Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 5 (2015), pp. 3687-3701

http://www.ripublication.com

Statistical Estimation Methods for Unconditional Finite Normal Mixture Distribution

Zetty Ain Kamaruzzaman and Zaidi Isa

School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
Telephone number: +603-89213595 Fax number: +603-89254519
E-mail: zettyain@yahoo.com; zaidiisa@ukm.my
Corresponding Author: Zetty Ain Kamaruzzaman, School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia. E-mail: zettyain@yahoo.com

Abstract

Normal mixture distribution (NM) is arguably the most important mixture models, and also the most challenging technically. It has been successfully applied in many fields where the application is still expending. In this paper, we provide a tutorial exposition on expectation–maximization (EM) algorithm and Gibbs sampler for parameter estimation of unconditional finite Normal mixture distribution. Both methods are extremely useful for solving difficult computation problems especially in Normal mixture distribution case. Practical issues that arise in the use of EM algorithm and Gibb sampler are discussed, as well as variants of algorithm and programming that help to deal with these challenges. The purpose of this paper is to provide a good conceptual explanation of the statistical estimation methods with illustrative example so the reader can have a grasp of some of the basic principles and techniques.

Keywords: Normal mixture distribution; EM algorithm; Gibbs sampler

1. Introduction

Normal mixture distribution allows for great flexibility in capturing various density shapes; however, this same flexibility also turns out to lead to some estimation problems in practice. Fitting the parameters of Normal mixture distribution is one of the oldest estimation problems in statistical literature. However, attempts have been made to solve this problem. In particular, various estimation methods have been proposed to estimate the parameters in mixture models. Maximum Likelihood (ML) estimation, by far, is the most popular methods. Another important estimation method is the Bayesian approaches. Other techniques are graphical methods, method of moments and minimum-distance methods (see [1] for an exhaustive review of these methods).

In this paper, we provide a tutorial exposition on statistical estimation methods using the expectation–maximization (EM) algorithm and Gibbs sampler. Practical issues that arise in the use of both methods are discussed, as well as variants of algorithm and programming that help to deal with these challenges. The purpose of this paper is to provide a good conceptual explanation of the basic principles and techniques for the statistical estimation methods with illustrative example. We focus on these two themes: (1) statistical estimation methods and (2) modelling with unconditional finite Normal mixture distribution.

The paper is structured as follows. We present basic definitions and concepts of Normal mixture distribution and explore its distribution properties in Section 2. We briefly outline statistical software packages in Section 3. In Section 4, we discuss how to determine the number of components k in the mixtures. Next, in Section 5, we review two methods used to estimate the parameters of Normal mixture distribution. In section 6, we present a case study as the application. Lastly, Section 7 concludes.

2. Unconditional Finite Normal Mixture Distribution

Normal mixture distribution has gain increasing attention in various disciplines of knowledge. The earliest recorded application of Normal mixture distribution was undertaken by Simon Newcomb in his study in Astronomy in 1886 followed by Karl Pearson in his classic work on Method of Moments in 1894. A good introduction to the theory and applications of Normal mixture distribution can be found in [1-5].

Some attractive property of Normal mixture distribution is that it is flexible to accommodate various shapes of continuous distributions by adjusting its component weights, means and variances [6]. Other advantage is they maintain the tractability of Normal, have finite higher order moments, plus can capture excess kurtosis [7]. Besides, Normal mixture distribution can capture the structural change both in the mean and variance and it can be asymmetric [8]. Also, they are able to capture leptokurtic, skewed and multimodal characteristics.

In general, the cumulative distribution function (cdf) of a mixture of k Normal random variable X can be represented by

$$F(x) = \sum_{i=1}^{K} \pi_i \Phi\left(\frac{x-\mu_i}{\sigma_i}\right) \tag{2.1}$$

where Φ is the cdf of $N(0,1)$. Therefore its probability density function (pdf) is

$$f(x) = \sum_{i=1}^{K} \pi_i \phi\left(x; \mu_i, \sigma_i\right) \tag{2.2}$$

where, for $i = 1, ..., K$

$$\phi(x;\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{-(x-\mu_i)^2}{2\sigma_i^2}}$$

$$\sum_{i=1}^{K} \pi_i = 1 \quad and \quad 0 \le \pi_i \le 1$$

If X is a mixture of K Normal with pdf in (2.2), then its mean, variance, skewness and kurtosis are

$$\mu = \sum_{i=1}^{K} \pi_i \mu_i$$

$$\sigma^2 = \sum_{i=1}^{K} \pi_i \left(\sigma_i^2 + \mu_i^2\right) - \mu^2$$

$$\tau = \frac{1}{\sigma^3} \sum_{i=1}^{K} \pi_i \left(\mu_i - \mu\right)\left[3\sigma_i^3 + \left(\mu_i - \mu\right)^2\right] \tag{2.3}$$

$$\kappa = \frac{1}{\sigma^4} \sum_{i=1}^{K} \pi_i \left[3\sigma_i^4 + 6\left(\mu_i - \mu\right)^2 \sigma_i^2 + \left(\mu_i - \mu\right)^4\right]$$

In Normal mixture distribution, each component has its unique mean μ_i, standard deviation σ_i and weight (sometimes also called as probability or mixing parameter) π_i [9].

