IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: (3007-1658) Online

On Inference of Finite Mixture of Rayleigh Distribution by Gibbs Sampler and Metropolis-Hastings

Fadhaa Ali

Department of Statistics, College of Administration and Economics, University of Baghdad, 10011 Baghdad, Iraq

ARTICLE INFO

Article history:
Received 02 October 2024
Revised 07 October 2024,
Accepted 20 October 2024
Available online 21 October 2024

Keywords:
Mixture of Rayleigh Distribution
Bayesian inference
Gibbs Sampler
Metropolis-Hastings
Bayesian Information Criteria(BIC)

ABSTRACT

Inferential methods of statistical distributions have reached a high level of interest in recent years. However, in real life, data can follow more than one distribution, and then mixture models must be fitted to such data. One of which is a finite mixture of Rayleigh distribution that is widely used in modelling lifetime data in many fields, such as medicine, agriculture and engineering. In this paper, we proposed a new Bayesian frameworks by assuming conjugate priors for the square of the component parameters. We used this prior distribution in the classical Bayesian, Metropolis-hasting (MH) and Gibbs sampler methods. The performance of these techniques were assessed by conducting data which was generated from two and three-component mixture of the Rayleigh distribution according to several scenarios and comparing the results of the scenarios by calculating the mean of classification successful rate (MCSR) and the mean of mean square error(MMSE). The results showed that Gibbs sampler algorithm yields a better computation results than the others in terms of MMSE and MCSR.

1. Introduction

Mixture models have been widely used in several fields of research. Its applications have been considered as a proper tackling tool for many problematic issues with data of interest when it does not belong to a specific distribution. When several subpopulations are merged in unknown weights to construct a larger population, the existing distribution will be invalid in terms of fitting them to such a population. As an example, some biological, agricultural and electrical items may be formed from more than one subpopulation depending on failure time, gender, age, class or any other categories. To model and infer some aspects of such a population, a mixture model may be fitted to data collected from engineering, biological and physical fields. Several studies have been conducted to fit a mixture model to several types of real-world data. Mixture of Laplace and normal distributions was an earlier study that has been proposed to model wind shear samples [1-2]. Justice and crime data was inferred by a mixture of normal distribution [3]. In a lifetime, the population may be divided into several subpopulations depending upon device failure that is required to be modelled by a mixture of exponential distributions [4]. Image compression and pattern recognition studies are analysed by using the Gaussian mixture model to improve the maximum likelihood of the model by showing the impact of split and merge operations [5]. Mixture of t-distribution is also proposed to merge sub-sets within one large set of observations with long normal tails [6]. In classifying individuals in the experiments of blood science based on their density function of the mixture models, a good classifier was built to classify new histograms with an unknown label into iron and normal deficient groups of

* Corresponding author: E-**mail** address: ali_fadhaa@coadec.uobaghdad.edu.iq
https://doi.org/10.62933/93v3c985

red blood (RBCs)[7]. The key importance of fitting a mixture model to data is inferring some aspects of the number of underlying components and their parameters. Theoretically, mixture models may consist of components that represent the same distribution family or different ones, which makes them an important predicting or confirmatory tool in several analytical studies. Many techniques have been proposed in the literature to infer a mixture model. [8-11] employed maximum likelihood(ML) as a classical method that can be used in inferring mixture of multivariate normal distributions. However, researchers can come across some drawbacks that can significantly undermine the accuracy of inferences, such as knowing the exact number of components and membership of observations. Many techniques have been proposed to tackle these issues. One of which is by examining the actual underlying structure in terms of how many components they involved [12-13]. Several studies considered turning the incomplete-data likelihood into a complete one by adding a latent variable which can be an optimal way of copping uncertainty in the number of components [14]. A famous technique was employing the EM algorithm via maximum likelihood to infer the model parameters and observation memberships [15-16]. However, some technical issues have been well-documented within the progress of the EM algorithm, such as sluggish convergence and being stuck in the local optima [17-18]. To overcome the latter issue, many studies have come to light by proposing sampling algorithms from the whole space of the parameters by constructing the posterior distribution of parameters via Bayesian inference [19-23]. One of the Bayesian inferential algorithms is Gibbs sampler in which the posterior distribution of a sub-vector can be drawing from the conditional density of one parameter given the remaining ones [24]. Moreover, the Metropolis-Hastings is another Bayesian algorithm that depends on simulating samples from a mixture of posterior densities to estimate parameters [25-27]. In recent studies, several Bayesian approaches were proposed to infer the optimal number of components in a Gaussian mixture model [28]. [29] proposed Approximate Bayesian Computation–Population Monte Carlo (ABC–PMC) algorithm as an alternative framework for inference on finite mixture models. [30] studied the use of Gibbs sampler as a clustering algorithms based on finite mixture and infinite mixture models of exponential approximation to several multinomial generalized Dirichlet distributions.

This study is to infer the Rayleigh mixture distribution so that it can be used to model lifetime data sets. This can be done by the following motivations:

a. To consider a conjugate prior distribution for the model parameters and derive the posterior distribution of weight and component parameters.
b. To employ the K-means algorithm and Bayesian information criteria (BIC) to determine the underlying number of components.
c. To construct Gibbs sampler and Metropolis-Hasting algorithms to estimate model parameters.
d. To estimate observation memberships and calculate the successful rate of clustering observations into two or three components.
e. To compare the three methods to find their advantages and disadvantages.

The rest of the study is organized as follows: Section 2 provides the pdf of the Rayleigh mixture distribution. The inferential methods are studied in Section 3. In Section 4, simulation data according to various scenarios were used to evaluate the three methods by mean square error and classification successful rate. Advantages and disadvantages of the three methods are given in Section 5. Conclusion is provided in Section 6.

