JIRSS (2017)
Vol. 16, No. 2, pp 97-110
DOI:10.22034/jirss.2017.16.06

Skew Laplace Finite Mixture Modelling

Kheirolah Okhli[1], **Mahdieh Mozafari** [2], **Mehrdad Naderi**[3]

[1]Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.
[2]Department of Statistics, Faculty of Mathematics and Computing, Higher Education Complex of Bam, Bam, Iran.
[3]Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

Received: 01/01/2017, Revision received: 22/04/2017, Published online: 27/11/2017

Abstract. This paper presents a new mixture model via considering the univariate skew Laplace distribution. The new model can handle both heavy tails and skewness and is multimodal. Describing some properties of the proposed model, we present a feasible EM algorithm for iteratively computing maximum likelihood estimates. We also derive the observed information matrix for obtaining the asymptotic standard error of parameter estimates. The finite sample properties of the obtained estimators as well as the consistency of the associated standard error of parameter estimates are investigated by a simulation study. We also demonstrate the flexibility and usefulness of the new model by analyzing real data example.

Keywords. EM algorithm, Finite mixture model, Mean-variance mixture distribution, Skew laplace distribution.

MSC: 62E10; 62E15.

Kheirolah Okhli (okhli.kh@mail.um.ac.ir)
Mahdieh Mozafari (mozafari@bam.ac.ir)
Corresponding Author: Mehrdad Naderi (mehrdad.naderi@ymail.com)

1 Introduction

Finite mixture (FM) models are analytical paradigm aiming to provide a statistical tool for analyzing data sets which come from a heterogeneous population (McLachlan and Basford, 1988; Titterington *et al.*, 1985). Formally, the FM models are weighted sum of g combination of distribution functions and, therefore, they are useful in a cluster or classification analysis and pattern recognition problems. Let X be a random variable having a finite mixture distribution. Then, the probability density function (pdf) of X can be formulated by

$$f(x;\boldsymbol{\Theta}) = \sum_{i=1}^{g} p_i\, f(x;\boldsymbol{\theta}_i), \qquad x \in \mathbb{R},$$

where p_is′ are mixing proportions, such that $\Sigma_{i=1}^{g} p_i = 1$, $f(.;\boldsymbol{\theta}_i)$ represents the pdf of class i, parametrized with $\boldsymbol{\theta}_i$, and $\boldsymbol{\Theta} = (p_1, ..., p_{g-1}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_g)$. More details of FM models can be found in McLachlan and Peel (2004), Bishop (2006) and Frühwirth-Schnatter (2006).

Due to usefulness of FM models in the study of heterogeneity and clustering, many kinds of mixture models have been recently introduced by considering non-Gaussian distribution as a mixing component. For instance, see the work on the multivariate t distribution considered by Lin *et al.* (2014), the skew-t distribution proposed by Vrbik and McNicholas (2012), Vrbik and McNicholas (2014) and Lee and McLachlan (2014). Also, Basso *et al.* (2010), Doğru *et al.* (2017) and Ho *et al.* (2012) provide considerable works via proposing FM models of the scale mixtures of skew-normal distributions, skew Laplace normal distributions and skew-t-normal distributions, respectively.

Another viewpoint of constructing skewed distribution is the normal mean-variance mixture (NMV) method. Because of asymmetric properties, the class of NMV distributions was considered in several works in the literature of clustering. Moreover, McNeil *et al.* (1977) showed that the family of generelized hyperbolic (GH) distributions can be constructed by the NMV model. The GH family includes symmetric distributions such as normal, Student t, Laplace, elliptically symmetric distributions and asymmetrical distributions as skewed t, skew Laplace and hyperbolic distributions as special cases. Arslan (2010) proposed an alternative skew Laplace (SL) distribution and showed that all maximum likelihood (ML) estimators of the parameters in this family have an explicit form and it provides better fit on the peaked data set. Recently, Franczak *et al.* (2014) have proposed finite mixture of shifted asymmetric Laplace distributions. Although the shifted asymmetric Laplace distribution is also constructed via the NVM model (considering exponential distribution with mean 1 as a mixing variable), its pdf is not same as the SL distribution.

In this paper, we propose a finite mixture version of the skew Laplace (FM-SL) model. Some properties of the new model are studied and the ML estimate of model parameters are obtained. To show asymptotic properties of the ML estimators, a simulation study is used. Finally, we fit the FM-SL model on a real data set and compare this model with the FM model based on the family of the scale mixture of skew normal distributions.