3. Statistical Software Packages

In this section, we briefly outline a few statistical software packages that capable of performing analysis of mixture models. There are many packages which specialize in fitting certain mixture models, but the packages we mention here provide a little more versatility with respect to the selection of functions they offer.

The R programming language has a few packages available for analyzing and implementing mixture models [10]. The “mclust” package [11] is pretty much standard for mixture models. One of the most powerful is “mixtools” [12], which, in addition to classic mixtures of parametric densities, handles mixtures of regressions and non-parametric mixtures. The “FlexMix” package [13] is very good at handling complicated situations. Otherwise, user can also do their programming and write own coding using R. The ability to read, understand, modify and write simple pieces of code is an essential skill for nowadays modern data analysis. Fortunately, writing code is not actually very hard, especially in R.

4. Number of Components

Determining the number of components k is a major issue in mixture modelling. Two commonly employed techniques in determining the numbers of components k are the information criterion and parametric bootstrapping of the likelihood ratio test statistic

values [14]. Majority of the estimation techniques assume that the number of components k, in the mixture is known at a priori where it is known before the estimation of parameters is attempted [2].
However, according to [12], these two techniques, likelihood approaches and Bayesian approaches can be used in assessing the number of components k when it is not known at a priori. In this case, there are few Bayesian procedures; one method is the Dirichlet process [15].
Other techniques are calibration checking, cross-validation and hypothesis testing [16]. For cross-validation, we need to do simple data-set splitting, where a randomly-selected half of the data is used to fit the model, and half to test. The basic idea is to split a data set into train and test. We fit the model using the training points, and then calculate the log-likelihood of the test points under the model. We pick the number of component which maximizes the likelihood of the data.
We should choose the minimum number of component. For example, if a two-component Normal mixture distribution seems good, we should not consider using more components as by going to three, four, etc. components, we improve the in-sample likelihood but we could expose ourselves to the danger of over-fitting. Besides, having so many parameters is not always desirable. It can lead to estimation problems and over-fitting the data can lead to specification problems.

5. Parameter Estimation

There are varieties of method for parameter estimation of finite mixture models. In this paper, we focus on the EM algorithm and Gibbs sampler.

5.1 EM Algorithm

The maximum likelihood method (MLE) is the most commonly preferred method for the estimation problem of Normal mixture distribution. Unfortunately, the MLEs have no closed forms; hence they have to be computed iteratively. However, the computation becomes straightforward using the expectation-maximization (EM) algorithm.
The earliest literature related to an EM-type algorithm appears in [17] with estimation of parameters of a mixture of two univariate Normal distribution. The formulation of EM algorithm is first introduced by [18]. The convergence and other basic properties of the EM algorithm under general conditions were established in their literature.
The EM algorithm is widely used as it is an easy and implementable method as well as a popular tool for simplifying difficult maximum likelihood problems plus has shown great performance in practice where it has the ability to deal with missing data, unobserved variables and mixture density problems. The EM algorithm will find the expected value as well as the current parameter estimates at the E step and maximizes the expectation at the M step. By repeating the E and M step, the algorithm will converge to a local maximum of the likelihood function. Various EM-type algorithms can be found in the literature (for example, see [5] and [19] for references).
Denote θ the parameters of the function to be optimized. The algorithm consists of iterating between two steps, the E-step and the M-step. In the Expectation (E) step, the

current estimates of the parameters are used to assign responsibilities according to the relative density of the training points under each model. Next, in the Maximization (M) step, these responsibilities are used in weighted maximum-likelihood fits to update the estimates of the parameters. The E-step is repeated, updated with a new value as the current value of θ and then the M-step again provides a further updated value for θ. Thus, the algorithm proceeds, iterating between the E-step and the M-step until convergence is achieved.

[20] introduce a simple procedure of the EM algorithm for the special case of Normal mixture distribution.

