Gazi University Journal of Science
GU J Sci
29(2):335- 341(2016)

Finite Mixtures of Matrix Variate *t* Distributions

Fatma Zehra DOĞRU[1,♠], Y. Murat BULUT[2], Olcay ARSLAN[3]

[1]*Giresun University, Faculty of Arts and Sciences, Department of Statistics, 28100 Giresun/Turkey*

[2]*Eskisehir Osmangazi University, Faculty of Science and Letters, Department of Statistics, 26480 Eskisehir/Turkey*

[3]*Ankara University, Faculty of Science, Department of Statistics, 06100 Ankara/Turkey*

Received: 06/01/2016 *Accepted: 20/04/2016*

ABSTRACT

Finite mixtures of multivariate *t* distributions (Peel and McLachlan (2000)) were introduced as an alternative to the finite mixtures of multivariate normal distributions to model data sets with heavy tails. In this study, we define the finite mixtures of matrix variate *t* distributions as an extension of finite mixtures of multivariate *t* distributions. Mixtures of matrix variate *t* distributions can provide an alternative robust model to the mixtures of matrix variate normal distributions (Viroli (2011)) for modeling matrix variate data sets with heavy tails. We give an Expectation Maximization (EM) algorithm to find the maximum likelihood (ML) estimators for the parameters of interest. We also provide a small simulation study to illustrate the performance of the proposed EM algorithm for finding estimates.

Key Words: *Finite mixtures, matrix variate t, matrix variate normal, ML estimator, EM algorithm.*

1. INTRODUCTION

Finite mixture models are widely used in several areas such as classification, clustering, data mining, image analysis, pattern recognition, machine learning and etc. (see Titterington et al. (1985), McLachlan and Basford (1988), McLachlan and Peel (2000), Frühwirth-Schnatter (2006) for more details). There are several studies about finite mixtures of multivariate distributions in literature. Some of these studies can be summarized as follows. The parameters estimation of multivariate normal mixture model was studied by McLachlan and Basford (1988), Peel and McLachlan (2000) proposed finite mixtures of multivariate *t* distributions, Lin (2009) studied multivariate skew normal mixture models, Pyne et al. (2009) and Lin (2010) introduced finite mixtures of multivariate skew *t* distributions, Cabral et al. (2012) proposed multivariate mixture modeling based on the skew-normal independent distributions and an EM-type algorithm for the mixture of multivariate normal inverse Gaussian distribution, which is a variance-mean mixture of multivariate Gaussians, was given by O'Hagan et al. (2015).

Matrix variate distributions model random matrices and have an important role to explore the theory of the multivariate analysis (see Dawid (1981), De Wall (1988) and Rowe (2003)). Matrix variate normal distribution is a generalization of the multivariate normal distribution. Some classical works about parameter estimation of matrix normal distribution have been done (Srivastava and Khatri (1979), Dawid (1981), Dutilleul (1999), Srivastava et al. (2008)). Gupta et al. (2013) introduced matrix variate *t* distribution as a

♠Corresponding author, e-mail: *fatma.dogru@giresun.edu.tr*

special case of Pearson type VII distribution which is an alternative to matrix variate normal distribution. Bulut and Arslan (2016) studied some distributional properties and gave parameter estimation of matrix variate *t* distribution. They used the EM algorithm to obtain estimates of parameters.

Concerning the finite mixtures of matrix variate distributions, Viroli (2011) introduced finite mixtures of matrix normal distributions for classifying three-way data and proposed an EM algorithm to estimate the parameters. However, when data sets have longer than normal tails or outliers, the estimators obtained using finite mixtures of matrix normal distributions will be influenced. To deal with the problem of heavy-tailedness and outliers in the data, we propose finite mixtures of the matrix variate *t* distributions.

The paper is organized as follows. In Section 2, we give some details about the matrix variate *t* distribution. In Section 3, we define the finite mixtures of matrix variate *t* distributions. In Subsection 3.1, the parameter estimation for the proposed model is carried out using the EM algorithm. Section 4 is devoted to the simulation study to demonstrate the performance of the proposed EM algorithm. The paper is finalized with a conclusion section.

