Statistics and Applications {ISSN 2454-7395 (online)}
Volume 20, No. 2, 2022 (New Series), pp 219–237

Clustering using Skewed Data via Finite Mixtures of Multivariate Lognormal Distributions

Deepana R. and Kiruthika C.
Department of Statistics, Pondicherry University, India

Received: 10 July 2021; Revised: 21 October 2021; Accepted: 26 December 2021

Abstract

Model-based clustering techniques are based on the finite mixture models. In this paper, an attempt is made to explore effect of the skewness in heterogeneous data using finite mixture models to clustering. In particular, this paper deals with model-based clustering using finite mixtures of multivariate lognormal distributions which can deal with skewness effectively. The Expectation Maximization (EM) algorithm is used for computing maximum likelihood estimates for model parameters. To examine the performance of clustering multivariate log normal mixtures models, some simulation studies are presented for heterogeneous data with asymmetric behavior. A real dataset is also used to illustrate the use of finite mixtures of multivariate lognormal distributions to clustering.

Key words: Multivariate log normal distribution; Finite mixture model; Model based clustering; EM algorithm.

AMS Subject Classifications: 62K99, 62J05

1. Introduction

Clustering is an unsupervised learning technique. Clustering is grouping of a set of data objects into several clusters so that objects within a cluster have high level of similarity, but they are dissimilar to the objects in other clusters. Clustering is also defined in a probabilistic approach, where the notion of clusters is formalized through their probability distributions. One of the main advantages of this probabilistic approach is that it can be interpreted from a statistical point of view for the obtained clusters. In the model-based clustering methods, the observations are generated from a mixture of probability distributions, in which each component represents a different cluster. An extensive review of finite mixture models and their clustering applications are given by Everitt and Hand (1981), Titterington *et al.* (1985) and McLachlan and Peel (2000). Finite mixtures of multivariate Gaussian distribution are widely used in model-based clustering. One may refer to McLachlan and Basford (1988), McNicholas and Murphy (2008), Beak and McLachlan (2010) and among others. Melnykov and Semhar (2016) have discussed about the challenges of model-based clustering such as initialization techniques, dimension reduction and variable selection. However, clustering based on Gaussian mixture models is not capable of reasonably fittings for heavy tails, asymmetric and outliers to the heterogeneous data.

Corresponding Author: Deepana R
Email: deepana192007@gmail.com

Model-based clustering using finite mixture models with non-normal distributions have received increasing attention and showed advantages in modeling heterogeneous data with heavy tails, asymmetric and outliers. Non-normal finite mixture distribution plays an important role in clustering applications when the component densities are skewed and heavy tailed. Karlis *et al.* (2002), Lin *et al.* (2007), Pyne *et al.* (2009), Soltyk and Gupta (2011) have given application of univariate and multivariate finite mixtures of skew-normal and skew-t distributions to clustering. Schnatter *et al.* (2010) have proposed Bayesian approach for finite mixture models of univariate and multivariate skew-t and skew normal distributions. The estimation of parameters in these mixture models is carried out by EM algorithm. Lee and McLachlan (2013a) have provided finite mixture models with skew normal and skew-t distributions and it has increased importance in modeling data withequal asymmetry and heavy tails simultaneously. Also, they have classified multivariate skew distributions into four types namely, 'restricted', 'unrestricted', 'extended' and 'generalized' forms. Lee and McLachlan (2013b) have compared the clustering performance of mixture in multivariate skew normal and skew-t distributions with other non-normal mixture distributions like generalized hyperbolic distributions, multivariate inverse-Gaussian distributions and shifted asymmetric Laplace distributions. Lee and McLachlan (2014) have provided some recent developments of mixtures in multivariate skew-t distributions. Also, they have discussed about various characterizations of multivariate skew-t distribution. Further, they have used existing EM algorithms for estimating the parameters of the restricted and unrestricted forms of multivariate skew-t mixture models. Sanjeena *et al.* (2014) have considered univariate and multivariate normal inverse Gaussian distribution for model-based clustering approach in finite mixture models and parameter estimation is carried out by the EM algorithm. A shifted asymmetric Laplace distribution is considered for model-based clustering by Franczak *et al.* (2014). A multivariate generalized hyperbolic mixture model was proposed by Browne and McNicholas (2015). Adrian *et al.* (2016) proposed clustering using multivariate normal inverse Gaussian distribution for heavy tails and asymmetric data. Melnykov *et al.* (2018) have developed finite mixture modeling with components that can handle skewness in matrix-valued data.

Although many non-symmetric distributions are available, model-based clustering using finite mixtures of multivariate lognormal distribution is considered in this paper. A finite mixture of multivariate lognormal distribution is useful in modeling heterogeneous data with asymmetric behaviour. In the present study, an attempt is made to obtain clusters for skewed data based on model-based clustering using finite mixtures of multivariate lognormal distribution. A parsimonious family of finite mixtures of multivariate lognormal distribution is also developed. Algorithms for model parameter estimation and initialization technique are presented in this paper. Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) are used for model selection. The clustering performance is evaluated using Adjusted Rand Index (ARI) and Misclassification Rate (MR). The performance of multivariate lognormal mixture models in clustering for real and simulated data are studied. The proposed initialization method to determine the initial value for the component parameters using EM algorithm is presented in the next section. The methodology for initialization technique considered in this paper overcomes the issue of initial values in EM algorithm by using K-means clustering with Mahalanobis distance measures.

The rest of this paper is organized as follows. Section 2 presents the initialization techniques for model-based clustering approach. Section 3 describes the multivariate lognormal mixture models using EM algorithm. In Section 4 real and simulated datasets are applied for multivariate lognormal mixture models to clustering and are compared to some well-known existing methods in Section 5. In Section 6, some concluding remarks are given.

2. Initialization Technique for Multivariate Lognormal Mixture Models

The EM algorithm relies on the specified starting values for component parameters. However, it is difficult to specify good starting values. Several research works have been done for initialization for component parameters in EM algorithm. Mahalanobis distance measure is used to capture the covariance structures of clusters. Mahalanobis distance measure is used to identify and correctly classify non-spherical clusters for non-homogeneous data. Mahalanobis distance measure overcomes the variable standardization by yielding scale invariant classification. The proposed algorithm is presented below.

