World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Bayesian inference approach for a mixture of normal regression models

Abdelbaset A.Sh Abdalla, Nasir Ibrahim Elmesmari * and Mohammed M. Elnazzal

Department of Statistics, University of Benghazi. Libya.

World Journal of Advanced Research and Reviews, 2024, 24(03), 214–227

Publication history: Received on 18 October 2024; revised on 29 November 2024; accepted on 02 December 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.24.3.3622

Abstract

Mixture distributions are widely used to model data with distinct groups, providing a flexible approach to estimating density. However, Bayesian approaches for mixture models pose challenges, such as label switching in the Gibbs sampler output due to the non-identifiability of component parameters. We review advanced methods for Bayesian analysis, including the Markov chain Monte Carlo (MCMC) reversible jump algorithm and model comparison based on joint measures of fit and complexity. We also present a Bayesian regression model based on a two-component mixture model, implemented using the Gibbs sampler algorithm and applied to a dataset of time measurement differences between two clocks. Our theoretical investigation highlights the importance of latent variables in implementing the Bayesian normal mixture model with two components. When applied to the dataset, our model effectively assigned probabilities to the two states of the phenomenon under study and identified two processes with identical slopes, intercepts, and variances. Our findings demonstrate the power of Bayesian mixture models in uncovering hidden structures within complex datasets. In general, our review and application provide insight into the challenges and potential solutions for Bayesian mixture modeling and highlight the usefulness of these methods in various fields.

Keywords: Finite mixture regression models; Bayesian approaches for mixture models; Gibbs sampler; MCMC

1. Introduction

Mixture models provide a flexible framework for capturing heterogeneity in data by assuming distinct sub-populations each following their regression relationships. Bayesian methods offer a powerful approach for parameter estimation and inference in mixture regression models. This paper presents a Bayesian inference approach for a mixture of normal regression models. Mixture models have gained widespread use in modeling heterogeneous data in various fields. Each mixture component can represent a distinct subgroup in the population. The most commonly used type of mixture model consists of Gaussian components as noted by early researchers [1] and [2]. Overviews of mixture models and their applications are provided in seminal texts [3]. Recent works have discussed advances and challenges in this area [4]; [5]; [6]. Typically, maximum likelihood via EM algorithm has been employed for parameter estimation in mixture models [7]. Mixture regression models extend simple mixture models to allow for regression-type relationships within components. This flexibility enables the modeling of heterogeneous regression behaviors between subpopulations. Bayesian methods provide an effective framework for inference in such complex models. The work [8] introduced Gibbs sampling for normal mixture regression estimation. Since then, Bayesian analysis of increasingly sophisticated mixtures has been enabled via MCMC techniques. This paper presents a Bayesian approach for a novel mixture of normal regression models that incorporate latent variables. Latent variables are important for class identification, probability estimation, and individual classification [9]. A Gibbs sampler algorithm is implemented for parameter estimation, drawing inspiration from previous research [10]. The proposed model is applied to a real dataset. Latent variables are shown to play a key role in constructing meaningful components. The remainder of the paper is organized as follows. Section 2 introduces the proposed mixture regression model and priors. Section 3 outlines the Gibbs sampling steps. Section 4 applies the model and discusses the findings. Section 5 concludes and suggests future work.

* Corresponding author: Nasir Elmesmari

2. The Statistical Model

This section provides an overview of the essential concepts related to finite mixture models, including parameter estimation methods such as the Expectation-Maximization (EM) algorithm. It also delves into the Bayesian estimation approach for finite mixture regression models.

2.1. Finite Mixture of Gaussian Regression Models

Suppose a random sample $\{(x_i, y_i), i = 1, \dots, n\}$ of independent identically distributed (iid) observations are drawn from a finite mixture of normal regression models. Then the probability distribution function is given by

$$g(y_i|x_i: \Psi) = \sum_{k=1}^{K} \alpha_k \phi(y_i|x_i\beta_k; \sigma_k^2) \quad \text{.....(1)}$$

where K is the total number of mixture regression components, $\phi(y_i|x_i\beta_k; \sigma_k^2)$ is a Gaussian density function of the K^{th} component with mean $x_i\beta_k$ and variance σ_k^2. The mixing proportions $\alpha_k, k = 1, \dots, K$ have the following restrictions: $0 < \alpha_k \leq 1$ and $\sum_{k=1}^{K} \alpha_k = 1$. Therefore, the parameter vector Ψ contains $\{\alpha_1, \dots, \alpha_k, \beta_1, \dots, \beta_k, \sigma_1^2, \dots, \sigma_k^2\}$, where $\beta_1, \dots, \beta_k, \sigma_1^2, \dots, \sigma_k^2$ are the component-specific regressions coefficients and variances, respectively.

The model presented here is a two-component normal mixture model. The density function for a random variable y is believed to come from one of two simple linear regression equations are

$$g(y|x) = \propto \phi(y|\beta_{01} + \beta_{11}x, \sigma_1^2) + (1-\propto)\phi(y|\beta_{02} + \beta_{12}x, \sigma_2^2) \quad \text{.....(2)}$$

Where $\propto \in [0,1]$. Equation 2 suggests that a situation with two possible outcomes can be explained by two distinct linear processes that do not overlap in terms of their starting point, rate of change, or variability [11]. One of these processes occurs with a certain probability $\propto$, while the other occurs with the remaining probability $(1-\propto)$.

The complete-data setup is given iid samples from $f(y|x)$; we define the latent variable z_i such that

$$z_i = \begin{cases} 1 \; if \; the \; i^{th} \; observation \in k^{th} \; component \\ 0 \; otherwise \end{cases}$$

The joint distribution of the dependent variables $y_1, \dots, y_n$ given the parameter Ψ is represented as:

$$p(y|\Psi) = \prod_{i=1}^{n} [\propto \phi(y|\mu_1, \sigma_1^2)]^{1-z_i} \times [(1-\propto)\phi(y|\mu_2, \sigma_2^2)]^{z_i} \quad \text{..... (3)}$$

Where $\Psi = \{z, \alpha, \beta_{01}, \beta_{11}, \sigma_1^2, \beta_{02}, \beta_{12}, \sigma_2^2\}$ and $\phi(y|\beta_{0k} + \beta_{1k}x, \sigma_k^2), k = 1,2$ is a normal probability density functions with mean $\mu_1 = \beta_{01} + \beta_{11}x$ and $\mu_2 = \beta_{02} + \beta_{12}x$ respectively, and variances σ_1^2 and σ_2^2. The parameter $\propto$ represents the probability of belonging to the first group, and z_i is the binary indicator variable for the i^{th} observation. The common goal of statistical inference in this setting is to estimate the parameters of the model. The traditional maximum likelihood approach using the EM algorithm [7]. A Bayesian approach is used to estimate the parameters of the model. This method entails evaluating the probability distribution for the parameters, which is determined by the likelihood of the data and the prior probability for those parameters. In particular, Jeffreys priors, developed by [12], are specified for this paper.

