

Research Article

Estimation of Parameters of Finite Mixture of Rayleigh Distribution by the Expectation-Maximization Algorithm

Noor Mohammed and Fadhaa Ali 

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

Correspondence should be addressed to Fadhaa Ali; ali_fadhaa@coadec.uobaghdad.edu.iq

Received 22 September 2022; Revised 16 November 2022; Accepted 2 December 2022; Published 30 December 2022

Academic Editor: Ding-Xuan Zhou

Copyright © 2022 Noor Mohammed and Fadhaa Ali. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the lifetime process in some systems, most data cannot belong to one single population. In fact, it can represent several subpopulations. In such a case, the known distribution cannot be used to model data. Instead, a mixture of distribution is used to modulate the data and classify them into several subgroups. The mixture of Rayleigh distribution is best to be used with the lifetime process. This paper aims to infer model parameters by the expectation-maximization (EM) algorithm through the maximum likelihood function. The technique is applied to simulated data by following several scenarios. The accuracy of estimation has been examined by the average mean square error (AMSE) and the average classification success rate (ACSR). The results showed that the method performed well in all simulation scenarios with respect to different sample sizes.

1. Introduction

Fitting an appropriate distribution of data is one of the recent issues that challenge research worldwide. One of which arises from the date when it belongs to more than one population. Therefore, finite mixture models are employed to tackle such issues. In many applications, one large population can be formed from several subpopulations that are mixed in an unknown proportion. In such a case, one can come across a population of biological or electrical elements that may be divided into two or more subpopulations based on the possible causes of failure or any other categories such as male, female, and age categories. Many researchers have introduced several inferential methods to infer some aspects of data following heavy-tailed distributions [1, 2].

Many methods have been proposed to apply the mixture model to data in physical, biological, engineering sciences, and other fields as a valid technique of machine learning methods. A mixture of normal and Laplace distributions is fitted to wind shear data [3, 4]. In addition, the mixture model was used in modeling the crime and justice data [5]. A mixture of exponential distributions was considered a valid tool to infer device failure by dividing the population into

several subpopulations [6, 7]. Split-and-merge operations were used to improve the likelihood via the mixture of the t -distribution model to study image compression and pattern [8]. The mixture model was used as a classifier tool to cluster individuals based on estimated density functions for each individual based on blood cell data [9].

The mixture model can be structured from densities following the same families of distributions or different ones. Either way, it is useful as a confirmatory or predicting tool. Indeed, the maximum likelihood (ML) is the popular method for inferring the model [10, 11]. However, the challenge is the unknown number of allocations and observation memberships, through which the accuracy of estimation is undermined. These issues have been tackled by many proposed techniques over years. Therefore, determining the exact number of components by performing a relevant test is considered an important starting point in applying the mixture model [12, 13]. In the same manner, the most common way of handling the above issues is by using a latent variable that leads to a complete-data log likelihood rather than using the incomplete one [14], and then, the expectation-maximization (EM) algorithm is employed to estimate model parameters [15–17].

Moreover, Markov hidden model can be employed by adding components-indicator labels to the mixture model to determine observations memberships. This is an extension of Gaussian mixture model to model k -components of stationary or non-stationary auto-regressive [18], or to model time series data by non-Gaussian mixture [19]. Bayesian inference can also be used to infer the parameters of mixture models by assuming prior distribution for the parameters which in turn result in the posterior distribution of the parameters. This framework can be conducted by Markov chain Monte Carlo (MCMC) [20, 21]. This framework of estimation can avoid many difficulties in its applications that involve (ML) estimation [22–24]. The authors in [25] introduced a general form of inverse Rayleigh distribution, which is known as exponentiated inverse Rayleigh distribution (EIRD), which extends a more flexible distribution for modeling life data.