Algorithm

Input: Data X and the number of groups G

Output: Cluster Indicator $z_1, z_2, ..., z_n$

1. Randomly select the mean vector according to G groups from the dataset X.
2. Compute Euclidean distance based on the mean vectors. Assigning each observation nearest to the group mean vector. Compute the new mean vector $\underline{c}_k; k = 1, 2, ..., G$ and the covariance matrix $S_k; k = 1, 2, ..., G$ based on the assignments.
3. While for $1, 2, ..., G$ do
4. Compute the Mahalanobis distance measure based on the new mean vector $\underline{c}_k$ and the covariance matrix S_k
$$D(\underline{x}_i, \underline{c}_k) = \sqrt{(\underline{x}_i, \underline{c}_k^{(q)^{(t)}}) S_k^{-1^{(q)}} (\underline{x}_i, \underline{c}_k^{(q)})}$$
5. Assignment: Assign each observation nearest to cluster center $\underline{z}_{ik} = 1$ if $D(\underline{x}_i, \underline{c}_k)$
6. Update: Recalculate the mean and covariance matrix for (k = 1, 2, . . . , G) based on the assignments.
$$\underline{c}_k^{(q+1)} = \frac{\sum_{i=1}^n z_{ik}\underline{x}_i}{\sum_{i=1}^n z_{ik}}$$
$$S_k^{(q+1)} = \frac{\sum_{i=1}^n z_{ik}(\underline{x}_i, \underline{c}_k^{(q+1)})(\underline{x}_i, \underline{c}_k^{(q+1)})^{(t)}}{\sum_{i=1}^n z_{ik}}$$
where, q is the iteration number and t represents the transpose.
7. end While

Based on the cluster indicators $z_1, z_2, ..., z_n$ the initial component parameter values $\pi_k^{(0)}, \underline{mu}_k^{(0)}, \Sigma_k^{(0)}$. The initial values are used to initiate the EM algorithm for Multivariate Lognormal (MLN) mixture models to clustering. The parameter estimation procedure is derived in the following section.

3. Parameter Estimation Procedure for Multivariate Lognormal Mixture Models

Let data X be a d-dimensional random variable which follows a multivariate lognormal distribution with mean vector $\underline{\mu}_k$ and the covariance matrix Σ_k. The G-component finite mixture model of multivariate lognormal distributions is given by

$$f(\underline{x}_i|\Theta) \;= \sum_{k=1}^{G} \pi_k \; \frac{1}{(2\pi)^{1/2}|\Sigma_k|^{1/2}|\underline{x}_i} e^{-\frac{1}{2}(ln(\underline{x}_i)-\underline{\mu}_k)^t \Sigma_k^{-1}(ln(\underline{x}_i)-\underline{\mu}_k)} \tag{1}$$

where π_k represents the mixing proportion with $\sum_{k=1}^{G} \pi_k = 1, 0 < \pi_k < 1$. The unknown parameter Θ is $\{\pi_1, \pi_2, ..., \pi_{G-1}, \underline{x}_1, \underline{x}_2, ..., \underline{x}_G, \Sigma_1, \Sigma_2, ..., \Sigma_G\}$.

Consider the random sample of size n from multivariate Lognormal mixture models defined the probability density function given in (1). EM algorithm [Dempster *et al.* 1977] is used for the parameter estimation. The complete data in EM algorithm is written as (X, Z). The observed data vector $X = (x_1, x_2, ..., x_n)^T$ is viewed as incomplete. The component label vector is defined as $Z = z_1, z_2, ..., z_n$. The likelihood of complete data of multivariate lognormal mixture model is given by

$$L(\Theta; X, Z) = \prod_{i=1}^{n}\prod_{k=1}^{G} [\pi_k f(\underline{x}_i; \underline{\mu}_k, \Sigma_k)]^{z_{ik}} \tag{2}$$

$$= \prod_{i=1}^{n}\prod_{k=1}^{G} [\pi_k \; \frac{1}{(2\pi)^{1/2}|\Sigma_k|^{1/2}|\underline{x}_i} e^{-\frac{1}{2}(ln(\underline{x}_i)-\underline{\mu}_k)^t \Sigma_k^{-1}(ln(\underline{x}_i)-\underline{\mu}_k)}]^{z_{ik}}$$

The log-likelihood of complete data of multivariate lognormal mixture models is given by

$$l(\Theta; X, Z) = \sum_{i=1}^{n}\sum_{k=1}^{G} z_{ik}[log\pi_k + log[\; \frac{1}{(2\pi)^{1/2}|\Sigma_k|^{1/2}\underline{x}_i}] + -\frac{1}{2}(ln(\underline{x}_i) - \underline{\mu}_k)^t \Sigma_k^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)] \tag{3}$$

The conditional expectation of the log-likelihood of multivariate lognormal mixture models is given by

$$E_{Z|X} l(\Theta; X, Z) = \sum_{i=1}^{n}\sum_{k=1}^{G} \tau_{ik}[log\pi_k + f(\underline{x}_i; \underline{\mu}_k, \Sigma_k)]$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{G} \tau_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\Sigma_k|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t \Sigma_k^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]]$$

E-step:

The expectation of $l(\Theta; X, Z)$ over $Z|X$ based on current parameter choice Θ^s is $Q(\Theta, \Theta^{(s)})$

$$Q(\Theta, \Theta^{(s)}) = E_{Z|X}[l(\Theta; X, Z); \Theta^{(s)}] \tag{4}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{G} \hat{\tau}_{ik} log\pi_k - \frac{nd}{2}log(2\pi) + \sum_{i=1}^{n}\sum_{k=1}^{G} \hat{\tau}_{ik} log(\underline{x}_i) - \sum_{i=1}^{n}\sum_{k=1}^{G} \frac{\hat{\tau}_{ik}}{2} log|\Sigma_k|^{-1}$$

$$- \sum_{i=1}^{n}\sum_{k=1}^{G} \frac{\hat{\tau}_{ik}}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t \Sigma_k^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]$$

where $\hat{\tau}_{ik}$ is the probability of observation i belonging to the group k based on the current parameter choice $\Theta^{(s)}$. It can be calculated by

$$\hat{\tau}_{ik}^{(s)} = \frac{\pi_k^{(s)} f(\underline{x}_i; \underline{\mu}_k^{(s)}, \Sigma_k^{(s)})}{\sum_{k=1}^{G} \pi_k^{(s)} f(\underline{x}_i; \underline{\mu}_k^{(s)}, \Sigma_k^{(s)})} \tag{5}$$