2. Methodology

The time of two consecutive occurrences can be defined as a random variable following the Rayleigh distribution. It may refer to the time of an item which stopped functioning or the time between two failures. Therefore, inference of such a model is required regularly. The probability density function(pdf) is written as

$$h(x;\mu) = \frac{x}{\mu^2}\, e^{\frac{-x^2}{2\mu^2}}, where\ x, \mu > 0 \qquad (1).$$

In many scenarios of lifetime data, a sample may come from several subpopulations to form a large one, which leads to the impossibility of applying the known distribution. In such cases, the mixture model is used to deal with this issue. One of these models that is used for this purpose is a finite mixture of Rayleigh distribution. The (pdf) function of such a model based on n components is

$$f(x_i;\ \varphi) = \sum_{j=1}^{n} \alpha_j \frac{x_i}{\mu_j^2}\, e^{\frac{-x_i^2}{2\mu_j^2}}, \qquad (2)$$

Where $\sum_{j=1}^{n} \alpha_j = 1, 0<\alpha_j < 1, \varphi = (\alpha_1, \alpha_2, \dots, \alpha_{n-1}, \mu_1, \mu_2, \dots, \mu_n)^T, x_i, \mu > 0.$

3. Inferential Methods

The traditional method of inferring the mixture models is by forming maximum likelihood (ML). As the observation memberships are unknown at this stage, the resulting (ML) is called incomplete-data maximum likelihood, and it is defined as

$$L(\varphi; X) = \prod_{i=1}^{m} \sum_{j=1}^{n} \alpha_j \frac{x_i}{\mu_j^2}\, e^{\frac{-x_i^2}{2\mu_j^2}}, \qquad (3)$$

Where $X = (x_1, x_2, \dots, x_m)^T$.

Here, the component number is unknown and can be considered as hidden states, which is defined by a latent or indicator vector $Z = (z_1, z_2, \dots, z_m)^T$, where $z_i = (z_{i1}, z_{i2}, \dots, z_{in})^T$ is a missing variable and it defines as

$$z_{ij} = \begin{cases} 1, & if\ subject\ i\ belong\ to\ class\ j \\ 0, & otherwise \end{cases}$$

Hence, the likelihood function of complete-data is

$$L(\varphi; X, Z) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[\alpha_j \frac{x_i}{\mu_j^2}\, e^{\frac{-x_i^2}{2\mu_j^2}} \right]^{z_{ij}} \qquad (4)$$

Let A_j be the set of all items of the component j, where

$A_j = \{x_i\ ;\ z_{ij} = 1\ , i = 1,2, \dots, m\}$

To simplify the above equation, we re-write it as

$$L(\varphi|X, Z) = \alpha_1^{m_1} \alpha_2^{m_2} \dots \alpha_n^{m_n}\, \frac{\prod_{i=1}^{m} x_i^{z_{i1}} x_i^{z_{i2}} \dots x_i^{z_{in}}}{\mu_1^{2m_1} \mu_2^{2m_2} \dots \mu_m^{2m_n}}\, e^{\frac{-\sum_{i=1}^{m} z_{i1} x_i^2}{2\mu_1^2}} . e^{\frac{-\sum_{i=1}^{m} z_{i2} x_i^2}{2\mu_2^2}} \dots e^{\frac{-\sum_{i=1}^{m} z_{in} x_i^2}{2\mu_n^2}}, \qquad (5)$$

Where $m_j = \#\ A_j$

To estimate the parameters of eq. (5), several methods were developed. In this work, Bayesian techniques are conducted to estimate the parameters. To do so, conjugate priors were used to evaluate posterior distributions of the model parameters by three different methods. The next subsections show the details of these methods.

3.1 Determine the number of components.

The above methods depend on known allocation A_j. Therefore, we may use a clustering algorithm to provide us with the potential number of components and their estimated memberships. In this work, we consider the K-means algorithm [31] to specify the initial allocation, where k=1, 2,…, n. Then, we assess the suitable number of components according to DIC [32] which has the formula

$DIC = \overline{D} + p_D,$

Where

$\overline{D} = \frac{\sum_{i=1}^{m} D_i}{m}$, and $D_i = -2 \log f(x_i;\ \varphi)$,

p_D is the number of parameters.

We use the K-means to cluster data into K=1, 2, 3, 4,… ,10, and calculate DIC for every k. Then, the estimated number of components would correspond to the smallest DIC value. This will also determine the observation memberships, so that we can apply the posterior estimation of the first method straightforward. However, in the second and the third methods, observation memberships will be updated at each iteration until achieving convergence which gives privilege over the first method.

3.2 Bayesian Framework

To construct a Bayesian framework, conjugate priors are considered for $\propto, \mu_j^2$. In so doing, we let
$\propto \sim \mathrm{Dir}\,(\gamma_1, \gamma_2, \dots, \gamma_n)$, where $\propto = (\alpha_1, \alpha_2, \dots, \alpha_n)^T$, and the (pdf) is written as

$$g\left(\propto; \underline{\gamma}\right) = \frac{\Gamma(\sum_{j=1}^{n} \gamma_j)}{\prod_{j=1}^{n} \Gamma\gamma_j} \prod_{j=1}^{n} \alpha_j^{\gamma_j - 1} \quad (6)$$

We also let $\mu_j^2 \sim \mathrm{InvGamma}(\eta, \beta)$, where $\eta > 0, \beta > 0$, and its pdf is written as

$$q\left(\mu_j^2; \lambda, \beta\right) = \frac{\beta^{\lambda}}{\Gamma\lambda\, \mu_j^{2(\lambda+1)}} e^{\frac{-\beta}{\mu_j^2}} \quad (7)$$

Note that, unlike many studies in the literature, the novelty of study that we assume a conjugate prior for the square of the component parameters μ_j^2. This will simplify deriving the posterior distribution.

