مجله‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار
دوره‌ی ۱۳، شماره‌ی ۲، پاییز و زمستان ۱۳۹۵، صص ۱۹۷–۲۱۴

J. Statist. Res. Iran **13** (2016): 197–214
DOI: 10.18869/acadpub.jsri.13.2.197

Mixture of Normal Mean-Variance of Lindley Distributions

Mehrdad Naderi*, Alireza Arabpour and Ahad Jamalizadeh

Shahid Bahonar University of Kerman

Received: 8/14/2016 Approved: 4/29/2017

Abstract. In this paper, a new mixture modelling using the normal mean-variance mixture of Lindley (NMVL) distribution has been considered. The proposed model is heavy-tailed and multimodal and can be used in dealing with asymmetric data in various theoretic and applied problems. We present a feasible computationally analytical EM algorithm for computing the maximum likelihood estimates. The behavior of the obtained maximum likelihood estimators is studied with respect to bias and mean squared errors through conducting a simulation study. Two examples with flow cytometry data are used to illustrate the applicability of the proposed model.

Keywords. Finite mixture model; Mean-variance mixture distribution; Lindley distribution; EM algorithm.

MSC 2010: 62H30; 62H10; 62E15.

1 Introduction

Known as a finite mixture of distributions, a finite convex linear combinations of distribution functions is used in various scientific areas. It is proposed as a tool for modelling population heterogeneity as well as to approximate

* Corresponding author

complicated probability densities in presenting multimodality, skewness and heavy tails. In these models, each distribution function is known as a mixture component and comprehensive surveys of them can be found in Böhning (2000), McLachlan and Peel (2000) and from a Bayesian point of view in Frühwirth-Schnatter (2006).

Not only in the applied statistics but also in methodological contexts the mixture of normal distributions (N-MIX) has been well recognized as a useful model. When data has some degrees of skewness, the N-MIX model may not provide a suitable model. In attempting to appropriately model a set of data arising from a class or several classes with asymmetric observations, Lin et al. (2007a,b, 2014) introduced new mixture models with components followed by the skew-normal (SN-MIX), skew-t (ST-MIX) and skew-t-normal (STN-MIX) distributions, respectively, and found that ST-MIX and STN-MIX fitted data better than SN-MIX. Although these models are attractive, the maximum likelihood (ML) estimator of degree of freedoms have not an explicit form and should be obtained numerically.

In this paper, we present a finite mixture version of the NMVL (NMVL-MIX) model. Some properties of the new model have been studied and the ML estimates of the parameters are computed by the Expectation conditionally maximization (ECM) algorithm. By fitting NMVL-MIX model on a real data set, we compare this model with N-MIX and some members of the finite mixture of scale mixture of skew-normal distribution family. Finally, we investigate the finite sample properties of the ML estimates via conducting a simulation study.

The rest of the paper is organized as follows. In Section 2, we briefly review the NMV and Lindley distributions. Subsection 2.3 describes the NMVL distribution, and mentions some of its properties. Finite mixture of NMVL distributions and ECM procedure for computing the parameter estimates are studied in Section 3. In Section 4, we finally check out the performance of the proposed model and the obtained ML estimates using a real data example and simulation study, respectively.

2 Preliminaries

2.1 Normal Mean-Variance Mixture Model

Let X be a random variable represented as

$$X \stackrel{d}{=} \mu + W\lambda + \sqrt{W}Y, \tag{1}$$

where $\stackrel{d}{=}$ denotes equality in distribution, $\mu, \lambda \in \mathbb{R}$, Y is distributed by the normal distribution with mean zero and variance σ^2 ($Y \sim N(0, \sigma^2)$), and W is a non-negative independent random variable with cumulative distribution function (cdf) $H(\cdot; \boldsymbol{\theta})$ which parametrized by the vector parameter $\boldsymbol{\theta}$. Then, X is said to have a univariate normal mean-variance mixture distribution. The cdf of X can be easily obtained as

$$F_X(x; \mu, \lambda, \sigma^2, \boldsymbol{\theta}) = \int_0^\infty \Phi((x - \mu - w\lambda)w^{-1/2}; \sigma^2)\, dH(w; \boldsymbol{\theta}), \qquad x \in \mathbb{R},$$

where $\Phi(\cdot; \sigma^2)$ denotes the cdf of $N(0, \sigma^2)$. As a special case, if $H(\cdot; \boldsymbol{\theta})$ is absolutely continuous with probability density function (pdf) $h(\cdot; \boldsymbol{\theta})$, we can readily obtain the pdf of X as

$$f_X(x; \mu, \lambda, \sigma^2, \boldsymbol{\theta}) = \int_0^\infty \frac{1}{\sqrt{w}} \phi\left(\frac{(x - \mu - w\lambda)}{\sqrt{w}}; \sigma^2\right) h(w; \boldsymbol{\theta})\, dw, \quad x \in \mathbb{R} \tag{2}$$

where $\phi(\cdot; x)$ denotes the pdf of $N(0, \sigma^2)$. Provided the mixture variable, W, has finite variance, we have

$$E(X) = \mu + \lambda E[W], \qquad \text{and} \qquad Var(X) = E[W]\sigma^2 + Var(W)\lambda^2.$$