2.2. Estimation in Mixture Models

Several methods have been developed for estimating the parameters in finite-mixture models. We highlight four widely used methods: the method of moments, the minimum distance method, the maximum likelihood method, and the Bayesian method. Dating back to the work of Pearson (1984), the method of moments is one of the earliest techniques for estimating parameters in finite mixture models [13]. It was commonly used in applications where computing power was insufficient to maximize the log-likelihood function. Further advancements in moment estimators can be found in [14] and [15].

Even today, they remain valuable as initial values for iterative numerical methods to calculate maximum likelihood estimates [16]. The minimum distance estimation, first introduced by [17], presents another general approach to estimating Ψ in a finite mixture. This method aims to minimize the disparity between the empirical distribution and the mixture distribution or between the kernel density and the mixture density. It is worth noting that the maximum likelihood estimator (MLE) can be seen as a special case of minimum distance estimators, as it works to minimize the

distance between the empirical distribution and the mixture distribution [18]. With the feasibility of finding numerical solutions to likelihood equations, likelihood-based inference has experienced rapid development and played a crucial role in finite mixture models. Consider a data set in the form of a random sample of observations $X_1 = x_1, \dots, X_n = x_n$ where each $X's$ distribution is defined by a finite parametric mixture density of the form 1. The log-likelihood function of Ψ is provided, then, we can write the complete-data log-likelihood function as

$$\ell_c(\Psi) = \sum_{i=1}^{n} \sum_{k=1}^{K} I(Z_{ik} = 1)\{log\ \alpha_k + \log\ \phi(y_i|x_i\beta_k; \sigma_k^2)\} \dots\dots\dots\dots(4)$$

The maximum likelihood estimator of *Psi* is defined to be

$$\hat{\psi} = \underset{\psi \epsilon \Omega}{\operatorname{argmax}}\, \ell_c(\Psi)$$

When such situations arise, explicit expressions for the MLEs are often unavailable. Various numerical algorithms have been developed to maximize the log-likelihood function. Among these, the expectation-maximization (EM) algorithm is a popular choice. For further insights, refer to [19]. Ordinary MLEs in mixture models can be inconsistent or undefined due to factors such as unidentifiable parameters or non-regular likelihood functions. These issues can be addressed using alternative estimation methods. One solution is a Bayesian approach that incorporates prior knowledge or regularization terms to provide a more robust framework for estimating model parameters. Empirical Bayes methods can also be used to inform prior distributions and address MLE inconsistencies. Regularization techniques like penalized likelihood or constrained optimization can restrict the parameter space and improve the estimate stability. The choice of estimation method depends on the problem's characteristics and available data, and it is essential to evaluate the performance of each method and consider their assumptions and limitations. For example, in the case of the two-component Normal mixture 2, the MLE encounters challenges as $\ell_c(\Psi)$ approaches ∞ when $\mu_1 = x_1, \sigma_1^2$ approaches 0, and the other parameters are held fixed. To address this issue, two studies [20] and [21] proposed the use of constrained MLE, while another study [22] studied the properties of penalized MLE [23]. The fourth method for estimating Ψ is the Bayesian approach. Let $\ell_c(X_1, \dots, X_n|\Psi)$ be the likelihood function of Ψ. Assuming a prior distribution $p(\Psi)$ on Ψ is available, the posterior density $P(\Psi|X_1, \dots, X_n)$ can be obtained. The Bayesian approach can provide a coherent way of estimating Ψ by incorporating prior knowledge and uncertainty through the prior distribution. By Bayes' theorem, the posterior density can be expressed as

$$P(\Psi|X_1, \dots, X_n) \propto\ \ell_c(X_1, \dots, X_n|\Psi)p(\Psi)$$

which combines the likelihood of observing the data given Ψ and the prior distribution of Ψ. The proportionality constant can be determined by normalizing the product of the likelihood and prior, such that the posterior density integrates to one. In practice, Markov chain Monte Carlo (MCMC) algorithms, such as Gibbs sampling or Metropolis-Hastings, are often employed to sample from the posterior distribution and approximate the posterior density.

2.3. Parameter priors

The prior distributions of the parameters play a crucial role in Bayesian statistics. These priors capture the uncertainty regarding the true values of the parameters before observing the data. They are essential for making probabilistic inferences about the parameters and predictions about future observations. The choice of prior distributions reflects the available information about the parameters and can significantly impact the resulting posterior inferences. This section will introduce the prior distribution for each parameter.

$$p(\Psi) \propto \frac{1}{\sqrt{2\pi}\tau_0} \exp\left\{\frac{-1}{2\tau_0}\ (\beta_{01} - \mu_0)^2\right\}. \times \frac{1}{\sqrt{2\pi}\tau_1} \exp\left\{\frac{-1}{21}\ (\beta_{02} - \mu_1)^2\right\} \dots\dots\dots\dots(5)$$

Combining Equation 3 and 5 via multiplication gives the posterior distribution for Ψ

$$\pi(\Psi) \propto \pi(y|\Psi) \times p(\Psi) \dots\dots\dots\dots\dots\dots (6)$$

Which can be written as follows:

$$p(\Psi|.) \propto \frac{1}{\sqrt{2\pi}\tau_0} \exp\left\{\frac{-1}{2\tau_0}\ (\beta_{01} - \mu_0)^2\right\}.$$
$$\times \frac{1}{\sqrt{2\pi}\tau_1} \exp\left\{\frac{-1}{21}\ (\beta_{02} - \mu_1)^2\right\} \times \prod_{i=1}^{n} [\alpha\ \phi(y|\mu_1, \sigma_1^2)]^{1-z_i}\ \times\ [(1-\alpha)\phi(y|\mu_2, \sigma_2^2)]^{z_i}$$

After performing some algebra, it is possible to determine the conditional distributions of the parameters. In this instance, all the full conditionals have closed forms, allowing for the use of a Gibbs sampler to obtain draws from the joint posterior.

2.4. The Full Conditional Distributions for the Parameters

Given a multivariate posterior distribution for Ψ, it is easier to sample from a conditional distribution than to marginalize by integrating over a joint distribution.

