Research Journal of Applied Sciences, Engineering and Technology 10(2): 112-117, 2015
DOI:10.19026/rjaset.10.2562
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: May 06, 2012 Accepted: July 09, 2012 Published: May 20, 2015

Research Article

Testing Homogeneity of Mixture of Skew-normal Distributions Via Markov Chain Monte Carlo Simulation

[1]Rahman Farnoosh [2]Morteza Ebrahimi and [2]SetarehDalirian
[1]Department of Mathematics, South Tehran Branch, Islamic Azad University, Tehran,
[2]Department of Mathematics, Karaj Branch, Islamic Azad University, Karaj,
[3]School of Mathematics, Iran University of Science and Technology, Narmak, Tehran 16846, Iran

Abstract: The main purpose of this study is to intoduce an optimal penalty function for testing homogeneity of finite mixture of skew-normal distribution based on Markov Chain Monte Carlo (MCMC) simulation. In the present study the penalty function is considered as a parametric function in term of parameter of mixture models and a Baysian approach is employed to estimating the parameters of model. In order to examine the efficiency of the present study in comparison with the previous approaches, some simulation studies are presented.

Keywords: Homogeneity test, markov chain monte carlo simulation, penalty function, skew-normal mixtures

INTRODUCTION

Finite mixture model that is convex linear combination of densities, have been applied in various areas such as genetics, image processing, medicine and economics. These models have been used to model heterogeneous data (Bohning, 2000; McLachlan and Peel, 2000). Many works are also devoted to Baysian mixture modeling (Diebolt and Robert, 1994; Escobar and West, 1995; Richardson and Green, 1997; Stephens 2000).

Testing for homogeneity is one of the particular statistical problem in 2 component mixture models that is used to determine whether the data are from a mixture of 2 distributions or single distribution. But there are some limitation for testing in previous methods, such as Likelihood Ratio Test(LRT), Modified Likelihood Ratio Test (MLRT) and the test of Expectation- Maximization (EM). Since the mixture models don't have regularity conditions therefore limiting distribution of LRT is very complex (Liu and Shao, 2003). Although in MLRT (Chen, 1998; Chen and Chen, 2001; Chen *et al*., 2004) a penalty is added to the Log-likelihood function but dependence on establishment of several regularity conditions and chosen penalty function are restriction for this test. Li *et al*. (2008) have been proposed EM test based on another form of penalty function. Though this test is independent of some necessary conditions but like MLRT test is based on penalized likelihood function.

In the present study, to overcome the above mentioned disadvantages we consider penalty function as a parametric function and employ Metropolis-Hasting sampling as a MCMC method for estimation parameters of finite mixture of skew-normal distribution and parameter of determinative shape of the penalty function. The proposed method is based on the work of Farnoosh *et al*. (2012).

Recently, skewed distributions such as Skew-Normal (SN) and Skew-T (ST) are considered as component of mixture models in order to efficiently deal with population heterogeneity and skewness (Azzalini, 1985; Azzalini and Capitanio, 2003; Cabral *et al*., 2008). Literature review show taht Lin *et al.* (2007b) have been proposed a mixture model based on the skew-normal distribution. This study was extended by Lin *et al*. (2007a) where they considered mixtures of the student-t skewed distributions (Azzalini and Capitanio, 2003). Lin (2009) has proposed Maximum Likelihood Estimation (MLE) for multivariate skew-normal mixture models. Basso *et al*. (2010) considered estimation for univariate finite mixtures of flexible class of Skew-Normal Independent distribution (SNI) (Lachos *et al*., 2010) which is a subclass of the skew-elliptical class propossed by Branco and Dey (2001). In this study, we apply the methology of Farnoosh *et al*. (2012) to test homogeneity of mixture of skew-normal distribution.

FINITE MIXTURE OF SKEW-NORMAL DISTRIBUTION

Skew-Normal distribution: As defined by Azzalini (1985), a random variable Y has skew-normal

Corresponding Author: Rahman Farnoosh, Department of Mathematics, South Tehran Branch, Islamic Azad University, Tehran, Iran, Tel.: (+9821)73225427; Fax: (+9821)77240302

distribution with location parameter μ, scale parameter σ^2 and skewness parameter λ, if its density is given by:

$$\psi(y;\mu,\sigma^2,\lambda) = 2\phi(y;\mu,\sigma^2)\Phi(\frac{\lambda(y-\mu)}{\sigma}), \quad (1)$$

where $\emptyset(.;\mu,\sigma^2)$ denotes the density of the univariate normal distribution with mean μ and variance $\sigma^2 > 0$ and $\Phi(.)$ is the distribution function of the standard univariate normal distribution. It is denoted by $Y \sim SN(\mu, \sigma^2, \lambda)$.

