\boldsymbol{y}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_j}}{\sqrt{2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}}\frac{K_{2-d/2}\left(u_j\right)}{K_{1-d/2}\left(u_j\right)},$$

$$E\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}}{\sqrt{\boldsymbol{y}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_j}}\frac{K_{2-d/2}\left(u_j\right)}{K_{1-d/2}\left(u_j\right)} - \frac{2-d}{\boldsymbol{y}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_j}$$

where $u_j = \sqrt{\left(2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right)\left(\boldsymbol{y}_j^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_j\right)}$, K_ν is the modified Bessel function of the third kind with index ν.

Appendix C: Density function of Type I shifted multivariate asymmetric Laplace distribution

Assuming a d-dimensional random vector $\mathbf{y}$ that follows a Type I shifted multivariate asymmetric Laplace distribution, i.e.$\mathbf{y} \sim \mathcal{SAL}_d^{\mathrm{I}}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$, the density function of $\mathbf{y}$ is given by

$$f(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}) = \frac{2K_\nu(u)}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}\left(\frac{\delta(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\Sigma})}{2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}}\right)^{\nu/2}\exp\left\{\left(\mathbf{y} - \boldsymbol{\mu}\right)^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right\} \qquad (.2)$$

where $\nu = (2-d)/2$, $u = \sqrt{\left(2 + \boldsymbol{m}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{m}\right)\delta(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\Sigma})}$, $\delta(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\Sigma}) = \left(\mathbf{y} - \boldsymbol{\mu}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\mathbf{y} - \boldsymbol{\mu}\right)$, K_ν is the modified Bessel function of the third kind with index ν. $\boldsymbol{\Sigma}$ is a covariance matrix, $\boldsymbol{\mu} \in \mathbb{R}^d$ is a shift parameter and $\boldsymbol{m} \in \mathbb{R}^d$ represents the skewness in each dimension.

Appendix D: MLEs of parameters in Type I shifted multivariate asymmetric Laplace distribution via EM algorithm

MLEs of parameters in Type I shifted multivariate asymmetric Laplace distribution can be obtained via the EM algorithm. From the SR (1.3) of the Type I shifted multivariate asymmetric Laplace random vector, we obtain an equivalent mixture representation: $Z \sim$ Exp(1) and $\mathbf{y}|Z \sim N_d(\boldsymbol{\mu} + Z\boldsymbol{m}, Z\boldsymbol{\Sigma})$. Assume $\mathbf{y}_j \overset{\text{iid}}{\sim} \mathcal{SAL}_d^{\mathrm{I}}(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ for $j = 1, \ldots, n$ and $\boldsymbol{y}_j$ denote the corresponding realization of $\mathbf{y}_j$. For observations $Y_{\text{obs}} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$, we introduce latent variables $Z_1, ...Z_n \overset{\text{iid}}{\sim} Exp(1)$ and the missing data is $Y_{\text{mis}} = \{z_1, \ldots, z_n\}$. The complete data is $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\}$ and the complete-data likelihood function is given by

$$
\begin{aligned}
L\left(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}|Y_{\text{com}}\right) &= \prod_{j=1}^{n} \left\{ (2\pi)^{-\frac{d}{2}} \left|z_j \boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp\left[-\frac{\left(\boldsymbol{y}_j - \boldsymbol{\mu} - z_j \boldsymbol{m}\right)^{\top} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - z_j \boldsymbol{m}\right)}{2z_j} \right] \cdot \exp\left(-z_j\right) \right\} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\sum_{j=1}^{n} \frac{\left(\boldsymbol{y}_j - \boldsymbol{\mu} - z_j \boldsymbol{m}\right)^{\top} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_j - \boldsymbol{\mu} - z_j \boldsymbol{m}\right)}{2z_j} \right].
\end{aligned}
$$

Thus, for the M-step, the MLEs of $(\boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma})$ are given as

$$
\begin{aligned}
\hat{\boldsymbol{\mu}} &= \frac{\sum_{j=1}^{n} (1/z_j) \sum_{j=1}^{n} \boldsymbol{y}_j - n \sum_{j=1}^{n} \left(\boldsymbol{y}_j / z_j\right)}{\sum_{j=1}^{n} z_j \sum_{j=1}^{n} (1/z_j) - n^2} \\
\hat{\boldsymbol{m}} &= \frac{\sum_{j=1}^{n} z_j \sum_{j=1}^{n} \left(\boldsymbol{y}_j / z_j\right) - n \sum_{j=1}^{n} \boldsymbol{y}_j}{\sum_{j=1}^{n} z_j \sum_{j=1}^{n} (1/z_j) - n^2} \\
\hat{\boldsymbol{\Sigma}} &= \frac{1}{n} \sum_{j=1}^{n} \frac{\left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}} - z_j \hat{\boldsymbol{m}}\right)\left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}} - z_j \hat{\boldsymbol{m}}\right)^{\top}}{z_j} \\
&= \frac{1}{n} \sum_{j=1}^{n} \left(\left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}\right)^{\top} / z_j - \left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}\right) \hat{\boldsymbol{m}}^{\top} - \hat{\boldsymbol{m}} \left(\boldsymbol{y}_j - \hat{\boldsymbol{\mu}}\right)^{\top} + z_j \hat{\boldsymbol{m}} \hat{\boldsymbol{m}}^{\top} \right)
\end{aligned}
$$