Then the full conditional distributions for the parameters of Ψ our model are

$$A = \alpha\phi(y|\beta_{01} + \beta_{11}x, \sigma_1^2)$$

$$B = (1-\propto)\phi(y|\beta_{02} + \beta_{12}x, \sigma_2^2)$$

$$z_i \sim Ber(\tfrac{A}{A+B}) \text{ ..... (7)}$$

$$\alpha \sim Beta(1 + n_1, 1 + n_0) \text{ .....(8)}$$

$$\beta_{0k} \sim N(\bar{y}_i - \beta_{01}\bar{x}_i, \tfrac{\sigma_k^2}{n_i}) \text{ for } i = 0,1, \text{ and k = 1,2 .....(9)}$$

$$\text{Where } \bar{y}_i = \tfrac{1}{n_i}\textstyle\sum_{i|z_i=i} y_i \text{ , } \bar{x}_i = \tfrac{1}{n_i}\sum_{i|z_i=i} x_i \text{ , } i = 0,1$$

$$\beta_{1k} \sim N(\tfrac{\sum_{i|z_i=i} y_i x_i}{\sum_{i|z_i=i} x_i}_{i} - \beta_{0k}, \tfrac{\sigma_k^2}{\sum_{i|z_i=i} x_i^2}) \text{ for } i = 0,1, \text{ and k = 1,2 ..... (10)}$$

Where $\bar{y}_i = \tfrac{1}{n_i}\sum_{i|z_i=i} y_i \bar{x}_i = \tfrac{1}{n_i}\sum_{i|z_i=i} x_i \;\; i = 0,1, i|z_i = i$ is used to denote the set of i such that $z_i = 0$ and n_0 is the count of the z_i where $z_i = 0$. The same type of notation is used for $i|z_i = 0$ and n_0. And for σ_k^2

$$\sigma_k^2 \sim Inv - Gamma\left(\tfrac{n_i}{2}, \tfrac{1}{2}\textstyle\sum[y_i - (\beta_{0k} + \beta_{1k}x_i)]\right) \text{ , } for\ i = 0,1, and\ k = 1,2 \text{ ..... (11)}$$

2.5. The Gibbs Sampler

In this section, we present the posterior distribution for a two-component normal regression mixture model under a conjugate prior and introduce the Gibbs sampler for both general case and normal mixture application. The Gibbs sampler is a valuable simulation method that produces a sample from the posterior distribution.

The Gibbs sampler is an algorithm for generating samples from the joint probability distribution of multiple random variables. It was first described in a 1984 statistical paper by Stuart and Donald Geman, building on earlier work by [24], and [25] on Markov chain Monte Carlo methods. The paper by Stuart and Donald Geman had a big impact on Bayesian statistics, computational statistics, and stochastic processes. influenced it [26] that significantly boosted the use of Bayesian methods and Gibbs sampling in particular. Since then, the Gibbs sampler has become one of the most widely used techniques for approximating complex posterior distributions. While studies [27] and [28] proposed similar algorithms, their work did not receive as much attention from statisticians as Stuart and Donald Geman paper, which is largely credited with popularizing the Gibbs sampler and its applications in Bayesian analysis.

We used the Gibbs sampler framework Geman and Geman (1984) developed to sample from the conditional distributions presented in Section 2. The idea of the Gibbs sampler algorithm in this case is that: Given a multivariate distribution, it is simpler to sample from a conditional distribution than to marginalize by integrating over a joint distribution. Suppose we want to obtain k samples from $\theta = \{z, p, \beta_{01}, \beta_{11}, \sigma_1.\beta_{02}, \beta_{12}, \sigma_2\}$ form a joint distribution $p(z, p, \beta_{01}, \beta_{11}, \sigma_1.\beta_{02}, \beta_{12}, \sigma_2)$ denote the i^{th} sample by:

$$\theta^i = \left\{z^i, p^i, \beta_{01}^{(i)}, \beta_{11}^{(i)}, \sigma_1^{(i)}.\beta_{02}^{(i)}, \beta_{12}^{(i)}, \sigma_2^{(i)}\right\}$$

implementing Gibbs sampling steps can be shown below:

We begin with some initial value $\theta^{(0)}$ for each variable parameter.

For each sample $i = \{1, \dots, k\}$, sample each variable parameter $\theta_j^{(i)}$ from the conditional distribution $p\left(\theta_j|\theta_{-\theta_j}\right)$. Sample each variable parameter from the distribution conditioned on all other parameters, using the latest values and updating the variable with its new value once sampled. To illustrate $p(\beta_{01}^{(i)}|z^{(i)}, p^i, \beta_{11}^{(i)}, \sigma_1^{(i)}.\beta_{02}^{(i-1)}, \beta_{12}^{(i-1)}, \sigma_2^{(i-1)})$. The samples collected provide an estimation of the joint distribution of every model parameter. Furthermore, the samples can be used to estimate the distribution of any specific subset of parameters by focusing only on those samples, disregarding the irrelevant parameters. Additionally, the average of all samples can be utilized to estimate the expected value of any parameter in the posterior distribution.

3. Simulation Study

The computational aspect of the modeling was done in R, this section presents the ones from the joint posterior distribution for the parameters.

We aim to develop an algorithm using a simplified two-component normal mixture regression model. To estimate the model parameters, we employ a Bayesian method. This method views the parameters' posterior distribution as directly proportional to the likelihood multiplied by the combined prior for said parameters. In this paper, we define Jeffrey's priors.

We tend to focus on estimating the unknown parameters $\beta_{01}, \beta_{11}.\beta_{02}, \beta_{12}$ with variances σ_1^2, σ_2^2 and mixture proportion π via MCMC algorithm for various sample sizes $n = \{50, 100, 1000, 2000\}$ observations. Section 2.5 covered the usual approach of setting initial guesses for unknown parameters [29]. Various initial guesses lead to different iterative estimation results, with those achieving the highest maximized likelihood deemed the best. The vector of parameters (τ, β, σ^2) used to generate the mixture are reported in Table 1.