The rest of the paper is organized as follows. Section 2 briefly outlines the skew Laplace distribution and studies some of its properties. Section 3 describes the finite mixture of SL distributions, and then mentions some of its properties including the implementation of expectation maximization (EM) procedure, introduced by Dempster *et al.* (1977), for computing ML estimation. A simulation study is conducted to verify consistency of estimated standard errors and finite sample properties of the ML estimates in Section 4. Furthermore, the performance and advantages of the proposed methodology is illustrated through the analysis of a real data set in this section. The paper concludes with a summary and a short discussion on some possible directions for future researches in Section 5.

2 Methodology

2.1 Skew Laplace Distribution

Let X be a random variable taking real values. It follows a skew Laplace distribution, denoted by $X \sim SL(\mu, \sigma, \lambda)$, if its pdf is of form

$$f_{SL}(x;\mu,\sigma,\lambda) = \frac{1}{2\tau\sigma}\exp\left\{-\tau\frac{|x-\mu|}{\sigma} + \frac{\lambda(x-\mu)}{\sigma^2}\right\}, \qquad x \in \mathbb{R}, \tag{2.1}$$

where $\sigma > 0$, $\mu \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ are the scale, location and skewness parameters, respectively, and $\tau = \sqrt{1 + (\lambda/\sigma)^2}$. It can be easily shown that the SL distribution is obtained based on the normal mean-variance mixture model as

$$X \stackrel{d}{=} \mu + W\lambda + \sqrt{W}Z, \tag{2.2}$$

where $\stackrel{d}{=}$ denotes the equality in distribution, Z is a random variable that follows a normal distribution with mean zero and variance σ^2 ($Z \sim N(0, \sigma^2)$), and W is a random variable having an exponential distribution with mean 2 ($Exp(0.5)$).

Proposition 2.1. *If $X \sim SL(\mu,\sigma,\lambda)$, then the tail of the distribution is*

$$f_{SL}(x;\mu,\sigma,\lambda) \sim c\exp\Big\{-\frac{\tau}{\sigma}|x| + \frac{\lambda}{\sigma^2}x\Big\}, \quad as \quad x \longrightarrow \pm\infty,$$

where c is a constant. Also, if $\lambda > 0$, the right tail gets heavier and the heavier right tail degrades as

$$f_{SL}(x;\mu,\sigma,\lambda) \sim c\exp\Big\{\Big(\frac{\tau}{\sigma^2} - \frac{\lambda}{\sigma}\Big)x\Big\}, \quad as \quad x \longrightarrow +\infty.$$

Similarly, the heavier left tail degrades as

$$f_{SL}(x;\mu,\sigma,\lambda) \sim c\exp\Big\{\Big(\frac{\tau}{\sigma^2} + \frac{\lambda}{\sigma}\Big)x\Big\}, \quad as \quad x \longrightarrow -\infty \ \ if \ \ \lambda < 0.$$

The aforementioned proposition reveals that the tail of the SL distribution is an exponential function and it can be regarded as a semi-heavy-tailed distribution.

Using the definition of the generalized inverse Gaussian (GIG) distribution, we derive a proposition which is helpful for applying the EM algorithm. A random variable W with the pdf

$$f_{GIG}(w;\kappa,\chi,\psi) = \frac{(\psi/\chi)^{\kappa/2}w^{\kappa-1}}{2K_\kappa(\sqrt{\psi\chi})}\exp\Big\{\frac{-1}{2}\Big(w^{-1}\chi + w\psi\Big)\Big\}, \ \ w > 0, \tag{2.3}$$

has a GIG distribution, denoted by $W \sim GIG(\kappa,\chi,\psi)$ hereafter. In this pdf, the parameter $\kappa \in \mathbb{R}$ and two parameters χ and ψ are such that $\chi \geq 0, \psi > 0 \ if \ \kappa > 0$; $\psi \geq 0, \chi > 0 \ if \ \kappa < 0$ and $\chi > 0, \psi > 0 \ if \ \kappa = 0$. The GIG distribution was originally introduced by Good (1953) and is widely used for modeling and analyzing lifetime data. The pdf of GIG distribution is unimodal and contains the gamma and inverse gamma densities as special cases, corresponding to $\chi = 0$ and $\psi = 0$, respectively. In these cases, (2.3) must be interpreted as limiting cases, which can be evaluated using the asymptotic relations $K_\kappa(x) \sim \Gamma(\kappa)2^{\kappa-1}x^{-\kappa}$ as $x \downarrow 0$ for $\kappa > 0$, and $K_\kappa(x) \sim \Gamma(-\kappa)2^{-\kappa-1}x^{\kappa}$ as $x \downarrow 0$ for $\kappa < 0$.