M-step:

Find the estimate $\hat{\Theta}$, which maximizes $Q(\Theta, \Theta^{(s)})$ for fixed $\Theta^{(s)}$ subject to the equation $\sum_{k=1}^{G} \pi_k = 1$. Using Lagrangian method, we have

$$\Psi = Q(\Theta, \Theta^{(s)}) + \gamma(1 - \sum_{k=1}^{G}) \tag{6}$$

Maximizing the function Ψ with respect to π_j and equation them zero, we get

$$\hat{\pi_j} = \frac{\sum_{i=1}^{n} \hat{\tau}_{ij}}{n}; j = 1, 2, ..., G \tag{7}$$

Maximizing the function $Q(\Theta, \Theta^{(s)})$ with respect to $\underline{\mu}_j$ and equating them zero, we get

$$\frac{\partial(\Theta, \Theta^{(s)})}{\partial \underline{\mu}_j} = 0$$

$$\hat{\underline{\mu}}_j = \frac{\sum_{i=1}^{n} \hat{\tau}_{ij} \underline{l}n(x_i)}{\sum_{i=1}^{n} \hat{\tau}_{ij}} \tag{8}$$

To maximize the function $Q(\Theta, \Theta^{(s)})$ with respect to Σ_j

$$= -\frac{1}{2}[\sum_{i=1}^{n} \hat{\tau}_{ij} log|\Sigma_j| + tr\Sigma_j^{-1} \sum_{i=1}^{n} [(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t]]$$

So, maximizing the function $Q(\Theta, \Theta^{(s)})$ with respect to Σ_j is equivalent to maximizing the above expression with respect to Σ_j. Here, $\hat{\Sigma}_j$ is obtained by using the Lemma 3.2.2 of Anderson (1984) and we get

$$\hat{\underline{\Sigma}}_j = \frac{\sum_{i=1}^{n} \hat{\tau}_{ij} [(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t]}{\hat{\tau}_{ij}} \tag{9}$$

Another important objective of model-based clustering is to study the covariance structures. Fraley *et al.* (1998) have considered different covariance structures for Gaussian mixture models to clustering techniques. Different covariance structures for multivariate lognormal mixture models are developed in the following section.

4. Estimation via Geometric Decomposition

To provide easy and simple interpretable models, Banfield *et al.* (1993) have parameterized the covariance matrices in terms of the eigen-value decompositions for Gaussian mixture models. Fraley *et al.* (1998) considered an eigen-value decomposition of the cluster covariance matrices to provide a wide range of parsimonious covariance structures. Fraley *et al.* (2002) have provided an in-depth discussion of the eigen-value decomposition approach for finite mixture models to clustering. This work is implemented in the MCLUST package. MCLUST package consists of 14 mixture models that arise from the imposition of constraints upon the group of covariance matrix. MCLUST is the most well-established package for model-based clustering technique using Gaussian mixture models. Details of the constraints that can be imposed are summarized in Fraley *et al.* (2003, 2006) which is available in the R software. Fraley *et al.* (2012) summarized the covariance structures available in the MCLUST package, corresponding to geometric characteristics such as shape, volume and orientation. If the number of components is not specified, it assumes that the number of components lies between one to nine. Following this, EM algorithm is implemented corresponding to each initial classification and estimates for parameters are obtained. Then BIC is computed for each resulting mixture model. The model having highest BIC value is identified as the best model.

Browne *et al.* (2014) have pointed out that the covariance technique of Celeux *et al.* (1995) for the EVE and VVE models are computationally infeasible in higher dimensions. They have proposed an alternative algorithm for these two models, based on an accelerated line search on the orthogonal model. Browne *et al.* (2015) have developed another approach, using fast maximization-minimization algorithms, for the EVE and VVE models. This approach is implemented in the mixture packages for R. Several other approaches have been presented, and the excellent review of covariance structures is given by Bouveyron *et al.* (2007).

From the above existing procedures, it is observed that different covariance structures are important for multivariate non-normal mixture models. This paper considers the different covariance structures based on eigen-value decomposition techniques. Let us recall the conditional expectation of the log-likelihood for multivariate lognormal finite mixture models as given in the equation (4).

4.1. The Parsimonious MLN family of models

An eigen-value decomposition of the component covariance matrices is given by

$$\Sigma_k = \lambda_k D_k A_k D_k^t \tag{10}$$

where λ_k is a constant of proportionality, D_k is a orthogonal matrix of eigen vectors and A_k is a orthogonal matrix of eigen vectors and $\det A_k = 1$. Celeux *et al.* (1995) developed eight eigen-value decomposition of a component covariance matrix. The volume of the cluster is determined by λ_k. D_k determines the orientation of the clusters and A_k determines the shape of the density contours. d is the number of dimensions in the datasets. The parsimonious MLN mixture models, herein referred to as PMLN, whose density is given by

$$f(\underline{x}_i|\Theta) \;= \sum_{k=1}^{G} \pi_k f(\underline{x}_i; \underline{\mu}_k, \lambda_k D_k A_k D_k^t) \tag{11}$$

To fit the parsimonious MLN mixture models, EM algorithm is used. The details of parameter estimation methods are like those described in Section 3. To compute $\hat{\Sigma_k}$ To fit the parsimonious MLN mixture models, EM algorithm is used. The details of parameter estimation methods are like those described in Section 3. For the most general MLN family member (VVV model), the complete-data likelihood is given by

$$L(\Theta; X, Z) = \prod_{i=1}^{n}[\prod_{k=1}^{G}[\pi_k]^{z_{ik}}[\prod_{k=1}^{G} f(\underline{x}_i; \underline{\mu}_k, \lambda_k D_k A_k D_k^t)]^{z_{ik}}]] \tag{12}$$

where $f(\underline{x}_i; \underline{\mu}_k, \lambda_k D_k A_k D_k^t)$ is the density of multivariate lognormal distribution with mean vector $\underline{\mu}_k$ and covariance matrix $\Sigma_k = \lambda_k D_k A_k D_k^T$.The conditional expectation of the complete data log-likelihood Q is given by

$$Q(\Theta, \Theta^{(s)}) = E_{Z|X}[l(\Theta; X, Z); \Theta^{(s)}] \tag{13}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{G} \hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\Sigma_k|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t\Sigma_k^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]]$$

The E-step of sth iteration consists of the component membership labels with their conditional expected values is given by

$$\hat{\tau}_{ik}^{(s)} = \frac{\pi_k^{(s)} f(\underline{x}_i; \underline{\mu}_k^{(s)}, \lambda_k D_k A_k D_k^{T(s)})}{\sum_{k=1}^{G} \pi_k^{(s)} f(\underline{x}_i; \underline{\mu}_k^{(s)}, \lambda_k D_k A_k D_k^{T(s)})}$$

To perform the decomposition for MLN mixture models, we follow the procedures outlined in Celeux *et al.* (1995).