In this paper, we derived the expectation-maximization (EM) algorithm to estimate parameters of the mixture of Rayleigh distributions. This can be done through the maximum likelihood of the model, and then, we can use the EM algorithm to update estimators by iterative steps until reaching convergence. Afterward, the accuracy of the model inferred is examined by simulation based on different scenarios and the number of components. The success rate of clustering is also calculated accordingly. The rest of the paper is organized to give the theoretical details of the theoretical part in Section 2. In Section 3, the method is applied to the simulated data by using several scenarios. In Sections 4 and 5, the discussion and conclusion are involved to show the complete picture of the method and its results.

2. Materials and Methods

A random variable of Rayleigh distribution is defined as the time between two consecutive occurrences [26]. The probability density function is written as

$$f(x; \sigma) = \frac{x}{\sigma^2} e^{(-x^2/2\sigma^2)}, \text{ where } x, \sigma > 0. \quad (1)$$

In many lifetime scenarios, data can come from more than one component, and then, known distributions become inaccurate to fit the data. In such a case, the mixture model can be used to handle the problem. When it comes to lifetime data, a finite mixture of Rayleigh distribution is used to model the data. The probability density function (PDF) of the mixture of Rayleigh distribution with k components can be written as

$$f(x; \theta) = \sum_{j=1}^k \pi_j \frac{x}{\sigma_j^2} e^{(-x^2/2\sigma_j^2)}, \quad (2)$$

where $\sum_{j=1}^k \pi_j = 1, 0 < \pi_j < 1, \theta = (\pi_1, \pi_2, \dots, \pi_{k-1}, \sigma_1, \sigma_2, \dots, \sigma_k)^T, x, \sigma > 0$.

3. Inferential Method

The classical way of estimating the parameters of the mixture model is by the maximum likelihood (ML). The incomplete-data likelihood function of the k -component mixture model is defined as follows:

$$L(\theta; x) = \prod_{i=1}^n \sum_{j=1}^k \pi_j \frac{x_i}{\sigma_j^2} e^{(-x_i^2/2\sigma_j^2)}, \quad (3)$$

where $X = (x_1, x_2, \dots, x_n)^T$.

The allocations here are considered a hidden state that needs to be defined by a latent vector Z . The variable Z is a Markovian vector, and it is defined as $Z = (z_1, z_2, \dots, z_n)^T$, where $Z_i = (z_{i1}, z_{i2}, \dots, z_{ik})^T$. z_{ij} is defined as

$$z_{ij} = \begin{cases} 1, & \text{if item } i \text{ belong to component } j, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Hence, the complete-data likelihood function is written as

$$L(\theta; X, Z) = \prod_{i=1}^n \prod_{j=1}^k \left[\pi_j \frac{x_i}{\sigma_j^2} e^{(-x_i^2/2\sigma_j^2)} \right]^{z_{ij}}, \quad (5)$$

where $X = (x_1, x_2, \dots, x_n)^T$.

Taking the logarithm of Equation (2) results in

$$\ell(\theta; X, Z) = \sum_{i=1}^n \sum_{j=1}^k z_{ij} \left\{ \log \pi_j + \log x_i - \log \sigma_j^2 - \frac{x_i^2}{2\sigma_j^2} \right\}. \quad (6)$$

Many methods have been proposed to estimate parameters of Equation (3). One of which is the expectation-maximization (EM) algorithm that is used with some mixture models [27]. In this paper, the EM algorithm is derived for the k -component mixture of Rayleigh distribution. The next section consists of the details of such a method.

4. EM Algorithm

Equation (6) is the complete-data log-likelihood function that needs to be maximized given the data. The EM algorithm is used here to estimate the model parameters by considering z_{ij} missing data. It consists of two iterative steps, E (expectation step) and M (maximization step). For $m = 1, 2, \dots, M$, we repeat the following two steps.

