Appl. Math. J. Chinese Univ.
2024, 39(2): 343-362

A unified Minorization-Maximization approach for estimation of general mixture models

HUANG Xi-fen[1,*] LIU Deng-ge[1] ZHOU Yun-peng[2] ZHU Fei[1]

Abstract. The mixed distribution model is often used to extract information from heterogeneous data and perform modeling analysis. When the density function of mixed distribution is complicated or the variable dimension is high, it usually brings challenges to the parameter estimation of the mixed distribution model. The application of MM algorithm can avoid complex expectation calculations, and can also solve the problem of high-dimensional optimization by decomposing the objective function. In this paper, MM algorithm is applied to the parameter estimation problem of mixed distribution model. The method of assembly and decomposition is used to construct the substitute function with separable parameters, which avoids the problems of complex expectation calculations and the inversion of high-dimensional matrices.

§1 Introduction

Due to the existence of heterogeneity in real life data, the mixture model has been widely applied in different fields to identify the hidden, latent classes. The books [15] did a thorough introduction on finite mixture models and provided applications in different areas. [5] made a summary of the early applications of both continuous mixed model and discrete mixed model. For the continuous mixture of distributions, many recent researches in clustering adopted the setting of Gaussian mixture by [1]. Similarly, such clustering technique can be used in the field of computer vision, [23] applied this model to detect objects from image. In addition, there are many other applications using different mixed distributions like gamma distributions. In clinical practice, [11] applied the mixture of two different distributions to estimate the maximum likelihood estimators for parameters of interest. [16] used Weibull mixture distribution to

Received: 2022-04-14. Revised: 2022-10-09.
MR Subject Classification: 62-08.
Keywords: MM algorithm, mixed distribution model, parameter estimation, assembly decomposition technology, parameter separation.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-024-4736-x.
Supported by the National Natural Science Foundation of China(12261108), the General Program of Basic Research Programs of Yunnan Province(202401AT070126), the Yunnan Key Laboratory of Modern Analytical Mathematics and Applications(202302AN360007) and the Cross-integration Innovation team of modern Applied Mathematics and Life Sciences in Yunnan Province, China(202405AS350003).
*Corresponding author.

model the heterogeneous data. Also, the mixed exponential distribution is used to model the loss distributions in actuarial studies [10]. For the discrete mixture of distributions, Poisson distribution is commonly applied in different applications. [9] discussed the properties of Poisson mixtures. Following this, [8] also developed an EM type of algorithm for parameter estimation and provided its application on crime data. As for the computational inefficiency caused by the complicated mixed distribution, [18] proposed the Monte Carlo sampling to avoid the numerical problem. Also, mixed Poisson distribution can be applied for the analysis of count data [4] and high-dimensional noisy data [6]. Also, in recent studies, the mixture model can be applied in other fields like modeling the wind speed [21] and capturing the stock price volatility [13].

The most general way to estimate the parameters in a mixture model is EM algorithm developed by [19] and the computer assisted mixture analysis is given by [2]. However, the M-step is computationally inefficient when we consider complex objective functions or high-dimensional covariates. [17] introduced an ECM algorithm which replaces the M-step with a computationally simpler CM-step which shares similar convergence properties where each parameter is maximized conditionally with other parameters being fixed. However, ECM cannot deal with the general class of mixture models given complex objective functions. Since the EM algorithm belongs to the class of MM algorithm, some researchers adopted this method to solve the parameter estimation problem. [20] applied MM algorithm to the two-component mixture model. For a general class of Gaussian mixture model (GMM), [14] provided an MM scheme which had a closed form for GMM parameters and an efficient optimization scheme for latent variables. Similarly, [7] introduced a new surrogate function for General Mixed Multinomial Logit Models to achieve efficient estimation. However, these methods cannot be applied to the general mixture models. Different from their approaches, [3] modified the CM-step based on the logarithm properties which extend the ECM optimization to a more general case with various linear constraints on parameter spaces.

The methods mentioned above do not target on the general case of mixture model. Therefore, in order to achieve efficient estimation with general objective functions, based on the work of [12], we decompose the objective function using [22]'s approach and apply the MM algorithm to avoid the complex expectation calculation and achieve efficient parameter estimation. We introduce the general MM algorithm and use the Normal, T, Gamma, Poisson and Geometric models to illustrate the performance of our MM algorithm for both continuous and discrete distributions by conducting simulations on the property of convergence and computational efficiency. Also, a real dataset is applied where the number of latent variables are determined using the BIC criterion.

The paper is organized as follows. In Section 2, we give a brief overview of the principles of Minorization-Maximization principle (MM algorithm). In Section 3, a general method for parameter estimation of mixture models is constructed using the MM algorithm. In Section 4, the parameter estimation of mixed Normal, T, Gamma, Poisson and Geometric distributions are studied based on the MM algorithm. In Section 5, we illustrate the convergence properties of the MM algorithm under mildly regular conditions. In Section 6, we apply the BIC criterion to determine the order of the mixture distribution and perform numerical simulations. In Section

7, we apply the MM algorithm to fit a real data case. The final discussion is in Section 8.

§2 MM algorithm

Assuming that Y_{obs} is the observed data, $\ell(\boldsymbol{\theta} \mid Y_{obs})$ is the log-likelihood function of Y_{obs}. The unknown parameter is $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T$ and the corresponding maximum likelihood estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \operatorname{argmax} \ell(\boldsymbol{\theta} \mid Y_{obs})$. MM algorithm [24] includes a minimization step and a maximization step. The minimization step is to construct a substitution function through a series of inequality scaling which satisfies

$$\begin{cases} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}) & \leqslant \ell(\boldsymbol{\theta} \mid Y_{obs}) \\ Q(\boldsymbol{\theta}^{(t)} \mid \boldsymbol{\theta}^{(t)}) & = \ell(\boldsymbol{\theta}^{(t)} \mid Y_{obs}) \end{cases}, \tag{1}$$

where $\boldsymbol{\theta}^{(t)}$ is the approximate value of the t-th iteration of $\hat{\boldsymbol{\theta}}$. By formula (1), the function $Q(\cdot \mid \boldsymbol{\theta}^{(t)})$ is always below $\ell(\cdot \mid Y_{obs})$, $Q(\cdot \mid \boldsymbol{\theta}^{(t)})$ is the tangent of $\ell(\cdot \mid Y_{obs})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$. Then, $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ is the substitution function of the objective function $\ell(\boldsymbol{\theta} \mid Y_{obs})$ in the $(t+1)$-th iteration. For the maximization step, the $(t+1)$-th approximation of $\hat{\boldsymbol{\theta}}$ can be obtained by maximizing the substitution function $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ where $\boldsymbol{\theta}^{(t+1)} = \operatorname{argmax} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$.

