Iraqi Journal of Statistical Sciences

www.stats.mosuljournals.com

Estimating the Parameters of Mixture Gamma Distributions Using Maximum Likelihood and Bayesian Method

Nagham Ibrahim Abdulla Najm and Raya Salim Al_Rassam

Department of Statistics and Informatics, College of Computer Sciences and Mathematics, University of Mosul, Mosul, Iraq

Article information

Article history:
Received September 16, 2023
Accepted November 5, 2023
Available online June 1, 2024

Keywords:
Gamma distribution, Mixture distribution, Bayesian estimation, Likelihood function, Expectation Maximization Algorithm, Metropolis-Hastings

Correspondence:
Nagham Ibrahim Abdulla
nagham.21csp96@student.Uomosu.edu.iq.

Abstract

This paper focuses on the mixture Gamma distribution and uses the maximum likelihood and Bayesian techniques to estimate its parameters. This study uses Expectation Maximization Algorithm (EM) to find the maximum likelihood estimators and the random Metropolis-Hastings algorithm is used to simulate the Bayesian estimates of the parameters of mixture gamma distribution. then these estimates are compared by using the sum of the modulus of the bias (MBias), and the root-mean square error (RMSE). It has been shown that the Bayesian estimator is better than the maximum likelihood estimator.

DOI 10.33899/IQJOSS.2024.183254

1- Introduction

A random variable is always considered as a sample from a distribution. This may be well-known distribution or not. Some random variables are drawn from one single distribution , such as the normal distribution but this is not always so easy because in real-life the random variables might have been generated from a mixture of several distributions.

In studying mixture distributions the formula of this distribution have been difficult then it is used some algorithms to facilitate finding the estimators , where EM algorithm is used to find the maximum likelihood estimators and the metropolis Hastings algorithm to find the Bayesian estimators . if the distribution is an exponential family , with density $f(\theta) = C(\theta)h(x)exp(\emptyset(\theta)S(x))$,then a conjugate prior distribution for θ exists and the prior distribution $p(\theta) \propto C(\theta)EXP(\emptyset(\theta)b)$ is conjugate to the likelihood of the exponential family , see (Bernardo,2009).

Many authors considered estimating the parameters of the mixture distributions. For example, (Newcomb ,1886) suggested an iterative reweighting scheme that can be viewed as an application of the EM algorithm of (Dempster et al. ,1977) to compute the common mean of a mixture in known proportions of a finite number of univariate normal distributions with known variances. (Jewell ,1982) provided maximum likelihood estimates of mixture of exponential distributions using EM algorithm..(Li L.A., 1983) quoted several features of mixture models and defined two types of mixture models. If the component distributions of a mixture belong to same family, their mixture is known as a type-I mixture model. Whereas, a type-II mixture model is defined as the component distributions of a mixture belong to different families .. (Upadhyay et. al. ,2002) proposed Bayesian inference in life testing and reliability by using Markov Chain Monte Carlo (MCMC). (Pang et. al. ,2004) used MCMC techniques to carry out a Bayesian estimation procedure using Hirose's simulated data. (Chojogh,B,et al,2019) presented a research in which he clarified mixture distributions the research include model of the normal mixture distribution and Poisson mixture distribution for tow component and for k-components and estimating the parameters of these model using (EM) algorithm. ("A mixture model for determining SARS-COV-2 variant composition in pooled samples") presented a research includes a mixture model distributions and apply it to a set of variables SARS-COV-2 the model is built by looking at a pre-defined set of data ,the results showed that these models support these data well.

Gamma Distribution

It is a type of continuous probability distribution and is used in many fields such as Statistics, Economics, Physics, Computer Science and others, the Gamma distribution can be determined by two parameters, the shape parameter (α) and the scale parameter (β), and the probability density function (pdf) for this distribution is as follows: -

$$f(x) = \frac{\beta^{\alpha}}{\Gamma\alpha} x^{\alpha-1} e^{-\beta x} \qquad (1)$$

where $\alpha > 0$, $\beta > 0$ and $x > 0$.

2- Mixture Distribution Models

It is the process of analyzing data to determine the best mixture model that can be used to describe the observed data. Mixture models consist of several different probability distributions and are characterized by their ability to represent the distribution of data more accurately than single models.

Every random variable can be considered as a sample from a distribution, . Some random variables are drawn from one single distribution, such as a normal distribution. But life is not always so easy! Most of real-life random variables might have been generated from a mixture of several distributions and not a single distribution.

Random variables usually come from only one distribution, like (gamma distribution or normal distribution), but in real life there are some variables that come from several mixture distributions and these distribution may be from the same family, i.e. from one family, for example, all of them from the normal distribution, but with different parameters, or these distributions may be different, for example (gamma distribution and normal distribution) together.