4.1. E-Step. The unobservable data (here z_{ij}) are handled by the E -step, which takes the expectation of the complete-data log likelihood conditionally on the given observed data X , using the value of θ of the current iteration. Let $\theta^{(0)}$ be the value chosen initially for θ . The first iteration of the EM algorithm computes the conditional expectation of $\log L(\theta)$ given X by using $\theta^{(0)}$ for θ , which can be written as

$$Q(\theta; \theta^{(0)}) = E_{\theta^{(0)}}(\log L(\theta) | x). \quad (7)$$

As $\log \theta$ is linear in the missing data z_{ij} , the E -step at the $(m+1)^{th}$ iteration requires the current expectation of the random variable Z_{ij} , which corresponds to z_{ij} , to be calculated conditionally on the given observed data X .

$$E_{\theta^{(0)}}(Z_{ij} | x) = Pr_{\lambda^{(m)}}(Z_{ij} = 1 | x) = \tau_{ij}^m(x_i; \theta^{(m)}). \quad (8)$$

Then, we get

$$\tau_{ij}^{(m)}(x_i; \theta^{(m)}) = \frac{\pi_j^{(m)} f_j(x_i; \sigma_j^{(m)})}{\sum_{j=1}^k \pi_j^{(m)} f_j(x_i; \sigma_j^{(m)})}. \quad (9)$$

The quantity $\tau_j^{(m)}(x; \theta^{(m)})$ is the posterior probability, which implies that the i^{th} element of the sample with the observed value x_i belongs to the j^{th} component of the mixture. By substituting z_{ij} into Equation (3) with $\tau_j^{(m)}(x; \theta^{(m)})$, the Q function becomes

$$Q(\theta; \theta^{(m)}) = \sum_{i=1}^n \sum_{j=1}^k \tau_{ij}^{(m)}(x_i; \theta^{(m)}) \left\{ \log \pi_j + \log x_i - \log \sigma_j^2 + \frac{-x_i^2}{2\sigma_j^2} \right\}. \quad (10)$$

4.2. M-Step. To get the updated estimate $\theta^{(m+1)}$, the global maximization of $Q(\theta; \theta^{(m)})$ with respect to θ over the parameter space is required by the M-step on the $(m + 1)^{th}$ iteration. Given the updated $\tau_{ij}^{(m)}(x_i; \theta^{(m)})$ at the iteration $(m)^{th}$, the weighted parameter $\pi_j^{(m+1)}$ can be estimated by taking the first derivative of $Q(\theta; \theta^{(m)})$ with respect to π_j and equating it to 0, which results in

$$\hat{\pi}_j^{(m+1)} = \frac{\sum_{i=1}^n \tau_{ij}^{(m)}}{n}. \quad (11)$$

The component parameter σ_j at the iteration $(m + 1)^{th}$ can be computed by deriving $Q(\theta; \theta^{(m)})$ with respect to σ_j and equating it to 0, and we get

$$\begin{aligned} \frac{\partial Q}{\partial \sigma_j} &= \sum_{i=1}^n \left\{ \frac{-2\sigma_j \tau_{ij}}{\sigma_j^2} + \frac{\tau_{ij} x_i^2}{\sigma_j^3} \right\} \\ &\longrightarrow \frac{-2 \sum_{i=1}^n \tau_{ij}}{\sigma_j} + \frac{\sum_{i=1}^n \tau_{ij} x_i^2}{\sigma_j^3} = 0 \\ &\longrightarrow \frac{2 \sum_{i=1}^n \tau_{ij}}{\sigma_j} = \frac{\sum_{i=1}^n \tau_{ij} x_i^2}{\sigma_j^3} \end{aligned} \quad (12)$$

$$2\sigma_j^2 \sum_{i=1}^n \tau_{ij} = \sum_{i=1}^n \tau_{ij} x_i^2.$$

Hence, at the iteration $(m + 1)^{th}$, the component parameter can be estimated as

$$\hat{\sigma}_j^{(m+1)} = \sqrt{\frac{\sum_{i=1}^n \tau_{ij} x_i^2}{2 \sum_{i=1}^n \tau_{ij}}}. \quad (13)$$

The algorithm is terminated as soon as we get convergence or after the prespecified number of iterations. Note that the final iteration \hat{z}_{ij} can be estimated by generating it from multinomial distribution as

$$\hat{z}_{i \sim \text{Mult}}(1, \tau_i). \quad (14)$$

The estimated value of \hat{z}_i can be used as a classifier of component members based on the definition mentioned in Equation (1).