From multiplying eq.(5), eq.(6) and eq.(7), we can formulate the joint pdf as
$p(\varphi|\, X, Z) \propto$

$$L(\varphi|X, Z) \frac{\Gamma(\sum_{j=1}^{n} \gamma_j)}{\prod_{j=1}^{n} \Gamma\gamma_j} \prod_{j=1}^{n} \alpha_j^{\gamma_j - 1} \prod_{j=1}^{n} \frac{\beta^{\lambda}}{\Gamma\lambda\, \mu_j^{2(\lambda+1)}} e^{\frac{-\beta}{\mu_j^2}} \quad (8)$$

Simplifying the above equation will lead to

$$p(\varphi|\, X, Z) \propto \sigma \left[\frac{\Gamma\left(\sum_{j=1}^{n} \gamma_j + m_j\right)}{\prod_{j=1}^{n} \Gamma\left(\gamma_j + m_j\right)} \prod_{j=1}^{n} \alpha_j^{m_j + \gamma_j - 1} \right] \prod_{j=1}^{n} \left[\frac{\left(\beta + \frac{1}{2}\sum_{i=1}^{m_j} x_i^2\right)^{(m_j+\lambda)}}{\Gamma(m_j+\lambda)\, \mu_j^{2(m_j+\lambda+1)}} e^{\frac{-(\beta+\frac{1}{2}\sum_{i=1}^{m_j} x_i^2)}{\mu_j^2}} \right] \quad (9)$$

Here, σ is the quantity of all remaining terms that can be ignored when taking the integral with respect to the parameters of interest. At this point, the posterior distribution of $\mu_j^2|.$ is
$\mu_j^2|. \sim \mathrm{INVGamma}\left(m_j + \lambda, \beta + \frac{1}{2}\sum_{i=1}^{m_j} x_i^2\right)$ (10).

The expected value of $\mu_j^2|.$ is calculated according to the formula

$$E(\mu_j^2|.) = \frac{\beta + \frac{1}{2}\sum_{i=1}^{m_j} x_i^2}{m_j + \lambda},$$

Which leads to

$$\hat{\mu}_j = \sqrt{\frac{\beta + \frac{1}{2}\sum_{i=1}^{m_j} x_i^2}{m_j + \lambda}} \quad (11)$$

In the same manner, the posterior distribution of $\propto$ is
$\propto \sim \mathrm{Dir}\,(m_1 + \gamma_1, m_2 + \gamma_2, \dots, m_n + \gamma_n)$ (12)

and the estimated value of $\propto$ will be equal to

$$\hat{\alpha}_j = E(\alpha_j|.) = \frac{m_j + \gamma_j}{\sum_{j=1}^{n}(m_j + \gamma_j)} \quad (13)$$

The probability that the object i comes from class j can be computed by the formula

$$\hat{p}_{ij} = P\left(Z_{ij} = 1 \middle| \hat{\mu}_j^2, \hat{\alpha}_j, X\right) = \frac{\hat{\alpha}_j\, f(x_i, \hat{\mu}_j)}{\sum_{j=1}^{n} \hat{\alpha}_j\, f(x_i, \hat{\mu}_j)} \quad ,(14)$$

where $\hat{p}_i = (\hat{p}_{i1}, \hat{p}_{i2}, \dots, \hat{p}_{in})$. The observation memberships can be drawn from multinational distribution according to $\hat{Z}_i \sim \mathrm{Mult}(1, \hat{p}_i)$. A_j is updated with respect to the estimated $\hat{Z}_i$.
Observation memberships are drawn from $\mathrm{Mult}(1, \hat{p}_i)$ after evaluating the Bayesian estimators $\hat{p}_i$.

3.3 Gibbs Sampler algorithms

Bayesian estimation for a mixture of Rayleigh distribution can be constructed through data augmentation by using the Gibbs sampler scheme for performing MCMC. This scheme estimates the augmented parameter (A, φ) by drawing it from the posterior distribution $p(A, \varphi|X)$ of the complete-data. The posterior distribution of the model parameters is proportional to the joint pdf which is defined in (8). The Gibbs sampler algorithm starts with some initial allocations $A^{(0)}$ that can be found from K-means algorithms and fixed values for the parameters α, η, β, where $j = 1,2, \dots, n$
and $A^{(0)} = \{A_1^{(0)}, A_2^{(0)}, \dots, A_n^{(0)}\}$. Then, we repeat the following steps for $r = 1,2, \dots, R$.

a) The parameters $(\alpha_1^{(r)}, \alpha_2^{(r)}, \dots, \alpha_n^{(r)})$ can be drawn from the Dirichlet distribution $Dir\left(\alpha_1 + m_1^{(r-1)}, , \alpha_2 + m_2^{(r-1)}, \dots, \alpha_n +\right.$

$\mathrm{m}_n^{(r-1)}\Big)$, Conditional on the allocations $A^{(r-1)}$, and α_j, where $\mathrm{m}_j^{(r-1)} = \#\ \mathrm{A}_j^{(r-1)}$.

b) Draw the components parameter $(\mu_j^2)^r$ from $\mathrm{Inv}Gamma\left(\mathrm{m}_j^{(r-1)} + \lambda, \beta + \frac{1}{2}\sum_{i=1}^{\mathrm{m}_j^{(r-1)}} x_i^2\right)$, provided that $x_i \in \mathrm{A}_j^{(r-1)}$.