Sperical Family

In spherical family, the shape of the clusters is spherical. The shape of the covariance matrix is always diag(1,1). Two spherical families are considered here.

(1) Fitting of EII model ($\Sigma = \lambda I$)

First consider the simplest structure where every component has spherical shape and equal volume. Substitute the $\Sigma_k = \Sigma = \lambda I$ in equation (13). The complete data log-likelihood for the EII model is given by

$$l(\lambda I) = \sum_{i=1}^{n}\sum_{k=1}^{G} \hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\lambda I|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t\lambda I^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]] \tag{14}$$

$$= K - \sum_{i=1}^{n}\sum_{k=1}^{G} \frac{\hat{\tau}_{ik}}{2} log \det \lambda I - \sum_{i=1}^{n}\sum_{k=1}^{G} \frac{\hat{\tau}_{ik}}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^T(\lambda I)^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]$$

$$= \lambda^{-1}\sum_{k=1}^{G} tr(W_k) + dlog\sum_{k=1}^{G}\sum_{i=1}^{n} \hat{\tau}_{ik}$$

$$= \lambda^{-1}\sum_{k=1}^{G} tr(W) + dlog\lambda$$

where K is the constant with respect to model parameters $\underline{\mu}_k$ and λ. Maximizing the equation (13) with respect to λ, we get

$$\hat{\Sigma} = \hat{\lambda} = \frac{tr(W)}{nd} = \frac{\sum_{k=1}^{G}\sum_{i=1}^{n}\hat{\tau}_{ik}[(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t]}{nd}$$

where $n = \sum_{k=1}^{G}\sum_{i=1}^{n}\hat{\tau}_{ik}$

(2) Fitting of VII model ($\Sigma_k = \lambda_k I$)

This is the second simplest model where the component has spherical shape and different volume. Substitute in the equation (14) $\Sigma_k = \lambda_k I$ in equation (13). The complete data log-likelihood for the EII model is given by

$$l(\lambda_k I) = \sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\lambda_k I|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t\lambda_k I^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]] \quad (15)$$

$$= K - \sum_{i=1}^{n}\sum_{k=1}^{G}\frac{\hat{\tau}_{ik}}{2}log\det\lambda_k I - \sum_{i=1}^{n}\sum_{k=1}^{G}\frac{\hat{\tau}_{ik}}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t(\lambda_k I)^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]$$

$$= \lambda^{-1}\sum_{k=1}^{G} tr(W_k) + dlog\sum_{k=1}^{G}\sum_{i=1}^{n}\hat{\tau}_{ik}$$

$$= \lambda^{-1}\sum_{k=1}^{G} tr(W) + d\sum_{k=1}^{G} log\lambda_k\sum_{i=1}^{n}\hat{\tau}_{ik}$$

where K is the constant with respect to model parameters $\underline{\mu}_k$ and λ_k. Maximizing the equation (15) with respect to λ_k, we get

$$\hat{\Sigma_k} = \hat{\lambda}_k = \frac{\sum_{k=1}^{G}\sum_{i=1}^{n}\hat{\tau}_{ik}[(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t]}{\tau_k d}; k = 1, 2, ..., G$$

where $\tau_k = \sum_{i=1}^{n}\hat{\tau}_{ik}$

General Family

(3) Fitting an EVV model ($\Sigma_k = \lambda D_k A_k D_k^T$)

This is generalized model and the component has the same volume but different shape and orientation. Substitute in the equation (13) $\Sigma_k = \lambda D_k A_k D_k^T$ and $C_k = D_k A_k D_k^T; \Sigma_k = \lambda C_k$. The complete data log-likelihood for the EVV model is given by

$$l(\lambda C_k) = \sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\lambda C_k|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t\lambda C_k^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]] \quad (16)$$

$$= K - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}2log|\lambda C_k| + \sum_{k=1}^{G}tr(W_k)(\lambda C_k)^{-1}$$

where K is the constant with respect to model parameters C_k and λ. The equation (16) is maximizing with respect to C_k and λ and equating them zero. We get,

$$\hat{C}_k = \frac{\sum_{i=1}^{n}\hat{\tau}_{ik}(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^T}{|\sum_{i=1}^{n}\hat{\tau}_{ik}(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t|^{\frac{1}{d}}}$$

and

$$\hat{\lambda}_k = \frac{|\sum_{i=1}^{n}\hat{\tau}_{ik}(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t|^{\frac{1}{d}}}{n}$$

$$\hat{\Sigma_k} = \hat{\lambda}\hat{C}_k$$

(4) Fitting an EEE model ($\Sigma_k = \Sigma = \lambda DAD^t$)

This model is a common model for all components and it considers same size, volume and orientation. Substitute in the equation (13) $\Sigma_k = \Sigma = \lambda DAD^t$. The complete data log-likelihood for the EEE model is given by

$$l(\lambda DAD^t) = \sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|(\lambda DAD^t)|^{-1} - \frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t(\lambda DAD^T)^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]] \quad (17)$$

$$= K - \sum_{i=1}^{n}\sum_{k=1}^{G}\frac{\hat{\tau}_{ik}}{2}log|\lambda DAD^T| - \sum_{i=1}^{n}\sum_{k=1}^{G}\frac{\hat{\tau}_{ik}}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t(\lambda DAD^t)^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]$$

$$= K - \frac{1}{2}[tr(W\Sigma^{-1}) + nlog|\Sigma|]$$

where k is the constant with respect to the model parameters $\mu_k, \lambda, DandA$.