The elements of the component j are represented by S_j , where

$$S_j = (\hat{z}_{ij} = 1, i = 1, 2, \dots, n). \quad (15)$$

5. Assessing the Number of Components

The abovementioned EM algorithm is based on known allocation S_j . Therefore, we assess the appropriate number of components by calculating the Bayesian information criteria according to the formula [28]:

$$\text{BIC} = -2 \log L(\theta) + d \log n, \quad (16)$$

where $L(\theta)$ is the maximum likelihood function in (5) and $d = 2k - 1$. Observation memberships can be updated after getting Bayesian estimators of \hat{z}_{ij} as previously mentioned.

6. Simulations

To examine the performance of the EM algorithm on the finite mixture of Rayleigh distribution, we generate data by following two-component and three-component scenarios of the mixture model. The following scenarios were used to generate x_j according to predefined parameters $\theta = (\pi_1, \theta_1, \theta_2)^T$, where $(\pi_1, \theta_1, \theta_2) = (0.35, 1, 4), (0.35, 0.5, 2), (0.25, 2, 3)$. And $\theta = (\pi_1, \pi_2, \theta_1, \theta_2, \theta_3)^T$, where $((\pi_1, \pi_2, \theta_1, \theta_2, \theta_3) = (0.3, 0.6, 1, 4, 7), (0.3, 0.6, 0.5, 2, 4), (0.2, 0.5, 1, 2, 3)$. The sample sizes $n = 90, 120, 200$. A comparison was made by calculating the average mean square error (AMSE) of the estimated model and the average classification success rate (ACSR) for the number of replications equal to R by using the following formulas:

$$\text{MSE} = \frac{\sum_{i=1}^n f(x_i; \theta) - f(x_i; \hat{\theta})^2}{n}, \quad (17)$$

$$\text{AMSE} = \frac{\sum_{r=1}^R \text{MSE}_r}{R}. \quad (18)$$

Here, MSE_r is the MSE of the model on the replication r^{th} .

For calculating ACSR, we let A_r be the number of items with the correctly estimated membership divided by the sample size n on the replication r^{th} .

$$\text{ACSR} = \frac{\sum_{r=1}^R A_r}{R}. \quad (19)$$

7. Results

The results in Table 1 represent the data generated from the two-component mixture model and show that the EM algorithm is good at estimating the parameters of the model with respect to AMSE, which means the estimated parameters are close to the real ones used in the generated data. It can be seen that the AMSE is decreased as sample sizes are

TABLE 1: The results of AMSE and ACSR of applying the EM algorithm to simulated data from the two-component mixture model for three different scenarios and three different sample sizes with 1000 replications.

N	$\pi_1 = 0.35, \pi_2 = 0.65,$ $\theta_1 = 1, \theta_2 = 4$		$\pi_1 = 0.35, \pi_2 = 0.65,$ $\theta_1 = 0.5, \theta_2 = 2$		$\pi_1 = 0.25, \pi_2 = 0.75,$ $\theta_1 = 2, \theta_2 = 3$	
	AMSE	ACSR	AMSE	ACSR	AMSE	ACSR
90	0.002112	0.8290	0.007457	0.7724	0.002109	0.6781
120	0.001640	0.8404	0.006523	0.7834	0.001923	0.6963
200	0.00155	0.875	0.006312	0.7932	0.001722	0.7261