c) Computing $\mathrm{p}_{ij}^{(r)}$ by the formula shown in (14):

$$\mathrm{p}_{ij}^{(r)} = \frac{\alpha_j^{(r-1)}\, f(x_i, (\mu_j^2)^{(r-1)})}{\sum_{j=1}^{n} \alpha_j^{(r-1)}\, f(x_i, (\mu_j^2)^{(r-1)})}.$$

d) Updating the allocation of all observations x_i conditional on estimated $\hat{p}_i{}^{(r)}$ by sampling $\mathrm{z}_j^{(r)}$ independently for each $i = 1,2,\dots,m$ from the conditional distribution $\mathrm{z}_i^{(r)} \sim Mult(1, \mathrm{p}_i^{(r)})$, which results in the new allocations

$$\mathrm{A}_j^{(r)} = \left\{x_i;\ \mathrm{z}_{ij}^{(r)} = 1\ , i = 1,2,\dots,m\right\}$$

Increase r by one, and return to step (a).

3.4 Metropolis-Hastings algorithm

One of the importing tools is drawing samples from the posterior distribution by the use of a Metropolis-Hastings algorithm [25-27]. Here, the posterior distribution of φ as in eq.11 and eq.13 is used. The MH algorithm, starts with some initial parameters $\varphi^{(0)}$, and initial allocation $\mathrm{A}_j^{(0)}$ that can be found from K-means algorithms. We then repeat the following steps for r=1, 2, …R.

a) i) Propose a new parameter $\mu_j^{2\,(new)}$ by drawing from a proposal density $q_1(\mu_j^2|\mu_j^{2\,(r-1)})$. Here q_1 will be the density of the normal distribution. For sampling from normal distribution is quite critical in terms of specifying mean and standard deviation. The suitable choice of standard deviation is to be calculated for each component from ${x_i}^2$ to ensure that we will cover all the area of parameter space which leads to accurate estimation.

ii) Propose a new parameter $\alpha^{(new)}$, where $\alpha^{(new)} = (\alpha_1^{(new)}, \alpha_2^{(new)}, \dots, \alpha_k^{(new)})^T$ by drawing from a proposal density $q_2(\alpha|\alpha^{(r-1)})$, such that $\sum_{j=1}^{n-1} \alpha_j^{(new)} < 1$, where $j = 1,2,\dots,n-1$. We then calculate $\alpha_n^{(new)} = 1 - \sum_{j=1}^{n-1} \alpha_j^{(new)}$. Here q_2 will be the density of uniform distribution.

b) i) Move the sampler to $\mu_j^{2\,(new)}$ with probability $\min(1, B_j)$, where

$$B_j = \frac{p(\mu_j^{2\,(new)}; X, \lambda, \beta) q_1(\mu_j^{2\,(r-1)} | \mu_j^{2\,(new)})}{p(\mu_j^{2\,(r-1)}; X, \lambda, \beta) q_1(\mu_j^{2\,(new)} | \mu_j^{2\,(r-1)})}.$$

If $\mathrm{u}_1 < \min(1, B_j)$, where u_1 is a random number from the uniform distribution U(0,1), then accept $\mu_j^{2\,(new)}$ and set $\mu_j^{2\,(m)} = \mu_j^{2\,(new)}$, otherwise reject $\mu_j^{2\,(new)}$ and set $\mu_j^{2\,(r)} = \mu_j^{2\,(r-1)}$. Note that the quantity $p(\mu_j^{2\,(new)}; X,\ \lambda, \beta)$ is the pdf of inverse gamma distribution with parameters shown in eq.(11).

ii) Move the sampler to $\alpha^{(new)}$ with probability $\min(1, V)$, where

$$V = \frac{P(\alpha^{(new)}; \alpha) q_2(\alpha^{(r-1)} | \alpha^{(new)})}{P(\alpha^{(r-1)}; \alpha) q_2(\alpha^{(new)} | \alpha^{(r-1)})}. \qquad (15)$$

If $\mathrm{u}_2 < \min(1, V)$, where u_2 is a random number sampled from the uniform distribution U(0,1), then accept $\alpha^{(new)}$ and set $\alpha^{(r)} = \alpha^{(new)}$, otherwise reject $\alpha^{(new)}$ and set $\alpha^{(new)} = \alpha^{(r-1)}$. Here, $p(\alpha^{(new)}, \alpha)$ represents the pdf of the posterior distribution shown in eq. (13) and q_2 is the pdf of U(0,1) which is usually equal to one.

c) Calculating $\mathrm{p}_{ij}^{(r)}$ by eq. (14) as shown below:

$$\mathrm{p}_{ij}^{(r)} = \frac{\alpha_j^{(r-1)}\, f(x_i, (\mu_j^2)^{(r-1)}}{\sum_{j=1}^{n} \alpha_j^{(r-1)}\, f(x_i, (\mu_j^2)^{(r-1)})}.$$

d) Updating the classification of all observations x_i conditional on estimated

$\hat{p}_i^{(r)}$ by drawing $\mathrm{Z}_i^{(r)}$ independently for each $i = 1,2,\dots,m$ from the multinomial distribution $\mathrm{z}_i^{(r)} \sim Mult(1, \hat{p}_i^{(r)})$, which results in the new classification $\mathrm{A}_j^{(r)} = \left\{x_i;\ \mathrm{z}_{ij}^{(r)} = 1\ , i = 1,2,\dots,m\right\}$.

Increase r by one and return to step (a).

4. Simulation

Data can be generated from two- components of the Rayleigh mixture distribution as follows:

a. Set the sample size to be of m observations and $\varphi = (\alpha_1, \mu_1, \mu_2)^T$.
b. For the observation $i = 1, 2, \dots, m$, we generate $u_i \sim unif(0,1)$.
c. If the $u_i < \alpha_1$, then generate $x_i \sim h(x; \mu_1)$ and set observation membership to be $z_i = (1,0)^T$. Otherwise, generate $x_i \sim h(x; \mu_2)$ and set observation membership to be $z_i = (0,1)^T$.