Let $X_1, X_2, X_3, \dots, X_n$ be independent random variables and $x_1, x_2, x_3, \dots, x_n$ the observations of the random variable and the probability density function for the mixture distribution (pdf) containing k of the components can be expressed as follows:-

$$f(x) = \sum_{j=1}^{k} \lambda_j f_j(x|\theta_j) \quad (2)$$

where λ_j represents the mixture weights and is $0 < \lambda_j < 1$ and $\sum_{j=1}^{k} \lambda_j = 1$ and $f_j(x|\theta_j)$ represents the probability density function of the variable (x) and $\theta = (\theta_1, \theta_2, \dots \theta_k)$ represents the parameters vector of the mixture distribution, and it is worth noting that the parameter θ is treated as a random variable rather than a constant (Tahir & et al, 2016). The mixture gamma distribution of k of components is written as follows:-

$$f(x, \alpha, \beta, \lambda) = \sum_{j=1}^{k} \lambda_j \frac{{\beta_j}^{\alpha_j}}{\Gamma \alpha_j} x^{\alpha_j - 1} e^{-\beta_j x_i} \quad (3)$$

$$\alpha_j > 0 \ , \ \beta_j > 0 \ , \times > 0 \ , k > 1 \ .$$

3- SOME METHODS OF ESTIMATE THE PARAMETERS OF MIXTURE DISTRBUTION

Mixture distributions are common statistical distributions, which are used in many fields such as data analysis, machine learning, and others, and these distributions depend on the idea of collecting several simple distributions together to produce a complex distribution. And these distributions need to estimate a set of parameters that determine the distribution of mixture data.

When we have a sample size n $(x_1, x_3, x_2, \dots x_n)$ are randomly drawn from a known distribution but the distribution parameters are unknown, for example a sample drawn from the normal distribution with unknown parameters (mean and variance), the main objective is to estimate the parameters of this distribution. In this study, we will discuss two methods for estimating parameters of mixture distribution.

A- Maximum Likelihood Estimation (MLE):

This method is one of the most important methods of point estimation and was proposed by the famous statistician Fisher in 1920, as it assumes that the parameters to be estimated for a particular population is an unknown fixed quantity which estimated based on the sample data.

Assume we have a sample with size n, i.e., $(x_1, x_3, x_2, \dots x_n)$ Also assume that we know the distribution from which this

sample has been randomly drawn but we do not know the parameters of that distribution. The principle of this method is to find an estimate such as $\hat{\theta}$ for the parameter θ which makes the likelihood function at its maximum value.

If $x_1, x_3, x_2, \dots x_n$ are random variables and these variables have an independent and identically distributed (iid) and size (n) and drawn from a population with a probability density function $f(x|\theta)$, the estimator of the likelihood function that makes the likelihood function at its maximum value can be obtained by deriving the likelihood function and equating it to zero. The likelihood function will be as follows:-

$$L(\theta_j) = \prod_{i=1}^{n} f(x_i|\theta_j) \quad (4)$$

by used (2)

$$L(\theta_j) = \prod_{i=1}^{n} \sum_{j=1}^{k} \lambda_j f(x_i|\theta_j) \quad (5)$$

by given log:

$$lnL(\theta_j) = \sum_{i=1}^{n} ln[\sum_{j=1}^{k} \lambda_j f(x_i|\theta_j)] \quad (6)$$

Then we take the partial derivative of θ_j once and for λ_j again to get the equation for each parameter, but it will be difficult to solve the equations formed directly because of the presence of addition operations inside the logarithm, so it is necessary to rely on numerical methods and algorithms that use iterative operations in order to reach the maximum likelihood estimator (Friedman & *et al*, 2009).

Expectation Maximization Algorithm (EM):

The expectation maximization algorithm (EM) was proposed by (Dempster, Laird & Rubin ,1977) and still to this day, and it is one of the most important methods to find the maximum likelihood estimators in the case of latent variables or missing values. And this algorithm is used in statistics and machine learning to solve problems related to statistical analysis of data such as classification, aggregation and factor analysis (Filho, 2008).

For example, assuming the collection of data about a particular disease, where the severity of the disease was not recorded, but the presence or absence of the disease was recorded, i.e. the absence of the disease was expressed by zero, and the presence of the disease was expressed in

$x> 0$, in this case we do not know the values of x, is it 100 or 5 ?, in this case, we cannot use the method of maximum likelihood because there are missing values.