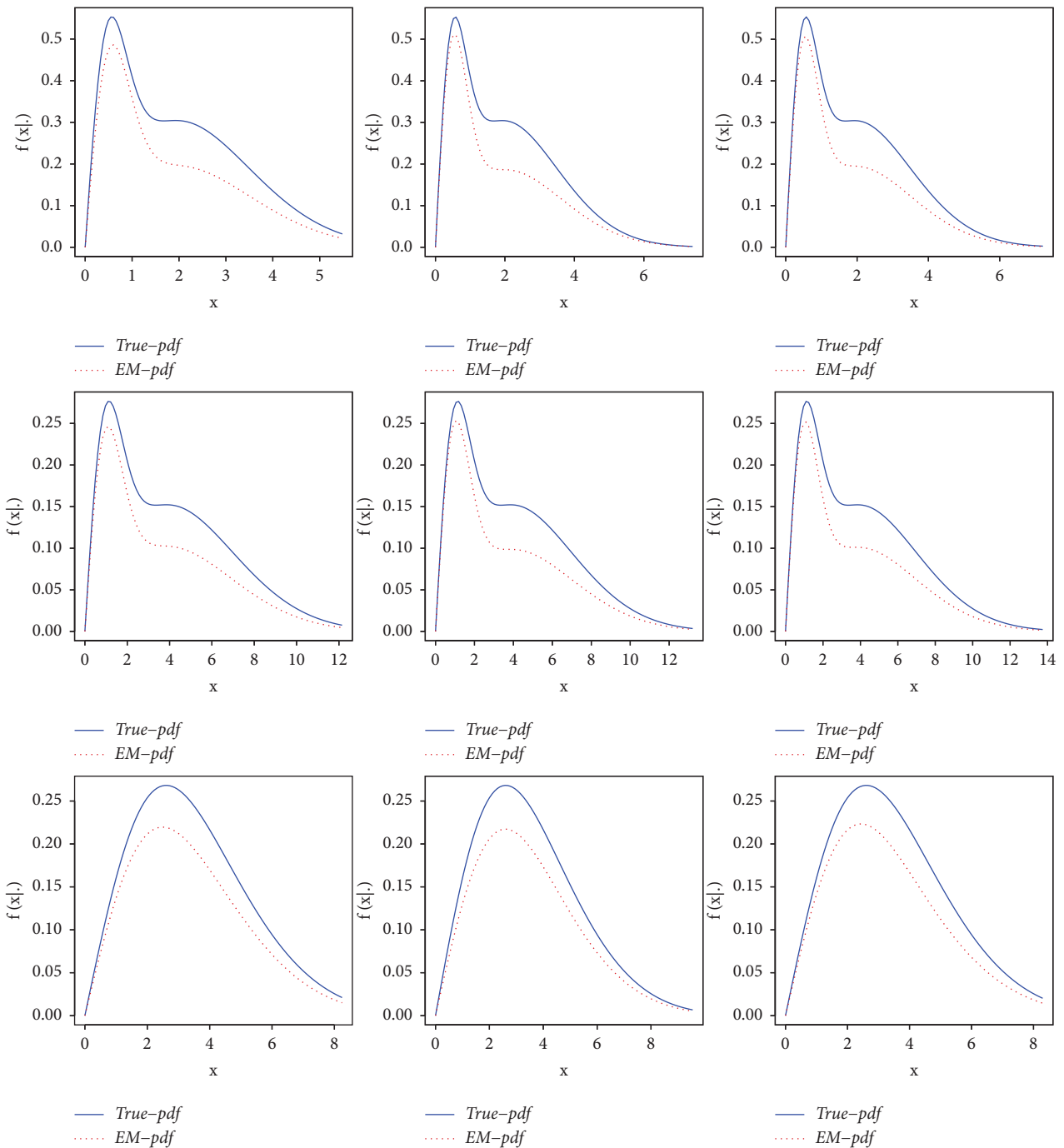


FIGURE 1: The graphs of the PDF of the model for true parameters and estimated ones by the EM algorithm. The columns of graphs represent the sample sizes $n = 90, 120,$ and $200,$ respectively. The rows of graphs represent the parameter scenarios $(\pi_1, \theta_1, \theta_2) = (0.35, 1, 4), (0.35, 0.5, 2),$ and $(0.25, 2, 3),$ respectively.