§3 The MM estimation method for Mixed Model

Assuming that the random variable X comes from a mixed distribution composed of m distributions $g_k(\cdot)$ with a proportional π_k, $(k = 1, \ldots, m)$ where $\sum_{k=1}^{m} \pi_k = 1$. The density function of this mixed distribution can be expressed as

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k g_k(x \mid \boldsymbol{\theta}_k), \tag{2}$$

$\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{\boldsymbol{\theta}_k\}_{k=1}^m\} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ is the parameter space. $g_k(x \mid \boldsymbol{\theta}_k)$ is density function from the k-th distribution with parameter $\boldsymbol{\theta}_k$. Then (2) is called m-order mixed distribution model. Assuming that $Y_{obs} = \{y_i\}_{i=1}^n$ is the observed data from the m-order mixed distribution model (2), the log-likelihood function can be decomposed into the following form

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k g_k(\boldsymbol{\theta}_k \mid y_i) = \ell_0(\boldsymbol{\nu}) + \sum_{i=1}^{n_1} \ell_{1i}(a_i^T h_i(\boldsymbol{\nu})) + \sum_{i=1}^{n_2} \ell_{2i}(r_i(\boldsymbol{\nu})), \tag{3}$$

where $\ell_0(\boldsymbol{\nu}) = \sum_{i=1}^{q} \ell_{0i}(\boldsymbol{\nu}_i)$ is completely additively separable, each $\ell_{0i}(\cdot)$ is a univariate function; $\ell_{1i}(\cdot)$ is a univariate concave function, $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ip})^T$ and $\{h_{ij}(\boldsymbol{\nu})\}_{j=1}^{p_i}$ may be nonlinear; $\ell_{2i}(\cdot)$ is a univariate convex function, and each $r_i(\cdot)$ is a linear combination of several low-dimensional functions. Under this setting, the objective function is decomposed into three parts and the appropriate inequality is chosen to scale each part according to its concavity or convexity in order to obtain the final substitution function. For mixed continuous distribution model or mixed discrete distribution model, the log-likelihood function can be written as

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log(a_i^T h_i(\boldsymbol{\nu} \mid y_i)) = \sum_{i=1}^{n} \ell_{1i}(a_i^T h_i(\boldsymbol{\nu} \mid y_i)).$$

Then we have $\ell_{1i}(\cdot) = \log(\cdot)$; $\boldsymbol{a_i} = (a_{i1}, \ldots, a_{im})^T$ and for all $k = 1, \ldots, m$, $a_{ik} = 1$; $\boldsymbol{h_i}(\boldsymbol{\nu} \mid y_i) = (h_{i1}(\pi_1, \boldsymbol{\nu}_1 \mid y_i), \ldots, h_{im}(\pi_m, \boldsymbol{\nu}_m \mid y_i))^T$, for all $k = 1, \ldots, m$, we have $h_{ik}(\pi_k, \boldsymbol{\theta}_k \mid y_i) = \pi_k g_k(y_i \mid \boldsymbol{\theta}_k)$. It can be seen that $\ell(\boldsymbol{\nu} \mid Y_{obs})$ only contains the second term in (3), and the substitution function that satisfies (1) can be obtained through Jensen's inequality. That

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m}\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log(\pi_k g_k(y_i \mid \boldsymbol{\theta}_k)) + c^{(t)},$$

where

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} g_k(y_i \mid \boldsymbol{\theta}_k^{(t)})$$

is the weight function, $w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \geqslant 0$ and $\sum_{i=1}^{m} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = 1$, $c^{(t)} = -\sum_{i=1}^{n}\sum_{i=1}^{m} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$ is a constant term that does not depend on $\boldsymbol{\nu}$. To solve the maximum likelihood estimation(MLE) of the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$, the function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ can be written as

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(\boldsymbol{\theta}_k \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)}, \tag{4}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k, and

$$Q_{2k}(\boldsymbol{\theta} \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log g_k(y_i \mid \boldsymbol{\theta}_k)$$

is a function of the other unknown parameters. Then, solving the MLE for $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ is equivalent to solving the MLE for $Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)})$ and $Q_{2k}(\boldsymbol{\theta} \mid \boldsymbol{\nu}^{(t)})$. Then the following iterative algorithm can be obtained.

Step 1. Let $\boldsymbol{\nu}^{(0)}$ be the initial value of $\boldsymbol{\nu}$.

Step 2. Update the estimation of π_k by maximizing $Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)})$. Update the estimation of $\boldsymbol{\theta}_k$ via maximizing $Q_{2k}(\boldsymbol{\theta}_k \mid \boldsymbol{\nu}^{(t)})$.

Step 3. Repeat step 2, until $\frac{1}{|\ell(\boldsymbol{\nu}^{(t)}|Y_{obs})| + 1} |\ell(\boldsymbol{\nu}^{(t+1)} \mid Y_{obs}) - \ell(\boldsymbol{\nu}^{(t)} \mid Y_{obs})| < \varepsilon$, where ε is a sufficiently small value.

§4 Applications

4.1 Mixed Normal Distribution

Assume that Y_{obs} is from the m-order mixed Normal distribution with a density function of

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}},$$

where $\pi_1, \ldots, \pi_m$ is the mixing ratio and $\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{\mu_k\}_{k=1}^m, \{\sigma_k\}_{k=1}^m\}$. Then the log-likelihood function of Y_{obs} is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(y_i-\mu_k)^2}{2\sigma_k^2}}.$$

The MLE of $\boldsymbol{\nu}$ can be obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$. Using the MM algorithm, via (4), the objective function of the $(t+1)$-th iteration is given by

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)}, \tag{5}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k,

$$Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[-\frac{\log \sigma_k^2}{2} - \frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right]$$

is a function of μ_k and σ_k^2, $c^{(t)}$ is a constant term that does not depend on $\boldsymbol{\nu}$, the weight function

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} \frac{1}{\sqrt{2\pi \sigma_k^{2(t)}}} e^{-\frac{(y_i - \mu_k^{(t)})^2}{2\sigma_k^{2(t)}}}.$$

Then, the maximization of the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$, namely solving the MLE of $\boldsymbol{\nu}$, can be turned to the maximization of the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$. By (5), the unknown parameters π_k, μ_k and σ_k^2 have been separated, where $Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)})$ is a function of π_k, and $Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)})$ is a function of μ_k and σ_k^2. We can obtain the $(t+1)$-th iteration of π is $\pi_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$, then we get the $(t+1)$-th iteration of μ_k and σ_k^2 by letting $\frac{\partial Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)})}{\partial \mu_k} = 0$ and $\frac{\partial Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)})}{\partial \sigma_k^2} = 0$ for $k = 1, \cdots, m$. Therefore, the final parameter iteration formula is

$$\begin{cases} \pi_k^{(t+1)} & = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \\ \mu_k^{(t+1)} & = \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) y_i \\ (\sigma_k^2)^{(t+1)} & = \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})(y_i - \mu_k^{(t)})^2 \end{cases} \tag{6}$$

4.2 Mixed T Distribution

Assume that Y_{obs} is from the m-order mixed T distribution with a density function of

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k \frac{\Gamma\left(\frac{1+v_k}{2}\right)}{\Gamma\left(\frac{v_k}{2}\right)} \left(\frac{1}{\pi \sigma_k^2 v_k} \right)^{\frac{1}{2}} \left[1 + \frac{(x - u_k)^2}{\sigma_k^2 v_k} \right]^{-\frac{1+v_k}{2}},$$

where $\pi_1, \ldots, \pi_m$ is the mixing ratio and $\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{u_k\}_{k=1}^m, \{\sigma_k\}_{k=1}^m, \{v_k\}_{k=1}^m\}$. We obtain a mixed T distribution with location parameter u_k, scale parameter σ_k and degrees of freedom v_k. Then the log-likelihood function of Y_{obs} is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k \frac{\Gamma\left(\frac{1+v_k}{2}\right)}{\Gamma\left(\frac{v_k}{2}\right)} \left(\frac{1}{\pi \sigma_k^2 v_k} \right)^{\frac{1}{2}} \left[1 + \frac{(y_i - u_k)^2}{\sigma_k^2 v_k} \right]^{-\frac{1+v_k}{2}}.$$

The MLE of $\boldsymbol{\nu}$ can be obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$. Using the MM algorithm, via (4), the objective function of the $(t+1)$-th iteration is given by