Proposition 2.2. *Let random variables X and W follow $SL(\mu,\sigma,\lambda)$ and $Exp(0.5)$, respectively. Then by the Bayes' rule, for any $x \in \mathbb{R}$, distribution of W given $X = x$ is GIG distribution with parameters $(0.5, ((x-\mu)/\sigma)^2, \tau^2)$. Also,*

$$E[W^n|X = x] = \Big(\frac{x-\mu}{\sigma\tau}\Big)^n R_{(0.5,n)}(\tau(x-\mu)/\sigma), \qquad n = \pm 1, \pm 2, ..., \tag{2.4}$$

where $R_{(\kappa,a)}(c) = K_{\kappa+a}(c)\Big/K_\kappa(c)$.

3 Finite Mixture of Skew Laplace Distributions

This section displays the structure of the FM-SL model and demonstrates the way of employing the EM algorithm to find the ML estimate of parameters.

3.1 Model Formulation and Parameter Estimation

Consider n independent random variables $X_1, ..., X_n$, which are taken from a mixture of SL distributions. The pdf of a g-component FM-SL model is

$$f(x; \boldsymbol{\Theta}) = \sum_{i=1}^{g} p_i \, f_{SL}(x; \boldsymbol{\theta}_i), \qquad j = 1, 2..., n, \tag{3.1}$$

where p_is′ are mixing proportions, $f(.; \boldsymbol{\theta}_i)$ is the SL pdf defined in Equation (2.1) with $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \lambda_i)$ and $\boldsymbol{\Theta} = (p_1, ..., p_{g-1}, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_g)$. Given the set of observed data $\boldsymbol{x} = (x_1, ..., x_n)$, the observed data log-likelihood function for $\boldsymbol{x}$ is

$$\ell(\boldsymbol{\Theta}|\mathbf{X} = \mathbf{x}) = \sum_{j=1}^{n} log\left(\sum_{i=1}^{g} p_i \, f_{SL}(x_j; \boldsymbol{\theta}_i)\right). \tag{3.2}$$

Remark 1. The main problem occurred in the log-likelihood function (3.2) is that it is unbounded when one of the observations x_j equals μ_i and σ_i tends to 0, since the FM-SL is related to the normal mixture model with unequal variances via representation (2.2). As a result, the corresponding global ML estimator of FM-SL is undefined. Following Chen *et al.* (2008), one way to solve this problem is to put a constraint on the parameter space such that the likelihood is bounded. It is also noted that identifying an appropriate parameter space is difficult. Moreover, in many cases, different choices of the parameter space may give different constrained global ML estimators. To overcome this deficiency more accurately, Yao (2010) suggested to find the constraint parameter via proposing a profile log-likelihood of the likelihood function. More details of the profile log-likelihood method can be found in Yao (2010).

A useful and accurate approach to obtain the ML estimate in the FM model is the EM algorithm. The EM algorithm is an iterative way to facilitate the producer of obtaining ML estimations when data are incomplete. The main idea for getting ML estimators in this algorithm is to solve a difficult incomplete-data problem by solving tractable complete data problems. More details on the EM algorithm are given by McLachlan and Krishnan (2008). To apply this approach to our proposed model, on supporting the interpretation of incomplete data, we consider a set of group membership variables

$\mathbf{V}_j = (V_{1j}, ..., V_{gj})$ for $j = 1, ..., n$, taking $V_{ij} = 1$ if y_j belongs to the ith component and $V_{ij} = 0$ otherwise. We assume that all of the group membership labels are unknown in a clustering paradigm. This implies that $\mathbf{V}_j$ follows a multinomial distribution with probabilities $(p_1, ..., p_g)$ and one trial, denoted by $\mathbf{V}_j \sim M(1; p_1, ..., p_g)$. By (2.2), the hierarchical representation of the FM-SL model can be written as

$$\begin{aligned} X_j\Big|(W_j = w_j, V_{ij} = 1) &\sim N(\mu_i + w_j\lambda_i, w_j\sigma_i^2), \\ W_j\Big|V_{ij} = 1 &\sim Exp(\frac{1}{2}), \\ \mathbf{V}_j &\sim M(1, p_1, p_2, ..., p_g). \end{aligned}$$