Algorithm 1. *HTF EM-Algorithm [20]*

1. *Take initial guesses for the parameters* $\hat{\pi}, \hat{\mu}_1, \hat{\sigma}_1^2, \hat{\mu}_2, \hat{\sigma}_2^2$
2. *Expectation (E) Step: Compute the responsibilities*

$$\gamma_i = \frac{\pi\phi\left(x_i; \mu_2, \sigma_2^2\right)}{\pi\phi_1\left(x_i; \mu_1, \sigma_1^2\right) + (1-\pi)\phi_2\left(x_i; \mu_2, \sigma_2^2\right)} \tag{5.1}$$

for $i = 1, \ldots, N.$

3. *Maximization (M) Step: Compute the weighted means and variances*

$$\mu_1 = \frac{\sum_{i=1}^{N}(1-\gamma_i)x_i}{\sum_{i=1}^{N}(1-\gamma_i)}, \quad \sigma_1^2 = \frac{\sum_{i=1}^{N}(1-\gamma_i)(x_i-\mu_1)^2}{\sum_{i=1}^{N}(1-\gamma_i)},$$

$$\mu_2 = \frac{\sum_{i=1}^{N}\gamma_i x_i}{\sum_{i=1}^{N}\gamma_t}, \quad \sigma_2^2 = \frac{\sum_{i=1}^{N}\gamma_i(x_i-\mu_2)^2}{\sum_{i=1}^{N}\gamma_i}, \tag{5.2}$$

$$\pi = \sum_{i=1}^{N}\frac{\gamma_i}{N}$$

4. *Iterate steps 2 and 3 until convergence*

For initial guess for the parameters, usually we take $\hat{\pi} = 0.5$, take two x_i randomly as the initial guesses for $\hat{\mu}_1$ and $\hat{\mu}_2$, and take $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \sum_{i=1}^{N}(x_i - \bar{x})^2 / N$. Details refer [20].

5.2 Gibbs Sampler

Gibbs sampler is a useful simulation method which generates sample from the posterior distribution. First, we present the posterior distribution for Normal mixture distribution under a conjugate prior. Next, we present the Gibbs sampler for both general case and Normal mixture application.

Calculating the posterior under conjugate prior for Normal mixture distribution is very complicated. One way is to introduce the zero-one component indicator vector variable into the model to make the calculation easier and the result more straightforward. However, there is no difference in parameter estimation result among original Normal mixture distribution model, latent model and indicator Normal mixture model through the EM algorithm.

Choosing a proper prior distribution is very important for Bayesian method, since the improper prior may not lead to an analytical tractable form of posterior. Specify a conjugate prior can guarantee the easily calculable form of the posterior.

Theorem 1. *(Conjugacy of the Normal, inverse Gamma and Beta prior for two-component Normal mixture distribution) In the Normal mixture distribution, a Normal prior along with a Normal mixture joint likelihood function produced a Normal posterior for the mean parameter; an inverse Gamma prior with the same Normal mixture likelihood produced an inverse Gamma posterior for the variance parameter; a Beta prior with the same Normal mixture likelihood produced a Beta posterior for the mixing parameter. [21]*

For two-component Normal mixture distribution, if suppose the prior distributions are

$$\mu_i \left| \sigma_i^2 \sim N\left(\xi_i, \frac{\sigma_i^2}{n_i}\right)\right.$$

$$\sigma_i^2 \sim IG\left(\frac{\nu_i}{2}, \frac{s_i^2}{2}\right) \tag{5.3}$$

$$\pi \sim Be(\alpha, \beta)$$

with density functions

$$p\left(\mu_i \left| \xi_i, \frac{\sigma_i^2}{n_i}\right.\right) \propto \left(\sigma_i^2\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma_i^2/n_i}\left(\mu_i - \xi_i\right)^2\right]$$

$$p\left(\sigma_i^2 \left| \frac{\nu_i}{2}, \frac{s_i^2}{2}\right.\right) \propto \left(\sigma_i^2\right)^{-\left(\frac{\nu_i}{2}+1\right)} \exp\left[-\frac{s_i^2}{2\sigma_i^2}\right]$$

$$p\left(\pi \left| \alpha, \beta\right.\right) \propto \pi^{\alpha-1}\left(1-\pi\right)^{\beta-1}$$

hence, the posterior distribution of $\left(\mu_i, \sigma_i^2, \pi\right)$ is

$$\mu_i \left| \sigma_i^2, \mathbf{x}, \mathbf{z} \sim N\left(\frac{n_i \xi_i + \overline{z}_i \overline{x}_i(z)}{n_i + \overline{z}_i}, \frac{\sigma_i^2}{n_i + \overline{z}_i}\right)\right.$$