We can also generate data from three-components of the Rayleigh mixture distribution as follows:

a. Set the sample size to be of m observations and $\varphi = (\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3)^T$.
b. For the observation $i = 1, 2, \dots, m$, we generate $u_i \sim unif(0,1)$.
c. If the $u_i < \alpha_1$, then generate $x_i \sim h(x; \mu_1)$ and set observation membership to be $z_i = (1,0,0)^T$. If $\alpha_1 < u_i < \alpha_2$, generate $x_i \sim h(x; \mu_2)$ and set observation membership to be $z_i = (0,1,0)^T$. Otherwise, generate $x_i \sim h(x; \mu_3)$ and set observation membership to be $z_i = (0, 0, 1)^T$.

Note that $h(x; \mu_j)$ is the pdf of Rayleigh distribution that is defined in Eq. 1. It is also important to mention that is at the end of step c, we know each observation come from which component.

To generate x_i from the mixture of two components, we use the following predefined values for the parameters $\varphi = (\alpha_1, \mu_1, \mu_2)^T$, where $(\alpha_1, \mu_1, \mu_2) = (0.35,1,4), (0.35,0.5,2), (0.25,2\ ,3)$, and sample sizes $m = 60, 90, 120, 200$. Whereas, to generate x_i from the mixture of three components, we use the following predefined values for the parameters $\varphi = (\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3)^T$,where $(\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3) = (0.25, 0.35,1,4, 7), (0.25, 0.35,0.5,2, 4), (0.25, 0.35, 2\ ,6, 10)$ and $m = 60, 90, 120, 200$. After getting $\hat{\varphi}$ of the three methods, we calculate MSE and CSR for each data set generated from a particular scenario. The results of various scenarios and methods were compared by the mean square error (MSE) and the classification successful rate (CSR) of T replications. At iteration t, MSE can be calculated according to the formulas

$$MSE_t = \frac{\sum_{i=1}^{m}\left(f(x_i;\varphi) - f\left(x_i;\hat{\varphi}^{(k)}\right)\right)^2}{m}.$$

For all replications, the mean of MSE was computed as

$MMSE = \frac{\sum_{t=1}^{T} MSE_t}{T}$ where MSE_t is the MSE of the mixture probability density function in the replication t^{th}.

For computing MCSR, we let SR_t be

$SR_t = \#\ \{z_i = \hat{\mathrm{z}}_i^{(t)}, i = 1, 2, \dots, m\}$.

Here, $\hat{Z}_i^{(t)}$ is denote the estimated membership of the observations in the replication t^{th}. The classification successful rate is computed as

$$CSR_t = \frac{SR_t}{m}\text{, for the replication } t^{th}.$$

The mean of classification successful rate (MCSR) can be computed as

$$MCSR = \frac{\sum_{t=1}^{T} CSR_t}{T},$$

Where T is the total number of replications which has been chosen to be 500 in this study.

To apply the methods, we first determine the number of components according to K-means algorithms that are mentioned in Section 3.4. We then apply the three methods to the same

data that was simulated from one particular scenario. After applying the methods of posterior expectation, Gibbs sampler and metropolis-hasting to the simulated data according to the three scenarios of predefined values and four sample sizes. It can be seen from Table 1 that the Gibbs sampler outperformed the other two methods in terms of both MMSE and MCSR for all cases. The table also shows that the metropolis-hasting method gives good results compared to the traditional posterior expectation. In Table 2, it can be noticed that estimation of the parameters is also good but not better than in the case of two components as shown from the values of MMSE and MCSR. The Gibbs sampler is also outperformed the others in all scenarios.

To apply the methods, we first determine the number of components according to K-means algorithms that are mentioned in Section 3.4. We then apply the three methods to the same data that was simulated from one particular scenario. After applying the methods of posterior expectation, Gibbs sampler and metropolis-hasting to the simulated data according to the three scenarios of predefined values and four sample sizes. It can be seen from Table 1 that the Gibbs sampler outperformed the other two methods in terms of both MMSE and MCSR for all cases. The table also shows that the metropolis-hasting method gives good results compared to the traditional posterior expectation. In Table 2, it can be noticed that estimation of the parameters is also good but not better than in the case of two components as shown from the values of MMSE and MCSR. The Gibbs sampler is also outperformed the others in all scenarios. Overall, all methods showed reasonable results and the improvement of results increased as the sample size increased.

The graphs of Figure 1 and Figure 2 show the curves of the mixture pdf based on the true values that we used to generate the data and the estimated values of the parameters by the posterior expectation, Gibbs sampler and metropolis-hasting. They show these curves with respect to the space of the random variable. It can be seen that the curves become obvious in terms of the number of model components when there is a large distance between the component parameters. The graphs also show that the curves of pdf estimated by Gibbs sampler and MH are close to the one based on the true values, which means that the estimations are good in all scenarios.

5. Conclusion and discussion

We proposed the conjugate priors of the inverse gamma distribution for the square of component parameters which leads to a posterior distribution of the same distribution. We also used the Dirichlet distribution as a conjugate prior for the weight parameters, which resulted in posterior distributions of the same distribution. We then used the derived posterior distributions in the popular Bayesian framework, Gibbs sampler and Metropolis-Hastings. These methods may be used with any number of components of a finite mixture. However, as the number of components increased, the accuracy of the estimation decreased unless we used large samples. The important finding that can be concluded from this study is that using conjugate priors can result in accurate posterior distributions that give the best estimators. We also conclude that Gibbs sampler and MH methods perform better than EM algorithms that has been discussed in a recent publication [16]. The possible reason is that the EM algorithm at the maximize step

may be stuck in the local maximum. Whereas Gibbs sampler and MH are based on sampling from the whole posterior distribution, which can lead to better estimation.