2.1.1 Generalized Hyperbolic Distribution

In most literature the density of generalized hyperbolic (GH) distribution is defined directly, such as Protassov (2004). The application and inference based on this definition are inconvenient since some important characterizing parameters are not invariant under linear transformations. On the other hand, McNeil et al. (2005) considered the GH distribution via proposing the generalized inverse Gaussian (GIG) distribution as a mixing random variable in the normal mean-variance mixture model. Specifically, let W in the stochastic representation (1) be a random variable followed by the GIG

distribution ($W \sim GIG(\kappa, \chi, \psi)$) with the following pdf,

$$g(w;\kappa,\chi,\psi) = (\frac{\psi}{\chi})^{\kappa/2}\frac{w^{\kappa-1}}{2K_\kappa(\sqrt{\psi\chi})}\exp\left\{\frac{-1}{2}\left(w^{-1}\chi + w\psi\right)\right\}, \quad w>0, \quad (3)$$

where $K_\kappa(\cdot)$ denotes the modified Bessel function of the third kind, $\kappa \in \mathbb{R}$ and two parameters χ, ψ are given, such that $\chi \geqslant 0$, $\psi > 0$ if $\kappa > 0$; $\psi \geqslant 0$, $\chi > 0$ if $\kappa < 0$ and $\chi > 0$, $\psi > 0$ if $\kappa = 0$. The density of GIG distribution actually contains the density of gamma distribution as a special limiting case. When $\chi = 0$ and $\lambda > 0$, the GIG distribution becomes to the so-called gamma distribution with parameter λ and $\psi/2$, $Gamma(\lambda, \psi/2)$. In this case (3) must be interpreted as a limit, which can be evaluated using the asymptotic relation $K_\kappa(x) \sim \Gamma(\kappa)2^{\kappa-1}x^{-\kappa}$ as $x \downarrow 0$ for $\kappa > 0$.

Remark 1. Let $W \sim GIG(\kappa, \chi, \psi)$. Then $W^{-1} \sim GIG(-\kappa, \psi, \chi)$ and

$$E[W^n] = (\frac{\chi}{\psi})^{n/2}R_{(\kappa,n)}(\sqrt{\chi\psi}),$$

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

Proposing $W \sim GIG(\kappa, \chi, \psi)$, the random variable X in (1) has a GH distribution. Therefore, the pdf of X, obtained from (2), is given by

$$f_{GH}(x;\mu,\lambda,\sigma^2,\kappa,\chi,\psi) = C\frac{K_{\kappa-0.5}(\sqrt{(\psi+\lambda^2/\sigma^2)(\chi+(x-\mu)^2/\sigma^2)})}{\left\{\sqrt{(\psi+\lambda^2/\sigma^2)(\chi+(x-\mu)^2/\sigma^2)}\right\}^{0.5-\kappa}} \times \exp\left\{\lambda(x-\mu)/\sigma^2\right\}, \quad x \in \mathbb{R},$$

where

$$C = \frac{(\psi/\chi)^{\frac{\kappa}{2}}(\psi+\lambda^2/\sigma^2)^{\frac{1}{2}-\kappa}}{\sqrt{2\pi}\sigma K_\kappa(\sqrt{\psi\chi})}.$$

Under this parameterization, Blæsild (1981) showed that the linear transformations of GH random variable remain in this family.

2.2 Lindley Distribution

A non-negative random variable W follows the Lindley distribution if it has the following pdf

$$f(w;\alpha) = \frac{\alpha^2}{1+\alpha}(1+w)e^{-\alpha w}, \qquad w > 0,\ \alpha > 0.$$

We denote this distribution by $Lindley(\alpha)$. The Lindley distribution, introduced by Lindley (1958, 1965), is positively skewed and it can be seen that its pdf is a mixture of exponential and Gamma distributions. i.e,

$$f(w;\alpha) = \frac{\alpha}{1+\alpha} f_{GIG}(w;1,0,2\alpha) + \frac{1}{1+\alpha} f_{GIG}(w;2,0,2\alpha).$$

2.3 The Normal Mean-variance Mixture of Lindley Distribution

Definition 1. A random variable X is said to have a NMVL distribution if in representation (1) $W \sim Lindley(\alpha)$.

The following theorem shows that the pdf of the NMVL is a mixture of two pdfs of GH distribution.

***Theorem* 1.** *Let $X \sim NMVL(\mu,\lambda,\sigma^2,\alpha)$. Then the pdf of X are given by*

$$\begin{aligned} f_X(x;\mu,\lambda,\sigma^2,\alpha) = &\frac{\alpha}{1+\alpha} f_{GH}(x;\mu,\lambda,\sigma^2,1,0,2\alpha) \\ &+ \frac{1}{1+\alpha} f_{GH}(x;\mu,\lambda,\sigma^2,2,0,2\alpha), \qquad x \in \mathbb{R}. \end{aligned}$$

Also, the mean, variance and characteristic function of X are

$$\begin{aligned} E(X) =& \mu + \frac{\alpha+2}{\alpha(\alpha+1)}\lambda, \\ Var(X) =& \frac{\alpha+2}{\alpha(\alpha+1)}\sigma^2 + \frac{\alpha^2+4\alpha+2}{\alpha^2(\alpha+1)^2}\lambda^2, \\ \varphi_X(s) =& \exp(is\mu) M_W\big(is\lambda - \frac{1}{2}s^2\sigma^2\big), \end{aligned}$$

where $M_W(\cdot)$ is the moment generating function of the Lindley distribution.