$$\sigma_i^2 \left| \mathbf{x}, \mathbf{z} \sim IG\left(\frac{\nu_i + \overline{z}_i}{2}, \frac{1}{2}\left[s_i^2 + \sum_{i=1}^{N} z_{ij}\left(x_j - \overline{x}_i(z)\right)^2 + \frac{n_i \overline{z}_i\left(\overline{x}_i(z) - \xi_i\right)^2}{n_i + \overline{z}_i}\right]\right)\right. \tag{5.4}$$

$$\pi \left| \mathbf{x}, \mathbf{z} \sim Be\left(\overline{z}_2 + \alpha, \overline{z}_1 + \beta_2\right)\right.$$

where $\overline{z}_j = \sum_{i=1}^{N} z_{ij}$ and $\overline{x}_j(z) = \frac{1}{\overline{z}_j}\sum_{i=1}^{N} z_{ij} x_i$.

Algorithm 2. *Gibbs sampler for two-component Normal mixture distribution [21]*

1. Take initial values $\theta^{(0)} = \left(\pi^{(0)}, \mu_1^{(0)}, \mu_2^{(0)}, \left(\sigma_1^2\right)^{(0)}, \left(\sigma_2^2\right)^{(0)}\right)$, *where these parameters come from the prior distributions (5.3)*

2. Repeat for $t = 1, 2, \ldots$

a) *For* $i = 1, 2, ..., N$ *, generate* $z_i^{(t)} \in \{0,1\}$ *with*

$$z_i^{(t)} \sim Ber\left(\frac{\pi^{(t)}\varphi_{\theta_2}(x_i)}{\left(1-\pi^{(t)}\right)\varphi_{\theta_1}(x_i)+\pi^{(t)}\varphi_{\theta_2}(x_i)}\right)$$

b) *For* $j = 1, 2$ *, generate parameters as follows*

$$\pi^{(t+1)} \sim Be\left(\overline{z}_2^{(t)}+\alpha, \overline{z}_1^{(t)}+\beta_2\right)$$

$$\left(\sigma_j^2\right)^{(t+1)} \sim IG\left(\frac{\nu_j+\overline{z}_j^{(t)}}{2}, \frac{1}{2}\left[s_j^2+\sum_{i=1}^{N} z_{ij}^{(t)}\left(x_j-\overline{x}_i(z)^{(t)}\right)^2+\frac{n_j\overline{z}_j^{(t)}\left(\overline{x}_j(z)^{(t)}-\xi_j\right)^2}{n_j+\overline{z}_j^{(t)}}\right]\right)$$

$$\mu_j^{(t+1)} \sim N\left(\frac{n_j\xi_j+\overline{z}_j^{(t)}\overline{x}_j(z)^{(t)}}{n_j+\overline{z}_j^{(t)}}, \frac{\left(\sigma_j^2\right)^{(t+1)}}{n_j+\overline{z}_j^{(t)}}\right)$$

3. *Continue Step 2 until the joint distribution of* $\left(z^{(t)}, \theta^{(t)}\right)$ *does not change.*

Algorithm 3. *Gibbs sampler for two-component Normal mixture distribution with known variance and mixing parameter [20]*

1. *Take initial values* $\theta^{(0)} = \left(\mu_1^{(0)}, \mu_2^{(0)}\right)$
2. *Repeat for* $t = 1, 2, ...$

a) *For* $i = 1, 2, ..., N$ *, generate* $\Delta_i^{(t)} \in \{0,1\}$ *with* $\Pr\left(\Delta_i^t = 1\right) = \hat{\gamma}_i\left(\theta^{(t)}\right)$

b) *Set*

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{N}\left(1-\Delta_i^{(t)}\right).x_i}{\sum_{i=1}^{N}\left(1-\Delta_i^{(t)}\right)}, \hat{\mu}_2 = \frac{\sum_{i=1}^{N}\Delta_i^{(t)}.x_i}{\sum_{i=1}^{N}\Delta_i^{(t)}}$$

and generate $\mu_1^{(t+1)} \sim N\left(\hat{\mu}_1, \hat{\sigma}_1^2\right)$, $\mu_2^{(t+2)} \sim N\left(\hat{\mu}_2, \hat{\sigma}_2^2\right)$

3. *Continue Step 2 until the joint distribution of* $\left(\Delta^{(t)}, \mu_1^{(t)}, \mu_2^{(t)}\right)$ *does not change.*

6. Application

For illustration, in this section, we present a case study as a tutorial exposition on expectation–maximization (EM) algorithm and Gibbs sampler for parameter estimation of unconditional finite Normal mixture distribution using R programme.