The expectation maximization algorithm consists of two steps (Chris & Raftery, 2017):

a-Step One: E-Step

This step aims to estimate probability distributions by taking the expectation of the logarithm of the likelihood function in order to find an appropriate estimate of the parameters.

$$Q(\theta) = E[ln\ ln\ l(\theta)]$$

here the missing values are treated as constants and not variables (Chojogh *et al*, *2019*).

b-Step Two: M-Step

This step aims to determine the optimal values of the parameters using the expectation function in the first step.

$$\hat{\theta} = argmaxQ(\theta)$$

To estimate mixture Gamma distribution we have the p.d.f of mixture Gamma distributions

$$f(x|\theta_j) = \sum_{j=1}^{k} \lambda_j f_j(x|\theta_j)$$

by used (3)

$$f(x_i|\alpha,\beta,\lambda) = \sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i{}^{\alpha_j-1} e^{-\beta_j x_i}$$

$$L(\alpha,\beta,\lambda) = \prod_{i=1}^{n}\sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i{}^{\alpha_j-1} e^{-\beta_j x_i}$$

Taking the logarithm to the above equation we get

$$ln \, ln \, L(\alpha,\beta,\lambda) = \sum_{i=1}^{n} ln[\sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} {x_i}^{\alpha_j-1} e^{-\beta_j x_i}] \quad (7)$$

Optimizing this log-likelihood is difficult because of the summation within the logarithm. However, we can use the indicator parameter z_i for each observation x_i as follows (Corduneanu and Bishop, 2001).

$$z_{ij} = \{1 \;\; if\ the\ observation\ x_i\ belong\ to\ the\ j^{th}\ component\ 0 \qquad otherwise.$$

And the probability is:

$$p(z_{ij} = 1) = \lambda_j$$
$$p(z_{ij} = 0) = 1 - \lambda_j$$

For fixed i , $\sum_{j=1}^{k} z_{ij} = 1$, $z_{ij} \sim Bernolli(\lambda_j)$and $z_i|\lambda \sim multinomial(1, \lambda_1, \lambda_2, , , \lambda_k)$, the probability density function for z_i as the following form:
(Saeed,2005)

$$(\lambda) = \frac{1!}{z_{i1}!\, z_{i2}! \ldots .\, z_{ik}!} \prod_{j=1}^{k} \lambda_j^{z_{ij}} = \prod_{j=1}^{k} \lambda_j^{z_{ij}}$$

Since $z_1, z_2, , , z_n$ Are independent, we write the joint indicator density as the following form:

$$f(\lambda) = \prod_{i=1}^{n} \prod_{j=1}^{k} \lambda_j^{z_{ij}} \quad (8)$$

where $(x_1, z_1), (x_2, z_2), \ldots, (x_n, z_n)$ denoted the complement data. Therefore we can write the joint pdf of the observation x_i and the indicator z_i as following form:

$$f(\alpha,\beta,\lambda) = \sum_{j=1}^{k} z_{ij} f_j(x_i|\theta_j)$$

$$= \prod_{j=1}^{k} [f_j(x_i|\theta_j)]^{z_{ij}}$$

$$= \prod_{j=1}^{k} (\lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} {x_i}^{\alpha_j-1} e^{-\beta_j x_i})^{z_{ij}} \quad (9)$$

and the complement data likelihood is given by:

$$L(\alpha,\beta,\lambda) = \prod_{i=1}^{n} f(\alpha,\beta,\lambda) \quad (10)$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{k} [\lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} {x_i}^{\alpha_j-1} e^{-\beta_j x_i}]^{z_{ij}}$$

$$= \prod_{j=1}^{k} \left[\left[\frac{\lambda_j {\beta_j}^{\alpha_j}}{\Gamma\alpha_j} \right]^{\sum_{i=1}^{n} z_{ij}} [\prod_{i=1}^{n} x_i^{z_{ij}}]^{\alpha_{j-1}} e^{-\beta_j \sum_{i=1}^{n} x_i z_{ij}} \right]$$

$$= \prod_{j=1}^{k} \left[\left[\frac{\lambda_j {\beta_j}^{\alpha_j}}{\Gamma\alpha_j} \right]^{\sum_{i=1}^{n} z_{ij}} p_j^{\alpha_j-1} e^{-\beta_j \sum_{i=1}^{n} x_i z_{ij}} \right] \quad (11)$$

Where $p_j = \prod_{i=1}^{n} x_i^{z_{ij}}$

The log of the complement data likelihood function is

$$lnL(\alpha,\beta,\lambda) = \sum_{j=1}^{k} \sum_{i=1}^{n} z_{ij} ln\lambda_j + \sum_{j=1}^{k} \alpha_j \sum_{i=1}^{n} z_{ij} ln\beta_j - \sum_{j=1}^{k} \sum_{i=1}^{n} z_{ij} ln\Gamma\alpha_j + \sum_{j=1}^{k} (\alpha_j - 1) \sum_{i=1}^{n} z_{ij} ln x_i - \sum_{j=1}^{k} \beta_j \sum_{i=1}^{n} x_i z_{ij} \quad (12)$$

The z_{ij} is latent or missing value because we do not know whether it be $z_{ij} = 0$ or $z_{ij} = 1$ therefore we used the Expectation Maximization(EM) to estimate the parameters (Sattayatham and Talangtam, 2012).