$$W = \sum_{k=1}^{G}W_k = \sum_{i=1}^{n}\hat{\tau}_{ik}[(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t)]$$

and

$$n = \sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}$$

EEE model is unconstrained model and it's considered common covariance matrix.

$$\hat{\Sigma_k} = \frac{W}{n} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t}{n}$$

(5) Fitting an VVV model ($\Sigma_k = \lambda_k D_k A_k D_k^t$)

This is the most generalized model. This is the model where every component has different shape, different volume and different orientation. VVV model is the unconstrained model. Substitute in the equation (13) $\Sigma_k = \lambda_k D_k A_k D_k^t$. The complete data log-likelihood for the VVV model is given by

$$l(\lambda_k D_k A_k D_k^t) = \sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}[log\pi_k - \frac{nd}{2}log(2\pi) + log(\underline{x}_i) - \frac{1}{2}log|\lambda_k D_k A_k D_k^t|^{-1} \\ -\frac{1}{2}[(ln(\underline{x}_i) - \underline{\mu}_k)^t(\lambda_k D_k A_k D_k^t)^{-1}(ln(\underline{x}_i) - \underline{\mu}_k)]] \tag{18}$$

$$= K - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{G}\hat{\tau}_{ik}2log|\lambda_k D_k A_k D_k^t| + \sum_{k=1}^{G} tr(W_k)(\lambda_k D_k A_k D_k^t)^{-1}$$

where K is the constant with respect to model parameters μ_k, D_k, A_k and λ_k.

$$\Sigma_k = \frac{\sum_{k=1}^{G} W_k}{n} = \frac{\sum_{i=1}^{n}\hat{\tau}_{ik}[(ln(\underline{x}_i) - \underline{\mu}_k)(ln(\underline{x}_i) - \underline{\mu}_k)^t)]}{\tau_k}; k = 1, 2, ..., G$$

where

$$\tau_k = \sum_{i=1}^{n}\hat{\tau}_{ik}$$

The summary of eigen-value decomposition covariance structures is given in the Table 1.

Table 1: Nomenclature, scale matrix structure and the number of free scale parameters for the eigen-decomposed family of models

Model	λ_k	A_k	D_k	Σ_k	$Number of Covariance Parameters$
EII	Equal	Spherical	-	λI	1
VII	Variable	Spherical	-	$\lambda_k I$	G
EVV	Equal	Variable	Variable	$\lambda D_k A_k D_k^T$	$\frac{Gd(d+1)}{2} - (G-1)d$
EEE	Equal	Equal	Equal	λDAD^T	$\frac{d(d+1)}{2}$
VVV	Variable	Variable	Variable	$\lambda_k D_k A_k D_k^T$	$\frac{Gd(d+1)}{2}$

Covariance Estimation

An alternative estimation method for covariance matrix is presented in this paper. The decomposed elements of the covariance matrix are updated according to the following algorithm. τ_{ik} represents the probability that observation i belongs to group k given the current component parameters

$$n_k = \tau_{ik} = \frac{\pi_k f(\underline{x}_i; \underline{\mu}_k, \Sigma_k)}{\sum_{j=1}^{G}\pi_j f(\underline{x}_j; \underline{\mu}_j, \Sigma_j)}; j \neq k$$

M-step involves the conditionally maximizing the parameters with respect to complete log-likelihood. The estimated mixing proportion and sample cross-product matrix for the kth component is given by

$$\hat{\pi_k} = \frac{n_k}{n}; k = 1, 2, ..., G$$

$$W_k = \sum_{i=1}^{n} n_k(\underline{x}_i - \underline{\mu}_k)(\underline{x}_i - \underline{\mu}_k)^t; k = 1, 2, ..., G$$

1. Iteration $q = 1$
2. Update
$$\lambda_k = \frac{\sum_{k=1}^{G} tr(n_k.W_k)}{nd}$$
where n is the number of observations and d is the dimension.
3. Update
$$A_k = \frac{diag(n_k.W_k)}{|n_k.W_k|^{\frac{1}{d}}}$$
4. Update
$$D_k = n_k W_k a_k$$
Where a_k is the largest eigen value of W_k
5. Update A_k, D_k, λ_k in Σ_k
6. Calculate $E_q = \frac{1}{\lambda} tr(n_k \lambda_k D_k A_k D_k^T + n * dlog(\lambda))$
7. If $t > 1, E_q - E_q - 1 > \epsilon$. If true $t = t + 1$ and return step 2 , or else end.

Five types of covariance structures are considered for finite mixtures of multivariate lognormal distributions to clustering. All covariance models based on eigen-value decomposition structures are used in the M-step of the EM algorithm. The description of the EM algorithm for MLN mixture models is given below.

EM Algorithm

1. Initialization: The initial values of $\pi_k^{(0)}, \underline{mu}_k^{(0)}, \Sigma_k^{(0)}$ are obtained using the algorithm in Section 2.

2. E-step: The conditional Expectation $(\hat{\tau_{ik}}^{(q)})$ of the group membership for each observation is obtained using the equation (5).

3. Mstep: Update the parameters $\hat{\pi_j}^{(q)}$ and $\hat{\mu_j}^{(q)}$ using the formula (7) and (8). Five parsimonious covariance models for MLN mixtures which are derived in the Section 4.1 are updated in the M-step.

4. Compute the log-likelihood $l_j^{(q)}$ and $l_j^{(q+1)}$ and Compare $l_j^{(q+1)}$ and $l_j^{(q)}$. —— $l_j^{(q+1)} - l_j^{(q)}|| < \epsilon$. STOP.

5. E-step and M-step are repeated till the same log-likelihood values are met.

After the convergence is reached, the $(\hat{\tau_{ik}}^{(q)})$ is the posterior probability of component membership for each observation and it is used to cluster the observation into groups. Predicated membership is obtained through Maximum A Posterior probability (MAP).

5. Experimental Results

In this section, the clustering performance of PLMN mixture models is assessed in terms of BIC, AIC, ARI and misclassification rate through simulated as well as real datasets. Numerical comparison of PMLN mixture models have been made with Multivariate Skew Normal (MSN) and Multivariate Normal (MN) mixture models. All numerical computations have been implemented through a program developed in R.