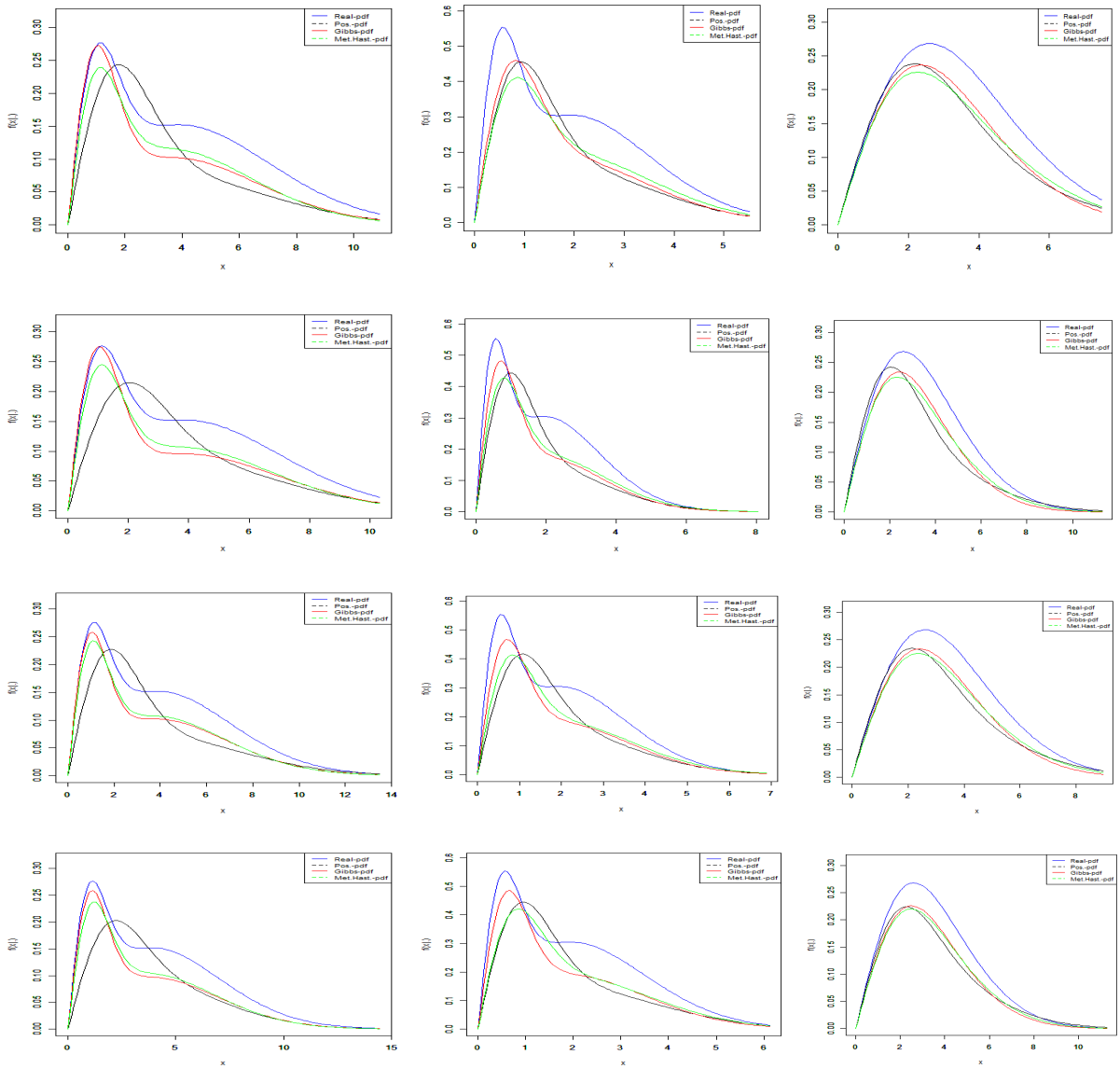

Figure 1: Figure shows the curves of the probability density function $f(x|.)$ of the two-components mixture model that is calculated according to Eq. 2 for data generated from scenarios $(\alpha_1, \mu_1, \mu_2) = (0.35, 1, 4), (0.35, 0.5, 2), (0.25, 2, 3)$, which represent the columns from left-to-right, respectively. The rows from top-to-bottom represent the sample sizes (60, 90, 120, 200), respectively.