Case 1:E-Step

$$E[x_i] = 0 \times P(x_i, \alpha, \beta, \lambda) + 1 \times P(x_i, \alpha, \beta, \lambda)$$

$$= P(x_i, \alpha, \beta, \lambda) = w_{ij} \quad (13)$$

$$(x_i) = \frac{f(z_{ij}=1)p(z_{ij}=1)}{\sum_{j=1}^{k} f(z_{ij}=1)p(z_{ij}=1)}$$

$$= \frac{\lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_{j-1}} e^{-\beta_j x_i}}{\sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_{j-1}} e^{-\beta_j x_i}} \quad (14)$$

The expected complete log-likelihood is

$$E[lnL(\alpha,\beta,\lambda)] = \sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} ln\lambda_j + \sum_{j=1}^{k}\alpha_j \sum_{i=1}^{n} T_{ij} ln\beta_i - \sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} ln\Gamma\alpha_j + \sum_{j=1}^{k}(\alpha_j - 1)\sum_{i=1}^{n} T_{ij} ln x_i - \sum_{j=1}^{k}\beta_j \sum_{i=1}^{n} x_i T_{ij} \quad (15)$$

$$\sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} = \sum_{j=1}^{k}\sum_{i=1}^{n} P(x_i)$$

$$= \sum_{i=1}^{n} \frac{\lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_{j-1}} e^{-\beta_j x_i}}{\sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_{j-1}} e^{-\beta_j x_i}}$$

$$= \sum_{i=1}^{n} 1 = n \quad (16)$$

Case 2:M-SteP

$$\boldsymbol{E[lnL(\alpha,\beta,\lambda)] = \sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} ln\lambda_j + \sum_{j=1}^{k}\alpha_j \sum_{i=1}^{n} T_{ij} ln\beta_i - \sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} ln\Gamma\alpha_j + \sum_{j=1}^{k}(\alpha_j - 1)\sum_{i=1}^{n} T_{ij} ln x_i - \sum_{j=1}^{k}\beta_j \sum_{i=1}^{n} x_i T_{ij} - \alpha\left(\sum_{j=1}^{k}\lambda_j - 1\right)} \quad (17)$$

$$\frac{\partial E[lnL]}{\partial \lambda_j} = \frac{\sum_{i=1}^{n} T_{ij}}{\lambda_j} - \alpha = 0$$

$$\Rightarrow \lambda_j = \frac{\sum_{i=1}^{n} T_{ij}}{\alpha}$$

$$\Rightarrow \sum_{j=1}^{k} = \sum_{j=1}^{k} \frac{\sum_{i=1}^{n} T_{ij}}{\alpha}$$

$$\Rightarrow \sum_{j=1}^{k} \frac{\sum_{i=1}^{n} T_{ij}}{\alpha} = 1$$

$$\Rightarrow \frac{1}{\alpha}\sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij} = 1$$

$$\Rightarrow \alpha = \sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij}$$

$$\Rightarrow \lambda_j = \frac{\sum_{i=1}^{n} T_{ij}}{\sum_{j=1}^{k}\sum_{i=1}^{n} T_{ij}}$$

$$\lambda_j = \frac{\sum_{i=1}^{n} T_{ij}}{n} \tag{18}$$

$$n \times \lambda_j = \sum_{i=1}^{n} T_{ij} = N_j$$

$$\frac{\partial E[lnL]}{\partial \alpha_j} = \sum_{i=1}^{n} T_{ij}\, ln\beta_j - \sum_{i=1}^{n} T_{ij}\frac{d ln\Gamma \alpha_j}{\partial \alpha_j} + \sum_{i=1}^{n} T_{ij}\, lnx_i = 0 \tag{19}$$

We solve this equation by Newton’s Raphson method

$$\frac{\partial E[lnL]}{\partial \beta_j} = \frac{\sum_{i=1}^{n} T_{ij}\alpha_j}{\beta_j} - \sum_{i=1}^{n} x_i T_{ij} = 0$$

$$\beta_j = \frac{\sum_{i=1}^{n} T_{ij}\alpha_j}{\sum_{i=1}^{n} x_i T_{ij}} \tag{20}$$

2-Bayesian Estimation Approach

In many cases, it is easy to find a suitable formula for the posterior distribution, but sometimes we may face difficulties in finding posterior distributions, which may require the integration of high- dimensional functions (high-grade functions), so it was necessary to develop methods that facilitate the process of finding posterior distributions and solve this problem, and the most important of these methods is the Markov Chain Monte Carlo (MCMC) where this method was used by researchers in the early 1990s and was widely applied to solve Bayes' problems as it relies on the idea of obtaining a random sample of conditional distributions of parameters .

The most commonly used methods of the Markov Chain Monte Carlo (MCMC) are the Gibbs Sampling Algorithm and the Metropolis-Hastings Algorithm, which we will use in this paper.