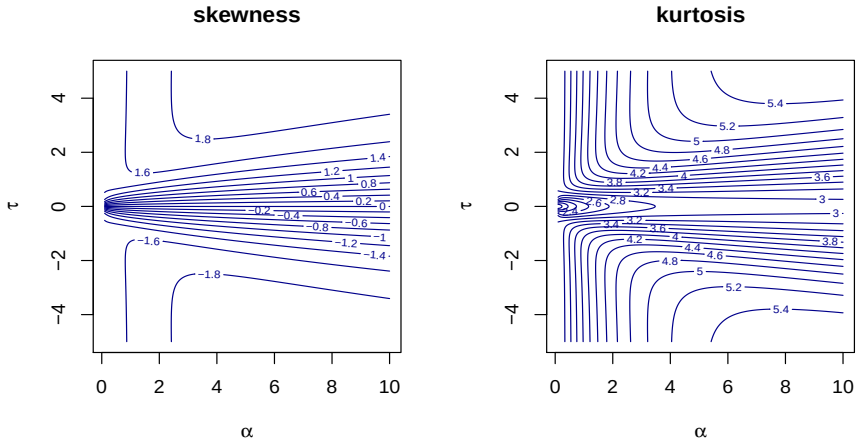

Figure 1. contour plot of skewness and kurtosis of NMVL.

By representation (1), the skewness and kurtosis of $X \sim NMVL(\mu, \lambda, \sigma^2, \alpha)$ can also be obtained as

$$\gamma_x = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{1.5}} \quad \text{and} \quad \kappa_x = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3,$$

where

$$\begin{aligned}
\mu_1 = E(X) &= \frac{\lambda(\alpha+2)}{\alpha(\alpha+1)}, \quad \mu_2 = E(X^2) = \frac{\alpha^2 + 2\alpha(\lambda^2+1) + 6\lambda^2}{\alpha^2(\alpha+1)}, \\
\mu_3 = E(X^3) &= \frac{6\lambda\big(\alpha^2 + \alpha(\lambda^2+3) + 4\lambda^2\big)}{\alpha^3(\alpha+1)}, \\
\mu_4 = E(X^4) &= \frac{6\Big\{\alpha^3 + \alpha^2 + 4\lambda^2\big(\alpha(\alpha+\lambda^2+6) + 5\lambda^2\big)\Big\}}{\alpha^4(\alpha+1)}.
\end{aligned}$$

Figure 1 shows the contour plots of the skewness and kurtosis of $NMVL(\mu, \lambda, \sigma^2, \alpha)$ as a function of α and λ. It can be observed that the range of asymmetry properties of the NMVL distribution is wider than SN distribution.

We establish the following proposition, which is useful for the calculation of some conditional expectations involved in the proposed EM algorithm discussed in the next section.

Proposition 1. Let X and W be the random variables with $NMVL(\mu, \lambda,$

$\sigma^2, \alpha)$ and $Lindley(\alpha)$, respectively. Then, for any $x \in \mathbb{R}$, the pdf of W given $X = x$ is a mixture of two GIG distribution, given as

$$f_{W\big|X=x}(w;\mu,\lambda,\sigma^2,\alpha) = p(x) f_{GIG}(w; 0.5, \chi(x,\mu,\sigma^2), \psi(\lambda,\sigma^2,\alpha)) + \{1-p(x)\} f_{GIG}(w; 1.5, \chi(x,\mu,\sigma^2), \psi(\lambda,\sigma^2,\alpha)), \quad w > 0,$$

where

$$p(x) = \frac{\alpha f_{GH}(x;\mu,\lambda,\sigma^2,1,0,2\alpha)}{\alpha f_{GH}(x;\mu,\lambda,\sigma^2,1,0,2\alpha) + f_{GH}(x;\mu,\lambda,\sigma^2,2,0,2\alpha)},$$

$\chi(x,\mu,\sigma^2) = (x-\mu)^2/\sigma^2$, $\psi(\lambda,\sigma^2,\alpha) = \lambda^2/\sigma^2 + 2\alpha$ and $f_{GIG}(.,\kappa,\chi,\psi)$ is the pdf of $GIG(\kappa,\chi,\psi)$. Furthermore, for $r = \pm 1, \pm 2, \ldots,$

$$E\big[W^r\big|X=x\big] = \left\{\frac{\chi(x,\mu,\sigma^2)}{\psi(\lambda,\sigma^2,\alpha)}\right\}^{r/2} \Big[p(x) R_{(0.5,r)}\{\sqrt{\psi(\lambda,\sigma^2,\alpha)\chi(x,\mu,\sigma^2)}\}, + \{1-p(x)\} R_{(1.5,r)}\{\sqrt{\psi(\lambda,\sigma^2,\alpha)\chi(x,\mu,\sigma^2)}\}\Big].$$

More details about NMVL distribution can be found in Naderi et al. (2017).