As a result, based on the observed data $\boldsymbol{x}$ and missing values $\boldsymbol{w} = (w_1, ..., w_n)$ and $\boldsymbol{V} = (\boldsymbol{V}_1, ..., \boldsymbol{V}_n)$, the log-likelihood of the complete data, with safely dropping out additive constant, is

$$\ell_c(\boldsymbol{\Theta}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{V}) = \sum_{j=1}^{n}\sum_{i=1}^{g} v_{ij}\Big[\log p_i - \frac{1}{2}(\log\sigma_i^2) - \frac{(x_j - \mu_i)^2}{2\sigma_i^2 w_j} - \frac{\lambda_i^2 w_j}{2\sigma_i^2} + \frac{\lambda_i(x_j - \mu_i)}{\sigma_i^2}\Big]. \quad (3.3)$$

At the iteration k of the EM algorithm, the expected value of complete-data log-likelihood (3.3) with respect to the conditional distribution of the missing values $(\boldsymbol{w}, \boldsymbol{V})$ given the observed data $\boldsymbol{x}$ is computed in the E-step at $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}^{(k)}$. This leads to the so-called Q-function as

$$Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) = \sum_{j=1}^{n}\sum_{i=1}^{g} \hat{v}_{ij}^{(k)}\Big[\log p_i - \frac{1}{2}\log(\sigma_i^2) - \frac{(x_j - \mu_i)^2}{2\sigma_i^2}\hat{t}_{ij}^{(k)} - \frac{\hat{w}_{ij}^{(k)}\lambda_i^2}{2\sigma_i^2} + \frac{\lambda_i(x_j - \mu_i)^2}{\sigma_i^2}\Big].$$

The conditional expectations $\hat{w}_{ij}^{(k)} = E(W_j|X = x_j, \boldsymbol{\theta}_i)$ and $\hat{t}_{ij}^{(k)} = E(W_j^{-1}|X = x_j, \boldsymbol{\theta}_i)$ are calculated by using (2.4) in Proposition 2.2. $\hat{v}_{ij}^{(k)}$ is also computed using $\hat{p}_i f(x_j; \hat{\boldsymbol{\theta}}_i^{(k)})\Big/f(x_j; \hat{\boldsymbol{\Theta}}^{(k)})$.

Let $n_i = \sum_{j=1}^{n}\hat{v}_{ij}^{(k)}$, $A_i = \sum_{j=1}^{n}\hat{v}_{ij}^{(k)}\hat{t}_{ij}^{(k)}$ and $C_i = \sum_{j=1}^{n}\hat{w}_{ij}^{(k)}\hat{v}_{ij}^{(k)}$. In the M-step, we maximize the Q-function with respect to the parameters to get the updates of the parameter estimates. The update of the mixing proportions is $\hat{p}_i^{(k+1)} = n_i/n$ and two parameters λ and μ can be updated as

$$\hat{\lambda}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{v}_{ij}^{(k)}(x_j - \hat{\mu}_i^{(k)})}{C_i}, \text{ and } \hat{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n}\hat{v}_{ij}^{(k)}(\hat{t}_{ij}^{(k)}x_j - \hat{\lambda}_i^{(k+1)})}{A_i}.$$

The update of the parameter σ_i^2 is also obtained by

$$\hat{\sigma}_i^{2(k+1)} = \frac{1}{n_i}\left[\sum_{j=1}^{n} \hat{v}_{ij}^{(k)}\Big(\hat{t}_{ij}^{(k)}(x_j - \hat{\mu}_i^{(k+1)})^2 + \hat{\lambda}_i^{2(k+1)}\hat{w}_{ij}^{(k)} - 2\hat{\lambda}_i^{(k+1)}(x_j - \hat{\mu}_i^{(k+1)})\Big)\right].$$

The above two steps are iterated until a suitable specified convergence rule is satisfied, for instance, if $|\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)})|$ is less than a desired tolerance. In this paper, we use the Aitken acceleration (Aitken, 1927) to estimate the asymptotic maximum of the log-likelihood at each iteration of the EM algorithm and hence to determine convergence of the algorithm (McNicholas *et al.*, 2010). The Aitken acceleration factor at iteration k is

$$aik^{(k)} = \frac{\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)})}{\ell(\hat{\boldsymbol{\theta}}^{(k)}) - \ell(\hat{\boldsymbol{\theta}}^{(k-1)})}.$$

At iteration $k+1$, the asymptotic estimate of the log-likelihood can be computed as

$$\ell_\infty(\hat{\boldsymbol{\theta}}^{(k+1)}) = \ell(\hat{\boldsymbol{\theta}}^{(k+1)}) + \frac{1}{1 - aik^{(k)}}\Big\{\ell(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)})\Big\},$$

and the procedure can be considered to have converged when $\ell_\infty(\hat{\boldsymbol{\theta}}^{(k+1)}) - \ell(\hat{\boldsymbol{\theta}}^{(k)}) < \varepsilon$ (Böhning *et al.*, 1994; Lindsay, 1995). We use this criterion with $\varepsilon = 10^{-5}$ in the next section.