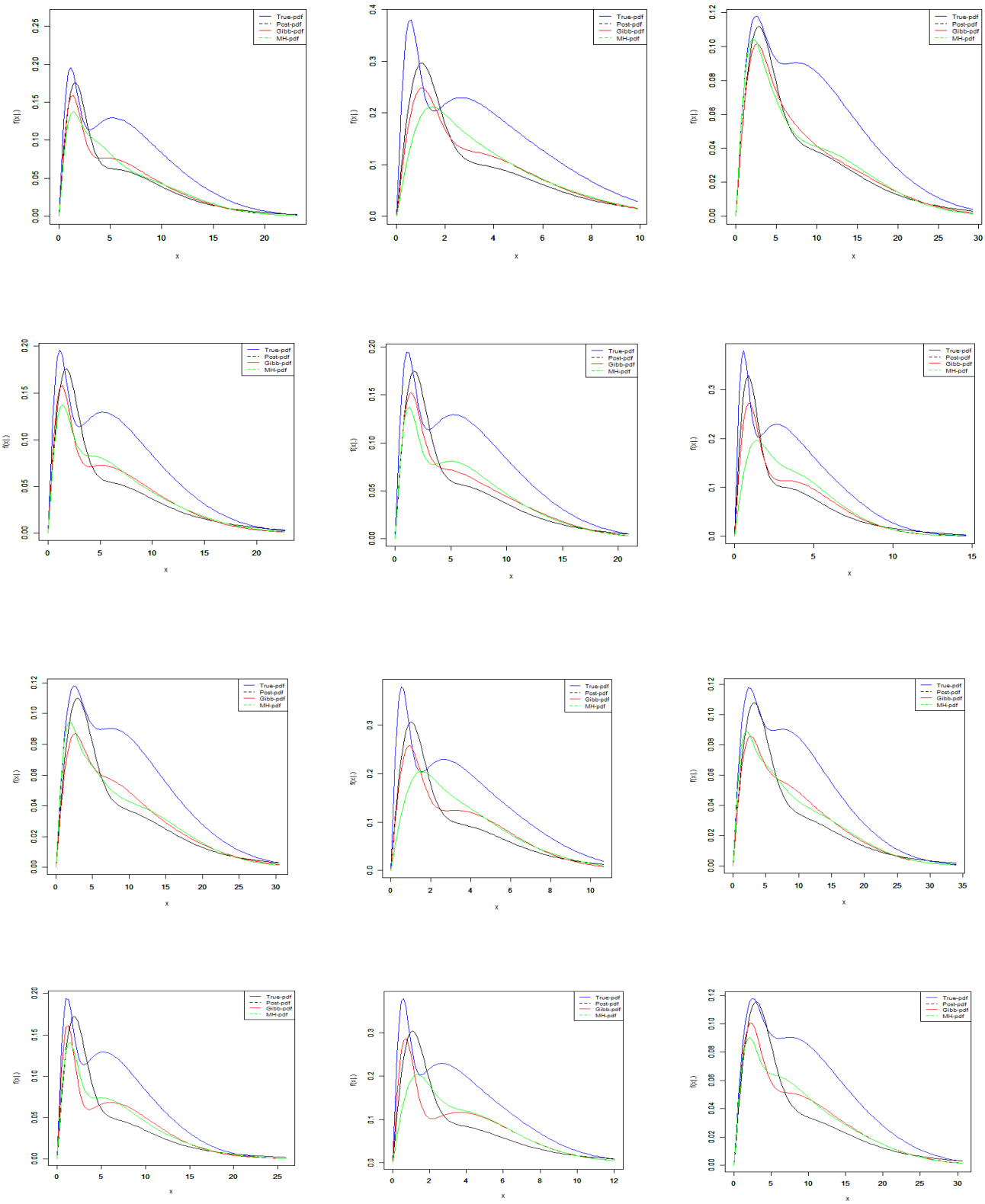

Figure 2: Figure shows the curves of the probability density function $f(x|.)$ of the three-components mixture model that is calculated according to Eq. 2. , according to the predefined parameters scenarios $(\alpha_1, \alpha_2, \mu_1, \mu_2, \mu_3) = (0.25, 0.35, 1, 4, 7), (0.25, 0.35, 0.5, 2, 4), (0.25, 0.35, 2, 6, 10)$ Which represent the columns from left-to-right, respectively. The rows from top-to-bottom represent the sample sizes (60, 90, 120, 200), respectively.

Table 1: The table shows the MMSE and MCSR of applying the methods of estimating parameters by posterior expectation, Gibbs sampler and metropolis-hasting based on three scenarios of parameters and four different sample sizes.

M	Method	$\alpha_1 = 0.35, \mu_1 = 1, \mu_2 = 4$		$\alpha_1 = 0.35, \mu_1 = 0.5, \mu_2 = 2$		$\alpha_1 = 0.25, \mu_1 = 2, \mu_2 = 3$	
		MMSE	MCSR	MMSE	MCSR	MMSE	MCSR
60	Post.	0.0062	0.76	0.0184	0.78	0.0029	0.56
	Gibbs	0.00211	0.85	0.0096	0.83	0.0016	0.70
	MH	0.00131	0.80	0.0118	0.78	0.0019	0.68
90	Post.	0.0055	0.78	0.0177	0.74	0.0033	0.55
	Gibbs	0.0018	0.87	0.0075	0.85	0.0015	0.71
	MH	0.0014	0.81	0.0104	0.80	0.0018	0.65
120	Post.	0.0051	0.76	0.0163	0.74	0.0033	0.62
	Gibbs	0.0016	0.87	0.0074	0.85	0.0015	0.73
	MH	0.0014	0.82	0.0101	0.81	0.0017	0.65
200	Post.	0.0045	0.79	0.0154	0.79	0.0029	0.62
	Gibbs	0.0014	0.87	0.0060	0.87	0.0015	0.74
	MH	0.0014	0.83	0.0097	0.82	0.0017	0.63

Table 2: The table shows the MMSE and MCSR of applying the methods of estimating parameters by posterior expectation, Gibbs sampler and metropolis-hasting based on three scenarios of parameters and four different sample sizes for data generated from three components mixture of Rayleigh distribution.