3 Finite Mixture of the NMVL Distributions

Consider n independent, random variable $X_1, \ldots, X_n$, which are taken from NMVL-MIX distributions. The density of a g-component MVNL-MIX model is

$$f(x_j;\boldsymbol{\Theta}) = \sum_{i=1}^{g} p_i\, f_{NMVL}(x_j;\boldsymbol{\theta}_i), \qquad j = 1,2,\ldots,n, \tag{4}$$

where p_i's are mixing proportions subject to $\Sigma_{i=1}^{g} p_i = 1$, $f_{NMVL}(.;\boldsymbol{\theta}_i)$ is a density of the NMVL distribution obtained in Theorem 1 with $\boldsymbol{\theta}_i = (\mu_i, \lambda_i, \sigma_i^2, \alpha_i)$ and $\boldsymbol{\Theta} = (p_1, \ldots, p_{g-1}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_g)$.

By observing data $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$, the observed data Log-likelihood function for x is

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

The ML estimator of the parameters can be obtained by maximizing $\ell(\boldsymbol{\Theta}|\boldsymbol{x})$ with respect to $\boldsymbol{\Theta}$. A direct maximization of this function is complicated, since its derivatives with respect to parameters are difficult to compute. Another approach for computing the ML estimator is the Expectation Maximization (EM) algorithm. In this approach, introduced by Dempster et al. (1977), the key idea is to solve a difficult incomplete Log-likelihood problem by repeatedly solving tractable complete Log-likelihood problems. The E-step of each iteration involves taking an expectation over complete-data log-likelihood given observed data, and then in the M-step of each iteration, the estimation of the parameter is obtained by maximization of this Expectation over the parameter space. For applying this approach to NMVL-MIX model, it is convenient to construct a complete Log-likelihood by introducing a set of allocation variables $\boldsymbol{Z}_j = (Z_{1j}, \ldots, Z_{gj})$ for $j = 1, \ldots, n$, taking $Z_{ij} = 1$ if y_j belongs to the ith component and $Z_{ij} = 0$ otherwise. This implies that the independent random variables $\boldsymbol{Z}_j$ follow a multinomial distribution with one trial and parameters $(p_1, \ldots, p_g)$, denoted as $\boldsymbol{Z}_j \sim M(1; p_1, \ldots, p_g)$. It also follows from (1) that the hierarchical representation of (4) can be represented by

$$\begin{aligned} X\big|(W_j = w_j, Z_{ij} = 1) &\sim N(\mu_i, w_j\lambda_i, w_j\sigma_i^2), \\ W_j\big|Z_{ij} = 1 &\sim Lindley(\alpha_i), \\ \boldsymbol{Z}_j &\sim M(1, p_1, p_2, \ldots, p_g). \end{aligned}$$

So, the complete-data Log-likelihood associated with the observed data $\boldsymbol{x}$ and hidden variables $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ and $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n)^\top$, omitting additive constants, is obtained as

$$\begin{aligned} \ell_c(\boldsymbol{\Theta}|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{Z}) = \sum_{j=1}^{n}\sum_{i=1}^{g} z_{ij}\Bigg[&\log(p_i) + \log\Big(\frac{\alpha_i^2}{1+\alpha_i}\Big) - \frac{1}{2}\log(\sigma_i^2) - w_j\alpha_i \\ &- \frac{(x_j-\mu_i)^2}{2w_j\sigma_i^2} - \frac{w_j\lambda_i^2}{2\sigma_i^2} + \frac{\lambda_i(x_j-\mu_i)}{\sigma_i^2}\Bigg], \end{aligned} \tag{5}$$

3.1 Parameter Estimation via ECM Algorithm

In this subsection, we apply ECM algorithm (Meng and Rubin, 1993) to estimate parameters of the NMVL-MIX model. The algorithm is iterated between the following steps.

E-step: Compute the conditional expectation of (5), known as a Q-function, at the kth iteration as

$$Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) = E\left[\ell_c(\boldsymbol{\Theta}|\boldsymbol{x},\boldsymbol{w},\boldsymbol{Z})|\boldsymbol{x},\hat{\boldsymbol{\Theta}}^{(k)}\right].$$