2. MATRIX VARIATE *T* DISTRIBUTION

A $n \times p$-variate random matrix X is said to have a matrix variate *t* distribution with mean matrix M and variance-covariance matrices Σ and Ψ, denoted by $Mt_{n,p}(M, \Sigma, \Psi, \nu)$, if it has the following probability density function (pdf)

$$f(X) = \frac{|\Sigma|^{-\frac{p}{2}}|\Psi|^{-\frac{n}{2}}\Gamma\left(\frac{np+\nu}{2}\right)}{(\pi\nu)^{\frac{np}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left[1+\frac{\delta_X(M,\Sigma,\Psi)}{\nu}\right]^{-\frac{np+\nu}{2}}, \tag{1}$$

where $\delta_X(M,\Sigma,\Psi) = tr\{\Sigma^{-1}(X-M)\Psi^{-1}(X-M)'\}$ which is the Mahalanobis distance from X to the center M with respect to Σ and Ψ (Gupta et. al (2013) and Bulut and Arslan (2015)).

Matrix variate *t* distribution can be obtained as a scale mixture of matrix variate normal and gamma distributions. Let $Z \sim N_{n,p}(0, I_n, I_p)$ and $U \sim Gamma\left(\frac{\nu}{2},\frac{\nu}{2}\right)$ be two independent random matrix and random variable, $M \in R^{n\times p}$, Σ and Ψ are positive definite symmetric matrices and $\Sigma^{\frac{1}{2}}$, $\Psi^{\frac{1}{2}}$ are the positive definite square roots of Σ and Ψ, respectively. Then, the random matrix X defined as

$$X = M + \Sigma^{\frac{1}{2}}Z\Psi^{\frac{1}{2}}U^{-\frac{1}{2}} \tag{2}$$

will have matrix variate *t* distribution (Gupta et. al (2013) and Bulut and Arslan (2015)). The scale mixture representation given in (2) not only simplifies random number generating but also provides the implementation of EM algorithm to obtain ML estimates for the parameters of the matrix variate *t* distribution. A hierarchical representation of matrix variate *t* distribution can be given as follows

$$X|U=u \sim N_{n,p}(M, u^{-1}\Sigma, \Psi),$$
$$U \sim Gamma\left(\frac{\nu}{2},\frac{\nu}{2}\right). \tag{3}$$

Using the hierarchical representation given in (3), the conditional distribution of U given X can be obtained as

$$U|X \sim Gamma\left(\frac{np+\nu}{2}, \frac{1}{2}(\nu + tr\{\Sigma^{-1}(X-M)\Psi^{-1}(X-M)'\})\right). \tag{4}$$

This conditional distribution will be necessary in EM algorithm to compute the conditional expectations of the complete data log-likelihood function, which will be given in Subsection 3.1. Further, using the conditional distribution given in (4) the following conditional expectations can be easily obtained

$$E(U|X) = \frac{np+\nu}{\nu + tr\{\Sigma^{-1}(X-M)\Psi^{-1}(X-M)'\}}, \tag{5}$$

$$E(\log U|X) = DG\left(\frac{np+\nu}{2}\right) - \log\left(\frac{1}{2}(\nu + tr\{\Sigma^{-1}(X-M)\Psi^{-1}(X-M)'\})\right), \tag{6}$$

where $DG(x) = \frac{d}{dx}\log\Gamma(x)$ is the digamma function. Again, these conditional expectations will be used in EM algorithm given in Subsection 3.1.

3. FINITE MIXTURES OF MATRIX VARIATE *T* DISTRIBUTIONS

The pdf of finite mixtures of matrix variate *t* distributions can be defined as follows

$$f(X_j;\Lambda)=\sum_{i=1}^{g} w_i\, f(X_j;M_i,\Sigma_i,\Psi_i,\nu_i)\,, \quad (7)$$

where w_i's are mixing probabilities with $\sum_{i=1}^{g} w_i=1$, $0\le w_i<1$, $f(X_j;M_i,\Sigma_i,\Psi_i,\nu_i)$ represents the density of ith component (the pdf of matrix variate *t* distribution) given in (1) and $\Lambda=(w_1,\dots,w_g,M_1,\dots,M_g,\Sigma_1,\dots,\Sigma_g,\Psi_1,\dots,\Psi_g,\nu_1,\dots,\nu_g)$.