M	Method	$\boldsymbol{\alpha_1 = 0.25, \alpha_2 = 0.35, \mu_1 = 1, \mu_2 = 4, \mu_3 = 7}$		$\boldsymbol{\alpha_1 = 0.25, \alpha_2 = 0.35, \mu_1 = 0.5, \mu_2 = 2, \mu_3 = 4}$		$\boldsymbol{\alpha_1 = 0.25, \alpha_2 = 0.35, \mu_1 = 2, \mu_2 = 6, \mu_3 = 10}$	
		MMSE	MCSR	MMSE	MCSR	MMSE	MCSR
60	Post.	0.0035	0.53	0.0081	0.54	0.0019	0.51
	Gibbs	0.0027	0.755	0.0071	0.70	0.0011	0.65
	MH	0.0029	0.65	0.076	0.72	0.0013	0.64
90	Post.	0.0029	0.50	0.0070	0.55	0.0017	0.55
	Gibbs	0.0024	0.772	0.0062	0.76	0.0011	0.77
	MH	0.0026	0.67	0.0065	0.78	0.0013	0.75
120	Post.	0.0024	0.51	0.0067	0.57	0.0015	0.58
	Gibbs	0.0022	0.82	0.0054	0.78	0.0009	0.79
	MH	0.0023	0.70	0.0059	0.81	0.0010	0.77
200	Post.	0.0021	0.43	0.0065	0.59	0.0012	0.59
	Gibbs	0.0018	0.84	0.0049	0.79	0.0008	0.81
	MH	0.0019	0.74	0.0053	0.83	0.0010	0.78

References

[1] Jones, P. N., & McLachlan, G. J. (1990). Laplace-normal mixtures fitted to wind shear data. *Journal of Applied Statistics, 17*(2), 271–276.

[2] Kanji, G. K. (1985). A mixture model for wind shear data. *Journal of Applied Statistics, 12*(1), 49–58.

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

[4] Rao, G. S. (2012). Estimation of reliability in multicomponent stress-strength based on generalized exponential distribution. *Revista Colombiana de Estadistica, 35*(1), 67–76.

[5] Ueda, N., Nakano, R., Ghahramani, Z., & Hinton, G. E. (2000). SMEM algorithm for mixture models. *Neural Computation, 35*(1), 2109–2128.

[6] Peel, D., & McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing, 10*, 335–344.

[7] Cadez, I. V., McLaren, C. E., Smyth, P., & McLachlan, G. J. (1999). Hierarchical models for screening of iron deficiency anemia. In *Proceedings of the Sixteenth International Conference on Machine Learning* (pp. 77–86). Morgan Kaufmann.

[8] Lehmann, E. L. (1980). Efficient likelihood estimators. *The American Statistician, 34*, 233–235.

[9] Lehmann, E. L. (1983). *Theory of Point Estimation*. Wiley.

[10] Montuelle, L., & Le Pennec, E. (2014). Mixture of Gaussian regressions model with logistic weights, a penalized maximum likelihood approach. *Electronic Journal of Statistics, 8*(1), 1661–1695.

[11] Boldea, O., & Magnus, J. R. (2009). Maximum likelihood estimation of the multivariate normal mixture model. *Journal of the American Statistical Association, 104*(488), 1539–1549.

[12] Karlis, D., & Xekalaki, E. (1999). On testing for the number of components in finite Poisson mixtures. *Annals of the Institute of Statistical Mathematics, 51*(1), 149–162.

[13] Henna, J. (1985). On estimating the number of constituents of a finite mixture of continuous distributions. *Annals of the Institute of Statistical Mathematics, 37*, 235–240.

[14] Liu, C., & Rubin, D. B. (1998). Maximum likelihood estimation of factor analysis using the ECME algorithm with complete and incomplete data. *Statistica Sinica, 8*, 729–747.

[15] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological), 44*(2), 226–233.

[16] Liu, C., & Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica, 5*(1), 19–39.

[17] Ali, F., & Zhang, J. (2017). Mixture model-based association analysis with case-control data in genome wide association studies. *Statistical Applications in Genetics and Molecular Biology, 16*(3), 173–187.

[18] Mohammed, N., & Ali, F. (2022). Estimation of parameters of finite mixture of Rayleigh distribution by the expectation-maximization algorithm. *Journal of Mathematics, 2022*(1), 1–7.

[19] Ali, F. (2023). Bayesian methods for estimation of parameters of finite mixture of inverse Rayleigh distribution. *Mathematical Problems in Engineering, 2023*, 1–9.

[20] Wong, C. S., & Li, W. K. (2000). On a mixture autoregressive model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(1), 95–115. https://doi.org/[DOI]

[21] Le, N. D., Martin, R. D., & Raftery, A. E. (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association, 91*(436), 1504–1514. https://doi.org/[DOI]

[22] Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association, 82*(398), 528–550.

[23] Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*(410), 398–409.

[24] Cowles, M. K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*(434), 883–904.

[25] Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association, 95*(451), 957–970.

[26] Brooks, S. P. (2001). On Bayesian analysis and finite mixtures for proportions. *Statistics and Computing, 11*(3), 179–190.

[27] Viallefont, V., Richardson, S., & Green, P. J. (2002). Bayesian analysis of Poisson mixtures. *Journal of Nonparametric Statistics, 14*(1-2), 181–202.

[28] Pezoulas, V. C., Tachos, N. S., Gkois, G., Olivotto, I., Barlocco, F., & Fotiadis, D. I. (2022). Bayesian inference-based Gaussian mixture models with optimal components estimation towards large-scale synthetic data generation for in silico clinical trials. *IEEE Open Journal of Engineering in Medicine and Biology, 3*, 108–114.

[29] Simola, U., Cisewski-Kehe, J., & Wolpert, R. L. (2021). Approximate Bayesian computation for

finite mixture models. *Journal of Statistical Computation and Simulation, 91*(6), 1155–1174.

[30] Su, X., Zamzami, N., & Bouguila, N. (2022). A fully Bayesian inference with Gibbs sampling for finite and infinite discrete exponential mixture models. *Applied Artificial Intelligence, 36*(1).

[31] Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics, 28*(1), 100–108.

[32] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). CRC Press.