Lemma : A random variable $Y \sim SN(\mu, \sigma^2, \lambda)$ has a stochastic representation given by:

$$Y = \mu + \sigma\delta|T_0| + \sigma(1-\delta^2)^{1/2}T_1, \quad (2)$$

where $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$, T_0 and T_1 are independent standard normal variable and |.| denotes absolute value (Henze, 1986).

The FM-SN model: The Finite Mixture of Skew-Normal distributions model (FM-SN) is defined by considering a random sample $y = (y_1,\cdots,y_n)^T$ from a g-component mixture of skew-normal densities given by:

$$f(y_i;\mathbf{\Theta}) = \sum_{j=1}^{g} p_j\psi(y_i;\mathbf{\theta}_j), \quad p_j \geq 0, \quad \sum_{j=1}^{g} p_j = 1, \quad i=1,\ldots n, j=1,\ldots g, \quad (3)$$

where $\theta_j = (\mu_j, \sigma_j^2, \lambda_j)$ is the specific vector of parameters for the component j, $\psi(y_i; \theta_j)$ is the $SN(\theta_j)$ density, $p_1,\cdots,p_g$ are the mixing probabilities and $\Theta = ((p_1,\cdots,p_g)^T, \theta_1^T,\cdots,\theta_g^T)^T$ is the vector with all parameters.

The finite mixture of skew-normal distributions model in this study is a special case of Finite Mixture of Scale Mixtures Of Skew-Normal Distribution(FM-SMSN) when $\kappa(u) = 1$ (Basso *et al*., 2010).

As introduced by McLachlan and Peel (2000), we consider the latent indicator variable Z_{ij} (for each i and j) such that,

$$P(Z_{ij}=1) = 1 - P(Z_{ij}=0) = p_j, \quad \sum_{j=1}^{g} Z_{ij} = 1 \quad and \quad y_i | Z_{ij} = 1 \sim SN(\mathbf{\theta}_j), \quad (4)$$

The indicators $Z_1,\ldots,Z_n$ where $Z_i = (Z_{i1},\ldots, Z_{ig})^T$, $i = 1,\ldots n$ are independent and each one having a multinomial distribution with probability function:

$$f(\mathbf{Z}_i) = p_1^{z_{i1}} p_2^{z_{i2}} \ldots (1-p_1-\ldots-p_{g-1})^{z_{ig}}, \quad (5)$$

and it is denoted by $Z_i \sim M(1; p_1,\ldots, p_g)$.

So, the likelihood function and Log-likelihood function of complete data are:

$$L^c(\mathbf{\Theta}|\mathbf{y},\mathbf{z}) = \prod_{i=1}^{n}\prod_{j=1}^{g}\left(p_j\psi(y_i;\mathbf{\theta}_j)\right)^{z_{ij}} \quad (6)$$

$$\ell^c(\mathbf{\Theta}|\mathbf{y},\mathbf{z}) = \sum_{i=1}^{n}\sum_{j=1}^{g} z_{ij}\left\{log(p_j) + log(\psi(y_i;\mathbf{\theta}_j))\right\} \quad (7)$$

Parameter stimation via the EM algorithm: In this subsection we present the EM algorithm (Dempster *et al*., 1977) for maximu likelihood estimation of a FM-SN distribution. Let $y = (y_1,\cdots,y_n)^T$ be a random sample from a FM-SN distribution. From (4) and Lemma we consider the following hierarchical representation for y_i:

$$Y_i | T_i = t_i, Z_{ij} = 1 \sim N(\mu_j + \Delta_j t_i, \Gamma_j), \quad (8)$$

$$T_i | Z_{ij} = 1 \sim HN(0,1) \quad (9)$$

$$\mathbf{Z}_i \sim M(1; p_1,\ldots,p_g), \quad i = 1,\ldots,n, j = 1,\ldots,g, \quad (10)$$

where,

$$\Gamma_j = (1-\delta_j^2)\sigma_j^2, \quad \Delta_j = \sigma_j\delta_j \quad and \quad \delta_j = \frac{\lambda_j}{\sqrt{\lambda_j^2+1}} \quad (11)$$

By using (8)-(10) complete-data log-likelihood function is:

$$\ell^c(\mathbf{\Theta}) = c + \sum_{i=1}^{n}\sum_{j=1}^{g} z_{ij}\left(log(p_j) - \frac{1}{2}log(|\Gamma_j|) - \frac{1}{2\Gamma_j}(y_i - \mu_j - \Delta_j t_i)^2\right), \quad (12)$$

where c is a constant that is independent of the parameter vector Θ.