3.1. ML estimation of the matrix variate *t* mixture model

Let $\boldsymbol{X}=(X_1,X_2,\dots,X_l)$ be a random sample of matrices in $\mathcal{R}^{n\times p}$ and assume that this random sample has a distribution of g-component mixture of matrix variate *t* distributions with the pdf given in (7). The ML estimator of Λ can be found by maximizing the following log-likelihood function for the mixtures of matrix variate *t* distributions

$$\ell(\Lambda;\boldsymbol{X})=\sum_{j=1}^{l}\log\left(\sum_{i=1}^{g} w_i\, f(X_j;M_i,\Sigma_i,\Psi_i,\nu_i)\right). \quad (8)$$

However, because of the mixture structure of the pdf, the maximizer of the above log-likelihood function cannot be explicitly obtained. Therefore, an EM-type algorithm (Dempster et al. (1977)) should be used to obtain the estimate for Λ. To do so let Z_{ij} be the component label defined as

$$Z_{ij}=\begin{cases}1, if\ j^{th}\ observation\ is\ from\ i^{th}\ component\\ 0, otherwise\end{cases}$$

where, $j=1,\dots,l$ and $i,\dots,g$.

Further, $\boldsymbol{U}=(U_1,\dots,U_l)$ and $\boldsymbol{Z}=(Z_1,\dots,Z_l)$ with $Z_j=(Z_{1j},\dots,Z_{gj})$ for $j=1,\dots,l$ will be regarded as missing data and $(\boldsymbol{X},\boldsymbol{U},\boldsymbol{Z})$ be the complete data. Then, using the hierarchical formulation given in (3) we get

$$X_j|U_j=u_j,Z_{ij}=1\sim N_{n,p}(M_i,u_j^{-1}\Sigma_{\rm i},\Psi_{\rm i})\,,$$
$$U_j|Z_{ij}=1\sim Gamma\left(\frac{\nu_i}{2},\frac{\nu_i}{2}\right). \quad (9)$$

The complete data log-likelihood function for $(\boldsymbol{X},\boldsymbol{U},\boldsymbol{Z})$ can be written as

$$\ell_c(\Lambda;\boldsymbol{X},\boldsymbol{U},\boldsymbol{Z})=\sum_{j=1}^{l}\sum_{i=1}^{g} z_{ij}\Big\{\log w_i-\frac{np}{2}\log(2\pi)+\frac{np}{2}\log(u_j)-\frac{p}{2}\log|\Sigma_i|-\frac{n}{2}\log|\Psi_i| -tr\left(\frac{u_j}{2}\Sigma_i^{-1}(X_j-M_i)\Psi_i^{-1}(X_j-M_i)'\right)+\frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right)-\log\Gamma\left(\frac{\nu_i}{2}\right)+\left(\frac{\nu_i}{2}-1\right)\log(u_j)-\frac{\nu_i}{2}u_j\Big\}. \quad (10)$$

Maximizing this complete data log-likelihood function will give the estimates for the parameters of interest. However, since we have latent variables $(\boldsymbol{U},\boldsymbol{Z})$, the ML estimators cannot be used to deal with this latency. We have to replace these latent variables by their conditional expectations given the observed data $\boldsymbol{X}$. The conditional expectation of the complete data log-likelihood function given the observed data is