Since the choice of starting points in the EM algorithm plays an important role to achieve a closer estimation of the real parameter value, a simple procedure suggested by Basso *et al.* (2010) is used. At first, we separate the sample into the g groups using cluster algorithm, namely the k-means. Then, the initial value of mixing proportions are equal to the proportion of data points belonging to the same cluster j, and the starting points of μ_j, σ_j and λ_j for each cluster j are computed using the estimation method of FM-SL, described above, for g = 1.

3.2 Estimation of Standard Errors

To compute the asymptotic covariance of the ML estimates, we use the information-based method suggested by Basford *et al.* (1997). By Meilijson's (Meilijson, 1989) formula, the empirical information matrix is obtained as

$$I_e(\boldsymbol{\theta}|x) = \sum_{j=1}^{n} \mathbf{s}(x_j|\boldsymbol{\theta})\mathbf{s}^\top(x_j|\boldsymbol{\theta}) - \frac{1}{n}\mathbf{S}(x|\boldsymbol{\theta})\mathbf{S}^\top(x|\boldsymbol{\theta}), \tag{3.4}$$

where $\mathbf{S}(x|\boldsymbol{\theta}) = \sum_{j=1}^{n} \mathbf{s}(x_j|\boldsymbol{\theta})$ and the individual score $\mathbf{s}(x|\boldsymbol{\theta})$ can be determined (Louis, 1982) as

$$\mathbf{s}(x_j|\boldsymbol{\theta}) = \frac{\partial f(x_j|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E\Big(\frac{\partial \ell_c(\boldsymbol{\theta}|x_j, w_j)}{\partial \boldsymbol{\theta}}\Big| x_j, \boldsymbol{\theta}\Big).$$

Substituting the ML estimates ($\hat{\boldsymbol{\theta}}$), (3.4) reduces to

$$I_e(\boldsymbol{\theta}|X) = \sum_{j=1}^{n} \hat{\mathbf{s}}_j \hat{\mathbf{s}}_j^{\top}, \tag{3.5}$$

where $\hat{\mathbf{s}}_j$ is an individual score vector whose elements are $(\hat{\mathbf{s}}_{j,p_1}, ..., \hat{\mathbf{s}}_{j,p_{g-1}}, \hat{\mathbf{s}}_{j,\mu_1}, ..., \hat{\mathbf{s}}_{j,\mu_g}, \hat{\mathbf{s}}_{j,\sigma_1}, ..., \hat{\mathbf{s}}_{j,\sigma_g} \hat{\mathbf{s}}_{j,\lambda_1}, ..., \hat{\mathbf{s}}_{j,\lambda_g})$ which, for $r = 1, ..., g$ can be computed as follows.

$$\begin{aligned}
\hat{\mathbf{s}}_{j,p_r} =& \frac{\hat{v}_{rj}}{\hat{p}_r} - \frac{\hat{v}_{gj}}{\hat{p}_g}, \\
\hat{\mathbf{s}}_{j,\mu_r} =& \hat{v}_{rj}\hat{\sigma}_r^{-2}\Big[(x_j - \hat{\mu}_r)\hat{t}_{rj} - \hat{\lambda}_r\Big], \\
\hat{\mathbf{s}}_{j,\lambda_r} =& \hat{v}_{rj}\hat{\sigma}_r^{-2}\Big[(x_j - \hat{\mu}_r) - \hat{\lambda}_r\hat{w}_{rj}\Big], \\
\hat{\mathbf{s}}_{j,\sigma_r} =& \hat{v}_{rj}\hat{\sigma}_r^{-3}\Big[\hat{t}_{rj}(x_j - \hat{\mu}_r)^2 + \hat{\lambda}_r^2\hat{w}_{rj} - 2\hat{\lambda}_r(x_j - \hat{\mu}_r) - \hat{\sigma}_r^2\Big].
\end{aligned}$$

So, the standard error of the estimator $\hat{\boldsymbol{\theta}}$ can be obtained by calculating the square roots of the diagonal elements of the inverse of (3.5).