Metropolis - Hastings Algorithm

The Metropolis-Hastings algorithm is one of the main methods of the (MCMC) the main methods to estimate the parameters of mixture distributions and is used in many scientific and engineering applications, especially in the fields of Statistics and Physics.

Let x_1, x_3, x_2, … x_n be identically distributed random (iid) and have a probability density function $f(x|\theta)$ and we do not know the posterior distribution of the parameters of this function and suppose that $q(\theta|\theta^{'})$ is a candidate distribution with the parameter θ', the steps of this algorithm are: (al-masri,2020)

Metropolis-Hastings Algorithm Steps:

1-Choose an initial value for the parameter $\theta^{(0)}$ so that it is close to the parameter values of the real data.

2-Choose the default sample sizes for random variable observations *x*

3-We make a repetition from i=1,2,…,N

a- We generate a suggested value θ' followed the proposed distribution (we use the prior distribution for each model).

b- We calculate the acceptance probability:

$$p(\theta^{i-1}, \theta') = min[1, \frac{f(x_1, x_2, \dots, x_n) q(\theta', \theta^{i-1})}{f(x_1, x_2, \dots, x_n) q(\theta^{i-1}, \theta')}$$

where the numerator represents the value of the proposed parameter compensated in the equation for a conditional distribution. The denominator represents the value estimated by the equation of a conditional distribution..

c- Generate random numbers *ui* of uniform (0,1).

d- If $u_i < \alpha(\theta^{i-1}, \theta')$, we assume that $\theta^i = \theta'$ and if $u_i \geq \alpha(\theta^{i-1}, \theta')$, we assume that $\theta^i = \theta^{i-1}$

4- We repeat the previous steps each time by making *i=i +1* and go to step 1.*1-*z_i*Posterior*

When the indicator parameter z_i is unknown, for all observation x_i, $i = 1, 2, \dots, n$ and the scale parameter *a*, the shape parameters β and the weight parameter λ are known. The conjugate prior $p(z_i)$ of z_i is multinomial with hyper parameters $(1, \lambda_1, \lambda_2, \dots, \lambda_k)$.

By using Bayes' theorem, the posterior distribution:

$$(z_{ij} = 1|x_i, \alpha, \beta, \lambda) = \frac{f(x_i, \alpha, \beta, \lambda|z_{ij}=1)p(z_{ij}=1)}{\sum_{j=1}^{k} f(x_i, \alpha, \beta, \lambda|z_{ij}=1)p(z_{ij}=1)} = \frac{\lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_j - 1} e^{-\beta_j x_i}}{\sum_{j=1}^{k} \lambda_j \frac{\beta_j^{\alpha_j}}{\Gamma\alpha_j} x_i^{\alpha_j - 1} e^{-\beta_j x_i}} = w_{ij} \tag{21}$$

Since each z_{ij} takes two values only 1 or 0, then

$$p(x_i, \alpha, \beta, \lambda) = 1 - (x_i, \alpha, \beta, \lambda) = 1 - w_{ij} \tag{22}$$

Therefore, the posterior distribution $p(x_i, \alpha, \beta, \lambda)$ has a multinomial distribution $(1, w_{i1}, w_{i2}, \dots, w_{ik})$, where $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, k$.

2- λ_j Posterior

When the weight parameter λ_j is unknown and the scale parameter *a* and the shape parameters β are known. By ignoring terms that contain α , β in (11) the complete data likelihood function is given by:

$$L(\lambda) \propto \prod_{j=1}^{k} (\lambda_j)^{\sum_{i=1}^{n} z_{ij}}$$

$$\propto \prod_{j=1}^{k} (\lambda_j)^{N_j} \tag{23}$$

Where N_j is the number of the observations

$$N_j = n \times \lambda_j$$

$$= \sum_{i=1}^{n} E(z_{ij}|x_i)$$

By using (13)

$= \sum_{i=1}^{n}[0 \times \pi(x_i, \alpha, \beta, \lambda) + 1 \times \pi(z_{ij} = 1|x_i, \alpha, \beta, \lambda) = \sum_{i=1}^{n} \pi(z_{ij} = 1|x_i, \alpha, \beta, \lambda) = \sum_{i=1}^{n} w_{ij}$ (24)

The conjugate prior $p(\lambda)$ is a Dirichlet distribution with hyperparameters $\mu = (\mu_1, \mu_2, \ldots, \mu_k)$ is given

$$p(\lambda) = \frac{\Gamma \sum_{j=1}^{k}(M_j)}{\prod_{j=1}^{k} \Gamma M_j} \prod_{j=1}^{k} \lambda_j^{M_j - 1}$$