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)}, \tag{7}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k,

$$Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[\log \Gamma(\frac{1+v_k}{2}) - \log \Gamma(\frac{v_k}{2}) \right.$$
$$\left. -\frac{1}{2} \log \pi \sigma_k^2 v_k - \frac{1+v_k}{2} \log(1 + \frac{(y_i - u_k)^2}{\sigma_k^2 v_k}) \right]$$

is a function of u_k , σ_k^2 and v_k, $c^{(t)}$ is a constant term that does not depend on $\boldsymbol{\nu}$, the weight function

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} \frac{\Gamma(\frac{1+v_k^{(t)}}{2})}{\Gamma(\frac{v_k^{(t)}}{2})} (\frac{1}{\pi \sigma_k^{2(t)} v_k^{(t)}})^{\frac{1}{2}} [1 + \frac{(y_i - u_k^{(t)})^2}{\sigma_k^{2(t)} v_k^{(t)}}]^{-\frac{1+v_k^{(t)}}{2}}.$$

Then, the maximization of the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$, namely solving the MLE of $\boldsymbol{\nu}$, can be turned to the maximization of the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$. By (7), the unknown parameters π_k, u_k, σ_k^2 and v_k have been separated, where $Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)})$ is a function of π_k, and $Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)})$ is a function of u_k , σ_k^2 and v_k. We can obtain the $(t+1)$-th iteration of π is $\pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$, then we get the $(t+1)$-th iteration of u_k, σ_k^2 and v_k by letting $\frac{\partial Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)})}{\partial u_k} = 0$, $\frac{\partial Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)})}{\partial \sigma_k^2} = 0$ and $\frac{\partial Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)})}{\partial v_k} = 0$, $k = 1, \cdots, m$.

For v_k, we get

$$\frac{\partial Q_{2k}(u_k, \sigma_k^2, v_k \mid \boldsymbol{\nu}^{(t)})}{\partial v_k} = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[\frac{d\left(\log \Gamma\left(\frac{1+v_k}{2}\right)\right)}{dv_k} - \frac{d\left(\log \Gamma\left(\frac{v_k}{2}\right)\right)}{dv_k} \right.$$
$$\left. -\frac{1}{2} \log(1 + \frac{(y_i - u_k)^2}{\sigma_k^2 v_k}) + \frac{1}{2} \frac{(y_i - u_k)^2 - \sigma_k^2}{\sigma_k^2 v_k + (y_i - u_k)^2} \right] = 0$$

is nonlinear. Here we take the approach of most authors and only perform parameter estimation on the mixed T distribution with known degrees of freedom, and the MM estimation method for mixed T distribution model has been well represented. Let $u_{ik}^{(t)} = \frac{1+v_k^{(t)}}{v_k^{(t)} (\sigma_k^2)^{(t)} + (y_i - u_k^{(t)})^2}$, then, the final parameter iteration formula is

$$\begin{cases} \pi_k^{(t+1)} & = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \\ u_k^{(t+1)} & = \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) u_{ik}^{(t)}} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) u_{ik}^{(t)} y_i \\ (\sigma_k^2)^{(t+1)} & = \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) (\sigma_k^2)^{(t)} u_{ik}^{(t)} (y_i - u_k^{(t)})^2 \end{cases}.$$

4.3 Mixed Gamma Distribution

Assume that Y_{obs} is from the m-order mixed Gamma distribution with a density function of

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} x^{\alpha_k - 1} e^{-\beta_k x},$$

where $\pi_1, \ldots, \pi_m$ is the mixing ratio and $\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{\alpha_k\}_{k=1}^m, \{\beta_k\}_{k=1}^m\}$. The log-likelihood function of Y_{obs} is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} y_i^{\alpha_k - 1} e^{-\beta_k y_i}.$$

The MLE of $\boldsymbol{\nu}$ can be obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$. Using the MM algorithm, via (4), then the objective function of the $(t+1)$-th iteration can be given by

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \left(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(\alpha_k, \beta_k \mid \boldsymbol{\nu}^{(t)}) \right) + c^{(t)}, \tag{8}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k,

$$Q_{2k}(\alpha_k, \beta_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[\alpha_k \log \beta_k - \log \Gamma(\alpha_k) + (\alpha_k - 1) \log y_i - \beta_k y_i\right]$$

is a function of α_k and β_k, $c^{(t)}$ is a constant term that does not depend on $\boldsymbol{\nu}$, the weight function

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} \left(\frac{1}{\Gamma(\alpha_k^{(t)})} (\beta_k^{(t)})^{\alpha_k^{(t)}} y_i^{\alpha_k^{(t)} - 1} e^{-\beta_k^{(t)} y_i} \right).$$

Similar to the mixed Normal distribution, the maximization of $\ell(\boldsymbol{\nu} \mid Y_{obs})$ with solving the MLE of $\boldsymbol{\nu}$ can be turned to solve the MLE of $\boldsymbol{\nu}$ by maximizing the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$. By (8), where π_k, α_k and β_k have been separated. We can obtain the $(t+1)$-th iteration of π is $\pi_k^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$, then we get the $(t+1)$-th iteration of α_k and β_k by setting $\frac{\partial Q_{2k}(\alpha_k, \beta_k \mid \boldsymbol{\nu}^{(t)})}{\partial \alpha_k} = 0$ and $\frac{\partial Q_{2k}(\alpha_k, \beta_k \mid \boldsymbol{\nu}^{(t)})}{\partial \beta_k} = 0$ for $k = 1, \cdots, m$.

Since there is no explicit solution for α_k, Newton's method is applied to solve these equations. The final parameter iteration formula is

$$\begin{cases} \pi_k^{(t+1)} & = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \\ \alpha_k^{(t+1)} & = \alpha_k^{(t)} + \dfrac{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[(\Gamma(\alpha_k^{(t)}))^2 (\log \beta_k^{(t)} + \log y_i) - \Gamma(\alpha_k^{(t)}) \Gamma(\alpha_k^{(t)})' \right]}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[\Gamma(\alpha_k^{(t)}) \Gamma(\alpha_k^{(t)})'' - (\Gamma(\alpha_k^{(t)})')^2 \right]} \\ \beta_k^{(t+1)} & = \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) y_i} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \alpha_k^{(t)} \end{cases}.$$

4.4 Mixed Poisson Distribution

Suppose that Y_{obs} is the m-order mixed Poisson distribution from the density function

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k \frac{\lambda_k^x}{x!} e^{-\lambda_k},$$

where $\pi_1, \ldots, \pi_m$ is the mixing ratio and $\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{\lambda_k\}_{k=1}^m\}$. The log-likelihood function of Y_{obs} is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k \frac{\lambda_k^{y_i}}{y_i!} e^{-\lambda_k}.$$

The MLE of $\boldsymbol{\nu}$ can be obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$. Using the proposed MM algorithm to estimate the unknown parameters, via (4), the objective function of the $(t+1)$-th iteration is

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(\lambda_k \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)}, \tag{9}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k,

$$Q_{2k}(\lambda_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})(y_i \log \lambda_k - \lambda_k)$$

is a function of λ_k, $c^{(t)}$ is a constant term that does not depend on $\boldsymbol{\nu}$, the weight function

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} \frac{(\lambda_k^{(t)})^{y_i}}{y_i!} e^{-\lambda_k^{(t)}}.$$