$$E(\ell_c(\Lambda;\boldsymbol{X},\boldsymbol{U},\boldsymbol{Z})|X_j)=\sum_{j=1}^{l}\sum_{i=1}^{g} E(Z_{ij}|X_j)\Big\{\log w_i-\frac{np}{2}\log(2\pi)-\frac{p}{2}\log|\Sigma_i|-\frac{n}{2}\log|\Psi_i| +\frac{np}{2}E(\log(U_j)|X_j,Z_{ij}=1)-tr\left(\frac{E(U_j|X_j,Z_{ij}=1)}{2}\Sigma_i^{-1}(X_j-M_i)\Psi_i^{-1}(X_j-M_i)'\right) +\frac{\nu_i}{2}\log\left(\frac{\nu_i}{2}\right)-\log\Gamma\left(\frac{\nu_i}{2}\right)+\left(\frac{\nu_i}{2}-1\right)E(\log(U_j)|X_j,Z_{ij}=1)-\frac{\nu_i}{2}E(U_j|X_j,Z_{ij}=1)\Big\}. \quad (11)$$

The conditional expectations $E(U_j|X_j,Z_{ij}=1)$ and $E(\log(U_j)|X_j,Z_{ij}=1)$ can be computed using the equations given in (5) and (6) and the conditional expectation $E(Z_{ij}|X_j)$ can be calculated using the classical theory of mixture modeling. Now, we will give the steps of the EM algorithm.

EM algorithm:

♠Corresponding author, e-mail: *fatma.dogru@giresun.edu.tr*

1. Take initial parameter estimate $\Lambda^{(0)}$ and fix a stopping rule Δ.

2. E-step: Given the current parameter values $\widehat{\Lambda}^{(k)}$and the observed data, compute the conditional expectations $\hat{z}_{ij}^{(k)}, \hat{u}_{1ij}^{(k)}$ and $\hat{u}_{2ij}^{(k)}$ using the following equations for $k = 0,1,2, \ldots$ iteration

$$\hat{z}_{ij}^{(k)} = E\left(Z_{ij} \middle| X_j, \widehat{\Lambda}^{(k)}\right) = \frac{w_i^{(k)} f\left(X_j; \widehat{M}_i^{(k)}, \widehat{\Sigma}_i^{(k)}, \widehat{\Psi}_i^{(k)}, \hat{v}_i^{(k)}\right)}{f\left(X_j; \widehat{\Lambda}^{(k)}\right)}, \quad (12)$$

$$\hat{u}_{1ij}^{(k)} = E\left(U_j \middle| X_j, Z_{ij} = 1, \widehat{\Lambda}^{(k)}\right) \quad (13)$$

$$= \frac{np + \hat{v}_i^{(k)}}{\hat{v}_i^{(k)} + tr\left\{\widehat{\Sigma}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)\widehat{\Psi}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)'\right\}},$$

$$\hat{u}_{2ij}^{(k)} = E\left(\log(U_j) \middle| X_j, Z_{ij} = 1, \widehat{\Lambda}^{(k)}\right) \quad (14)$$

$$= DG\left(\frac{np + \hat{v}_i^{(k)}}{2}\right) - \log\left(\frac{1}{2}\left(\hat{v}_i^{(k)} + tr\left\{\widehat{\Sigma}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)\widehat{\Psi}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)'\right\}\right)\right).$$

Then, after finding these conditional expectations we can form the following objective function

$$Q\left(\Lambda; \widehat{\Lambda}^{(k)}\right) = E\left(\ell_c(\Lambda; \boldsymbol{X}, \boldsymbol{U}, \boldsymbol{Z}) | X_j, \widehat{\Lambda}^{(k)}\right)$$

$$= \sum_{j=1}^{l}\sum_{i=1}^{g} \hat{z}_{ij}^{(k)} \{\log w_i + -\frac{np}{2}\log(2\pi) + \frac{np}{2}\hat{u}_{2ij}^{(k)} - \frac{n}{2}\log|\Psi_i| - tr\left(\frac{\hat{u}_{1ij}^{(k)}}{2}\Sigma_i^{-1}(X_j - M_i)\Psi_i^{-1}(X_j - M_i)'\right) - \frac{p}{2}\log|\Sigma_i| + \frac{v_i}{2}\log\left(\frac{v_i}{2}\right) - \log\Gamma\left(\frac{v_i}{2}\right)$$

$$+ \left(\frac{v_i}{2} - 1\right)\hat{u}_{2ij}^{(k)} - \frac{v_i}{2}\hat{u}_{1ij}^{(k)}\}. \quad (15)$$