4 Data Analyses

4.1 Simulation Study

In this subsection, we carry out a simulation study to illustrate the performance of our method. We check the asymptotic properties of the ML estimators and show that the method proposed in Section 3 to approximate the standard error of the ML estimation of the model parameters has good asymptotic properties. To proceed the experimental study, we generate 500 Monte Carlo samples from the FM-SL model with $\boldsymbol{\Theta} = (0.3, 2, 1, 1, 4, 2, 2)$ and sample size $n = 100$, 200, 500, 1000 and 2000.

For each sample size n, we investigate the bias and mean square error (MSE) as two asymptotic properties of the estimates obtained using the suggested EM algorithm. For each parameter θ, they are defined as

$$\text{Bias}(\theta) = \frac{1}{500}\sum_{i=1}^{500}\hat{\theta}_i - \theta, \quad \text{and} \quad \text{MSE}(\theta) = \frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta)^2,$$

where $\hat{\theta}_i$ is the estimation of θ_i in replication i.
Also, for each replication, we obtained the ML estimates of $\boldsymbol{\Theta}$ and the estimates of their standard error using (3.5). Then, the sample standard errors of θ (MC SE) and average estimated standard errors (A.SE) are computed to investigate consistency of the estimate of the standard errors. They are given, respectively, by

$$\text{MC SE} = \frac{1}{499}\left[\sum_{i=1}^{500}\hat{\theta}_i^2 - \frac{1}{500}\Big(\sum_{i=1}^{500}\hat{\theta}_i\Big)^2\right], \text{ and } \text{A.SE} = \Big(\frac{1}{500}\sum_{i=1}^{500} se_i(\hat{\theta})\Big)^2.$$

Table 1 presents the results of the simulation. As a general rule, it is evident that the bias as well as the MSE of estimators decreases when the sample size increases. It can also be seen that the values of MC SE and A.SE are close to each other and they agree by increasing in the sample size ($n > 500$). It is confirmed that the estimates of the standard error for ML estimators of the parameters are consistent.

4.2 Real Data Analyses

We use a real data set to test the performance of our algorithm. The new model is compared with FM models based on the family of the scale mixture of skew normal distributions. This class contains the finite mixture of skew-normal (FM-SN), the finite mixture of skew-*t* (FM-ST), the finite mixture of skew contaminated normal (FM-SCN) and the finite mixture of skew slash (FM-SSL) distributions. We use Akaike information (AIC) (Akaike, 1973) and Bayesian information criteria (BIC) (Schwarz, 1978) to compare the fitted models. For the accurate comparison, the misclassification rate (MCR) and two popular measures of the clustering agreement between two clusters, namely, the rank index (RI) and the adjusted rank index (ARI) (Hubert and Arabie, 1985) are also used. A lower misclassification rate represents a closer match between the 'true' class and the cluster class given by the candidate algorithm. Also, the more closeness of RI and ARI to 1 corresponds to the more exact match between the two sets of classes.

The diabetes data (Reaven and Miller, 1979), which can be obtained from ***mclust*** package in R (Fraley and Raftery, 1999, 2003), consists of three measurements on 145 subjects from three classes: chemical (36 observations), normal (76 observations), and

Table 1: The result of simulation study.

sample size	Measure	p	μ_1	μ_2	σ_1	σ_2	λ_1	λ_2
	Bias	-0.0321	-0.1429	-0.2091	0.0878	0.2175	0.1318	-0.0242
100	MES	0.0071	0.0821	0.1381	0.0828	0.2149	0.1045	0.1169
	MC SE	0.0071	0.0637	0.1114	0.0685	0.1547	0.0857	0.1144
	A.SE	0.0085	0.0654	0.1237	0.0769	0.1693	0.1020	0.1409
	Bias	0.0286	-0.1395	-0.1777	0.0445	0.1534	0.1259	-0.0093
200	MES	0.0046	0.0684	0.1009	0.0428	0.1326	0.0645	0.0692
	MC SE	0.0039	0.0387	0.0629	0.0427	0.1004	0.0527	0.0587
	A.SE	0.0043	0.0395	0.0637	0.0441	0.1013	0.0540	0.0601
	Bias	-0.0195	-0.0831	-0.1328	0.0142	0.0705	0.0813	-0.0067
500	MES	0.0026	0.0320	0.0558	0.0222	0.0646	0.0413	0.0294
	MC SE	0.0022	0.0256	0.0366	0.0224	0.0524	0.0273	0.0287
	A.SE	0.0022	0.0259	0.0368	0.0224	0.0527	0.0278	0.0291
	Bias	-0.0141	-0.0538	-0.0713	0.0050	0.0435	0.0478	-0.0045
1000	MES	0.0018	0.0187	0.0304	0.0146	0.0346	0.0333	0.0176
	MC SE	0.0016	0.0152	0.0245	0.0122	0.0349	0.0222	0.0166
	A.SE	0.0016	0.0152	0.0245	0.0122	0.0349	0.0222	0.0166
	Bias	-0.0087	-0.0300	-0.0410	-0.0025	0.0125	0.0072	-0.0015
2000	MES	0.0011	0.0079	0.0152	0.0078	0.0215	0.0204	0.0109
	MC SE	0.0010	0.0070	0.0135	0.0078	0.0198	0.0191	0.0107
	A.SE	0.0010	0.0070	0.0135	0.0078	0.0198	0.0191	0.0107