Similar to the continuous mixed distribution in previous subsections, the original problem is transformed into solving the MLE of $\boldsymbol{\nu}$ by maximizing the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$. By (9), π_k and λ_k have been separated. We can obtain the $(t+1)$-th iteration of π is $\pi_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$, then we get the $(t+1)$-th iteration of λ_k by letting $\frac{\partial Q_{2k}(\lambda_k \mid \boldsymbol{\nu}^{(t)})}{\partial \lambda_k} = 0$, $k = 1, \cdots, m$. Finally the iterative formula is given by

$$\begin{cases} \pi_k^{(t+1)} &= \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \\ \lambda_k^{(t+1)} &= \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) y_i \end{cases}.$$

4.5 Mixed Geometric Distribution

Suppose that Y_{obs} is from the m-order mixed Geometric distribution with the density function

$$f(x \mid \boldsymbol{\nu}) = \sum_{k=1}^{m} \pi_k (1-p_k)^x p_k,$$

where $\pi_1, \ldots, \pi_m$ is the mixing ratio and $\boldsymbol{\nu} = \{\{\pi_k\}_{k=1}^m, \{p_k\}_{k=1}^m\}$. Thus, the log-likelihood function of Y_{obs} is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k (1-p_k)^{y_i} p_k.$$

The MLE of $\boldsymbol{\nu}$ can be obtained by maximizing the log-likelihood function $\ell(\boldsymbol{\nu} \mid Y_{obs})$. Using the MM algorithm to estimate the unknown parameters, the objective function of the $(t+1)$-th iteration from (4) is

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(p_k \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)}, \tag{10}$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k$$

is a function of π_k,

$$Q_{2k}(p_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[y_i \log(1-p_k) + \log p_k \right]$$

is a function of p_k, $c^{(t)}$ is a constant term that does not depend on $\boldsymbol{\nu}$, the weight function

$$w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) = \pi_k^{(t)} \frac{1}{f(y_i \mid \boldsymbol{\nu}^{(t)})} (1-p_k^{(t)})^{y_i} p_k^{(t)}.$$

Similar to continuous mixed distribution, here we maximize the substitution function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ with unknown parameters π_k and p_k separated by (10). We can obtain the $(t+1)$-th iteration of π is $\pi_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})$, then we get the $(t+1)$-th iteration of p_k by letting $\frac{\partial Q_{2k}(p_k \mid \boldsymbol{\nu}^{(t)})}{\partial p_k} = 0, \quad k = 1, \cdots, m$. Therefore, we have the iterative formula

$$\begin{cases} \pi_k^{(t+1)} &= \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \\ p_k^{(t+1)} &= \dfrac{1}{\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)})(y_i+1)} \displaystyle\sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \end{cases}.$$

§5 Convergence Properties of the Proposed MM Algorithms

In this section, we first denote $\ell(\boldsymbol{\nu} \mid Y_{obs})$ be the log-likelihood function to maximize and $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ be the corresponding surrogate function of $\ell(\boldsymbol{\nu} \mid Y_{obs})$, where $\boldsymbol{\nu}$ is the parameter vector and $\boldsymbol{\nu}^{(t)}$ is its current estimate. Let $M(\boldsymbol{\nu})$ be the maximizer of $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$. Following [25], we present the convergence properties of the MM algorithms in Section 4 based on the following regularity conditions.

C1. The parameter space Ω is an open set.

C2. $\ell(\boldsymbol{\nu} \mid Y_{obs})$ is differentiable, with continuous derivative $\ell'(\boldsymbol{\nu} \mid Y_{obs})$.

C3. The level set $\Omega_c = \{\boldsymbol{\nu} \in \Omega : \ell(\boldsymbol{\nu} \mid Y_{obs}) \geq c\}$ is compact.

C4. $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ is continuous in both $\boldsymbol{\nu}$ and $\boldsymbol{\nu}^{(t)}$, and differentiable in $\boldsymbol{\nu}$.

C5. All the stationary points of $\ell(\boldsymbol{\nu} \mid Y_{obs})$ are isolated.

C6. There exists a unique global maximum of $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$.

Then we provide a lemma ([26]) which gives general conditions for proving the convergence of an MM sequence.

Lemma 1 ([26]). Let $\boldsymbol{\nu}^{(t)}$, $t = 0, 1, 2, \cdots$ denote an MM sequence.

(i) If **C6** holds, then $M(\cdot)$ is continuous at $\boldsymbol{\nu}^{(t)}$.

(ii) If **C1-C6** hold, then for any starting value $\boldsymbol{\nu}^{(0)}$, $\boldsymbol{\nu}^{(t)} \to \boldsymbol{\nu}^{(*)}$ when $t \to \infty$, for some stationary point $\boldsymbol{\nu}^{(*)}$. Moreover, $M(\boldsymbol{\nu}^{(*)}) = \boldsymbol{\nu}^{(*)}$, and if $\boldsymbol{\nu}^{(t)} \neq \boldsymbol{\nu}^{(*)}$ for all t, the sequence of likelihood values $\ell(\boldsymbol{\nu}^{(t)} \mid Y_{obs})$ strictly increases to $\ell(\boldsymbol{\nu}^{(*)} \mid Y_{obs})$.

For all the mixed distributions of Section 4, it is easy to verify that the conditions **C1**, **C2** and **C4** are all satisfied. To verify **C3**, taking the mixed Normal distribution for example, we further denote its parameter space as $\mathcal{G} = \{\boldsymbol{\nu} = (\pi_1, \cdots, \pi_m, \mu_1, \cdots, \mu_m, \sigma_1, \cdots, \sigma_m) \mid \sum_{i=1}^{m} \pi_k = 1, \pi_k \geqslant 0, \sigma_k > 0, \text{ for } i = 1, \cdots, m\}$, and provide an extra condition A as follows.

Condition A.

(i) Define the parameter constraint as $\Psi = \{\boldsymbol{\nu} \in \mathcal{G} : |\mu_k| \leqslant a \text{ and } 0 < b \leqslant \sigma_k \leqslant d < \infty, k = 1, \cdots, m\}$, for some constants a, b, d, a=max $\{|y_1|, \cdots, |y_n|\}$.

Note that:

(i) If for any $\sigma_k \to 0$ or ∞, $\ell(\boldsymbol{\nu} \mid Y_{obs}) \to -\infty$, however our $\ell(\boldsymbol{\nu} \mid Y_{obs})$ is bounded, so σ_k can not go to 0 or ∞.

(ii) If $\boldsymbol{\nu} \in \mathcal{G}$ satisfies $\mu_k > a$, then we can find $\boldsymbol{\nu}' \in \mathcal{G}$ from $\boldsymbol{\nu}$ by setting the k-th mean component equal to a, *s.t.* $\ell(\boldsymbol{\nu} \mid Y_{obs}) \leqslant \ell(\boldsymbol{\nu}' \mid Y_{obs})$. Similarly, if $\mu_k < -a$, an analogous result holds by letting the k-th mean component equal to $-a$. It follows that

$$\sup_{\boldsymbol{\nu} \in \mathcal{G}} \ell(\boldsymbol{\nu} \mid Y_{obs}) = \sup_{\boldsymbol{\nu} \in \Psi} \ell(\boldsymbol{\nu} \mid Y_{obs}).$$

Theorem 1. For the mixed Normal distributions in Section 4, if Condition A, **C5** and **C6** hold at a moderate m, for any starting value $\boldsymbol{\nu}^{(0)}$, the sequences $\{\pi^{(t)}, \mu^{(t)}, \sigma^{2(t)}\}$ generated by the MM algorithm that updates the estimates by (6) are convergent.