Thus, the EM algorithm for maximum likelihood estimation of Θ is defined as follow:

E-step: Given a current estimate $\widehat{\Theta}(k)$ and compute $\hat{Z}_{ij}$ for i = 1,..., n, j = 1,..., g as follow:

$$\hat{Z}_{ij}^{(k)} = E[Z_{ij} | \hat{\Theta}^{(k)}, y_i] = \frac{\hat{p}_j^{(k)}\psi(y_i;\hat{\mathbf{\theta}}_j^{(k)})}{\sum_{j=1}^{g} \hat{p}_j^{(k)} \psi(y_i;\hat{\mathbf{\theta}}_j^{(k)})} \quad (13)$$

M-step: Update $\widehat{\Theta}(k)$ by maximizing the expected complete-data function or Q-function over Θ given by:

$$Q(\mathbf{\Theta}|\hat{\mathbf{\Theta}}^{(k)}) = c + \sum_{i=1}^{n}\sum_{j=1}^{g} \hat{z}_{ij}^{(k)}\left(log(p_j) - \frac{1}{2}log(|\Gamma_j|) - \frac{1}{2\Gamma_j}(y_i - \mu_j - \Delta_j t_i)^2\right), \quad (14)$$

which leads to the following closed form expressions:

$$\hat{p}_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n} z_{ij}^{(k)} \tag{15}$$

$$\hat{\mu}_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\Delta}_j^{(k)}) \tag{16}$$

$$\hat{\Delta}_j^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu}_j^{(k+1)}) \tag{17}$$

$$\hat{\Gamma}_j^{(k+1)} = \frac{\sum_{i=1}^{n}(y_i - \hat{\mu}_j^{(k+1)} - \hat{\Delta}_j^{(k+1)})^2}{\sum_{i=1}^{n}\hat{z}_{ij}^{(k)}} \tag{18}$$

Note that $(\hat{\sigma}_j^2)^{(k+1)}$ and $\hat{\lambda}_j^{(k+1)}$ can be obtained by using (11), that is:

$$(\hat{\sigma}_j^2)^{(k+1)} = (\hat{\Delta}_j^{(k+1)})^2 + \hat{\Gamma}_j^{(k+1)} \tag{19}$$

$$\hat{\lambda}_j^{(k+1)} = \frac{\hat{\Delta}_j^{(k+1)}}{\sqrt{\hat{\Gamma}_j^{(k+1)}}} \tag{20}$$

This process is iterated untill a suitable convergence rule is satisfied, e.g., if $\| \hat{\theta}^{(k+1)} - \hat{\theta}^{(k)} \|$ is sufficiently small.

METHODOLOGY

LRT and MLRT: Suppose $Y_1,\ldots,Y_n$ is a random sample from a mixture of skew-normal distribution by g = 2 component:

$$f(y_i;\mathbf{\Theta}) = p\psi(y_i;\boldsymbol{\theta}_1) + (1-p)\psi(y_i;\boldsymbol{\theta}_2) \tag{21}$$

we wish to test,

$$\begin{cases} H_0 : p(1-p)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) = \mathbf{0} \\ H_1 : p(1-p)(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) \neq \mathbf{0} \end{cases} \tag{22}$$

where null hypothesis means homogeneity of population and alternative hypothesis means consisting population from 2 heterogeneous subpopulation as (3). If $\hat{\theta}_0$ and $(\hat{p}, \hat{\theta}_1, \hat{\theta}_2)$ are maximization of the likelihood function under null and alternative hypothesis, respectively, then the statistic that obtained from LRT is as follow:

$$R_n = 2\left\{log\,\ell_n(\hat{p},\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2) - log\,\ell_n(\hat{p},\hat{\boldsymbol{\theta}}_0,\hat{\boldsymbol{\theta}}_0)\right\} \tag{23}$$

Large value of this statistic leads to rejecting of null hypothesis.