Table 1 True parameter values for simulation study

$\boldsymbol{\Psi}$	$\boldsymbol{\alpha_1}$	$\boldsymbol{\alpha_2}$	$\boldsymbol{\beta_{01}}$	$\boldsymbol{\beta_{02}}$	$\boldsymbol{\beta_{11}}$	$\boldsymbol{\beta_{21}}$	$\boldsymbol{\sigma_1^2}$	$\boldsymbol{\sigma_2^2}$
True Parameter	0.3	0.7	2	-3	1.5	-2	0.5	1

Table 2 Estimated parameters and standard errors for the two mixture regression

n=50			
Component 1 $\boldsymbol{\Psi}$	**Estimation**	**Lower Tail**	**Upper Tail**
β_{01}	1.99	1.90	2.07
β_{02}	1.51	1.36	1.67
Component 2 β_{02}	-3.01	-3.08	-2.95
β_{12}	-3.005	3.08	-2.93
n=100			
Component 1 β_{02}	1.52	1.37	1.68
β_{12}	-3.005	3.08	-2.93
Component 2 β_{02}	-3.007	-3.08	-2.93
β_{12}	-3.005	3.08	-2.93

n=1000			
Component 1 β_{02}	1.51	1.36	1.69
β_{12}	-3.005	3.08	-2.93
Component 2 β_{02}	-3.005	-3.07	-2.94
β_{12}	-3.005	3.08	-2.93
n=2000			
Component 1 β_{02}	1.99	1.90	2.09
β_{12}	-3.005	3.08	-2.93
Component 2 β_{02}	-3.005	-3.08	-2.93
β_{12}	-3.005	3.08	-2.93

We employed the Gibbs sampler, as described in Section 2.5, and stored 2000 MCMC draws after a burn-in period of 3000 draws. Table 2 provided table contains estimated parameters and standard errors for two mixture regression models based on different sample sizes *n*. Here are the key components: Considering Component 1; For each sample size, we have estimates for β_{01}(intercept) and β_{11} (slope). The estimates are accompanied by their corresponding lower and upper tail values. For example, when (n = 50), the estimated values are: β_{01}: 1.99 (Estimation), 1.90 (Lower Tail), 2.07 (Upper Tail) β_{11}: 1.51 (Estimation), 1.36 (Lower Tail), 1.67 (Upper Tail). Conversely, for Component 2: Similar to Component 1, we have estimates for β_{02} and β_{12}. For example, when (n=1000), the estimates are: β_{02}: 1.51 (Estimation), 1.36 (Lower Tail), 1.69 (Upper Tail) β_{12}: -3.005 (Estimation), 3.08 (Lower Tail), -2.93 (Upper Tail) The sample sizes considered are *n* = {50,100,1000*, and* 2000}.

3.1. Statistical Analysis

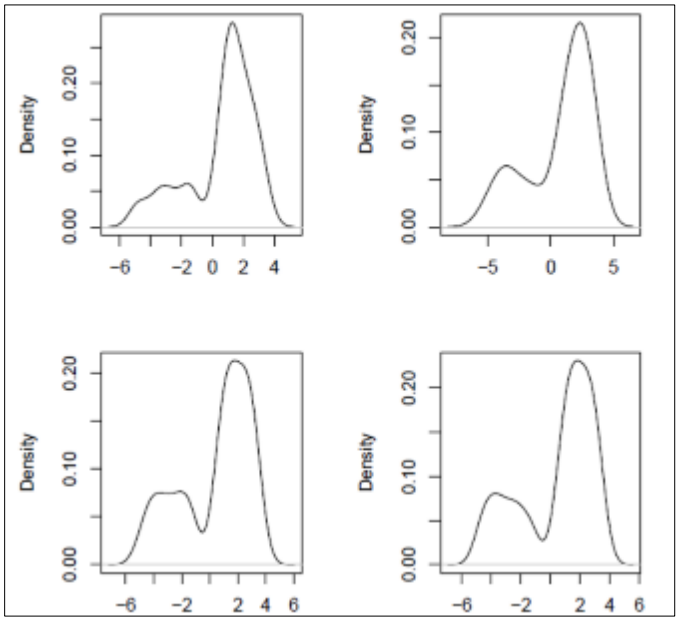

Figure 1 Density plots for data set for different sample size 50, 100, 1000 and 2000

Figure 1 shows that the predictive density based on the 2-component Normal mixture density function of the response using the same Gibbs sampler, revealing distinct behaviors of the two mixture response densities depending on the considered sample sizes, namely n = {50,100,1000, and 2000}.

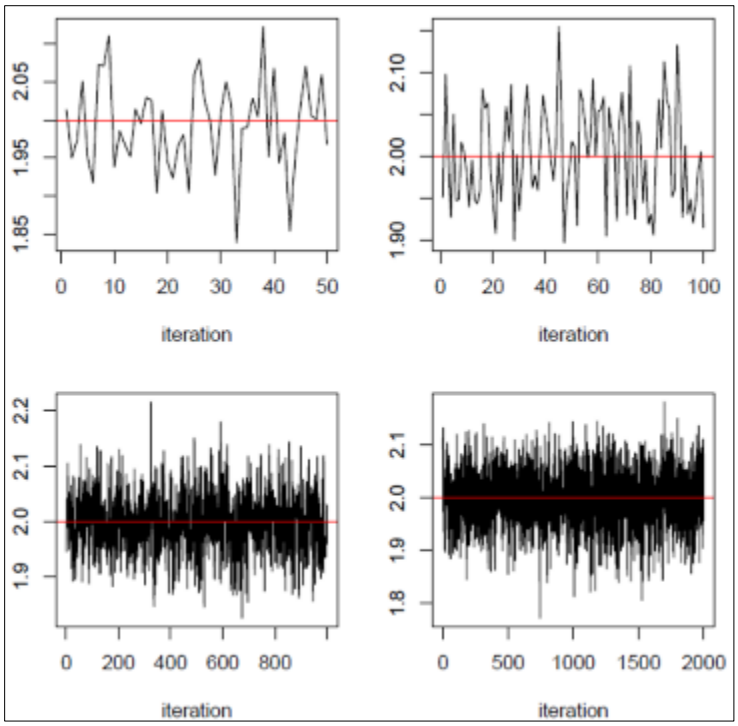

Figure 2 Convergence of the β_{01} for different sample sizes n = 50,100,1000, and 2000

Figure 2 shows the convergence of posterior distributions of the model parameters, resulting from 1000 steps of MCMC simulations. The estimated posterior parameters converge to true parameters which are displayed in (Table 1) as the sample size increases. However, for most of these distributions, in particular, Convergence is effective in increasing the sample size and becomes more stable as the sample size increases. For instance, when the sample size reaches 1000 (or 2000), the parameter distribution exhibits greater convergence.