$M_j > 0$, $0 < \lambda_j < 1$

By ignoring terms that contain $\alpha, \beta$the posterior distribution is a Dirchlet with hyperparameters $(M_1 + \sum_{i=1}^{n} w_{i1}, M_2 + \sum_{i=1}^{n} w_{i2}, \ldots, M_k + \sum_{i=1}^{n} w_{ik})$ is given by

$$p(\lambda|\alpha, \beta, x, z) \propto L(\lambda)\pi(\lambda)$$

$$\propto \frac{\Gamma \sum_{j=1}^{k}(M_j)}{\prod_{j=1}^{k} \Gamma M_j} \prod_{j=1}^{k} \lambda_j^{M_j - 1} \left(\lambda_j\right)^{N_j}$$

$$\propto \frac{\Gamma \sum_{j=1}^{k}(M_j)}{\prod_{j=1}^{k} \Gamma M_j} \prod_{j=1}^{k} \lambda_j^{M_j - 1 + N_j}$$

$$\propto \prod_{j=1}^{k} \lambda_j^{M_j - 1 + \sum_{i=1}^{n} w_{ij}} \tag{25}$$

3-a_j Posterior

When the shape parameter a_j is unknown, for some $j = 1, 2, \ldots, k$ and both the weight parameter λ and the scale parameters $\beta$are known. By ignoring terms that contain $\beta, \alpha_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_k$ in (11), the complete data likelihood function is given by:

$$l(\alpha_j) \propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j}\right)^{\sum_{i=1}^{n} z_{ij}} \left(\prod_{i=1}^{n} x_i^{\alpha_j}\right)^{z_{ij}}$$

$$\propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j}\right)^{\sum_{i=1}^{n} z_{ij}} (p_j)^{\alpha_j} \tag{26}$$

Where $p_j = \prod_{i=1}^{n}(x_i)^{z_{ij}}$

The conjugate prior $p(a_j)$ is an exponential family with hyper parameters (s_j, t_j) is given by

$$p(\alpha_j) \propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j}\right)^{s_j} t_j^{\alpha_j} \tag{27}$$

The posterior distribution $p(x, z, \lambda, \beta_j, \alpha_1, \alpha_2, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_k)$ with hyper parameters $(s_j' = \sum_{i=1}^{n} z_{ij} + s_j$, $t_j' = p_j t_j)$ is given by$p(x, z, \lambda, \beta_j, \alpha_1, \alpha_2, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_k) \propto l(\alpha_j) p(\alpha_j)$

$$\propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j}\right)^{\sum_{i=1}^{n} z_{ij} + s_j} \left(p_j t_j\right)^{\alpha_j}$$

$$\propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j}\right)^{s_j'} (t_j')^{\alpha_j} \tag{28}$$

4-β_j Posterior

When the scale parameter β_j is unknown, for some j = 1, 2, . . . , k and the shape parameter a, and the weight parameter λ are known. By ignoring terms that contain $(\beta_1, \beta_2, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k, \alpha_j, \lambda)$ in (11), the complete data likelihood function is given by:

$$l(\beta_j) \propto (\beta_j^{\alpha_j})^{\sum_{i=1}^n z_{ij}} e^{-\beta_j \sum_{i=1}^n x_i z_{ij}}$$

The conjugate prior $p(\beta_j)$ is the gamma distribution with hyper parameters (v_j, f_j)

$$p(\beta_j) = \frac{(f_j)^{v_j}}{\Gamma v_j} \beta_j^{v_{j-1}} e^{-f_j \beta_j} \tag{29}$$

The posterior distribution $p(x, z, \lambda, \alpha_j, \beta_1, \beta_2, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k)$ is the gamma distribution with hyper parameters (v'_j, f'_j)

$$p(x, z, \lambda, \alpha_j, \beta_1, \beta_2, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots, \beta_k) \propto l(\beta_j) p(\beta_j)$$

$$\propto (\beta_j^{\alpha_j})^{\sum_{i=1}^n z_{ij}} e^{-\beta_j \sum_{i=1}^n x_i z_{ij}} \beta_j^{v_{j-1}} e^{-f_j \beta_j}$$

$$\propto \beta_j^{\alpha_j \sum_{i=1}^n z_{ij} + v_{j-1}} e^{-\beta_j \sum_{i=1}^n x_i z_{ij} + f_j}$$

$$\propto \beta_j^{v'_{j-1}} e^{-\beta_j f'_j} \tag{30}$$

Where $v'_j = \alpha_j \sum_{i=1}^n z_{ij} + v_j$, $f'_j = \sum_{i=1}^n x_i z_{ij} + f_j$

5-Joint Posterior of α, β

When the weight parameter λ is known and the shape parameters α and the scale parameter β are unknown. By ignoring terms that contain λ in (11), the complete data likelihood function is given by:

$$L(\alpha_j, \beta_j) = \prod_{j=1}^{k} \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j} \right) \sum_{i=1}^{n} z_{ij} \quad p_j^{\alpha_j - 1} e^{-\beta_j \sum_{i=1}^n x_i z_{ij}}$$