The necessary conditional expectations to compute the Q-function include $\hat{z}_{ij}^{(k)} = E\big[Z_{ij}|x_j,\hat{\boldsymbol{\Theta}}_i^{(k)}\big]$, $\hat{w}_{ij}^{(k)} = E\big[W_j|x_j,\ Z_{ij}=1,\hat{\boldsymbol{\Theta}}_i^{(k)}\big]$ and $\hat{t}_{ij}^{(k)} = E\big[W_j^{-1}|x_j,Z_{ij}=1,\hat{\boldsymbol{\Theta}}_i^{(k)}\big]$. By Proposition 1, we can obtain these expectations as

$$\begin{aligned}
\hat{z}_{ij}^{(k)} &= \frac{\hat{p}_i f_{MNSB}(x_j;\hat{\boldsymbol{\theta}}_i^{(k)})}{f(x_j;\hat{\boldsymbol{\Theta}}_i^{(k)})},\\
\hat{w}_{ij}^{(k)} &= \left(\frac{\chi_{ij}}{\psi_i}\right)^{0.5}\Big[p_{ij}R_{(0.5,1)}(\sqrt{\psi_i\chi_{ij}})\\
&\quad + (1-p_{ij})R_{(1.5,1)}(\sqrt{\psi_i\chi_{ij}})\Big],\\
\hat{t}_{ij}^{(k)} &= \left(\frac{\psi_i}{\chi_{ij}}\right)^{0.5}\Big[p_{ij}R_{(-0.5,1)}(\sqrt{\psi_i\chi_{ij}})\\
&\quad + (1-p_{ij})R_{(-1.5,1)}(\sqrt{\psi_i\chi_{ij}})\Big],
\end{aligned} \tag{6}$$

where $p_{ij} = p_i(x_j)$, $\chi_{ij} = \chi(x_j,\mu_i,\sigma_i^2)$ and $\psi_i = \psi(\lambda_i,\sigma_i^2,\alpha_i)$, defined in proposition (1). So, the Q-function can be written as

$$\begin{aligned}
Q(\boldsymbol{\Theta}|\hat{\boldsymbol{\Theta}}^{(k)}) = \sum_{j=1}^{n}\sum_{i=1}^{g}\hat{z}_{ij}^{(k)}\Bigg[&\log p_i - \log\alpha_i - \frac{1}{2}(\log(\sigma_i) - \alpha_i\hat{w}_{ij}^{(k)}\\
&- \frac{(x_j-\mu_i)^2}{2\sigma_i^2}\hat{t}_{ij}^{(k)} - \frac{\hat{w}_{ij}^{(k)}\lambda_i^2}{2\sigma_i^2} + \frac{(x_j-\mu_i)\lambda_i}{\sigma_i^2}\Bigg].
\end{aligned} \tag{7}$$

M-step: Let $n_i = \sum_{j=1}^{n}\hat{z}_{ij}^{(k)}$, $A_i = \sum_{j=1}^{n}\hat{z}_{ij}^{(k)}\hat{t}_{ij}^{(k)}$, $B_i = \sum_{j=1}^{n}x_j\hat{z}_{ij}^{(k)}\hat{t}_{ij}^{(k)}$ and $C_i = \sum_{j=1}^{n}\hat{w}_{ij}^{(k)}\hat{z}_{ij}^{(k)}$. Maximizing the Q-function and update parameter at the $(k+1)$th iteration by the following CM-steps:

CM-step 1: Calculate

$$\begin{aligned}\hat{p}_i^{(k+1)} =& \frac{n_i}{n},\\ \hat{\lambda}_i^{(k+1)} =& \frac{A_i \sum_{j=1}^{n} x_j \hat{z}_{ij}^{(k)} - n_i B_i}{A_i C_i - n_i^2},\\ \hat{\alpha}_i^{(k+1)} =& \frac{-(C_i - n_i) + \sqrt{(C_i - n_i)^2 + 8 n_i C_i}}{2C_i}.\end{aligned}$$

CM-step 2: By maximizing (7) over μ_i, update $\mu_i^{(k)}$ as

$$\hat{\mu}_i^{(k+1)} = \frac{B_i - n_i \hat{\lambda}_i^{(k+1)}}{A_i}.$$

CM-step 3: Put $\mu_i = \hat{\mu}_i^{(k+1)}$, and update $\sigma_i^{2(k)}$ by maximizing (7) over σ_i^2. This leads to

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

4 Real Data Analysis

This section studies the performance of the proposed model and procedure of parameters' estimating which is discussed in the earlier section. To verify the beneficence of the NMVL-MIX, we use flow cytometry data set. The flow cytometry is a technique for scanning, outlining and sorting microscopic particles in a stream of water based on the laser. Because of using this technique in clinical research, it is used in a large number of biomedical applications such as molecular and cellular biology to measure the content DNA. It was recently shown that flow cytometric data is ideally suited for multimodal non-Gaussian mixture modelling (Pyne et al., 2004; Frühwirth-Schnatter and Pyne, 2010; Ho et al., 2012).

Glynn (2006) provided a working data set of flow cytometry in 'CC4-067-BM.fcs. The set consists of 5,634 cells which is related to the ten attributes measured. In this part, we analyze channels APC and FSC. Suggested by Hahne et al. (2009) in the BioConductor package flowCore, we prepare data by considering transformation $y/1000$, to compare the NMVL-MIX model

with the N-MIX, SN-MIX and ST-MIX models.