Proof of Theorem 1. The log-likelihood function of mixed Normal distribution is

$$\ell(\boldsymbol{\nu} \mid Y_{obs}) = \sum_{i=1}^{n} \log \sum_{k=1}^{m} \pi_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(y_i - \mu_k)^2}{2\sigma_k^2}},$$

its corresponding surrogate function is

$$Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}) = \sum_{k=1}^{m} \Big(Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) + Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)}) \Big) + c^{(t)},$$

where

$$Q_{1k}(\pi_k \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \log \pi_k,$$

and

$$Q_{2k}(\mu_k, \sigma_k^2 \mid \boldsymbol{\nu}^{(t)}) = \sum_{i=1}^{n} w_{ik}(y_i \mid \boldsymbol{\nu}^{(t)}) \left[-\frac{\log \sigma_k^2}{2} - \frac{(y_i - \mu_k)^2}{2\sigma_k^2} \right].$$

The surrogate function $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$ satisfies the conditions in (1), that is

$$\begin{aligned} \ell(\boldsymbol{\nu} \mid Y_{obs}) &\geqslant Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)}), \ \forall \boldsymbol{\nu} \\ \ell(\boldsymbol{\nu}^{(t)} \mid Y_{obs}) &= Q(\boldsymbol{\nu}^{(t)} \mid \boldsymbol{\nu}^{(t)}). \end{aligned}$$

From the forms of $\ell(\boldsymbol{\nu} \mid Y_{obs})$ and $Q(\boldsymbol{\nu} \mid \boldsymbol{\nu}^{(t)})$, it is easy to verify that conditions **C1**, **C2** and **C4** hold. For the level set $\Omega_c = \{(\pi, \mu, \sigma^2) : \ell(\pi, \mu, \sigma^2 \mid Y_{obs}) \geqslant c\}$, it follows from the continuity of $\ell(\pi, \mu, \sigma^2 \mid Y_{obs})$ that Ω_c is closed. By imposing a constraint Condition **A**, we have Ω_c is bounded, then **C3** holds. So, under conditions **C5** and **C6**, the MM sequences that update the estimates by (6) are convergent. Similarly, we can verify **C3** in the similar way and provide the convergence properties for the MM sequences of other mixed distributions in Section 4.

§6 Simulation study

Numerical simulation is conducted using R to solve the parameter estimation problem by MM algorithm with mixed distribution model. During each experiment, the iteration stops when $\varepsilon = 10^{-6}$. In Table (2, 4, 6, 8, 10), I is the average number of iterations; L is the average value of sample log-likelihood function; T is the average value of run times (seconds); Bias denotes the deviation of parameter estimation; MSE denotes the mean square error.

In order to determine the number of distributions (the m-order of the mixture distribution model) in the mixture distribution model with a known distribution, BIC criterion is applied with the following objective

$$\mathrm{BIC} = d_{\boldsymbol{\nu}} \log n - 2\ell(\hat{\boldsymbol{\nu}} \mid Y_{obs}),$$

where $d_{\boldsymbol{\nu}}$ is the number of parameters in the mixed model (2), $\ell(\hat{\boldsymbol{\nu}} \mid Y_{obs})$ is the log-likelihood function of the sample in the mixed model (2) and n is the sample size. In addition, for the following four examples, the accuracy of order detection using the BIC criterion is shown by Table (1, 3, 5, 7, 9).

EXAMPLE 1. We construct a 3-order mixed Normal distribution model: $0.3N(1, 0.5^2) + 0.4N(5, 0.8^2) + 0.3N(9, 1^2)$, generate 100, 150, and 200 random samples respectively and perform 500 repeated experiments for each sample size. The steps to generate samples are as follows,

Step 1. Generate random variables X_1, X_2 and X_3 that subject to $N(1, 0.5^2), N(5, 0.8^2)$ and $N(9, 1^2)$, respectively.

Step 2. Generate a random number U with a uniform distribution of $U(0, 1)$, if $U \leqslant 0.3$, $X = X_1$; if $0.3 < U \leqslant 0.7$, $X = X_2$; otherwise $X = X_3$.

Step 3. Repeat step 1 and 2 to generate n random samples.

We first conduct 500 simulations to estimate the order of this mixed model at different sample sizes. The BIC value of the mixed Normal distribution is calculated when $m = 1, \ldots, 6$ in order to select the best m where BIC is minimized. $P(\hat{m} = 1), \ldots, P(\hat{m} = 6)$ denote the empirical percentages of $\hat{m} = 1, \ldots, 6$ in these 500 simulations. Table 1 shows that $\hat{m} = 3$ has the highest empirical percentages. When the sample size is 100, there exists 1.4% that the smallest BIC value is not obtained at $m = 3$. Besides, we conduct an additional simulation with a sample size equaling to 50 and find that the percentage of error is 35%. We can conclude that a small sample size will lead to a large error which discriminates the true order of the mixed normal distribution. But as the sample size increases, the accuracy of the determination of the order of the mixed Normal distribution using the BIC criterion increases.

Table 1. BIC Results of 3-order Mixed Normal.

Sample size	$P(\hat{m}=1)$	$P(\hat{m}=2)$	$P(\hat{m}=3)$	$P(\hat{m}=4)$	$P(\hat{m}=5)$	$P(\hat{m}=6)$
n=100	0.014	0	0.986	0	0	0
n=150	0	0	1	0	0	0
n=200	0	0	1	0	0	0

Then, based on the order estimated result, we present the estimation results of the other parameters with 500, 800 and 1000 simulation results, respectively, to show that 500 simulations are sufficient, where $m = 3$. In Table 2, the true value of parameters of the mixed Normal distribution model is $\mu_1 = 1$, $\mu_2 = 5$, $\mu_3 = 9$, $\sigma_1^2 = 0.25$, $\sigma_2^2 = 0.64$ and $\sigma_3^2 = 1$. The mixing ratio is $\pi_1 = 0.3$, $\pi_2 = 0.4$ and $\pi_3 = 0.3$. From Table 2, Bias and MSE are relatively small which shows the accuracy of estimated parameters under mixed normal model. Also, the consistency of estimated parameters is also verified by the decreasing trend of Bias and MSE as the increase of sample size. We also conducted 800 and 1000 simulations and the results are shown in Table 2. From Table 2, it can be easily found that there are not much differences in their results by running 500, 800 and 1000 replications. That means 500 replications is enough.

EXAMPLE 2. We construct a 3-order mixed T distribution model: $0.3T(2, 1^2, 3) + 0.4T(7, 1^2, 3) + 0.3T(11, 1^2, 3)$, generate 100, 150, and 200 random samples respectively and perform 500 repeated experiments for each sample size. The steps to generate samples are as follows.

Step 1. Generate random variables X_1, X_2 and X_3 that subject to $T(2, 1^2, 3), T(7, 1^2, 3)$ and $T(11, 1^2, 3)$, respectively.

Step 2. Generate a random number U with a uniform distribution of $U(0, 1)$, if $U \leqslant 0.3$, $X = X_1$; if $0.3 < U \leqslant 0.7$, $X = X_2$; otherwise $X = X_3$.

Step 3. Repeat step 1 and 2 to generate n random samples.

Analogous to example 1, we conduct 500 simulations to estimate the order of this mixed model. The BIC value of the mixed T distribution is calculated when $m = 1, \ldots, 6$ in order to select the best m where BIC is minimized. $P(\hat{m} = 1), \ldots, P(\hat{m} = 6)$ denote the empirical percentages of $\hat{m} = 1, \ldots, 6$ in these 500 simulations. Table 3 shows that $\hat{m} = 3$ has the highest empirical percentages. When the sample size is 100, there exists 3.8% that the smallest BIC value is not obtained at $m = 3$. When the sample size is 150, the percentage is 0.4% where the value of BIC is not the smallest at $m = 3$. Also, as the sample size increases, the accuracy of the determination of the order of the mixed T distribution using BIC criterion increases.