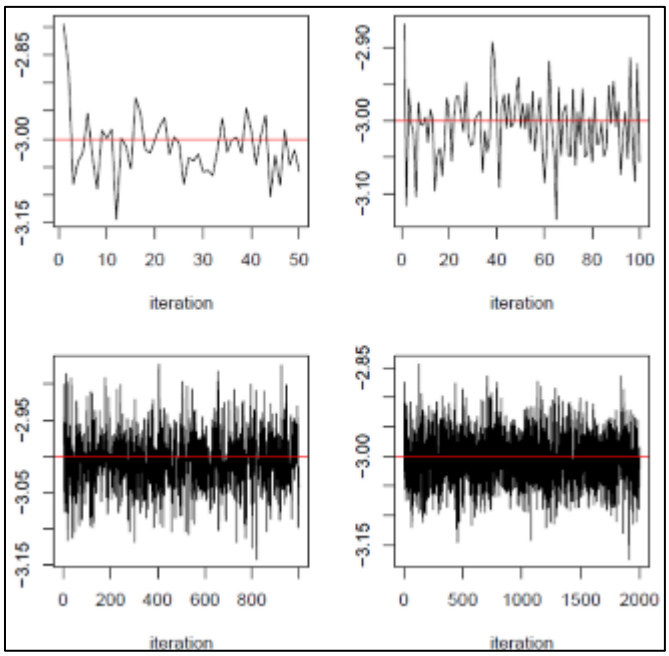

Figure 2 Convergence of the B_{02} for different sample size n=50, 100, 1000 and 2000

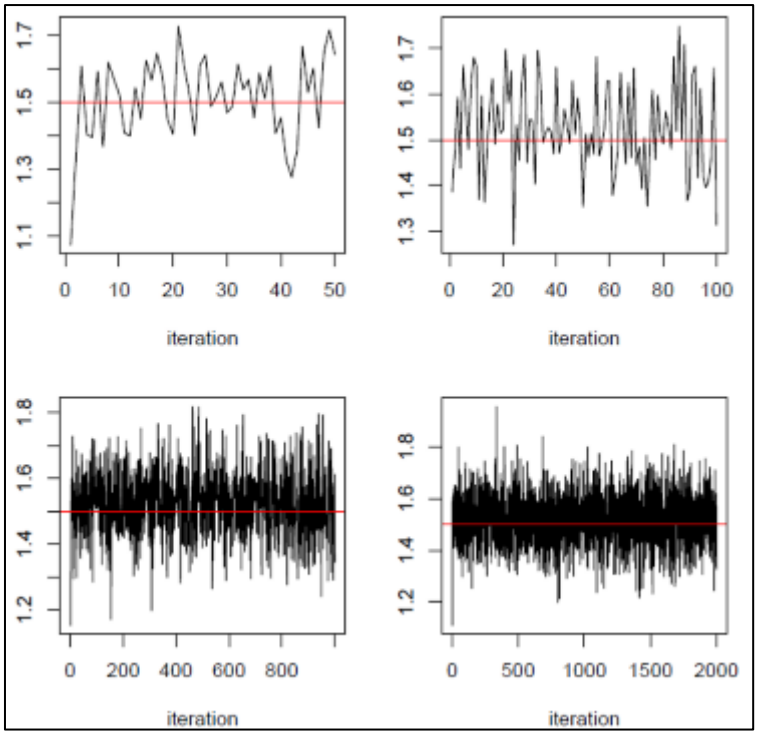

Figure 4 Convergence of the B11 for different sample size n=50, 100, 1000 and 2000

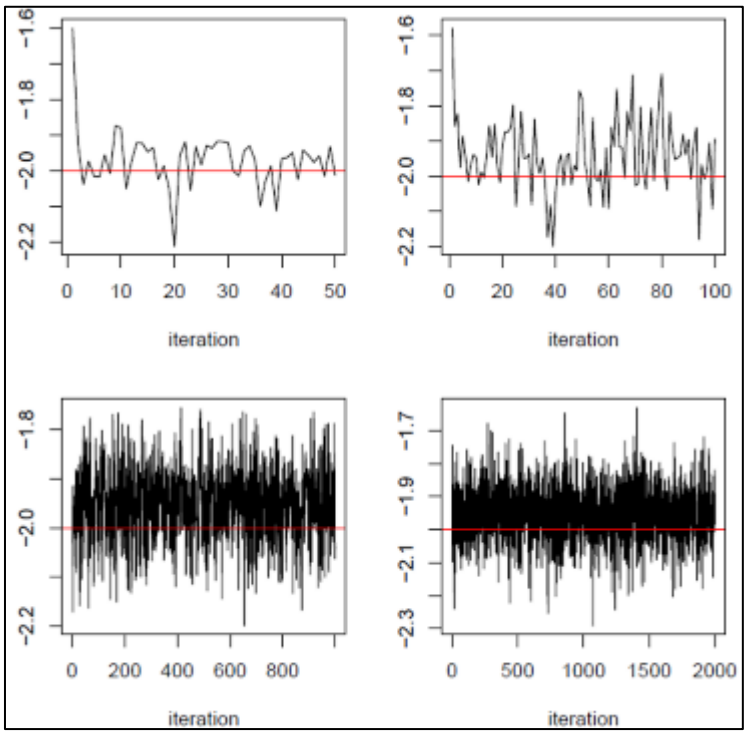

Figure 5 Convergence of the B_{02} for different sample size n=50, 100, 1000 and 2000

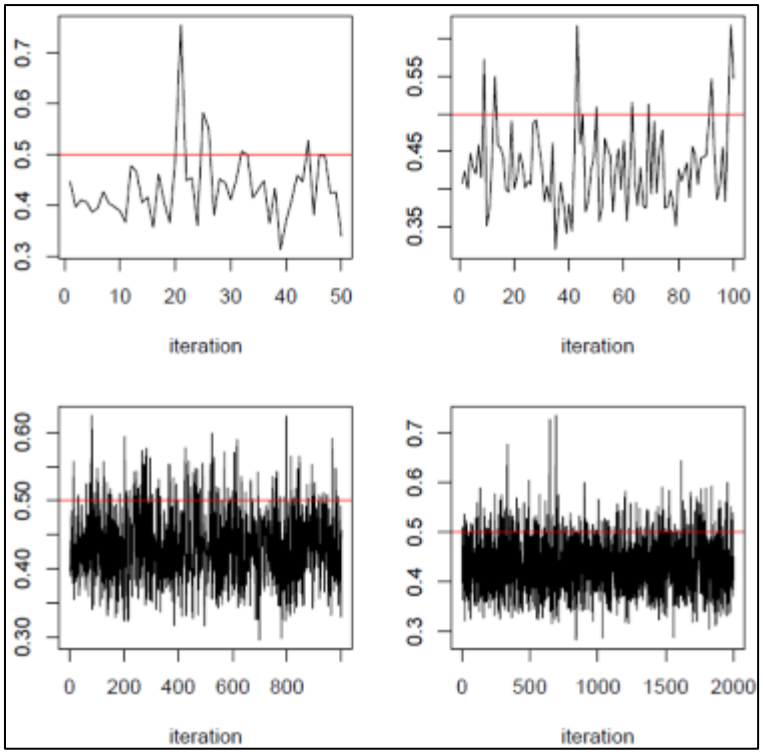

Figure 6 Convergence of the σ_1^2 for different sample size n=50, 100, 1000 and 2000