For overcoming the boundary problem and non-identifiablity, Chen and Chen (2001) proposed a penalty function in terms of p as T(p) add to LRT statistic, such that:

$$\lim_{p\to 0 or 1} T(p) = -\infty, \qquad arg\,max\,T(p)_{p\in[0,1]} = 0.5 \tag{24}$$

and penalized Log-likelihood function is as follows:

$$T\ell_n(p,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \ell_n(\hat{p},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) + T(p) \tag{25}$$

Use of this penalty function lead to the fitted value of p under the modified likelihood that is bounded away from 0 or 1. As showed Chen and Chen (2001) under some regularity conditions asymptotic null distribution of MLRT statistic is the mixture of the χ_1^2 and χ_0^2 (χ_0^2 is a degenerate distribution with all its mass at 0) whit the same weights, i.e.,

$$0.5\chi_1^2 + 0.5\chi_0^2$$

By choosing different penalty function, the estimation of p in EM algorithm (Section 2.3) will be changed. Chen and Chen (2001) proposed a penalty function as follow:

$$T(p) = Clog(4p(1-p)) \tag{26}$$

where, C is a positive constant. Li *et al.* (2008) also used a penalty function as the following form:

$$T(p) = C^* log\left(1 - |1-2p|\right) \tag{27}$$

where, (26) and (27) hold in condition (24). It is simply shown that by using penalty functions (26) and (27), the value of p in M-step of EM algorithm can be implemented as follows, respectively:

$$\hat{p}^{(k+1)} = \frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)} + C}{2C+n} \tag{28}$$

and

$$\hat{p}^{(k+1)} = \begin{cases} min\{\frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)} + C^*}{n+C^*}, 0.5\} & if \ \frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)}}{n} < 0.5, \\ o.5 & if \ \frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)}}{n} = 0.5, \\ max\{\frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)}}{n+C^*}, 0.5\} & if \ \frac{\sum_{i=1}^{n}\hat{z}_{i1}^{(k)}}{n} > 0.5. \end{cases} \quad (29)$$

Optimal proposed penalty functoin: The penalty function (26) puts too much penalty on the mixing proportion when p is close to 0 or 1, then in spite of clear observation of mixture distribution, MLRT statictics can't reject null hypothesis. For overcoming this problem the penalty function (27) has been proposed. For p = 0.5 the inequality $log(1 - |1- 2p|)$ $log(1- |1 - 2p|^2) = log(4p(1- p))$ convert to $log(1- | 1- 2p|) \approx - |1- 2p|$. It means penalty functions (26) and (27) are almost equivalent for values of p that are close to 0.5 but have considerable difference for values of p that are close to 0 or 1. Farnoosh *et al*. (2012) proposed the penalty function that is able to overcome the disadvanyages of the penalty functions (26) and (27) as follows:

$$g(h,p) = Clog(1-|1-2p|^h), \quad 0 < h\ 2 \quad (30)$$

The penalty functions (26) and (27) are obtained by substituting h = 1 and h = 2 in (30), respectively. How to choose the value of h effects on the shape of penalty function and on the obtained inferences. Due to difficulty of classical approach in determining parameters of model and penalty function, these parameters will be estimated via the Baysian approach. Rang of parameter h in (30) is interval (0,2) but because the values of h in interval (0,1) may improve the power of MLRT the uniform distribution has been used as prior distribution for h.

Baysian inference: In this subsection, we implement the Baysian methodology using MCMC techniques for the FM-SN and penalty function. Let $y = (y_1,\cdots,y_n)^T$ is a random sample from a mixture:

$$pSN(\mu_1,\sigma_1^2,\lambda_1) + (1-p)SN(\mu_2,\sigma_2^2,\lambda_2)$$

with an unknown mixing proportion $p = (p_1 = p, p_2 = 1-p)$. Therefor from (8)-(10) penalized likelihood function of $\Theta = (p, \theta_1 = (\mu_1, \sigma_1^2, \lambda_1), \theta_2 = (\mu_2, \sigma_2^2, \lambda_2))$ associated with (y, t, z) is given by:

$$L^c(\mathbf{\Theta}, h \mid \mathbf{y},\mathbf{t},\mathbf{z}) = \prod_{i=1}^{n}\prod_{j=1}^{2}\left(p_j\psi(y_i \mid \mathbf{\theta}_j)\right)^{z_{ij}}$$