4.1 Model Selection Criteria

The Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978) are computed to identify the best selected model. These measures are given by

$$-2\ell(\hat{\boldsymbol{\Theta}}) + m\, c_n,$$

where $\ell(\hat{\boldsymbol{\Theta}})$ and m represent the maximized log-likelihood and the number of estimated free parameters related to the model, respectively, and the penalty term c_n is a convenient sequence of positive numbers. The term c_n is chosen 2 for AIC and $\log(n)$ for BIC. As an alternative criterion, Biernacki et.al. (2000) proposed a measure based on the integrated completed likelihood (ICL) for estimating the proper number of mixing component. The ICL criteria is obtained by a BIC-like approximation as

$$ICL = BIC + \sum_{j=1}^{n}\sum_{i=1}^{g} \hat{z}_{ij}\log(\hat{z}_{ij}),$$

where $\hat{z}_{ij}$ is computed by equation (6) evaluated at $\boldsymbol{\Theta} = \hat{\boldsymbol{\Theta}}$.

4.2 Kolmogorov-Smirnov Test

Denote the order statistics of the random samples by $X_{(1)} \leqslant X_{(2)} \leqslant \ldots \leqslant X_{(n)}$. The Classical test statistics is defined as a some measure of cdf distance $F_n(x) - F(x,\hat{\theta})$ where $F_n(x) = n^{-1}\sum_{i=1}^{n} I\{X_i \leq t\}$ represents the empirical cdf, $F(\cdot,\hat{\theta})$ is the cdf fitted by the data in which $\hat{\theta}$ denotes the estimation of θ.

In particular, the popular Kolmogorov-Smirnov (KS) statistic is defined by

$$KS = \max\left\{K^+, K^-\right\},$$

where $\mathrm{K}^+ = \max\left[\frac{i}{n} - F(X_{(i)},\hat{\theta})\right]$ and $\mathrm{K}^- = \max\left[F(X_{(i)},\hat{\theta}) - \frac{i-1}{n}\right]$ over $1 \leqslant i \leqslant n$.

To compute the estimated p-value of the KS test, n random numbers are generated from standard uniform distribution and we order them as $u_{(1)}^{(i)} \leqslant$

$u_{(2)}^{(i)} \leqslant \ldots \leqslant u_{(n)}^{(i)}$. Compute,

$$d^{(i)} = \max_{j=1,\ldots,n} \left\{ \frac{j}{n} - u_{(j)}^{(i)}, u_{(j)}^{(i)} - \frac{j-1}{n} \right\}.$$

Now, Let $I_i = 1$ if $d^{(i)} \geqslant$ KS and 0 otherwise and repeat this producer N times to get $I_1, \ldots, I_N$. As a result, the estimated p-value is obtained by $\sum_{i=1}^{N} = I_i/N$.

4.3 Data Fitting

As a first example, we use FSC data to compare N-MIX, SN-MIX, ST-MIX and NMVL-MIX models. The result of fitting these models with series of mixture components (g = 2–4) are summarized in Table 1. It is worthwhile to note that the smaller value of AIC, BIC or ICL model has, the better fit on the data is provided.

It can be seen that the best number of component based on the three criteria is vary. But, the values clearly show that not only for the AIC criterion but also for the BIC and the ICL criteria the NMVL-MIX distributions fits data better than other competitors.

For the channel APC, we also compare N-MIX, SN-MIX, ST-MIX and NMVL-MIX models with different mixture components (g=2-4), as a second example. We found that the best number of the component is g=2 in all models. Table 2 shows the ML estimates with the associated standard errors for the best fitted NMVL-MIX model and the corresponding values for the other three competing 2-component mixture models. Also, the values of AIC, BIC and ICL are reported in this table, which show that the NMVL-MIX model is the best fit. This result can also be seen from the histogram of the data and estimated pdf of models, plotted in Figure 1. Furthermore, The results of the KS test are listed in Table 2. Of the four mixture models, the best fit is the NMVL-MIX model with a p-value of 0.402 which suggests that the APC data follow a mixture of the NMVL distributions.

4.4 Simulation Study

In order to analyze the performance of the estimates obtained using our proposed ECM algorithm, we investigate bias and mean square error as two

Table 1. model selection criteria for FSC data.

model	group	AIC	BIC	ICL	group	AIC	BIC	ICL
N-MIX	1	14974.72	14987.99	14987.99	2	**12234.03**	**12267.21**	**12285.01**
	3	12398.79	12451.88	12772.51	4	12434.61	12507.47	14997.52
SN-MIX	1	14922.17	14942.08	14942.08	2	12385.13	12431.58	12479.32
	3	11765.81	**11838.01**	**12158.27**	4	**11747.50**	11852.89	13827.60
ST-MIX	1	14583.07	14602.98	14602.98	2	11994.19	12051.42	12216.12
	3	11654.62	11745.70	**12173.16**	4	**11534.57**	**11634.63**	13573.19
NMVL-MIX	1	14501.38	14527.93	14527.93	2	11779.81	11839.54	11986.45
	3	11397.82	11490.62	**11784.73**	4	**11168.25**	**11294.34**	12764.76

Table 2. Parameter estimates of APC data.