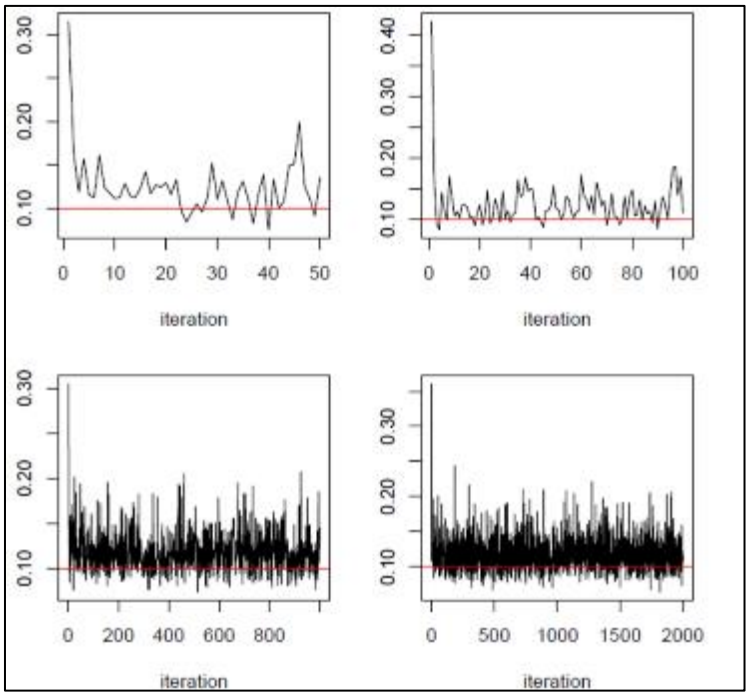

Figure 7 Convergence of the σ_2^2 for different sample size n=50, 100, 1000 and 2000

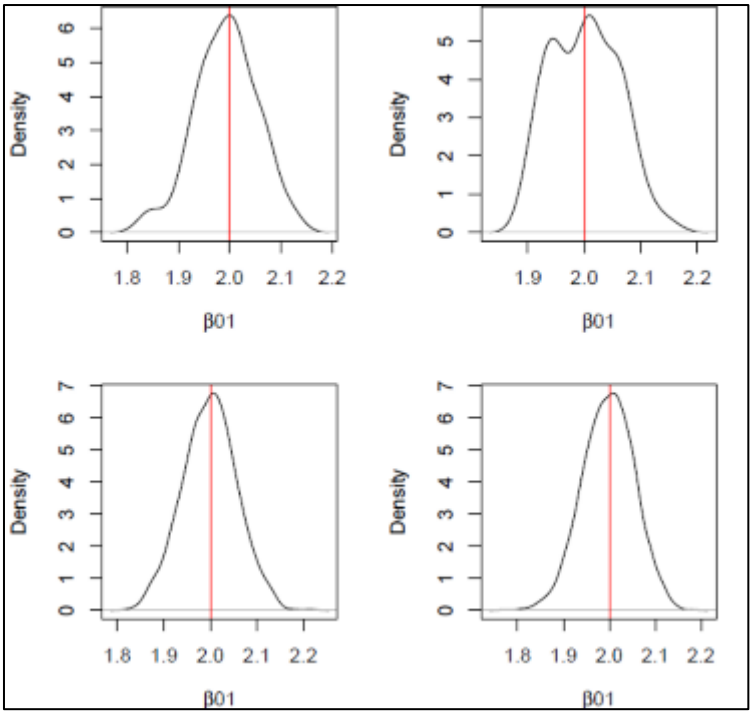

Figure 8 Posterior probability density of $\boldsymbol{\beta_{01}}$ for different sample size n=50, 100, 1000 and 2000

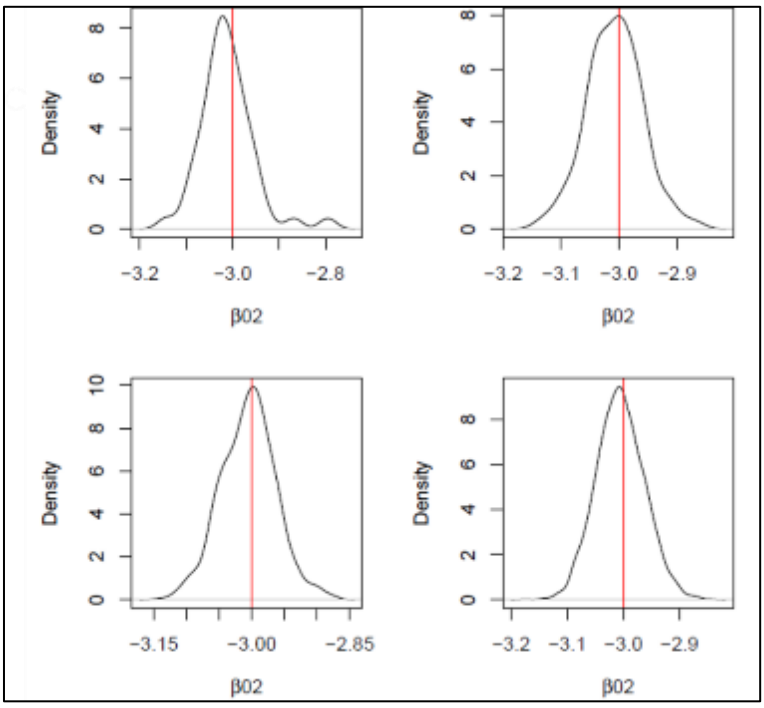

Figure 9 Posterior probability density of $\boldsymbol{\beta_{02}}$ for different sample size n=50, 100, 1000 and 2000