The conjugate prior $p(\alpha_j, \beta_j)$ with hyper parameters (s_j, m_j) is given by

$$p(\alpha_j, \beta_j) \propto \prod_{j=1}^{k} \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j} \right)^{s_j} \quad t_j^{\alpha_j - 1} e^{-\beta_j m_j} \tag{31}$$

By ignoring terms that contain λ, the joint posterior distribution

$$p(\alpha_j, \beta_j | x, \lambda, z) \propto L(\alpha_j, \beta_j)\, p(\alpha_j, \beta_j)$$

$$\propto \prod_{j=1}^{k} \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j} \right)^{\sum_{i=1}^n z_{ij} + s_j} t_j p_j^{\alpha_j - 1} e^{-\beta_j (\sum_{i=1}^n x_i z_{ij} + m_j)}$$

$$\propto \left(\frac{\beta_j^{\alpha_j}}{\Gamma \alpha_j} \right)^{s'_j} (t'_j)^{\alpha_j - 1} e^{-\beta_j m'_j} \tag{32}$$

with the hyper parameters

$$m'_j = \sum_{i=1}^{n} (x_i z_{ij} + m_j) , s'_j = \sum_{i=1}^{n} z_{ij} + s_j, t'_j = t_j p_j$$

4- **Simulation study**

In this section, a simulation study using Monte Carlo methods in Bayesian method of estimation and EM algorithm in maximum likelihood estimation and compare the efficiency of MLE method with Bayesian method of estimation using by computing the mean of the sum of the modulus of the bias (MBias), and the root-mean square error (RMSE), The general form of tow-component mixture gamma distribution is given by

$$f(x,\alpha,\beta,\lambda) = \lambda \frac{{\beta_1}^{\alpha_1}}{\Gamma\alpha_1} x^{\alpha_1-1} e^{-\beta_1 x} + (1-\lambda)\frac{{\beta_2}^{\alpha_2}}{\Gamma\alpha_2} x^{\alpha_2-1} e^{-\beta_2 x}$$

The simulation study was written using R language. The simulation study included the following basic stages:
First stage: choosing the initial vales as follows:
1-choosing the initial values for the parameters ($\alpha_1 = 3, \alpha_2 = 4,\ \beta_1 = 6,\ \beta_2 = 7, \lambda = 0.5$) , the λ and (1- λ) selected randomly from the first and the second component density.
2- choose different sample size(50, 100, 150) to generate the data set of tow-component mixture gamma distribution with parameters.
3-Repeat the experiment 1000 repetitions for each experiment.
4-choose values for the random variable.
Second stage: data generation :
A random variable is generated depending on the type of distribution
Third stage: estimating the parameters according to the mixture distributions using the estimation methods.
Fourth stage: the results compare the efficiency of MLE method with Bayesian method of estimation using by computing the mean of the sum of the modulus of the bias (MBias), and the root-mean square error (RMSE), where the smaller RMSE and MBias indicates a better overall quality of the estimates.

$$MBias = N^{-1}\sum_{i=1}^{N} |\hat{\alpha}-\alpha| + |\hat{\beta}-\beta|$$

$$RMSE = \sqrt{N^{-1}\sum_{i=1}^{N} (\hat{\alpha}-\alpha)^2 + (\hat{\beta}-\beta)^2}$$

To find the MLE estimators, the Newton Raphson method was adopted. The parameters (α, β) are estimated with Metropolis method (MT) of estimation using the joint prior in (31) with hyperparameters (s = 1; m = 1; t = 1) where the simulation study was carried out 1000 times. Table 1 present the estimates (Est.) and the RMSE and MBias values by MLE and MT method. The smaller RMSE and MBias for each sample size is highlighted in bold . Looking at these tables we observe that: we obtained that Metropolis method is uniformly better than MLE in all cases.

Table 1: MBias and RMSE of the MLE estimates and the MT estimators for two component mixture Gamma distribution

Sample size	Method	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\lambda}$	RMSE	MBise
50	EM	5.2594	2.3304	9.1590	3.8302	0.4065	2.3633	2.0702
	McMc	2.5164	3.3312	4.1121	5.8485	0.5114	**1.0555**	**0.8401**
100	EM	2.8220	6.3777	5.8096	10.8190	0.4999	2.0152	1.3130
	McMc	2.6802	6.2251	5.5243	10.6079	0.5091	**1.9129**	**1.3075**
150	EM	5.2594	2.3304	9.1590	3.8302	0.4065	2.3633	2.0702
	McMc	2.5164	3.3312	4.1121	5.8485	0.5114	**1.0555**	**0.8401**

5- Discussion

The parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda}$ are estimated with Metropolis method and the Expectation Maximization algorithm(EM) from the simulation results, it is observed that Bayes estimator better than maximum likelihood $\boldsymbol{L(\alpha,\beta,\lambda)L(\alpha,\beta,\lambda)}$estimation in all cases

6- Conclusion

1- Mixture distributions (in the case of similar and different components) the distributions formula becomes more complex, to make it easier to find a maximum likelihood estimator.it uses the EM algorithm . and MT algorithm to find the Bayesian estimators estimator in all cases.