2. M-step 1: To obtain the $(k+1)$ th parameter estimates maximize $Q\left(\Lambda; \widehat{\Lambda}^{(k)}\right)$ with respect to the unknown parameters $(w_i, M_i, \Sigma_i, \Psi_i)$, fixing v_i at $v_i^{(k)}$. This maximization gives

$$\widehat{w}_i^{(k+1)} = \frac{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}}{l}, \quad (16)$$

$$\widehat{M}_i^{(k+1)} = \frac{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}X_j}{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}}, \quad (17)$$

$$\widehat{\Sigma}_i^{(k+1)} = \frac{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}\left(X_j - \widehat{M}_i^{(k)}\right)\widehat{\Psi}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)'}{p\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}}, \quad (18)$$

$$\widehat{\Psi}_i^{(k+1)} = \frac{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}\hat{u}_{1ij}^{(k)}\left(X_j - \widehat{M}_i^{(k)}\right)\widehat{\Sigma}_i^{-1(k)}\left(X_j - \widehat{M}_i^{(k)}\right)'}{n\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}}. \quad (19)$$

3. M step 2: Using the new values of $\left(\widehat{w}_i^{(k+1)}, \widehat{M}_i^{(k+1)}, \widehat{\Sigma}_i^{(k+1)}, \widehat{\Psi}_i^{(k+1)}\right)$ given in M step 1, solve the following equation to obtain $(k+1)$ th estimate of v_i

$$\log\left(\frac{v_i}{2}\right) + 1 - DG\left(\frac{v_i}{2}\right) + \frac{1}{\sum_{j=1}^{l}\hat{z}_{ij}^{(k)}}\left(\hat{u}_{2ij}^{(k)} - \hat{u}_{1ij}^{(k)}\right) = 0. \quad (20)$$

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

4. SIMULATION STUDY

In this section, we provide a small simulation study to show the performance of the proposed algorithm. Here, we consider the simulation design as follows. We generate data using the stochastic representation given in (2) from a two-component and a three-component matrix variate *t* distributions

$$f(X_j; \Lambda) = w_1 f(X_j; M_1, \Sigma_1, \Psi_1, v) + (1 - w_1) f(X_j; M_2, \Sigma_2, \Psi_2, v)$$

and

$$f(X_j; \Lambda) = w_1 f(X_j; M_1, \Sigma_1, \Psi_1, v) + w_2 f(X_j; M_2, \Sigma_2, \Psi_2, v) + (1 - w_1 - w_2) f(X_j; M_3, \Sigma_3, \Psi_3, v)$$

where $j = 1, \dots, l$, respectively.

The parameters of two-component mixture model are taken according to the following formulation

$M_1(i,j) = i + j - 1$, $M_2(i,j) = 5i + 5j - 1$,
$i = 1, \dots, n$, $j = 1, \dots p$,
$\Sigma_1 = \Sigma_2 = 2I_n$, $\Psi_1 = \Psi_2 = I_p$.

The mixing probability is $w_1 = 0.5$ with $n = 2, 3$, $p = 2$. The degrees of freedom of the component distributions are taken as $\nu = 3$ and 5. Similarly, the parameters of three-component mixture model are

$M_1(i,j) = i + j - 1$, $M_2(i,j) = 5i + 5j - 1$,
$M_3(i,j) = 10i + 10j - 1$, $i = 1, \dots, n$, $j = 1, \dots, p$,
$\Sigma_1 = \Sigma_2 = \Sigma_3 = 2I_n$, $\Psi_1 = \Psi_2 = \Psi_3 = I_p$,

with the mixing probabilities $w_1 = 0.3$ and $w_2 = 0.3$. The dimensions are taken as $n = 2, 3$, $p = 2$. Again the degrees of freedom are taken as 3 and 5. Note that, for the sake of simplicity, we have the same degrees of freedom for all the components in the mixture model.