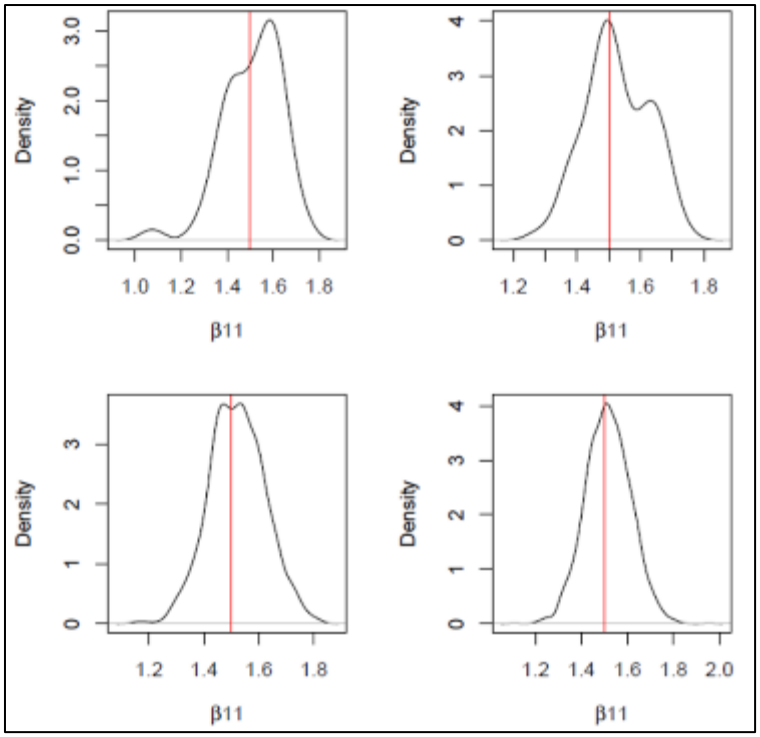

Figure 10 Posterior probability density of $\boldsymbol{\beta_{11}}$ for different sample size n=50, 100, 1000 and 2000

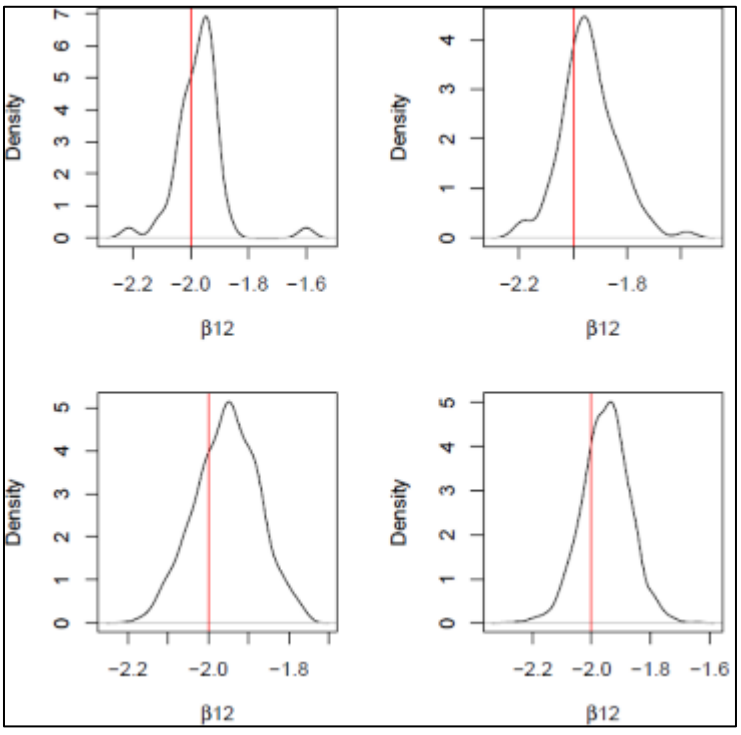

Figure 11 Posterior probability density of $\boldsymbol{\beta_{12}}$ for different sample size n=50, 100, 1000 and 2000

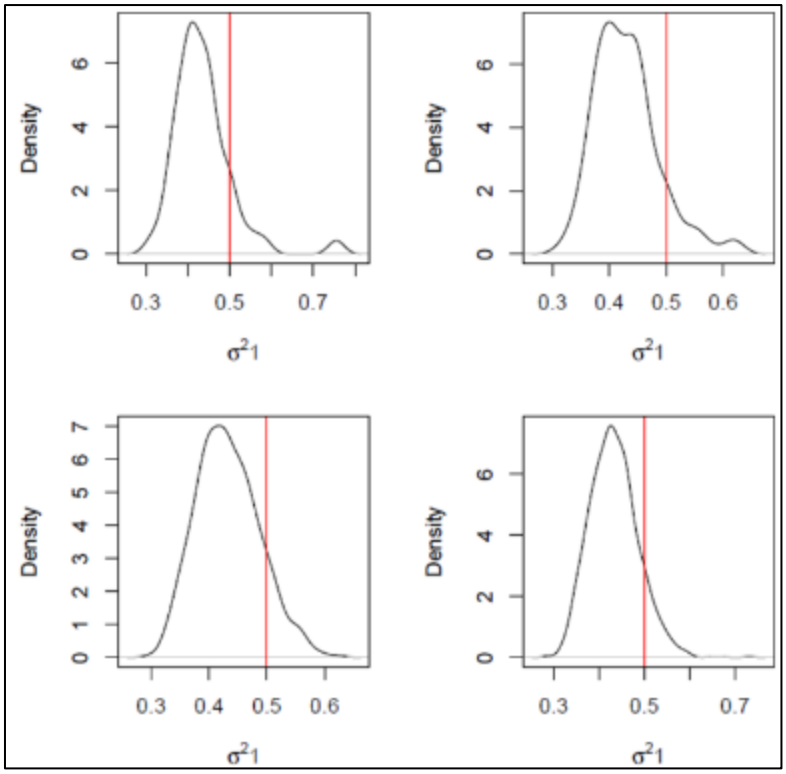

Figure 12 Density of $\boldsymbol{\sigma_1^2}$ for different sample size n=50, 100, 1000 and 20000