overt (33 observations). For our purpose, we consider Steady State Plasma Glucose (sspg) variable. Table 2 reports the parameter estimates together with the associated standard errors, Log-likelihood, AIC, BIC, MCR, IR and AIR for the fitted models. It can be seen that the FM-SL model has the smallest values of the Log-likelihood, AIC and BIC. This reveals that the new model provides a better fit on the data. Also, the summary of clustering measures shows that the FM-SL model achieved the lowest misclassification rate and the highest RI and ARI representing a close match with the true clustering.

5 Conclusion

In this paper, the finite mixture of skew Laplace distributions, called the FM-SL model, has been introduced. Studying some properties of the new model, we show that the FM-SL model is beneficial for modelling data exhibiting patterns of asymmetry, multimodality and fat tails. The ML estimator of the parameters are computed by

Table 2: Parameter estimates of diabetes data.

	FM-SL		FM-ST		FM-SN		FM-SCN		FM-SSL	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE	MLE	SE
p_1	0.4960	1.6826	0.6471	2.0847	0.6149	2.4593	0.6061	4.3659	0.5978	2.4570
p_2	0.2772	1.3584	0.2689	1.9345	0.3058	2.3436	0.3240	2.7833	0.3293	2.4190
μ_1	1.3100	0.9967	1.7239	1.0859	1.7222	0.9929	1.6383	1.8705	1.5778	1.4411
μ_2	1.2271	2.3389	2.0222	3.6414	1.8329	2.3421	1.8029	3.4803	1.8612	2.2025
μ_3	0.4100	1.8916	4.1481	9.6302	3.8423	2.0433	4.1011	3.2272	4.1353	2.1712
σ_1	0.4017	0.4963	0.6874	0.5624	0.7025	0.8914	0.5545	9.4806	0.5244	0.8027
σ_2	0.4516	1.4372	0.7564	3.6126	0.8478	2.9125	0.7474	6.1024	0.6868	1.8071
σ_3	0.4056	1.0266	1.5235	9.5239	1.6801	4.0588	1.3367	3.9983	1.3057	2.8796
λ_1	0.2089	0.9658	-1.4266	2.121	-1.6016	2.4506	-1.2249	5.8186	-1.1223	2.7434
λ_2	0.7718	3.3849	0.7633	8.7146	1.0720	8.2012	1.1849	9.6069	1.1126	7.5619
λ_3	0.3276	2.7935	0.5111	11.2348	1.0749	7.1717	1.2015	6.4412	1.0022	4.8558
ν_1	–		10	11.6237	–		0.8273	13.6202	3.5519	4.1356
ν_2	–		–	–	–		0.6782	13.5700	–	
Log.like	-198.1097		-200.7957		-200.7815		-200.9097		-200.9097	
AIC	418.2194		425.5914		423.563		425.5914		425.8193	
BIC	450.9635		461.3123		456.307		461.3123		461.5401	
MCR	0.3103		0.4965		0.5241		0.4827		0.5103	
IR	0.6237		0.5406		0.5414		0.5386		0.5408	
AIR	0.2469		0.0935		0.0813		0.0920		0.0835	

a feasible EM algorithm via presenting a convenient hierarchical formulation of the model. We carry out a simulation to investigate asymptotic properties of the ML estimates and consistency of their standard error estimate. Results of real data analysis suggest that the FM-SL model outperforms other competitors in cluster analyses.

As feasible extensions of the current work, the methodology presented in this paper can be developed in areas of multivariate cases. It is worthwhile to consider modeling random effects in linear mixed models and the mixture of the factor analyzers model by the FM-SL model. Also, the learning-based EM algorithm can be conducted by considering the FM-SL distribution (Hung and Chang-Chien, 2016).

Appendix

The proposed algorithms have been coded and implemented in R which is available from the authors upon request.

Acknowledgements

The authors are so grateful to the anonymous reviewers and the editor for their insightful comments and suggestions that greatly improved this paper.