In this section, we present the tips and tricks on how to estimate the parameters of Normal mixture distribution. For illustration, we present a case study using these three techniques: (i) manually programming using R (ii) ‘mclust’ package in R and (ii) ‘mixtools’ package in R. Data set used in this case study is the fictitious data from [20], which contains 20 data (Appendix A).

(i) Manually programming using R

Different initial guesses will lead to different iterative estimate results where the highest maximized likelihood is the best estimates. In section 5.1, we mentioned the common way to choose the initial guess for the parameters. [20] run the EM algorithm and received a best group of estimate values. The final maximum likelihood estimates, according to [20] are

$\hat{\pi} = 0.546, \hat{\mu}_1 = 4.62, \hat{\mu}_2 = 1.06, \hat{\sigma}_1^2 = 0.87, \hat{\sigma}_2^2 = 0.77$

We choose the best estimate values above to evaluate the known parameters of σ_1^2, σ_2^2 and π and also to be the initial guesses for μ_1 and μ_2. Next, we run Algorithm 1 and Algorithm 3 through R programme with initial guesses $\mu_1^{(0)} = 4.62$ and $\mu_2^{(0)} = 1.06$.

Table 1 summarize the result of the unknown parameters for the fictitious data set. Figure 1 depicts the mean value estimators, while Figure 2 depicts the density plot from Gibbs Sampler for the data set. From Table 1, Figure 1 and Figure 2, we find these two algorithms get quite a similar iterative estimation results. From Figure 2, the EM algorithm estimator, the mean value of the Gibbs sampler estimator and the Gibbs sampler estimator with the highest density are very close. However, the EM algorithm is faster and stable compared to the Gibbs sampler for this data set. From Figure 1, the EM algorithm converged in less than 10 steps, while the Gibbs sampler is still fluctuating widely after 200 iterations.

Table 1. The unknown parameters for the fictitious data

Estimation method	Algorithm	Unknown parameter	
		μ_1	μ_2
EM algorithm	Algorithm 1	4.6379	1.0686
Gibbs sampler	Algorithm 3	4.4360	0.9130