5.1. Simulation Experiment

Here, we consider a finite mixture of multivariate Lognormal distribution with three components. Random sample of size n = 262, 270 and 268 are simulated with parameters $\mu_1 = (0.29, 0.685)$, $\mu_2 = (1.68, 0.69)$, $\mu_3 = (0.88, 1.71)$ with same covariance matrix $\Sigma = \begin{bmatrix} 0.1986 & 0.8876 \\ 0.8876 & 0.8876 \end{bmatrix}$. The mean vectors and covariance matrix are generated from the clusterGeneration package which is available in R. Figure 1 displays the scatter plot of the simulated dataset.

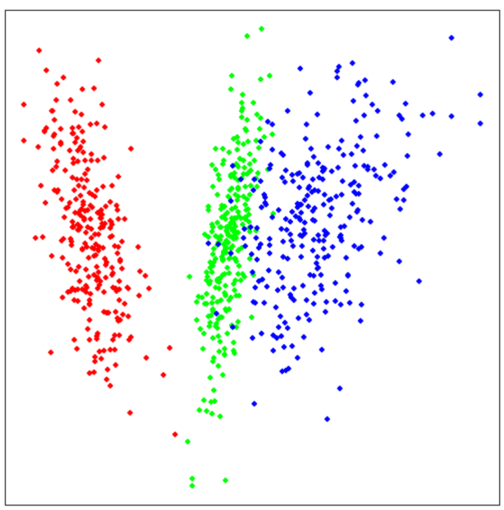

Figure 1: Scatter plot for simulated data

The initial component parameter values are obtained using the algorithm in Section 2. All the covariance models are initiated with the same initial values of the component parameters. The initial values are obtained iteratively till the same cluster membership labels are met. From the cluster membership labels, the initial mixing proportion, initial mean vector and initial covariance matrix are calculated. The initial values are presented in Table 2.

Table 2: Initial parameter values of three-component MLN mixture models

Component 1	$\pi_1 = 0.5211$	$\mu_1 = (0.29, 0.685)$	$\Sigma_1 = \begin{bmatrix} 0.1986 & 0.8876 \\ 0.8876 & 0.8876 \end{bmatrix}$
Component 2	$\pi_2 = 0.238$	$\mu_2 = (2.27, 0.57)$	$\Sigma_2 = \begin{bmatrix} 0.5392 & 1.0275 \\ 1.0275 & 6.2978 \end{bmatrix}$
Component 3	$\pi_3 = 0.2409$	$\mu_3 = (1.24, 2.65)$	$\Sigma_3 = \begin{bmatrix} 0.9786 & 2.9376 \\ 2.9376 & 9.2136 \end{bmatrix}$

Table 3: Clustering performance of various multivariate mixture models

Distributions	Model	BIC	AIC	MR	ARI	Log likelihood
MLN	EII	3380.15	3279.15	0.10	0.7169	-1523.851
MLN	VII	3256.14	3126.86	0.11	0.7328	-1503.132
MLN	EEE	3178.23	2814.25	0.09	0.8354	-1523.57
MLN	EVV	3445.76	3437.52	0.04	0.8369	-1529.57
MLN	VVV	3045.28	3012.19	0.07	0.7425	-1496.09
MSN	EEV	3389.461	3145.58	0.155	0.8269	-1467.04
MN	EEE	3193.09	3436.29	0.133	0.7932	-1498.96

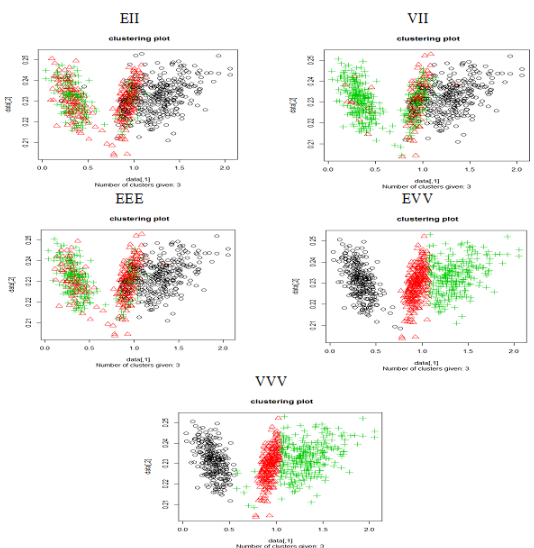

Figure 2: Scatter plot for five MLN mixture models

Different covariance structures in multivariate lognormal mixture models are considered. The clustering results of the simulated dataset are provided in Table 3. From Table 3, it is observed that EVV model gives lowest misclassification rate (0.04). The ARI is 83 % with BIC 3445.76 and AIC 3437.52. Among five covariance structures of MLN mixture models, EVV model achieved the highest ARI. The best model (EVV) is compared with other multivariate mixture models. The ARI value for MLN mixture model ranges from 0.71 to 0.83 which indicates that the dataset is classified with greater precision. EEV model gives better clustering performance for multivariate skew normal mixture models and EEE

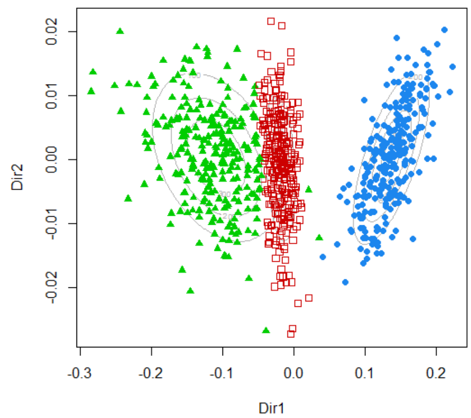

Figure 3: Contour plot for the EVV model

model provides better clustering results for multivariate normal mixture models. The results of both model are shown in Table 3. Table 4 provides the estimated parameter values of EVV model in case of MLN mixture models.