TABLE 2: The results of AMSE and ACSR of applying the EM algorithm to simulated data from the three-component mixture model for three different scenarios and three different sample sizes with 1000 replications.

N	$\pi_1 = 0.3, \pi_2 = 0.3, \pi_3 = 0.4,$ $\theta_1 = 1, \theta_2 = 4, \theta_3 = 7$		$\pi_1 = 0.3, \pi_2 = 0.3, \pi_3 = 0.4,$ $\theta_1 = 0.5, \theta_2 = 2, \theta_3 = 4$		$\pi_1 = 0.2, \pi_2 = 0.3, \pi_3 = 0.5,$ $\theta_1 = 1, \theta_2 = 2, \theta_3 = 3$	
	AMSE	ACSR	AMSE	ACSR	AMSE	ACSR
90	0.004427	0.69505	0.01316	0.63154	0.01388	0.56236
120	0.00416	0.7236	0.01287	0.65709	0.01363	0.57903
200	0.004007	0.74838	0.01239	0.67596	0.01311	0.5921

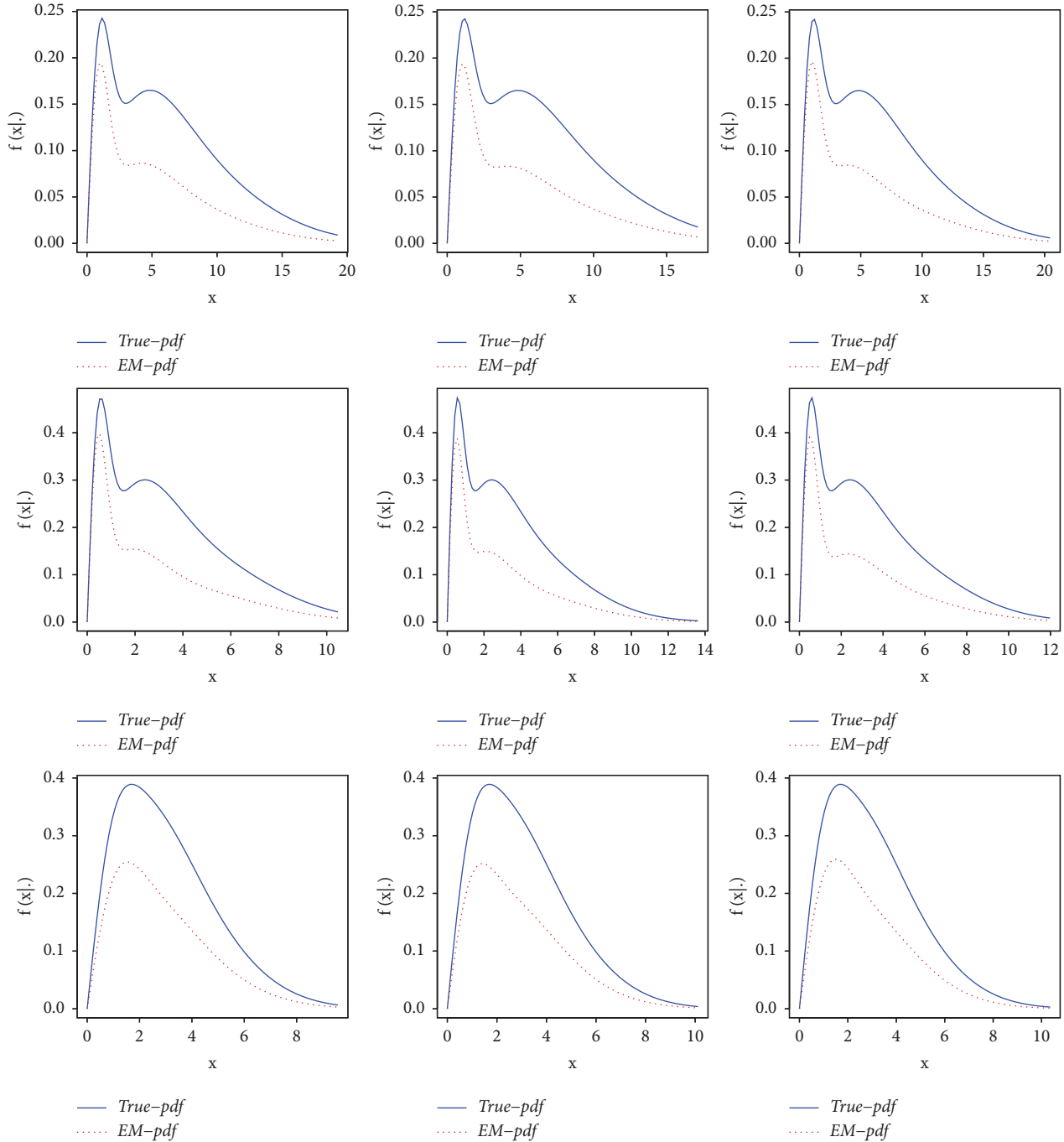


FIGURE 2: The graphs of the PDF of the model for true parameters and estimated ones by the EM algorithm. The columns of the graphs represent the sample sizes $n=90, 120,$ and $200,$ respectively. The rows of the graphs represent the parameter scenarios $(\pi_1, \pi_2, \theta_1, \theta_2, \theta_3) = (0.3, 0.3, 1, 4, 7), (0.3, 0.3, 0.5, 2, 4),$ and $(0.2, 0.3, 1, 2, 3),$ respectively.

increased. However, the ACSR is decreased when component parameters are close to each other. The graphs of the PDF of the model with respect to true parameters and estimators are shown in Figure 1, which shows the curves of probability distribution functions given the estimated parameters compared to the one of the true parameters values that has been used in the simulation. The results of applying the EM algorithm to the three-component model are shown in Table 1.

The results of applying the EM algorithm to the three-component model are shown in Table 2. It can be seen that the results of AMS of the large sample sizes are better than the small sample sizes. However, it is obvious that as the component parameters are close to each other, the ACSR is decreased, which means that the clustering of data becomes difficult. This is obvious in Figure 2, where the of the probability distribution functions of the third scenario does not show the distance between the tops.

8. Conclusion and Discussion

We derived the formulas for estimating the parameters of the finite mixture of Rayleigh distribution by the EM algorithm in a general form. The mathematical formulas of estimators are valid for the k -component mixture of Rayleigh distribution. However, the computations of the EM algorithm can be more accurate when sample sizes are increased. The underlying reason for this is the computation of the member variable as it is calculated with the E -step of the algorithm, given the observations. This in turn will result in great accuracy when it comes to estimating component parameters and weights. We apply the proposed framework to simulated data. The data were simulated according to two and three components as a special case from the general one. It can be seen from the results that the algorithm is performed in a good manner when sample sizes are large and parameters are not close enough to each other. We considered the AMSE and ACSR to be tools for measuring the accuracy of the estimation and clustering of data into their groups. The method was illustrated by graphs as well to show the estimated graph and compare it to one of the true parameters that are used in generating data for several scenarios.