References

Aitken, A.C. (1927), On Bernoulli's Numerical Solution of Algebraic Equations. *Proceedings of the Royal Society of Edinburgh*, **46**, 289–305.

Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle. *Proceeding of the Second International Symposium on Information Theory, B.N. Petrov and F. Caski, eds.*, 267–281, AkademiaiKiado, Budapest.

Arslan, O. (2010), An alternative multivariate skew Laplace distribution: properties and estimation. *Statistical Paper*, **51**, 865–887.

Basford, K.E., Greenway, D.R., McLachlan, G.J., and Peel, D. (1997), Standard errors of fitted component means of normal mixtures. *Computational Statistics*, **12**, 1–18.

Basso, R.M., Lachos, V.H., Cabral, C.R.B., and Ghosh, P. (2010), Robust mixture modeling based on scale mixtures of skew-normal distributions. *Computational Statistics and Data Analysis*, **54**, 2926–2941.

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

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P., and Lindsay, B. (1994), The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, **46**, 373–388.

Chen, J., Tan, X., and Zhang, R. (2008), Inference for normal mixtures in mean and variance. *Statistica Sinica*, **18**, 443–465.

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

Dog̀ru, F.Z., and Arslan, O. (2017), Parameter Estimation for Mixtures of Skew Laplace Normal Distributions and Application in Mixture Regression Modeling. Accepted for publication in *Communications in Statistics-Theory and Methods*.

Fraley, C., and Raftery, A.E. (1999), MCLUST: Software for model-based cluster analysis. *Journal of classification*, **16**, 297–306.

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

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

Frühwirth-Schnatter, S. (2006), *Finite mixture and Markov switching models*. Springer Science and Business Media.

Good, I.J., (1953), The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–260.

Ho, H.J., Pyne, S., and Lin, T.I. (2012), Maximum likelihood inference for mixtures of skew Student-t-normal distributions through practical EM-type algorithms. *Statistics and Computing*, **22**, 287–299.

Hubert, L., and Arabie, P. (1985), Comparing partitions. *Journal of classification*, **2**(1), 193–218.

Hung, W.L., and Chang-Chien, S.J. (2016), Learning-based EM algorithm for normal-inverse Gaussian mixture model with application to extrasolar planets. *Journal of Applied Statistics*, **44**, 1–22.

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

Lin, T.I., McNicholas, P.D., and Hsiu, J.H. (2014), Capturing patterns via parsimonious *t* mixture models. *Statistics and Probability Letters*, **88**, 80–87.

Lindsay, B.G., (1995), *Mixture models: theory, geometry and applications.* In NSF-CBMS regional conference series in probability and statistics, (pp. i-163). Institute of Mathematical Statistics and the American Statistical Association.

Louis, T.A. (1982), Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **44**, 226–233.

McLachlan, G.J., and Basford, K.E. (1988), *Mixture models: Inference and applications to clustering*. Marcel Dekker.

McLachlan, G.J. and Krishnan, T. (2008), *The EM Algorithm and Extensions*, 2nd edition. Wiley, New York.

McLachlan, G., and Peel, D. (2004), *Finite mixture models*. John Wiley and Sons.

McNeil, A.J., Frey, R., and Embrechts, P. (2005), *Quantitative risk management: Concepts, techniques and tools*. Princeton university press.

McNicholas, P.D., Murphy, T.B., McDaid, A.F., Frost, D. (2010), Serial and parallel implementations of modelbased clustering via parsimonious Gaussian mixture models. *Computational Statistics and Data Analysis*, **54**, 711–723.

Meilijson, I. (1989), A fast improvement to the EM algorithm on its own terms. *Journal of the Royal Statistical Society. Series B (Methodological)*, **51**, 127–138.

Reaven, G.M., and Miller, R.G. (1979), An attempt to define the nature of chemical diabetes using a multidimensional analysis. *Diabetologia*, **16**, 17–24.

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

Titterington, D.M., Smith, A.F., and Makov, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions*. Wiley.

Vrbik, I. and McNicholas P.D. (2012), Analytic calculations for the EM algorithm for multivariate skew-mixture models. *Statistics and Probability Letters*, **82**, 1169–1174.

Vrbik, I. and McNicholas P.D. (2014), Parsimonious skew mixture models for model-based clustering and classification. *Computational Statistics and Data Analysis*, **71**, 196–210.

Yao, W. (2010), A profile likelihood method for normal mixture with unequal variance. *Journal of Statistical Planning and Inference*, **140**, 2089–2098.