The E-step is to replace $\{z_j\}_{j=1}^{n}$ and $\{1/z_j\}_{j=1}^{n}$ by their conditional expectations, $E\left(Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$ and $E\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$ for $j = 1, \ldots, n$. From the density function, we can derive the conditional distributions of $Z_j|\mathbf{y}_j$ and find that they follow a generalized inverse Gaussian (GIG) distribution, i.e., $Z_j|\mathbf{y}_j \sim GIG(a, b, \nu)$, where $\nu = (2-d)/2$, $a \equiv 2 + \boldsymbol{m}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{m}$

and $b \equiv \delta(\mathbf{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma}) = (\mathbf{y}_j - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y}_j - \boldsymbol{\mu})$. Thus, the conditional expectations for $j = 1, \ldots, n$, are given by

$$E\left(Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{\delta(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}}{\sqrt{2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}} \frac{K_{2-d/2}(u_j)}{K_{1-d/2}(u_j)},$$

$$E\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}}{\sqrt{\delta(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}} \frac{K_{2-d/2}(u_j)}{K_{1-d/2}(u_j)} - \frac{2-d}{\delta(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}$$

where $u_j = \sqrt{(2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m})\, \delta(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})} = \sqrt{(2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}) \left[\left(\boldsymbol{y}_j - \boldsymbol{\mu}\right)^\top \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{y}_j - \boldsymbol{\mu}\right)\right]}$, K_ν is the modified Bessel function of the third kind with index ν.

Because of the $\left[\delta(\mathbf{y}, \boldsymbol{\mu}|\boldsymbol{\Sigma}) / \left(2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}\right)\right]^{\nu/2}$ term in the density of Type I SMAL distribution, as the EM algorithm iterates towards cnvergence, the value of $\hat{\boldsymbol{\mu}}$ will tend to an observation $\mathbf{y}_j$, which will create computational issues when trying to determine the other parameter values specifically the conditional expectation $E\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right)$, Brian and Ryan (2012) deal with this problem by taking the value of $\hat{\boldsymbol{\mu}}$ at the iteration before it becomes equal to any $\mathbf{y}_j$ (this value is denoted by $\hat{\boldsymbol{\mu}^*}$) as the estimate for $\hat{\boldsymbol{\mu}}$. And then update $\hat{\boldsymbol{m}}$ by

$$\hat{\boldsymbol{m}^*} = \frac{\sum_{j=1}^n \boldsymbol{y}_j - n\hat{\boldsymbol{\mu}^*}}{\sum_{j=1}^n z_j},$$

and update $\hat{\boldsymbol{\Sigma}}$ using $\hat{\boldsymbol{\mu}^*}$ and $\hat{\boldsymbol{m}^*}$.

If we test whether all of the non-diagonal elements in $\boldsymbol{\Sigma}$ are zero, i.e. $\sigma_{ij} = 0$ for $i, j = 1, \ldots, d$ and $i \neq j$. Under H_0, the components are still correlated and the restricted MLEs of parameters $\boldsymbol{\mu}$, $\boldsymbol{m}$ and $\{\sigma_{ii}\}_{i=1}^d$ are obtained via the following EM algorithm:
The M-step: the MLEs are given by

$$\hat{\boldsymbol{\mu}}_R = \frac{\sum_{j=1}^n (1/z_j) \sum_{j=1}^n \boldsymbol{y}_j - n \sum_{j=1}^n \left(\boldsymbol{y}_j/z_j\right)}{\sum_{j=1}^n z_j \sum_{j=1}^n (1/z_j) - n^2}$$

$$\hat{\boldsymbol{m}}_R = \frac{\sum_{j=1}^n z_j \sum_{j=1}^n \left(\boldsymbol{y}_j/z_j\right) - n \sum_{j=1}^n \boldsymbol{y}_j}{\sum_{j=1}^n z_j \sum_{j=1}^n (1/z_j) - n^2}$$

$$\hat{\sigma_{ii,R}} = \frac{1}{n} \sum_{j=1}^n \left((y_{ij} - \hat{\mu}_i)^2 / z_j - 2(y_{ij} - \hat{\mu}_i)\hat{m}_i + z_j \hat{m}_i^2\right)$$

The E-step is to replace $\{z_j\}_{j=1}^n$ and $\{1/z_j\}_{j=1}^n$ by their conditional expectations for $j = 1, \ldots, n$, which are given by

$$E_R\left(Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{\delta_R(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}}{\sqrt{2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}} \frac{K_{2-d/2}\left(u_{j,R}\right)}{K_{1-d/2}\left(u_{j,R}\right)},$$

$$E_R\left(1/Z_j|\boldsymbol{y}_j, \boldsymbol{\mu}, \boldsymbol{m}, \boldsymbol{\Sigma}\right) = \frac{\sqrt{2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}}{\sqrt{\delta_R(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}} \frac{K_{2-d/2}\left(u_{j,R}\right)}{K_{1-d/2}\left(u_{j,R}\right)} - \frac{2-d}{\delta_R(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}$$

where $u_{j,R} = \sqrt{\left(2 + \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}\right) \delta_R(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma})}$, $\delta_R(\boldsymbol{y}_j, \boldsymbol{\mu}|\boldsymbol{\Sigma}) = \sum_{i=1}^d \frac{(y_{ij}-\mu_i)^2}{\sigma_{ii}}$, K_ν is the modified Bessel function of the third kind with index ν. Similarly, we use the same solution to deal with the infinite likelihood problem as that we use in the EM algorithm for MLE.