Then, based on the order estimated result, where $m = 3$. We present a comparison of the estimation results of other parameters between the EM algorithm and the MM algorithm. In Table 4, the true value of parameters of the mixed T distribution model is $u_1 = 2$, $u_2 = 7$, $u_3 = 11$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$ and $\sigma_3^2 = 1$, degrees of freedom are the same and known $v_1 = v_2 = v_3 = 3$. The mixing ratio is $\pi_1 = 0.3$, $\pi_2 = 0.4$ and $\pi_3 = 0.3$. From Table 4, for different sample sizes, and under the same sample data, it can be seen from the comparison that the EM algorithm needs more iterations, and on the whole, the MSE generated by the EM algorithm is larger.

Table 2. Simulation Results for Example 1.

0.3 Normal(1 , 0.5^2)+0.4 Normal(5 , 0.8^2)+0.3 Normal(9 , 1^2)						
500 simulations						
	n=100		n=150		n=200	
I	15.142		14.734		14.588	
L	-214.543		-324.039		-432.933	
T	0.004		0.004		0.005	
	Bias	MSE	Bias	MSE	Bias	MSE
μ_1	0.002	0.008	0.000	0.006	-0.000	0.004
μ_2	-0.005	0.025	-0.002	0.013	0.001	0.010
μ_3	0.015	0.053	-0.005	0.028	-0.002	0.021
σ_1^2	0.006	0.005	0.002	0.003	0.005	0.002
σ_2^2	0.007	0.037	0.004	0.022	0.004	0.017
σ_3^2	0.004	0.137	0.025	0.074	-0.000	0.056
π_1	-0.002	0.002	0.001	0.001	-0.000	0.001
π_2	0.003	0.003	0.001	0.002	0.000	0.001
π_3	-0.002	0.002	-0.002	0.001	0.000	0.001
800 simulations						
	n=100		n=150		n=200	
I	15.049		14.942		14.386	
L	-214.202		-323.327		-432.445	
T	0.003		0.004		0.005	
	Bias	MSE	Bias	MSE	Bias	MSE
μ_1	-0.003	0.008	-0.003	0.006	0.002	0.004
μ_2	0.004	0.023	-0.001	0.014	0.002	0.010
μ_3	0.008	0.059	0.004	0.033	-0.001	0.023
σ_1^2	0.014	0.005	0.006	0.003	0.008	0.002
σ_2^2	0.005	0.042	0.005	0.025	0.007	0.016
σ_3^2	0.006	0.160	0.018	0.081	0.009	0.053
π_1	-0.002	0.002	0.001	0.001	0.002	0.001
π_2	0.005	0.003	-0.001	0.002	-0.001	0.001
π_3	-0.003	0.002	-0.002	0.001	-0.001	0.001
1000 simulations						
	n=100		n=150		n=200	
I	15.112		14.504		14.416	
L	-214.467		-323.677		-432.458	
T	0.003		0.004		0.005	
	Bias	MSE	Bias	MSE	Bias	MSE
μ_1	0.002	0.009	0.000	0.006	-0.002	0.005
μ_2	-0.004	0.021	0.003	0.014	0.003	0.009
μ_3	0.011	0.050	0.008	0.034	0.004	0.023
σ_1^2	0.007	0.005	0.003	0.003	0.004	0.002
σ_2^2	0.005	0.042	0.009	0.023	0.009	0.017
σ_3^2	0.007	0.124	0.003	0.085	0.008	0.060
π_1	0.002	0.002	-0.000	0.001	-0.000	0.001
π_2	0.000	0.003	0.000	0.002	0.001	0.001
π_3	-0.002	0.002	0.000	0.001	-0.000	0.001

Table 3. BIC Results of 3-order Mixed T.

Sample size	$P(\hat{m}=1)$	$P(\hat{m}=2)$	$P(\hat{m}=3)$	$P(\hat{m}=4)$	$P(\hat{m}=5)$	$P(\hat{m}=6)$
n=100	0.004	0.034	0.962	0	0	0
n=150	0	0.004	0.996	0	0	0
n=200	0	0	1	0	0	0

This also verifies the simplicity and stability of the MM algorithm.

Table 4. EM and MM Comparison of Simulation Results for Example 2.

0.3 $T(2\ ,\ 1^2,\ 3)$+0.4 $T(7\ ,\ 1^2,\ 3)$+0.3 $T(11\ ,\ 1^2,\ 3)$												
	EM						MM					
	n=100		n=150		n=200		n=100		n=150		n=200	
I	27.786		24.962		23.984		13.832		12.818		12.186	
L	-257.512		-389.211		-519.869		-258.303		-389.981		-520.584	
T	0.008		0.008		0.010		0.003		0.004		0.006	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
μ_1	0.003	0.074	0.003	0.049	0.011	0.033	0.002	0.059	0.005	0.042	0.016	0.031
μ_2	0.006	0.078	0.011	0.047	-0.003	0.033	0.014	0.056	0.017	0.041	0.002	0.027
μ_3	0.022	0.110	0.010	0.060	0.007	0.047	0.003	0.069	0.004	0.041	-0.003	0.033
σ_1^2	0.026	0.235	0.000	0.175	-0.016	0.120	0.039	0.060	0.020	0.039	0.012	0.035
σ_2^2	-0.147	0.773	-0.087	0.362	-0.031	0.275	0.030	0.054	0.015	0.033	0.019	0.032
σ_3^2	-0.045	0.441	-0.020	0.208	-0.028	0.192	0.029	0.064	0.015	0.038	0.013	0.035
π_1	0.007	0.003	0.001	0.002	0.003	0.001	0.003	0.002	-0.001	0.001	0.003	0.001
π_2	-0.011	0.007	-0.002	0.004	0.003	0.003	-0.005	0.003	0.001	0.002	0.004	0.002
π_3	0.004	0.004	0.001	0.003	-0.006	0.002	0.002	0.003	-0.001	0.001	-0.006	0.001

EXAMPLE 3. We construct a 2-order mixed Gamma distribution model: $0.4Ga(40,20)+0.6Ga(6,1)$, generate 100, 150, and 200 random samples respectively and perform 500 repeated experiments for each sample size. The steps for generating random samples are as follows,

Step 1. Generate random variables $X_1 \sim Ga(40,20)$ and $X_2 \sim Ga(6,1)$ respectively.

Step 2. Generate a random number U with a uniform distribution of $U(0,1)$, if $U \leqslant 0.4$, $X = X_1$; otherwise, $X = X_2$.

Step 3. Repeat step 1 and 2 to generate n random samples.

Similar to example 1, for different sample sizes, we first conduct 500 simulations to estimate the order of this mixed model. The BIC value of the mixed Gamma distribution is calculated when $m = 1, \ldots, 6$ in order to select the best m where BIC is minimized. $P(\hat{m}=1), \ldots, P(\hat{m}=6)$ denote the empirical percentages of $\hat{m} = 1, \ldots, 6$ in these 500 simulations. From Table 5, $\hat{m} = 2$ has the highest empirical percentages. When the sample size is 100 and 150, the percentage is 0.6% and 0.4% where the value of BIC is not the smallest at $m = 2$. When the sample size is larger, the BIC criterion is more accurate in determining the order of the mixed Gamma distribution.