$$\propto p^{n_1}(1-p)^{n_2}\prod_{j=1}^{2}\prod_{i,z_{ij}=1}\phi(y_i;\mu_j+\Delta_j t_i,\Gamma_j)\phi(t_i;0,1)I(0,\infty)\left(1-|1-2p|^h\right) \quad (31)$$

where, the n_j (j = 1,2) are the number of observations associates to the j components. Now we need to consider prior distribution to all the unknown parameters μ, Δ^2, Γ,p and h, as follows:

$$\mu_j \sim N(\mu_{\mu_j},\sigma_{\mu_j}^2) \quad (32)$$

$$\Delta_j \sim N(\mu_{\Delta_j},\sigma_{\Delta_j}^2) \quad (33)$$

$$\Gamma_j^{-1} \sim Gamma(\frac{\rho}{2},\frac{\rho}{2}), \qquad j = 1,2 \quad (34)$$

$$p \sim U(0,1) \quad (35)$$

$$h \sim \alpha U(0,0.5) + (1-\alpha)U(0.5,1), \qquad 0 < \alpha < 1 \quad (36)$$

We assume prior independence between the parameters, such that the complete prior setting can be written as follow:

$$\pi(\mathbf{\mu},\mathbf{\Delta},\mathbf{\Gamma},\mathbf{p},h) \propto \left(\prod_{j=1}^{2}\pi(\mu_j)\pi(\Delta_j)\pi(\Gamma_j)\right)\pi(p)\pi(h) \quad (37)$$

Combining the likelihood function (31) and the prior distribution (37), the joint posterior density is:

$$\pi(\mathbf{\mu},\mathbf{\Delta},\mathbf{\Gamma},\mathbf{p},h \mid \mathbf{y},\mathbf{t},\mathbf{z})) \propto L^c(\mathbf{\Theta},h \mid \mathbf{y},\mathbf{t},\mathbf{z}) \times \pi(\mathbf{\mu},\mathbf{\Delta},\mathbf{\Gamma},\mathbf{p},h)$$
(38)

Distribution (38) doesn't have a closed form but MCMC method can be used to draw samples. Two generation mechanism for production such Markov Chains are Gibbs and Metropolis-Hastings. Since the Gibbs sampler may fail to escape the atttraction of the local mode (Marin *et al*., 2005) a standard alternative i.e., Metropolis-Hastings is used for sampling from posterior destribution. This algorithm simulates a Markov Chain such that the stationary distribution of this chain coincides with the target distribution. Its iterative steps for k = 1, 2, ... are:

- Choose $P^{(0)}$, $\theta^{(0)}$ and $h^{(0)}$,

- Generate $(\tilde{\mu}, \tilde{\Delta}, \tilde{\Gamma}, \tilde{p}, \tilde{h})$ from $q(\mu, \Delta, \Gamma, p, h \mid \mu^{(k-1)}, \Delta^{(k-1)}, \Gamma^{(k-1)}, p^{(k-1)}, h^{(k-1)})$,
- Compute

$$r = \frac{\pi(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Delta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\mathbf{p}}, \tilde{h} \mid \mathbf{y}, \mathbf{t}, \mathbf{z}) q(\boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Delta}^{(k-1)}, \boldsymbol{\Gamma}^{(k-1)}, \mathbf{p}^{(k-1)}, h^{(k-1)} \mid \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Delta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\mathbf{p}}, \tilde{h})}{\pi(\boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Delta}^{(k-1)}, \boldsymbol{\Gamma}^{(k-1)}, \mathbf{p}^{(k-1)}, h^{(k-1)} \mid \mathbf{y}, \mathbf{t}, \mathbf{z}) q(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Delta}}, \tilde{\boldsymbol{\Gamma}}, \tilde{\mathbf{p}}, \tilde{h} \mid \boldsymbol{\mu}^{(k-1)}, \boldsymbol{\Delta}^{(k-1)}, \boldsymbol{\Gamma}^{(k-1)}, \mathbf{p}^{(k-1)}, h^{(k-1)})}$$

- Generate U from U(0,1)
- if $r < u$ then $(\mu^{(k)}, \Delta^{(k)}, \Gamma^{(k)}, p^{(k)}, h^{(k)}) = (\tilde{\mu}, \tilde{\Delta}, \tilde{\Gamma}, \tilde{p}, \tilde{h})$

else $(\mu^{(k)}, \Delta^{(k)}, \Gamma^{(k)}, p^{(k)}, h^{(k)}) = (\mu^{(k-1)}, \Delta^{(k-1)}, \Gamma^{(k-1)}, p^{(k-1)}, h^{(k-1)})$. where, q is proposal distribution that often is considered as the random walk Metropolis-Hastings (Marin *et al*., 2005).