Data Availability

All data were simulated, and no real data were used.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] S. Y. Huang, Y. L. Feng, and Q. Wu, "Learning theory of minimum error entropy under weak moment conditions," *Analysis and Applications*, vol. 20, no. 01, pp. 121–139, 2022.
- [2] F. S. Lv and J. Fan, "Optimal learning with Gaussians and correntropy loss," *Analysis and Applications*, vol. 19, no. 01, pp. 107–124, 2021.
- [3] P. N. Jones and G. J. McLachlan, "Laplace-normal mixtures fitted to wind shear data," *Journal of Applied Statistics*, vol. 17, no. 2, pp. 271–276, 1990.
- [4] G. K. Kanji, "A mixture model for wind shear data," *Journal of Applied Statistics*, vol. 12, no. 1, pp. 49–58, 1985.
- [5] C. M. Harris, "On finite mixtures of geometric and negative binomial distributions," *Communications in Statistics - Theory and Methods*, vol. 12, no. 9, pp. 987–1007, 1983.
- [6] G. S. Rao, "Estimation of reliability in multicomponent stress-strength based on generalized exponential distribution," *Revista Colombiana de Estadística*, vol. 35, no. 1, pp. 67–76, 2012.
- [7] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton, "SMEM algorithm for mixture models," *Neural Computation*, vol. 12, no. 9, pp. 2109–2128, 2000.
- [8] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and Computing*, vol. 10, pp. 335–344, 2000.
- [9] I. V. Cadez, C. E. McLaren, and P. Smyth, "Hierarchical models for screening of iron deficiency anemia," in *Proceedings of the Sixteenth International Conference on Machine Learning*, pp. 77–86, Morgan Kaufmann, San Francisco, CA, USA, June 1999.
- [10] E. L. Lehmann, "Efficient likelihood estimators," *The American Statistician*, vol. 34, no. 4, pp. 233–235, 1980.
- [11] E. L. Lehmann, *Theory of Point Estimation*, Wiley, New York, NY, USA, 1983.
- [12] D. Karlis and E. Xekalaki, "On testing for the number of components in finite Poisson mixtures," *Annals of the Institute of Statistical Mathematics*, vol. 5, pp. 149–162, 1999.
- [13] J. Henna, "On estimating of the number of constituents of a finite mixture of continuous distributions," *Annals of the Institute of Statistical Mathematics*, vol. 37, no. 2, pp. 235–240, 1985.
- [14] C. Liu and D. B. Rubin, "Maximum likelihood estimation of factor analysis using the ECME algorithm with complete and incomplete data," *Statistica Sinica*, vol. 8, pp. 729–747, 1998.
- [15] T. A. Louis, "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 44, no. 2, pp. 226–233, 1982.
- [16] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, pp. 19–39, 1995.
- [17] F. Ali and J. Zhang, "Mixture model-based association analysis with case-control data in genome wide association studies," *Statistical Applications in Genetics and Molecular Biology*, vol. 16, no. 3, pp. 173–187, 2017.
- [18] C. S. Wong and W. K. Li, "On a mixture autoregressive model," *Journal of the Royal Statistical Society: Series B*, vol. 62, no. 1, pp. 91–115, 2000.
- [19] N. D. Le, R. D. Martin, and A. E. Raftery, "Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models," *Journal of the American Statistical Association*, vol. 91, no. 436, pp. 1504–1514, 1996.
- [20] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, vol. 82, no. 398, pp. 528–540, 1987.
- [21] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 398–409, 1990.
- [22] M. K. Cowles and B. P. Carlin, "Markov chain Monte Carlo convergence diagnostics: a comparative review," *Journal of the*

- American Statistical Association*, vol. 91, no. 434, pp. 883–904, 1996.
- [23] G. Celeux, M. Hurn, and C. P. Robert, “Computational and inferential difficulties with mixture posterior distributions,” *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 957–970, 2000.
- [24] S. P. Brooks, “On Bayesian analysis and finite mixtures for proportions,” *Statistics and Computing*, vol. 11, no. 2, pp. 179–190, 2001.
- [25] G. S. Rao and S. Mbwambo, “Exponentiated inverse Rayleigh distribution and an application to coating weights of iron sheets data,” *Journal of Probability and Statistics*, vol. 2019, Article ID 7519429, 13 pages, 2019.
- [26] S. L. Miller and D. Childers, *Probability and Random Processes*, Academic Press, Cambridge, MA, USA, 2012.
- [27] J. M. G. Taylor, “Semi-parametric estimation in failure time mixture models,” *Biometrics*, vol. 51, no. 3, pp. 899–907, 1995.
- [28] S. L. Sclove, “Application of model-selection criteria to some problems in multivariate analysis,” *Psychometrika*, vol. 52, no. 3, pp. 333–343, 1987.