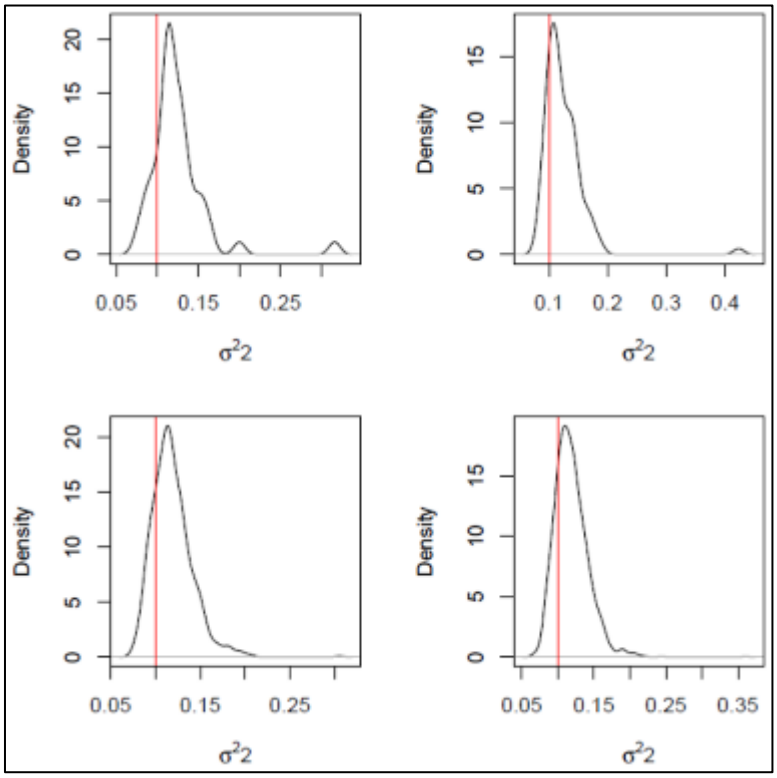

Figure 13 Density of $\boldsymbol{\sigma_2^2}$ for different sample size n=50, 100, 1000 and 20000

4. Discussion

This article introduces Bayesian inference for finite mixtures of regression models, focusing on mixtures with a predetermined number of components. Demonstrates the implementation of the Gibbs sampler, specifically applied to normal mixture models. Within the Bayesian framework, it explores several challenges, including trapping states, selection of priors, label switching, and convergence diagnostics.

Adopting a Bayesian approach for a finite mixture of regression models offers several advantages. First, incorporating proper priors introduces a smoothing effect on the mixture likelihood function, reducing the risk of spurious modes, especially in normal distributions. Second, unlike methods relying on asymptotic normality, Bayesian inference remains valid even when regularity conditions are violated, making it valuable for small sample sizes and low component weights. Third, the rapid development and straightforward implementation of Markov Chain Monte Carlo (MCMC) methods have made Bayesian estimation feasible. However, the Gibbs sampler, although natural, can suffer from label switching, posing challenges in parameter estimation. Empirical convergence diagnostics are essential, but assessing convergence behavior with limited realizations remains difficult.

5. Conclusion

Our findings highlight the effectiveness of Bayesian mixture models in revealing hidden structures within complex datasets. A key factor in the success of these models is initializing Markov Chain Monte Carlo (MCMC) algorithms with appropriate starting values [30]. Moving forward, future research should aim to extend the Gibbs sampling method to incorporate other widely used mixture distributions, including Poisson, Gamma, Weibull, and Lognormal distributions.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Day, N. E. (1969). Estimating the components of a mixture of normal distributions. Biometrika, 56(3), 463-474.

[2] Fraley, C. and Raftery, A. E. (2006), "MCLUST version 3 for R: Normal mixture modeling and Model-based clustering," Tech. rep., Citeseer.

[3] McLachlan, G. (2000). Finite mixture models. *A wiley-interscience publication*.

[4] Melnykov, V., Michael, S., and Melnykov, I. (2015), "Recent developments in model-based clustering with applications," Partitional clustering algorithms, 1–39.

[5] McNicholas, P. D. (2016), Mixture model-based classification, Chapman and Hall/CRC.

[6] McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. Annual review of statistics and its application, 6(1), 355-378

[7] Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: series B (methodological)*, *39*(1), 1-22.

[8] Diebolt, J. and Robert, C. (1994), "Estimation of finite mixture distributions through Bayesian sampling," Journal of the Royal Statistical Society: Series B (Methodological), 56, 363–375.

[9] Muth' en, B. O. (2001), "Latent variable mixture modeling," in New developments and techniques in structural equation modeling, Psychology Press, pp. 21–54.

[10] Geman, S. and Geman, D. (1984), "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Transactions on pattern analysis and machine intelligence, 721– 741.

[11] Frühwirth-Schnatter, S. (2006). Finite mixture and Markov switching models. Springer.

[12] Jeffreys, H. (1961). The theory of probability 3rd ed new york: Oxford university press; [google scholar].

[13] Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions* of the Royal Society of London. A, 185, 71-110.

[14] Lindsay, B. G., & Basak, P. (1993). Multivariate normal mixtures: a fast consistent method of moments. *Journal of the American Statistical Association*, *88*(422), 468-476..

[15] Furman, W. D. and Lindsay, B. G. (1994), "Testing for the number of components in a mixture of normal distributions using moment estimators," Computational Statistics & Data Analysis, 17, 473–492.

[16] Lindsay, B. G. (1995). Mixture models: theory, geometry, and applications. Ims.

[17] Wolfowitz, J. (1957), "The minimum distance method," The Annals of Mathematical Statistics, 75–88.

[18] Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. The annals of mathematical statistics, 22(1), 79-86.

[19] McLachlan, G. J., & Krishnan, T. (1997). Wiley series in probability and statistics. The EM algorithm and extensions.

[20] Hathaway, R. J. (1985), "A constrained formulation of maximum-likelihood estimation for normal mixture distributions," The Annals of Statistics, 795–800.

[21] Tan, X., Chen, J., and Zhang, R. (2007), "Consistency of the constrained maximum likelihood estimator in finite normal mixture models," Proceedings of the American Statistical Association, American Statistical Association, Alexandria, VA, 2113–2119.

[22] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of state calculations by fast computing machines," The journal of chemical physics, 21, 1087–1092.

[23] Chen, J., Tan, X., & Zhang, R. (2008). Inference for normal mixtures in mean and variance. *Statistica Sinica*, 443-465.

[24] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

[25] Peskun, P. H. (1973), "Optimum monte-carlo sampling using markov chains," Biometrika, 60, 607–612.

[26] Gelfand, A. E. and Smith, A. F. (1990), "Sampling-based approaches to calculating marginal densities", Journal of the American statistical association, 85, 398–409.

[27] Tanner, M. A. and Wong, W. H. (1987), "The calculation of posterior distributions by data augmentation," Journal of the American statistical Association, 82, 528–540.

[28] Besag, J., & Clifford, P. (1989). Generalized Monte Carlo significance tests. Biometrika, 76(4), 633-642.

[29] Elmesmari, N. (2021). Parameters Estimation Sensitivity of the Linear Mixed Model to Alternative Prior Distribution Specifications. Sch J Phys Math Stat, 9, 166-170.

[30] Elmesmari, N. (2021). Parameters Estimation Sensitivity of the Linear Mixed Model To Alternative Prior Distribution Specifications. Sch J Phys Math Stat, 9, 166-170.