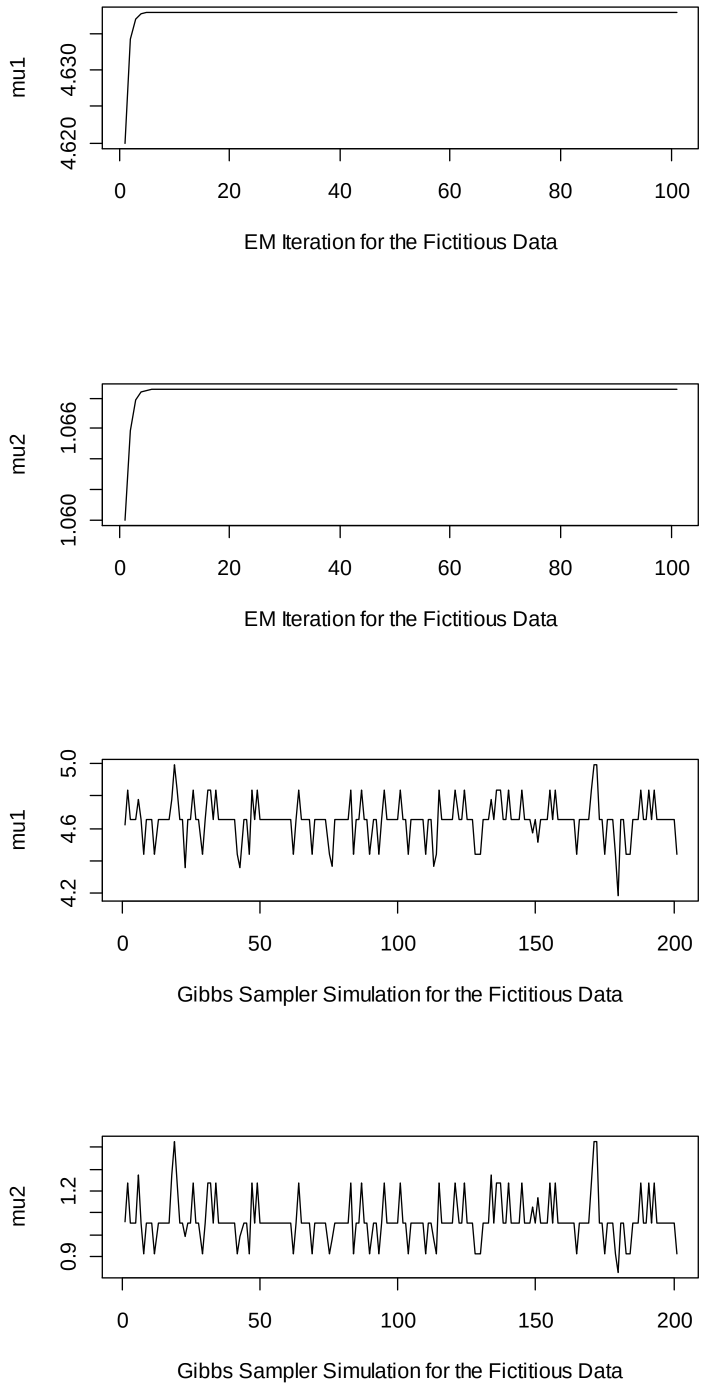

Figure 1. Mean value estimators for the fictitious data

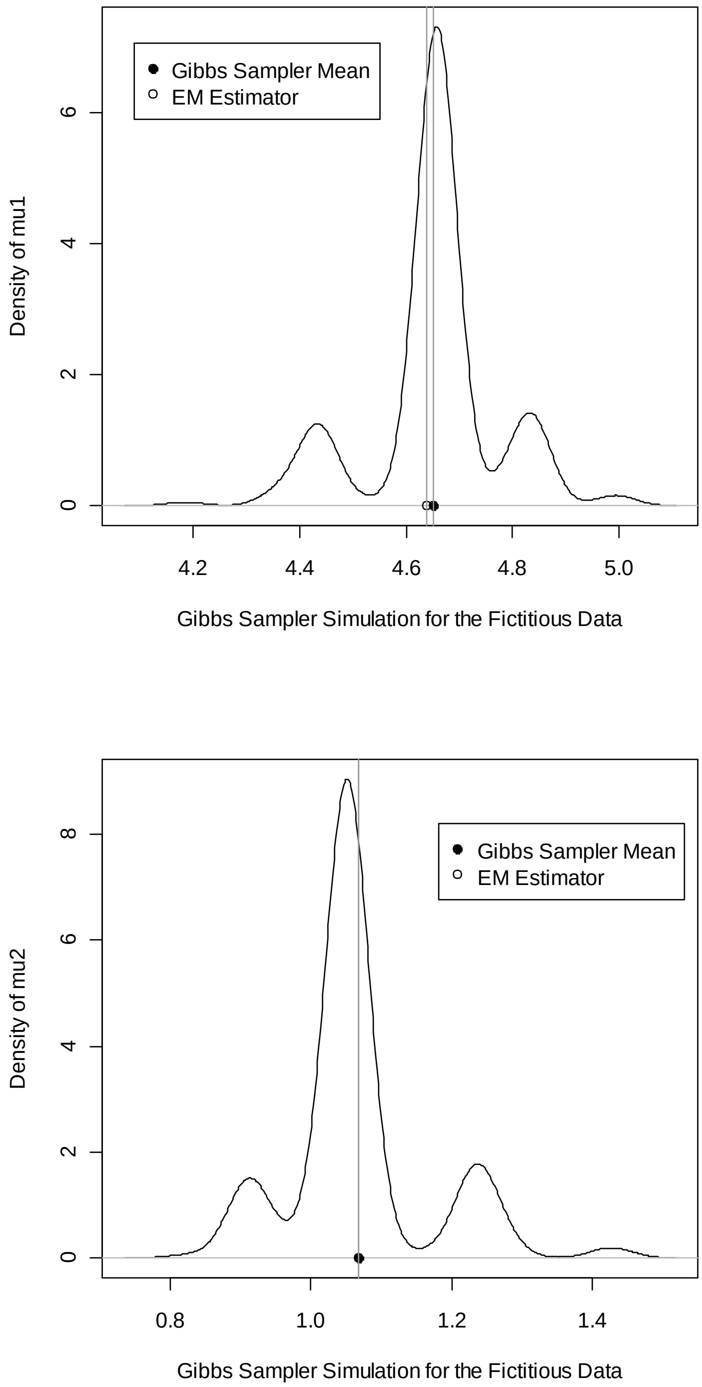

Figure 2. Density plots from Gibbs sampler for the fictitious data

Meanwhile, Table 2 reports the 95% confidence interval for μ_1 and μ_2 using the t-test to get the two-side confidence interval and Figure 3 depicts the box plot from Gibbs sampler for the fictitious data. Both show similar results.