However, different degrees of freedom can be taken. We conduct the computations using MATLAB (2013). Also, the simulation scenarios are repeated 100 times. For the simulation study, the stopping rule Δ is taken as 10^{-8}.

In our future work, we will use the mixture model in clustering. Thus, we are mainly interested in estimating the location parameters rather than the scatter matrices. Therefore, in our simulation study we mainly estimate the location parameters and the mixing probabilities and take the scatter parameters and the degrees of freedom as known.

In Tables 1-4, we present the Euclidean distance defined as

$$\|\hat{M} - M\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\hat{M}_{ij} - M_{ij}\right)^2\right)^{\frac{1}{2}}, \quad (21)$$

(Dutilleul (1999) and Bulut and Arslan (2015)) for the mean matrices. Also, the estimates for the mixing probabilities are given in the tables. The simulation results show that the proposed algorithm is working accurately to obtain the estimates for the location matrices. This can be observed from mean Euclidian distances that are getting smaller when the sample sizes increase. Further, the estimates for the mixing probabilities are very close to the true parameter values. For example, for the case $\nu = 3$ and $l = 100$ true mixing probability is 0.5 and the mean estimates over the 100 runs is 0.5037, which is very close to the true parameter value. We can observe the similar behavior for the other cases.

Table 1. Mean Euclidean distance for the mean matrices and $\hat{w}_1$ of the two-component mixture model for 2x2 matrix.

	$\nu = 3$			$\nu = 5$		
l	$\hat{w}_1$	$\|\hat{M}_1 - M_1\|$	$\|\hat{M}_2 - M_2\|$	$\hat{w}_1$	$\|\hat{M}_1 - M_1\|$	$\|\hat{M}_2 - M_2\|$
100	0.5037	0.3633	0.3780	0.4942	0.1781	0.1775
200	0.5007	0.2801	0.2902	0.5044	0.1228	0.1230
300	0.5023	0.2488	0.2647	0.4982	0.0999	0.1003
400	0.5082	0.2253	0.2311	0.5003	0.0863	0.0867

Table 2. Mean Euclidean distance for the mean matrices and $\hat{w}_1$ of the two-component mixture model for 3x2 matrix.

	$\nu = 3$			$\nu = 5$		
l	$\hat{w}_1$	$\|\hat{M}_1 - M_1\|$	$\|\hat{M}_2 - M_2\|$	$\hat{w}_1$	$\|\hat{M}_1 - M_1\|$	$\|\hat{M}_2 - M_2\|$
100	0.5043	0.4540	0.4726	0.5045	0.2016	0.2032
200	0.5007	0.3684	0.3738	0.5055	0.1427	0.1439

300	0.4947	0.3246	0.3362	0.5020	0.1185	0.1149
400	0.4986	0.3120	0.3081	0.5018	0.1067	0.1057

Table 3. Mean Euclidean distance for the mean matrices and $\hat{w}_1, \hat{w}_2$ of the three-component mixture model for 2x2 matrix.

$\nu = 3$					
l	$\hat{w}_1$	$\hat{w}_2$	$\Vert\hat{M}_1 - M_1\Vert$	$\Vert\hat{M}_2 - M_2\Vert$	$\Vert\hat{M}_3 - M_3\Vert$
100	0.2872	0.2949	0.5149	0.4362	0.4034
200	0.2858	0.2888	0.3924	0.3336	0.2933
300	0.2889	0.2919	0.3240	0.2425	0.2707
400	0.2871	0.2929	0.2962	0.2289	0.2466
$\nu = 5$					
l	$\hat{w}_1$	$\hat{w}_2$	$\Vert\hat{M}_1 - M_1\Vert$	$\Vert\hat{M}_2 - M_2\Vert$	$\Vert\hat{M}_3 - M_3\Vert$
100	0.2959	0.3017	0.2246	0.2273	0.1899
200	0.3030	0.2998	0.1562	0.1668	0.1385
300	0.3054	0.2979	0.1301	0.1317	0.1144
400	0.2960	0.3029	0.1104	0.1076	0.0945

Table 4. Mean Euclidean distance for the mean matrices and $\hat{w}_1, \hat{w}_2$ of the three-component mixture model for 3x2 matrix.