Table 1: MSE of maximum likelihood estimators

Parameters	p = 0.05	p = 0.1	p = 0.25	p = 0.5
p	0.2726	0.2340	0.1451	0.0204
$\hat{\mu}_1$	0.9338	0.5971	0.4019	0.2008
$\hat{\mu}_2$	11.2431	10.6838	9.0696	1.9269
$\widehat{\sigma_1^2}$	93.3313	84.1086	42.8225	12.8946
$\widehat{\sigma_2^2}$	43.3699	46.0698	40.2560	13.2444
$\hat{\lambda}_1$	2.9249	0.0058	0.0052	0.0071
$\hat{\lambda}_2$	0.0185	0.0172	0.0161	0.0091

Table 2: MSE of baysian estimators when $h \sim U(0,1)$

Parameters	p = 0.05	p = 0.1	p = 0.25	p = 0.5
$\hat{h}$	0.0006	0.0002	0.0036	0.0006
	(0.6761)	(0.3802)	(0.0807)	(0.6761)
$\hat{p}$	0.2650	0.1906	0.2068	0.0837
$\hat{\mu}_1$	0.0050	0.0054	0.0058	0.0055
$\hat{\mu}_2$	0.0031	0.0046	0.0052	0.0034
$\widehat{\sigma_1^2}$	0.4048	0.4137	0.3184	0.4058
$\widehat{\sigma_2^2}$	0.5136	0.4817	0.8792	0.5114
$\hat{\lambda}_1$	0.6049	0.5541	0.6560	0.6049
$\hat{\lambda}_2$	0.5242	0.4836	0.6328	0.5242

Table 3: MSE of baysian estimators when $h \sim 0.25U(0,0.5) + 0.75U(0.5,1)$

Parameters	p = 0.05	p = 0.1	p = 0.25	p = 0.5
$\hat{h}$	0.0005	0.0000007	0.0003	0.0011
	(0.4709)	(0.5060)	(0.6315)	(0.7428)
$\hat{p}$	0.3183	0.2464	0.1455	0.1064
$\hat{\mu}_1$	0.0049	0.0043	0.0060	0.0066
$\hat{\mu}_2$	0.0045	0.0061	0.0044	0.0044
$\widehat{\sigma_1^2}$	0.4164	0.5202	0.5129	0.3911
$\widehat{\sigma_2^2}$	0.3813	0.7096	0.6496	0.6604
$\hat{\lambda}_1$	0.8150	0.6936	0.8045	0.7792
$\hat{\lambda}_2$	0.4593	0.4638	0.6524	0.6678

Table 4: MSE of baysian estimators when $h \sim 0.75U(0,0.5) + 0.25U(0.5,1)$

Parameters	p = 0.05	p = 0.1	p = 0.25	p = 0.5
$\hat{h}$	0.00007	0.00005	0.0004	0.0012
	(0.5618)	(0.5518)	(0.401)	(0.2459)
$\hat{p}$	0.2852	0.2870	0.1454	0.0829
$\hat{\mu}_1$	0.0054	0.0051	0.0041	0.0052
$\hat{\mu}_2$	0.0062	0.0046	0.0052	0.0028
$\widehat{\sigma_1^2}$	0.4601	0.4546	0.5189	0.5430
$\widehat{\sigma_2^2}$	0.4803	0.4982	0.6287	0.6748
$\hat{\lambda}_1$	0.7856	0.7403	0.6868	0.7175
$\hat{\lambda}_2$	0.5166	0.4953	0.8476	0.4175

SIMULATION STUDY

In this section, we compare estimation of parameters using the penalty function (27) and penalty function that obtained from Baysian estimation. We simulate n = 500 samples from FM-SN distribution with parameters $\mu_1 = 15$, $\mu_2 = 20$, $\sigma_1^2 = 20$, $\sigma_2^2 = 16$, $\lambda_1 = 6$, $\lambda_2 = -4$ and use 4 different values p=0.05, 0.1, 0.25, 0.5 for p.