Table 2. The 95% confidence interval for unknown parameters

Unknown parameter	95% confidence interval
μ_1	(4.6356,4.6691)
μ_2	(1.0534,1.0807)

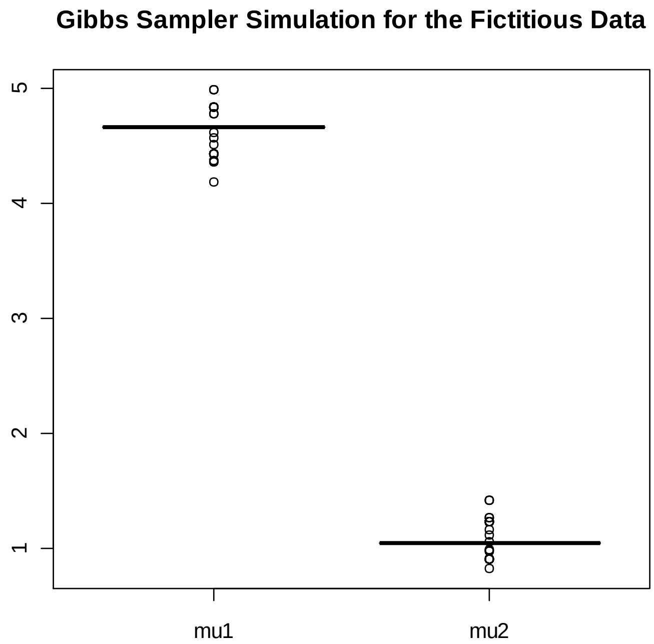

Figure 3. Box plot from Gibbs sampler for the fictitious data

(ii) 'mclust' package in R

This package implements the methodology by [22]. In this package, a Bayesian information criterion (BIC) is applied to choose the form of the mixture model. The best part is this package will report the best model for the data set instantaneously.

The best model using 'mclust' package for fictitious data set is an equal variance with two components. Table 3 reports the parameter estimates for the data set.

Table 3. Parameter estimates using 'mclust' package

Parameter	Estimates
π_1	0.5544
π_2	0.4456
μ_1	4.6552
μ_2	1.0827
$\sigma_1^2 = \sigma_2^2$	0.9026

Note: Best model: an equal variance with two components

(iii) 'mixtools' package in R

We could code up the EM algorithm for fitting mixture model from scratch, but instead we can just use 'mixtools' package. This package provides a set of functions for analyzing a variety of finite mixture models [23]. This package is based on EM algorithm or EM-like ideas.

Table 4 reports the parameters of Normal mixture distribution, number of iterations and overall log-likelihood using this package. From Table 4, the EM algorithm converged in 7 steps. Figure 4 depicts the calibration plot for two component Normal mixture distribution. It does look satisfactory.

Table 4. Estimation results using 'mixtools' package

Parameter	Component	
	1	2
π	0.5496	0.4504
μ	1.0635	4.6300
σ	0.8876	0.9228
Number of iterations	7	
Log-likelihood estimates	-38.9180	

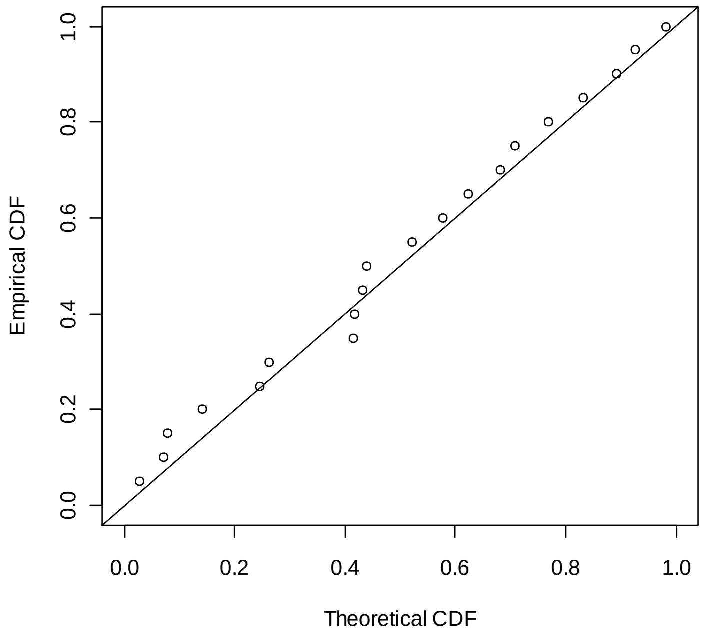

Figure 4. Calibration plot for the two component Normal mixture distribution

6.1 Discussion

The EM algorithm and Gibbs sampler are both good methods for parameter estimation. In our case study, they get similar results. The EM algorithm needs no prior information, faster and stable. Gibbs sampler is more complicated in computing and the selection of the prior parameter is important. However, the statistical software packages as illustrated in the case study can help user to deal with challenges in estimating and modelling using Normal mixture distribution.