	NMVL-MIX		ST-MIX		SN-MIX		N-MIX	
parameter	MLE	SE	MLE	SE	MLE	SE	MLE	SE
p	0.6854	0.0082	0.6826	0.0077	0.6497	0.0071	0.6285	0.0068
μ_1	3.1196	0.0089	3.2320	0.0085	3.2667	0.0097	2.8899	0.0174
μ_2	0.2660	0.0095	0.1626	0.0091	0.1179	0.0089	0.7880	0.0060
σ_1	0.7889	0.0158	0.4457	0.0131	0.5499	0.0133	0.3207	0.0194
σ_2	0.5835	0.0264	0.4740	0.0222	0.8036	0.0275	0.5678	0.0028
λ_1	-1.5927	0.1657	-3.1874	0.2104	-3.1353	0.1523	–	
λ_2	1.4250	0.0458	5.9648	0.7979	10.7986	1.3187	–	
α_1	5.4259	0.8451	–		–		–	
α_2	4.4339	0.7564	–		–		–	
ν	–		3.0475	0.3006	–		–	
$\ell(\hat{\boldsymbol{\Theta}})$	-5571.916		-5592.241		-5732.665		-6286.44	
AIC	11161.83		11200.48		11479.33		12582.88	
BIC	11221.56		11246.94		11525.79		12616.06	
ICL	11784.19		11798.84		11831.65		12813.01	
KS	0.0122		0.0159		0.0396		0.0613	
P.Value	0.402		0.1045		< 0.0001		< 0.0001	

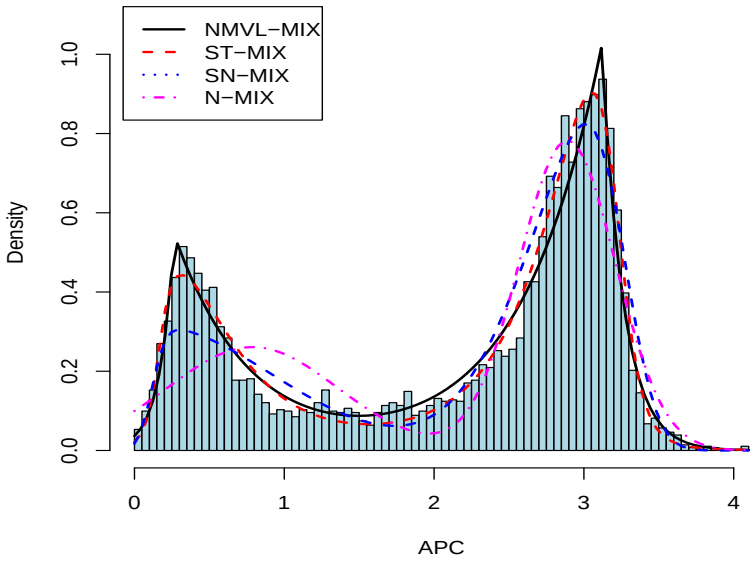

Figure 2. The histogram of data with four fitted models.

asymptotic properties of the estimates. We consider a set of the parameter values $\boldsymbol{\Theta} = (p, \mu_1, \lambda_1, \sigma_1, \alpha_1, \mu_2, \lambda_2, \sigma_2, \alpha_2) = (0.4, -2, -2, 3, 1, 2, 1, 4, 0.5)$ for this study.

For the proposed combination of parameter and sample size n =100, 200, 400, 800 and 1600, we generate 1000 samples from the NMVL-MIX model. Then, the absolute relative bias (R.Bias) and mean squared error (MSE) are computed over all samples. For each parameter θ, they are defined as

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

where $\hat{\theta}_i$ is the estimation of θ_i when the data is sample i.

Figure 3 presents the results of this simulation. Form this figure, a pattern of convergence to zero of the bias and MSE can be seen when n increases. As a general rule, we can say that R.Bias and MSE tend to approach to zero when the sample size increases indicating that the estimates based on the proposed ECM algorithm do provide good asymptotic properties.

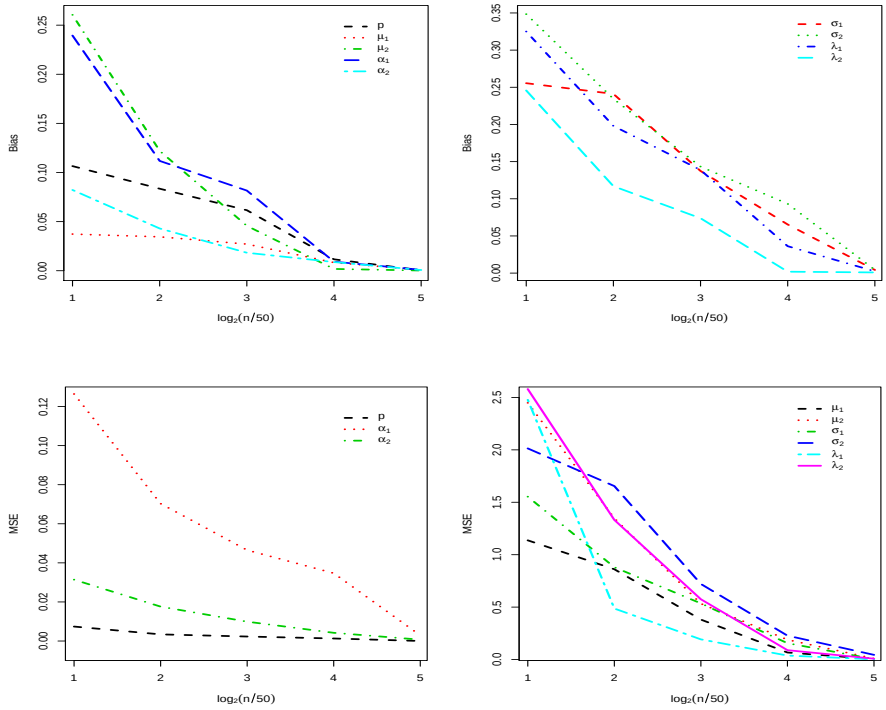

Figure 3. Relative Bias and MSE of the simulated data.