In Table 1 we present the Mean Square Error (MSE) of maximum likelihood estimation of mixture model parameters. Tables 2 to 4 present MSE of baysian estimation of parameters and estimated values of h; The distributions U(0,1), 0.25 U(0,0.5) + 0.75 U(0.5,1) and 0.75 U(0,0.5) + 0.25 U(0.5,1) are used as the prior distribution for h in Table 2 to 4, respectively.

The simulation results show that in FM-SN the mean square error of baysian estimator is less than MLE one, especially in estimation of scale parameters.

CONCLUSION

In this study, we propose an optimal penalty function for testing homogeneity of finite mixture of skew-normal distribution. Simulation study shows that Baysian approach is more effective than the classical approach in estimation of model parameters and determining of optimal penalty function.

REFERENCES

Azzalini, A., 1985. A class of distributions which includes the normal ones. Scand. J. Stat., 12: 171-178.

Azzalini, A. and A. Capitanio, 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. J. Roy. Stat. Soc. B, 65: 367-389.

Basso, R.M., V.H. Lachos, C.R. Cabral and P. Ghosh, 2010. Robust mixture modeling based on scale mixtures of skew-normal distributions. Comput. Stat. Data An., 54(12): 2926-2941.

Bohning, D., 2000. Computer-assisted Analysis of Mixtures and Applications. Chapman and Hall/CRC, Boca Raton.

Branco, M.D. and D.K. Dey, 2001. A general class of multivariate skew-elliptical distributions. J. Multivariate Anal., 79: 99-113.
Cabral, C.R.B., H. Bolfarine and J.R.G., Pereira, 2008. Bayesian density estimation using skew student-t-normal mixtures. Comput. Stat. Data An., 52: 5075-5090.
Chen, H. and J. Chen, 2001. The likelihood ratio test for homogeneity in the finite mixture models. Can. J. Stat., 29: 201-215.
Chen, H., J. Chen and J.D.K. Kalbfleisch, 2004. Testing for a finite mixture model with two components. J. Roy. Stat. Soc. B, 66: 95-115.
Chen, J., 1998. Penalized likelihood ratio test for finite mixture models with multinomial observations. Can. J. Stat., 26: 583-599.
Dempster, A., N. Laird and D. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B, 39: 1-38.
Diebolt, J. and C.P. Robert, 1994. Estimation of finite mixture distributions through Bayesian sampling. J. Roy. Stat. Soc. B, 56: 363-375.
Escobar, M.D. and M. West, 1995. Bayesian density estimation and inference using mixtures. J. Am. Stat. Assoc., 90: 577-588.
Farnoosh, R., M. Ebrahimi and A. Hajirajabi, 2012. Optimal penalty functions based on MCMC for testing homogeneity of mixture models. Res. J. Appl. Sci. Eng. Technol., 4(14): 2024-2029.
Henze, N., 1986. A probabilistic representation of the skew-normal distribution. Scand. J. Stat., 13: 271-275.
Lachos, V.H., P. Ghosh and R.B. Arellano-Valle, 2010. Likelihood based inference for skew normal independent linear mixed models. Stat. Sinica, 20: 303-322.
Li, P., J. Chen and P. Marrott, 2008. Non-finite Fisher information and homo-geneitty: The EM approach. Biometrika., 96: 411-426.
Lin, T.I., J.C. Lee and W.J. Hsieh, 2007a. Robust mixture modelling using the skew t distribution. Stat. Comput., 17: 81-92.
Lin, T.I., J.C. Lee and S.Y. Yen, 2007b. Finite mixture modelling using the skew normal distribution. Stat. Sinica, 17: 909-927.
Lin, T.I., 2009. Maximum likelihood estimation for multivariate skew normal mixture models. J. Multivariate Anal., 100: 257-265.
Liu, H.B. and X.M. Shao, 2003. Reconstruction of January to April mean temperature in the qinling mountain from 1789 to 1992 using tree ring chronologies. J. Appl. Meteorol. Sci., 14: 188-196.
Marin, J.M., K.L. Mengerson and C. Robert, 2005. Bayesian Modelling and Inference on Mixtures of Distributions. In: Dey, D. and C.R. Rao (Eds.), Handbook of Statistics Elsevier, 25: 459-507.
McLachlan, G.J. and D. Peel, 2000. Finite Mixture Models. Wiely, New York.
Richardson, S. and P.J. Green, 1997. On Bayesian analysis of mixtures with an unknown number of components. J. Roy. Stat. Soc. B, 59: 731-792.
Stephens, M., 2000. Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. Ann. Stat., 28: 40-74.