2- After creating the simulation by taking different sample sizes (50,100,150)and using comparison criteria RMSE and MBias show that the Bayesian estimators is the best.

7- Acknowledgment

The authors are sincerely grateful to the University of Mosul and College of Computer Sciences and Mathematics for their provided facilities, which helped me very much to improve this work's quality.

8- Conflict of interest

The authors have no conflict of interest.

9- References

1. Saieed, H,A.(2005)" Estimation 0f parameters of mixture distributions and Its application on Neonatal birth weight data in Nineveh Govemorate", Unpublished doctors thesis , college of computer Science and Mathematics **,** University of Mosul..
2. Al-masri, Hanin, s,. (2020)."Bayesian inference on the genralized gamma distrbution". Journal of natural studies, Islamic university of gaza, V(28), pp 01-18.
3. Jos´e M Bernardo and Adrian FM Smith.. (1982)" Bayesian theory". volume 405 of Wiley Series in Probability and Statistics. John Wiley & Sons, 2009.
4. Jewell NP. (1982)."Mixtures of exponential distributions". Ann Stat.; 10(2): 479-484.
5. Chris, F., & Raftery, A,. (2017). "Model-based clustering, discriminant analysis, and density estimation". *Journal of the American statistical Association*, *5*(1).
6. Corduneanu, A. & Bishop, C.M. (2001), "Variational bayesian model selection for mixture distrbution ", Artifitial intellegence and statistics, T. Jaakkola and T. Richardson (Eds) pp 27-34, Morgan Kaufmann.
7. Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B 39:1–38
8. Friedman, J., Hastie, T., and Tibshirani, R. (2009). " The elements of statistical learning,", Springer series in statistics , New york, Vol 2.
9. Filho, I. (2008). "Mixture Models for the Analysis of Gene Expression: Integration of Multiple Experiments and Cluster Validation. Berlin, Germany: Department of Mathematics and Computer Science, Free University of Berlin, GermanyPatwray, A. N.; Sriwastav, G. L.; Hazarika, J. "Inference of R= P (X< Y<Z) for n-Standby System: A Monte-Carlo Simulation Approach". *J. Math*, 2016, 12: 18-22.S
10. Ghojogh, B., Ghojogh, A., Crowley, M., and Karray, F. (2019). "Fitting a mixture distrbution to data" , Tutorial. Waterloo, canada.
11. Li L.A., Decomposition Theorems,.(1983), "Conditional Probability, and Finite Mixtures Distributions". PhD Thesis, State University, Albany, New York,.
12. Newcomb S. (1886)." A generalized theory of the combination of observations so as to obtain the best result).Am. J. Math. 8:343–66
13. SK Upadhyay, N Vasishta, and AFM Smith(2000). "Bayes inference in life testing and reliability via markov chain monte carlo simulation". Sankhy a: The Indian Journal of Statistics, Series A, pages 203{222.
14. Sattaytham, P. and Talangtam, T.(2012). "Fitting of finite mixture distrbutions to motor insurance claims". Journal of mathematics and statistics, 8(1), pp 49-56.
15. Valieris ,R. , Drummond,R.D., Defelicibns , A. , Dias-Neto , E., Rosales, R.A.& da Silva, I.T. "A mixture model for determining SARS-COV-2 varint composition in pooled samples", Universidade de sao paulo, Ribeirao, sao paulo 14040-901, Barazil.

تقدير معلمات توزيع كاما المختلط باستخدام دالة الامكان الاعظم واستدلال بيز

نغم ابراهيم عبدالله نجم و ريا سالم محمد علي الرسام
قسم الاحصاء والمعلوماتية ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل، الموصل، العراق

الخلاصة: يركز هذا البحث على توزيع كاما المختلط حيث تستخدم تقنيتي دالة الامكان الاعظم واسلوب بيز لتقدير معلماته. تستخدم هذه الدراسة خوارزمية تعظيم التوقع (Expectation Maximization Algorithm (EM لإيجاد مقدرات الامكان الاعظم كما تم استخدام خوارزمية ميتروبولس هاستنغ (Metropolis–Hastings Algorithm (MT المحاكاة التقديرات البيزية لمعلمات توزيع كاما المختلط ، ثم تتم مقارنة هذه المقدرات باستخدام مجموع معامل التحيز (MBise) والجذر التربيعي لمتوسط الخطأ ((MSE) . وقد تبين ان مقدر بيز هو افضل من مقدر الامكان الاعظم .

الكلمات المفتاحية: توزيع كاما ، التوزيعات المختلطة ، التقدير البيزي ، دالة الامكان ، خوارزمية تعظيم التوقع ، ميتروبولس هاستينغ.