5 Conclusion

In this paper, we have introduced a new mixture model via NMVL distribution. The new model which is called the NMVL-MIX, is heavy tail and has a wide range of skewness. As a result, the NMVL model is useful for modeling multimodal and heavy tails data and can be applicable for clustering and pattern recognition. To find the ML estimation of NMVL model, we have presented a convenient hierarchical representation and developed an ECM algorithm according to them. Real data results show that the proposed method performs reasonably well for the experimental data. We also conduct a simulation study to investigate the properties of the models parameter. The R code of the real data analysis can be found from the authors upon request.

An interesting extension of the current work that deserves attention in

future research concerns the multivariate case of NMVL-MIX (Naderi et al., 2017). Also, The use of mixtures of factor analyzers can be considered as a parsimonious modeling approach.

Acknowledgement

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

References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE transactions on automatic control*, **19**, 716-723.

Biernacki, C., Celeux, G. and Govaert, G. (2000). Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *IEEE transactions on pattern analysis and machine intelligence*, **22**, 719-725.

Blæsild, P. (1981). The Two-dimensional Hyperbolic Distribution and Related Distributions, with an Application to Johannsen's Bean Data. *Biometrika*, **68**, 251-263.

Böhning, D., (2000). *Computer-Assisted Analysis of Mixtures and Applications*. CRC, Chapman & Hall.

Dempster, A.P., Laird, N.M. and Rubin, D.B., (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society. Series B*, **39**, 1-38.

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

Frühwirth-Schnatter, S. and Pyne, S. (2010). Bayesian Inference for Finite Mixtures of Univariate and Multivariate Skew Normal and Skew-t Distributions. *Biostatistics*, **11**, 317-336.

Glynn, E.F. (2006). FCSExtract Utility. Stowers Institute for Medical Research. Online available at: http://research.stowers-institute.org/ efg/ScientificSoftware/Utility/FCSExtract/.

Hahne, F., LeMeur, N., Brinkman, R.R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E. and Gentleman, R. (2009). flowCore: a Bioconductor Package for High Throughput Flow Cytometry. *BMC Bioinform*, **10**, 106.

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. *Statistical Computing*, **20**, 343-356.

Lin, T.I., Ho, H.J. and Lee, C.R. (2014). Flexible Mixture Modelling Using the Multivariate Skew-t-Normal Distribution. *Statistical Computing*, **2**, 1-16.

Lin, T.I., Lee, J.C. and Hsieh, W.J. (2007a). Robust Mixture Modelling Using the Skew t Distribution. *Statistical Computing*, **17**, 81-92.

Lin, T.I., Lee, J.C. and Yen, S.Y. (2007b). Finite Mixture Modelling Using the Skew Normal Distribution. *Statistica Sinica*, **17**, 909-927.

Lindley, D.V. (1958). Fiducial Distributions and Bayes' Theorem. *Journal of the Royal Statistical Society. Series B*, **20**, 102-107.

Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference*, New York, Cambridge University Press.

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

McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Meng, X.L. and Rubin, D.B. (1993). Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika*, **80**, 267-278.

Naderi, M., Arabpour, A. and Jamalizadeh, A. (2017). Multivariate Normal Mean-variance Mixture Distribution based on Lindley Distribution. *Communications in Statistics-Simulation and Computation*, (just-accepted).

Protassov, R. (2004). EM-based Maximum Likelihood Parameter Estimation for Multivariate Generalized Hyperbolic Distributions with Fixed λ. *Statistical Computing*, **14**, issue 1.

Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.I., Maier, L., BaecherAllan, C., McLachlan, G.J., Tamayo, P., Hafler, D.A., De Jager, P.L. and Mesirov, J.P. (2004). Automated High-dimensional Flow Cytometric Data Analysis, Proc. Natl. Acad. Sci. USA, **106**, 8519-8524.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, **6**, 461-464.

Mehrdad Naderi
Department of Statistics,
Faculty of Mathematics and Computer,
and Young Researchers Society,
Shahid Bahonar University of Kerman,
Kerman, Iran.
email: *Mehrdad.Naderi@ymail.com*

Alireza Arabpour
Department of Statistics,
Faculty of Mathematics and Computer,
Shahid Bahonar University of Kerman,
Kerman, Iran.
email: *arabpour@uk.ac.ir*

Ahad Jamalizadeh
Department of Statistics,
Faculty of Mathematics and Computer,
Shahid Bahonar University of Kerman,
Kerman, Iran.
email: *a.jamalizadeh@uk.ac.ir*