Table 4: Estimated parameter values of three-component MLN mixture (EVV) model

component 1	$\pi_1 = 0.5901$	$\mu_1 = (1.07, 0.974)^t$	$\Sigma_1 = \begin{bmatrix} 0.09379 & 0.9396 \\ 0.9396 & 4.2789 \end{bmatrix}$
Component 2	$\pi_2 = 0.111$	$\mu_2 = (2.005, 0.772)^t$	$\Sigma_2 = \begin{bmatrix} 0.3327 & 2.0235 \\ 2.0235 & 5.1936 \end{bmatrix}$
Component 3	$\pi_3 = 0.3989$	$\mu_3 = (1.984, 2.728)^t$	$\Sigma_3 = \begin{bmatrix} 0.9726 & 2.9506 \\ 2.9506 & 8.9349 \end{bmatrix}$

From the table, correctly classified samples are presented here. That is, Almost 81% of samples are correctly classified for all models of MLN mixture models. The best model for MLN mixture gives 95% correct classification of the simulated dataset. For multivariate skew normal mixture, EEV model achieved 85 correct classification. Multivariate normal mixture models EEE model gives 87% correct classification. Figure 2 depicts the estimation of the cluster memberships into three clusters for the five models. In these figures, the clusters are indicated by three different characters (+, o and D). The volume of the five models is:

i) $\lambda I : \lambda = 0.2996$

ii) $\lambda_k I : \lambda_1 = 0.983, \lambda_2 = 2.371, \lambda_3 = 0.693$

iii) $\lambda DAD^t : \lambda = 3.2996$

iv) $\lambda D_k A_k D_k^t : \lambda = 5.2996$

v) $\lambda_k D_k A_k D_k^t : \lambda_1 = 1.283, \lambda_2 = 3.591 and \lambda_3 = 7.753$

The contour plot of the best (EVV) model is shown in Figure 3. The contour plot shows the different volume, size and orientation of the three clusters. The best fitted model is selected based on BIC and AIC value. It is also noticed that from the simulated dataset, general models perform better than spherical models.

5.2. Real Data (Old Faithful Dataset)

In this section, old faithful dataset is used for the PMLN mixture models. This dataset contains two variables (eruptions and waiting) and 275 observations. It is a bivariate dataset measuring the length of eruption and time to eruption, both variables are in millimeters. This dataset is available in R software. Many researchers have analyzed this dataset for model-based clustering approach. This dataset does not have true class labels. The original plot of the faithful dataset is shown in Figure4, where the observations are displayed into two clusters very clearly.

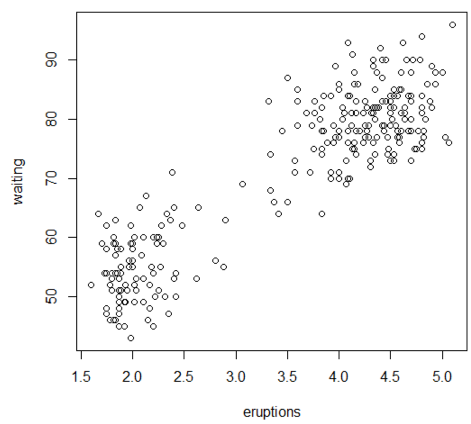

Figure 4: The bivariate Old faithful dataset

Table 5: Initial parameter values of two-component faithful dataset

Component 1	$\pi_1 = 0.5389$	$\mu_1 = (3.457, 70.794)^t$	$\Sigma_1 = \begin{bmatrix} 1.3899 & 14.3525 \\ 14.3525 & 182.461 \end{bmatrix}$
Component 2	$\pi_2 = 0.4611$	$\mu_2 = (3.518, 71)^t$	$\Sigma_2 = \begin{bmatrix} 1.2232 & 13.7003 \\ 13.7003 & 188.5333 \end{bmatrix}$

We compare the clustering performance of MLN, MSN and MN mixture models. The initial values of component parameters are calculated based on the algorithm as given in Section 2. Initial values of faithful datasets are presented in the Table 5.

For MSN and MN mixture models the best results are given in Table 6. The classification plot of each model for MLN mixture models are displayed in Figure 5. The clusters are

Table 6: Clustering performance of various multivariate mixture models

Distributions	Model	BIC	AIC	Log likelihood
MLN	EII	1876.912	1844.562	-796.53
MLN	VII	1887.072	1854.825	-769.94
MLN	EEE	1889.649	1883.544	-868.27
MLN	EVV	1825.195	1852.052	-893.82
MLN	VVV	1895.839	1869.302	-788.28
MSN	EVV	1892.361	1825.427	-834.25
MN	VVV	2371.702	2148.597	-919.29

Table 7: Estimated parameter values of two-component faithful dataset

Component 1	$\pi_1 = 0.653$	$\mu_1 = (3.093, 71.814)^t$	$\Sigma_1 = \begin{bmatrix} 1.2903 & 14.1739 \\ 14.1739 & 181.281 \end{bmatrix}$
Component 2	$\pi_2 = 0.347$	$\mu_2 = (2.948, 70.542)^t$	$\Sigma_2 = \begin{bmatrix} 1.2232 & 13.8103 \\ 13.8103 & 187.4933 \end{bmatrix}$

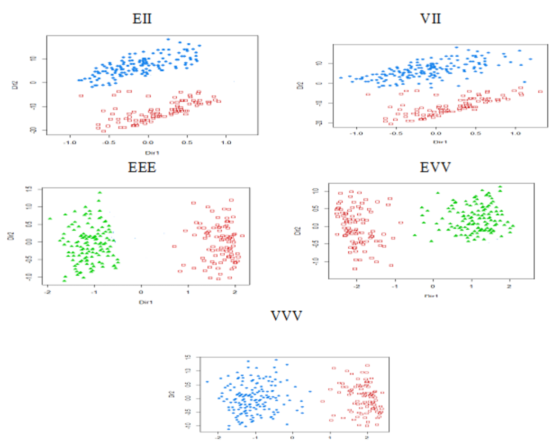

Figure 5: Scatter plot for five models using multivariate lognormal mixture models

represented by different symbols. VVV model gives good clustering results for multivariate normal mixture models. The parsimonious family of multivariate lognormal distributions shows that the clusters have different volume and size. The contour plot in Figure 6 shows different volume and size of clusters. Estimated parameters of VVV models for MLN mixtures are given in the Table 7.