Then, the result of other parameters when $m = 2$ is shown at Table 6. The true value of

Table 5. BIC Results of 2-order Mixed Gamma.

Sample size	$P(\hat{m}=1)$	$P(\hat{m}=2)$	$P(\hat{m}=3)$	$P(\hat{m}=4)$	$P(\hat{m}=5)$	$P(\hat{m}=6)$
n=100	0	0.994	0.006	0	0	0
n=150	0	0.996	0.004	0	0	0
n=200	0	1	0	0	0	0

Table 6. Simulation Results for Example 2.

0.4 Gamma(40,20)+0.6 Gamma(6,1)						
	n=100		n=150		n=200	
I	339.922		314.64		280.958	
L	-198.643		-299.934		-399.856	
T	0.079		0.082		0.088	
	Bias	MSE	Bias	MSE	Bias	MSE
α_1	-1.954	137.632	-0.798	89.773	-0.152	64.097
α_2	-0.678	3.296	-0.476	2.323	-0.254	1.195
β_1	-0.968	36.902	-0.367	23.973	-0.039	17.235
β_2	-0.106	0.086	-0.074	0.058	-0.037	0.029
π_1	-0.004	0.003	0.000	0.002	-0.004	0.002
π_2	0.004	0.003	-0.000	0.002	0.004	0.002

parameters of the mixed Gamma distribution model is $\alpha_1 = 40$, $\alpha_2 = 6$, $\beta_1 = 20$ and $\beta_2 = 1$. The mixing ratio is $\pi_1 = 0.4$, and $\pi_2 = 0.6$. For the mixed Gamma distribution model, we can also observe the decreasing trend of both Bias and MSE with the increase of sample size which shows the convergence of estimated parameters to true values.

EXAMPLE 4. We construct a 3-order mixed Poisson distribution model: $0.3P(5) + 0.3P(16) + 0.4P(59)$, generate 50, 100, and 150 random samples respectively and perform 500 repeated experiments for each sample size. The steps for generating random samples are as follows,

Step 1. Generate random variables X_1, X_2 and X_3 that subject to $P(5), P(16)$ and $P(59)$, respectively.

Step 2. Generate a random number U with a uniform distribution of $U(0,1)$, if $U \leqslant 0.3$, $X = X_1$; if $0.3 < U \leqslant 0.6$, $X = X_2$; otherwise, $X = X_3$.

Step 3. Repeat step 1 and 2 to generate n random samples.

Similar to previous examples, the BIC criterion is used. $P(\hat{m}=1), \ldots, P(\hat{m}=6)$ denote the empirical percentages of $\hat{m} = 1, \ldots, 6$, respectively. From Table 7, $\hat{m} = 3$ has the highest empirical percentages. And when the sample size is 50, the proportion is 1% that the value of BIC is not the smallest at $m = 3$. When the sample size is 100 and 150, the percentage of wrong order detection is 0.4% and 0.2%. Even for a small sample size, the BIC criterion can accurately determine the order of the mixed Poisson distribution.

Then, the simulation result for other parameter estimation given $m = 3$ is presented by Table 8. The true value of parameters of the mixed Poisson distribution model is $\lambda_1 = 5$,

Table 7. BIC Results of 3-order Mixed Poisson.

Sample size	$P(\hat{m}=1)$	$P(\hat{m}=2)$	$P(\hat{m}=3)$	$P(\hat{m}=4)$	$P(\hat{m}=5)$	$P(\hat{m}=6)$
n=50	0	0.004	0.99	0.006	0	0
n=100	0	0	0.996	0.004	0	0
n=150	0	0	0.998	0.002	0	0

Table 8. Simulation Results for Example 3.

0.3 Poisson (5)+0.3 Poisson (16)+0.4 Poisson (59)						
	n=50		n=100		n=150	
I	7.072		7.032		6.924	
L	-192.965		-389.167		-584.426	
T	0.001		0.002		0.002	
	Bias	MSE	Bias	MSE	Bias	MSE
λ_1	0.013	0.488	0.009	0.239	-0.013	0.173
λ_2	-0.035	1.425	-0.010	0.704	0.010	0.471
λ_3	0.067	2.747	0.033	1.311	0.000	0.958
π_1	-0.003	0.005	-0.001	0.002	-0.000	0.002
π_2	0.000	0.005	0.002	0.002	0.001	0.002
π_3	0.003	0.005	-0.001	0.002	-0.001	0.001

$\lambda_2 = 16$ and $\lambda_3 = 59$. The mixing ratio is $\pi_1 = 0.3$, $\pi_2 = 0.3$ and $\pi_3 = 0.4$. When the sample size is small, the MSE of λ_1, λ_2, and λ_3 appear to be slightly large, but all the Biases are relatively the small. However, as the sample size increases, MSE decreases. The estimated parameters for mixed Poisson model are accurate and consistent.

EXAMPLE 5. We construct a 2-order mixed Geometric distribution model: $0.4Ge(0.1) + 0.6Ge(0.7)$, generate 50, 100, and 150 random samples respectively and perform 500 repeated experiments for each sample size. The steps for generating random samples are as follows,

Step 1. Generate random variables $X_1 \sim Ge(0.1)$ and $X_2 \sim Ge(0.7)$ respectively.

Step 2. Generate a random number U with a uniform distribution of $U(0,1)$, if $U \leqslant 0.4$, $X = X_1$; otherwise, $X = X_2$.

Step 3. Repeat step 1 and 2 to generate n random samples.

For the 2-order mixed Geometric distribution model, we conduct simulations to estimate the order. For the sample sizes is 50, 100, and 150, BIC are calculated correspondingly to determine the $\hat{m}$ given smallest BIC. Table 9 shows that $\hat{m} = 2$ has the highest empirical percentages. When the sample size is 50, 3.8% of the smallest BIC is not the obtained at $m = 2$. When sample size is 100 and 150, The percentage of error is 0.2%. When the sample size is larger, the BIC criterion is more accurate in determing the order of the mixed Geometric distribution.

Then, the simulation result for other parameter estimation given $m = 2$ is presented by Table 10. The true value of parameters of the mixed Geometric distribution model is $p_1 = 0.1$ and $p_2 = 0.7$. The mixing ratio is $\pi_1 = 0.4$ and $\pi_2 = 0.6$. As the sample size increases, the Bias continues to decrease and finally stabilizes around a small value which denotes the parameter

Table 9. BIC Results of 2-order Mixed Geometric.

Sample size	$P(\hat{m}=1)$	$P(\hat{m}=2)$	$P(\hat{m}=3)$	$P(\hat{m}=4)$	$P(\hat{m}=5)$	$P(\hat{m}=6)$
n=50	0.038	0.962	0	0	0	0
n=100	0.002	0.998	0	0	0	0
n=150	0	0.998	0.002	0	0	0

Table 10. Simulation Results for Example 4.

0.4 Geometric (0.1)+0.6 Geometric (0.7)						
	n=50		n=100		n=150	
I	17.792		14.49		13.278	
L	-108.002		-217.421		-326.571	
T	0.002		0.002		0.002	
	Bias	MSE	Bias	MSE	Bias	MSE
p_1	-0.006	0.001	-0.003	0.000	-0.001	0.000
p_2	-0.023	0.015	-0.009	0.007	-0.004	0.005
π_1	-0.010	0.012	-0.004	0.006	-0.000	0.005
π_2	0.010	0.012	0.004	0.006	0.000	0.005

estimates are approaching the true value. The MSE is also relatively small with a decreasing trend.