$\nu = 3$					
l	$\hat{w}_1$	$\hat{w}_2$	$\Vert\hat{M}_1 - M_1\Vert$	$\Vert\hat{M}_2 - M_2\Vert$	$\Vert\hat{M}_3 - M_3\Vert$
100	0.2841	0.2953	0.6939	0.5781	0.5140
200	0.2823	0.2895	0.5018	0.4258	0.3853
300	0.2702	0.2937	0.4587	0.3217	0.3535
400	0.2870	0.2884	0.4078	0.2869	0.3210
$\nu = 5$					
l	$\hat{w}_1$	$\hat{w}_2$	$\Vert\hat{M}_1 - M_1\Vert$	$\Vert\hat{M}_2 - M_2\Vert$	$\Vert\hat{M}_3 - M_3\Vert$
100	0.2999	0.2952	0.2713	0.2792	0.2307
200	0.2968	0.2928	0.1954	0.1938	0.1584
300	0.2936	0.3008	0.1603	0.1597	0.1321
400	0.2963	0.2997	0.1354	0.1335	0.1158

5. CONCLUSIONS

In this paper, we have proposed the finite mixtures of matrix variate t distributions as a robust alternative to the finite mixtures of matrix variate normal distributions. We have given the EM algorithm to compute the estimates for the proposed mixture model. Also, we have provided a simulation study to show the estimation performance of the proposed mixture model. In the simulation study, we have observed that the mixture component and the location parameters are estimated accurately and mean Euclidean distance is getting smaller when the sample size is increasing.

CONFLICT OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Bulut, Y. M. and Arslan, O., “Matrix variate t-distribution: properties, parameter estimation and application to robust statistical analysis”, (2016) (submitted).

[2] Cabral, C., Lachos, V. and Prates, M., “Multivariate mixture modeling using skew-normal independent distributions”, Computational Statistics and Data Analysis, 56, 126-142, (2012).

[3] Dawid, A. P., “Some matrix-variate distribution theory: notational considerations and a Bayesian application”, Biometrika, 68, 265-274, (1981).

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

[5] DeWall, D. J., Matrix-variate distributions. In: Kotz, S., Johnson, N.L. (eds.): Encyclopedia of Statistical Sciences, vol. 5, pp. 326–333, Wiley, New York, (1988).

[6] Dutilleul, P., “The MLE algorithm for the matrix normal distribution”, Journal of Statistical Computation and Simulation, 64 (2), 105–123, (1999).

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

[8] Gupta, A. K., Varga, T. and Bodnar, T., Elliptically Contoured Models in Statistics and Portfolio Theory. Springer, New York, (2013).

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

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

[11] MATLAB version 8.1.0 (R2013a). The MathWorks Inc., Natick, Massachusetts, (2013).

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

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

[14] O’Hagan, A., Murphy, T. B., Gormley, I. C., McNicholas, P. and Karlis, D., “Clustering with the multivariate normal inverse Gaussian distribution”, Computational Statistics and Data Analysis, 93, 18-30, (2016).

[15] Peel, D. and McLachlan, G. J., “Robust mixture modelling using the t distribution”, Statistics and Computing, 10, 339-348, (2000).

[16] Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T. and Maier, L. M., “Automated high-dimensional flow cytometric data analysis”, Proceedings of the National Academy of Sciences USA 106, 8519-8524, (2009).

[17] Rowe, B. R., Multivariate Bayesian Statistics. Chapman and Hall/CRC, London/Boca Raton, (2003).

[18] Srivastava, M. and Khatri, C., An Introduction to Multivariate Statistics. North Holland, New York, USA, (1979).

[19] Srivastava, M., von Rosen, T. and von Rosen, D., “Models with a Kronecker product covariance structure: Estimation and testing”, Mathematical Methods of Statistics, 17(4), 357–370, (2008).

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

[21] Viroli, C., “Finite mixtures of matrix normal distributions for classifying three-way data”, Statistical Computing, 21, 511–522, (2011).