The number of observations in each cluster for MLN, MSN and MN mixture models are presented in Table 8. The volume of the clusters is given below:

i) $\lambda I : \lambda = 197.17$

ii) $\lambda_k I : \lambda_1 = 180, \lambda_2 = 69$

iii) $\lambda DAD^t : \lambda = 109.26$

Table 8: Clustering table for multivariate mixture models

	MLN Mixture Model					MSN	MN
Clusters	EII	VII	EEE	EVV	VVV	EVV	VVV
Cluster 1	177	99	174	170	178	175	168
Cluster 2	95	173	98	102	94	97	104

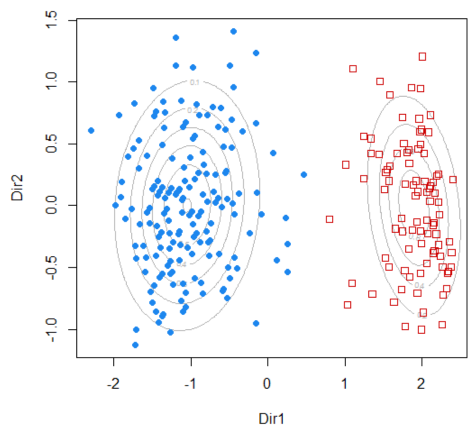

Figure 6: Contour plot for VVV model for multivariate lognormal mixture model

iv) $\lambda D_k A_k D_k^t : \lambda = 166.1296$

v) $\lambda_k D_k A_k D_k^t : \lambda_1 = 170, \lambda_2 = 110$

6. Conclusion

In this paper, a family of parsimonious MLN mixture models is introduced through an eigen-value decomposition of the components covariance matrix. From simulation experiments, the general (EVV) covariance model provides best clustering results than spherical models. The results of real dataset showed that all covariance model gives better clustering results according to BIC and AIC criteria. Proposed initialization techniques plays important role, because it gives reliable and true estimated parameter values for components. It is noticed that among general covariance models from numerical experiments, VVV gives good clustering results. VVV model allows with different size, volume, and orientation. Some parsimonious models give good clustering results, because those covariance models are close to the structure of the data.

References

Adrian, O. H., Thomas, B. M., Bredan, M., Isobel, C. G., Paul, D. M. and Dimitris, K. (2016). Clustering with the Multivariate Normal Inverse Gaussian Distribution. *Computational Statistics and Data Analysis*, **93**, 18–30.

Banfield, J. D. and Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian Clustering. *Biometrics*, **49(3)**, 803–821.

Baek, J. and McLachlan, G. J. (2010). Mixtures of factor analyzers with common factor loadings: applications to the clustering and visualization of high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **32(7)**, 1298–1309.

Bouveyron, C., Girard, S. and Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics and Data Analysis*, **52(1)**, 502–519.

Browne, R. P. and McNicholas, P. D. (2014). Orthogonal Stiefel manifold optimization for eigen-decomposed covariance parameter estimation in mixture models. *Statistics and Computing*, **24(2)**, 203–210.

Browne, R. P. and McNicholas, P. D. (2015). A mixture of generalized hyperbolic distributions. *Canadian Journal of Statistics*, **43(2)**, 176–198.

Bouveyron, C. and Brunet, C. (2012). Simultaneous model-based clustering and visualization in the Fisher discriminative subspace. *Statistics and Computing*, **22(1)**, 301–324.

Celeux, G. and Govaert, G. (1995). Gaussian parsimonious clustering model. *Pattern Recognition*, **28(5)**, 781–793.

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

Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall, London.

Fraley, C. and Raftery, A. E. (1998). How many clusters? Which clustering methods? Answers via model-based cluster analysis. *The Computer Journal*, **41(8)**, 578–588.

Fraley, C. and Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97(458)**, 611–631.

Fraley, C. and Raftery, A. E. (2003). Enhanced software for model-based clustering, density estimation, discriminant analysis: MCLUST. *Journal of Classification*, **20(2)**, 263–286.

Fraley, C. and Raftery, A. E. (2006). MCLUST Version 3 for R: Normal Mixture Modeling and Model-Clustering. *Technical Report 504, Department of Statistics, University of Washington*, First Published September 2006, Minor revisions January 2007 and November.

Fraley, C., Raftery, A. E., Murphy, T. B. and Scrucca, L. (2012). MCLUST Version 4 for R: Normal Mixture Modeling and Model-Clustering. *Technical Report 597*, Department of Statistics, University of Washington.

Schnatter, F. S. and Pyne, S. (2010). Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions, *Biostatistics*, **11**, 317—336.

Karlis, D. (2002). An EM type algorithm for maximum likelihood estimation of the normal-inverse Gaussian distribution. *Statistics and Probability and Letters*, **57(1)**, 43-–52.

Lin, T. I., Lee, J. C. and Yen, S. Y. (2007). Finite mixture modeling using skew normal distribution. *Statistica Sinica*, **17(3)**, 909–927.

Lee, X. L. and McLachlan, G. J. (2013a). On mixtures of skew normal and skew t-distributions. *Advances in Data Analysis and Classification*, **7(3)**, 241–266.

Lee, X. L. and McLachlan, G. J. (2013b). Model-based clustering and classification with non-normal mixture distributions. *Statistical Methods and Applications*, **22(4)**, 427–454.

Lee, X. L. and McLachlan, G. J. (2014). Finite mixtures of multivariate skew t- distributions: some recent and new results. *Statistics and Computing*, **24(2)**, 181–202

Melnykov, V. and Xuwen, Z. (2018). On Model-based clustering of skewed matrix data. *Journal of Multivariate Analysis*, **167(c)**, 181–194.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture Models: Inference and Applications.* Marcel Dekker, New York.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models.* John Wiley and Sons, Inc, New York.

McNicholas, P. D. and Murphy, T. B. (2008). Parsimonious Gaussian mixture models. *Statistics and Computing*, **18(3)**, 285–296.

Melnykov, V. and Semhar, M. (2016). Studying complexity of model-based clustering. *Communications in Statistics-Simulation and Computation*, **45(6)**, 2051–2069.

Pyne S., Hu X., Wang K., Rossin E., Lin, T. I., Maier et.al. (2009). *Automated High-Dimensional Flow Cytometric Data Analysis.* Proceedings of the National Academy of Sciences, USA.

Sanjeena, S. and McNicholas, P. D. (2014). Variational bayes approximations for clustering via mixtures of normal inverse gaussian distributions. *Advances in Data Analysis and Classification*, **8**, 167-193.

Soltyk, S. and Gupta, R. (2011). Application of the Multivariate Skew Normal Mixture Model with the EM Algorithm to Value-at-Risk. *MODSIM 2011 - 19th International Congress on Modelling and Simulation*, Perth, Australia, December 12-16, 2011.

Titterington, D., Smith, A. and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions.* John Wiley and Sons, New York.