§7 Case Analysis

Cohort Study Data in Northeast Thailand. This data set with the name "thai_cohort" can be obtained from the R package CAMAN. The data set comes from a cohort study in northeastern Thailand. It has records from June 1982 to September 1985 where 602 people are checked every two weeks. The health data for the health status of school-age children record the number of times that the children had a fever, cough, runny nose or all three symptoms during this period.

For this real data, we first attempt a simple model to fit the data. For the count data, Poisson distribution ($P(\lambda)$) is a commonly used statistical model. The MM algorithm is applied and the estimated parameter $\hat{\lambda} = 4.449$. Adding the fitted probability mass function curve to the frequency distribution histogram (Figure 1), it can be seen that a single Poisson distribution cannot fit the data well. An m-order mixed Poisson distribution model might be a better choice

for this data. Then, we use the BIC criterion to determine the order of the mixed distribution model. According to the BIC results of this data set from Table 11, when $m = 4$, the BIC results reach the smallest value.

Table 11. BIC Results of thai_cohort.

m	1	2	3	4	5	6
BIC	4277.244	3292.663	3174.983	3158.864	3171.681	3184.644

Therefore, a 4-order mixed Poisson distribution model should be fitted to this data set. Using the MM algorithm proposed in this paper to estimate the parameters of the data, the 4-order mixed Poisson distribution can be obtained as

$$f(x \mid \hat{\boldsymbol{\nu}}) = 0.19P(0.12) + 0.48P(2.76) + 0.27P(8.08) + 0.06P(16.08)$$

Adding the fitting probability mass function of the model to the frequency distribution histogram, it can be seen that $f(x \mid \hat{\boldsymbol{\nu}})$ can fit this data better.

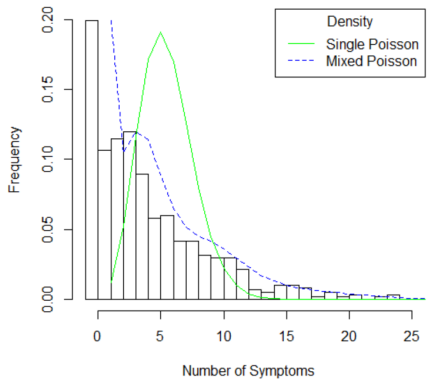

Figure 1. Frequency distribution histogram of thai_cohort.

§8 Discussion and Concluding Remarks

The mixed distribution model can accurately analyze the heterogeneous data, and the parameter estimation of mixed distribution models is usually based on maximum likelihood estimation. MM algorithm uses the assembly decomposition technique to separate the parameters of the objective function and construct the substitution function, which can deal with the parameter estimation of mixed distribution model very well. At the same time, the order of the mixed distribution model can be accurately found by the BIC criterion. The proposed method is suitable for both mixed continuous and mixed discrete distribution models. Through different numerical simulation experiments, we have verified that the MM algorithm has good results in the process of solving the parameter estimation of the mixed distribution model.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

[1] J D Banfield, A E Raftery. *Model-based Gaussian and non-Gaussian clustering*, Biometrics, 1993, 49(3): 803-821.

[2] D Bohning, P Schlattmann, B Lindsay. *Computer-assisted analysis of mixtures (CA MAN): statistical algorithms*, Biometrics, 1992, 48(1): 283-303.

[3] D Chauveau, D R Hunter. *ECM and MM algorithm for mixtures with constrained parameters*, 2011, https://hal.science/hal-00625285v1.

[4] R C Gupta, S H Ong. *Analysis of long-tailed count data by Poisson mixtures*, Communications in Statistics-Theory and Methods, 2005, 34(3): 557-573.

[5] C M Harris. *On finite mixtures of geometric and negative binomial distributions*, Communications in Statistics-Theory and Methods, 1983, 12(9): 987-1007.

[6] G Haro, G Randall, G Sapiro. *Translated poisson mixture model for stratification learning*, International Journal of Computer Vision, 2008, 80(3): 358-374.

[7] J James. *MM algorithm for general mixed multinomial logit models*, Journal of Applied Econometrics, 2017, 32(4): 841-857.

[8] D Karlis, L Meligkotsidou. *Finite mixtures of multivariate Poisson distributions with application*, Journal of Statistical Planning and Inference, 2007, 137(6): 1942-1960.

[9] D Karlis, E Xekalaki. *Mixed poisson distributions*, International Statistical Review/Revue Internationale de Statistique, 2005, 73(1): 35-58.

[10] C L Keatinge. *Modeling losses with the mixed exponential distribution*, In Proceedings of the Casualty Actuarial Society, 1999, 86: 654-698.

[11] S W Lagakos. *A stochastic model for censored-survival data in the presence of an auxiliary variable*, Biometrics, 1976, 32(3): 551-559.

[12] K Lange, D R Hunter, I Yang. *Optimization transfer using surrogate objective functions*, Journal of Computational and Graphical Statistics, 2000, 9(1): 1-20.

[13] R Liesenfeld. *A generalized bivariate mixture model for stock price volatility and trading volume*, Journal of Econometrics, 2001, 104(1): 141-178.

[14] K L Lim, H Wang, Z X Mou. *Learning Gaussian mixture model with a maximization-maximization algorithm for image classification*, In Proceedings of the 2016 12th IEEE International Conference on Control and Automation (ICCA), 2016, 887-891.

[15] B G Lindsay. *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, 1995, 5: 1-163.

[16] J M Marin, M Rodriguez-Bernal, M P Wiper. *Using weibull mixture distributions to model heterogeneous survival data*, Communications in Statistics-Simulation and Computation, 2005, 34(3): 673-684.

[17] X L Meng, D B Rubin. *Maximum likelihood estimation via the ECM algorithm: A general framework*, Biometrika, 1993, 80(2): 267-278.

[18] S H Ong, W J Lee, Y C Low. *A general method of computing mixed Poisson probabilities by Monte Carlo sampling*, Mathematics and Computers in Simulation, 2020, 170: 98-106.

[19] R A Redner, H F Walker. *Mixture densities, maximum likelihood and the EM algorithm*, SIAM Review, 1984, 26(2): 195-239.

[20] Z Shen, M Levine, Z Shang. *An MM algorithm for estimation of a two component semiparametric density mixture with a known component*, Electronic Journal of Statistics, 2018, 12(1): 1181-1209.

[21] J Y Shin, T B Ouarda, T Lee. *Heterogeneous mixture distributions for modeling wind speed, application to the UAE*, Renewable Energy, 2016, 91: 40-52.

[22] G L Tian, X F Huang, J Xu. *An assembly and decomposition approach for constructing separable minorizing functions in a class of MM algorithms*, Statistica Sinica, 2019, 29(2): 961-982.

[23] Z Zivkovic. *Improved adaptive Gaussian mixture model for background subtraction*, In Proceedings of the 17th International Conference on Pattern Recognition, 2004, 2: 28-31.

[24] D Hunter, K Lange. *Quantile Regression via an MM Algorithm*, Journal of Computational and Graphical Statistics, 2000, 9(1): 60-77.

[25] X F Huang, J Xu, G L Tian. *On profile MM algorithms for gamma frailty survival models*, Statistica Sinica, 2019, 29(2): 895-916.

[26] F Vaida. *Parameter convergence for EM and MM algorithms*, Statistica Sinica, 2005, 15(3): 831-840.

[1]School of Mathematics, Yunnan Normal University, Kunming 650092, China.
Email: 190004@ynnu.edu.cn
[2]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China.
Email: u3514104@connect.hku.hk