7. Conclusion

In this paper, we provide a tutorial exposition on statistical estimation methods using the expectation–maximization (EM) algorithm and Gibbs sampler. First, we present basic definitions, concepts and distributional properties of Normal mixture distribution. We also briefly outline some statistical software packages particularly in R that can be used to estimate the parameter of Normal mixture distribution and determine the number of component in the mixture. Lastly, we present an illustrative example using the basic principles through few techniques as well as discuss some practical issues that arise in the use of both statistical estimation methods.

Acknowledgment
The authors want to thank anonymous referees whose comments, suggestions and corrections led to a considerably improvement of this paper.

References:

[1] Titterington, D.M., Smith, A.F.M., and Makov, U.E., 1985, Statistical Analysis of Finite Mixture Distributions, John Wiley & Sons, New York.
[2] Everitt, B.S., and Hand, D.J., 1981, Finite Mixture Distributions, Chapman and Hall, London.
[3] McLachlan, G.J., and Basford, K.E., 1988, Mixture Models: Inference and Applications to Clustering, Marcel Dekker, New York.
[4] Lindsay, B.G., 1995, "Mixture Models: Theory, Geometry and Applications," NSF-CBMS Regional Conference Series in Probability and Statistics, Institute of Mathematical Statistics, Hayward, CA, 5.
[5] McLachlan, G., and Peel, D., 2000, Finite Mixture Models, John Wiley & Sons, New York.
[6] Tan, K., and Chu, M., 2012, "Estimation of Portfolio Return and Value at Risk Using a Class of Gaussian Mixture Distributions," Int. J. Bus. Finance Res., 6(1), pp.97-107.
[7] Tsay, R.S., 2005, Analysis of Financial Time Series, Wiley Series in Probability and Statistics, New Jersey, pp.15-16.
[8] Knight, J., and Satchell, S., 2001, Return Distributions in Finance, Butterworth-Heinemann, Quantitative Finance Series, Oxford.
[9] Subramanian, S., and Rao, U.S., 2007, "Sensex and Stylized Facts: An Empirical Investigation," Soc. Sci. Res. Net., id. 962828.
[10] Everitt, B.S., and Hothorn, T., 2010, A Handbook of Statistical Analyses Using R, Taylor & Francis, London, pp.139-159.
[11] Fraley, C., Raftery, A.E., and Wehrens, R., 2006, mclust: Model-based Cluster Analysis, URL http://www.stat.washington.edu/mclust.
[12] Young, D.S., 2008, "An Overview of Mixture Models," Statist. Surveys, 0, pp.1-24.
[13] Leisch, F., 2004, "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R," J. Statistical Software, 8, pp.1-18.
[14] McLachlan, G.J., 1987, "On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture," Appl. Statist., 36, pp.318-324.
[15] Ferguson, T.S., 1973, "A Bayesian Analysis of Some Nonparametric Problems," Ann. Statist., 1(2), pp.209-230.
[16] Shalizi, C., Lecture 20, Mixture Examples and Complements, Carnegie Mellon University.
[17] Newcomb, S., 1886, "A Generalized Theory of The Combination of Observations So As to Obtain the Best Result," Amer. J. Mathematics, 8, pp.343-366.

[18] Dempster, A.P., Laird, N.M., and Rubin, D.B., 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," J. Royal Statistical Society Series B, 39(1), pp.1-38.

[19] McLachlan, G.J., and Krishnan, T., 1997, The EM Algorithm and Extensions, Wiley, New York.

[20] Hastie, T., Tibshirani, R. and Friedman, J., 2001, The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer Verlag, New York, pp.272-282.

[21] Liang, L., 2009, "On Simulation Methods for Two Component Normal Mixture Models Under Bayesian Approach," U.U.D.M. Project Report 1009: 17.

[22] Fraley, C., and Raftery, A.E. 2002, "Model Based Clustering, Discriminant Analysis and Density Estimation," J. Amer. Statistical Assoc., 97, pp.611-631.

[23] Benaglia, T., Chauveau, D., Hunter, D.R, and Young, D.S., 2009, "Mixtools: An R Package for Analyzing Finite Mixture Models," J. Statistical Software, 32(6), pp.1-29.

APPENDIX A

Fictitious data set [22] pp.273

Data
-0.39
0.12
0.94
1.67
1.76
2.44
3.72
4.28
4.92
5.53
0.06
0.48
1.01
1.68
1.80
3.25
4.12
4.60
5.28
6.22