Research Article

Sequential Test for a Mixture of Finite Exponential Distribution

A.S. Al-Moisheer

Department of Mathematics, College of Science, Jouf University, P.O. Box 848, Sakaka 72351, Saudi Arabia

Correspondence should be addressed to A.S. Al-Moisheer; asalmoisheer@ju.edu.sa

Received 21 December 2020; Revised 15 March 2021; Accepted 30 March 2021; Published 19 April 2021

Academic Editor: Markos Koutras

Copyright © 2021 A.S. Al-Moisheer. 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.

Testing the number of components in a finite mixture is considered one of the challenging problems. In this paper, exponential finite mixtures are used to determine the number of components in a finite mixture. A sequential testing procedure is adopted based on the likelihood ratio test (LRT) statistic. The distribution of the test statistic under the null hypothesis is obtained using a resampling technique based on B bootstrap samples. The quantiles of the distribution of the test statistic are evaluated from the B bootstrap samples. The performance of the test is examined through the empirical power and application on two real datasets. The proposed procedure is not only used for testing the number of components but also for estimating the optimal number of components in a finite exponential mixture distribution. The innovation of this paper is the sequential test, which tests the more general hypothesis of a finite exponential mixture of k components versus a mixture of $k + 1$ components. The special case of testing an exponential mixture of one component versus two components is the one commonly used in the literature.

1. Introduction

The exponential distribution, which is analytically very simple, plays an important role in reliability and lifetesting analogues to the normal distribution in other areas. Consequently, the exponential distribution became a basic model to research associated with experiments on life expectancy. Applications of the exponential distribution include designing acceptance sampling plans [1], estimation of reliability in multicomponent stress-strength [2], and construction of multivariate control chart [3]. Also, neutrosophic statistics is applied when the data have uncertain parameters or values [4, 5]. A reason to study this distribution in mixtures is related to the lack of memory property.

Many authors have discussed mixture models such as Everitt and Hand [6]; Titterton et al. [7]; McLachlan and Basford [8]; Lindsay [9]; McLachlan and Krishnan [10]; and McLachlan and Peel [11].

Let $x_1, x_2, \dots, x_n$ be a random sample, arising from a mixture of finite exponential distribution (MFED), whose density is

$$f(x; \Theta) = \sum_{i=1}^k p_i f_i(x; \Theta_i), \quad 0 \leq p \leq 1, \quad (1)$$

where k is the number of exponential components, $\Theta_i = (p_i, \theta_i)$, $i = 1, 2, \dots, k$, and $f_i(x; \theta_i)$ represents the density function of the k^{th} component.

$$f_i(x; \theta_i) = \frac{1}{\theta_i} e^{-x/\theta_i}, \quad x \geq 0, \theta_i > 0, i = 1, 2, \dots, k. \quad (2)$$

Inference on the number of components can be conducted through statistical tests such as likelihood ratio tests. Some papers have dealt with bootstrapping the LRT such as McLachlan [12] and Feng and McCulloch [13] who used the bootstrap resampling for the number of one normal distribution against a mixture of two normal distributions. Also, Feng and McCulloch [13] who noted the bootstrap resampling for the number of normal mixture with difference variances is a preferred method. Seidel, Mosler and Alker [14, 15], who used a mixture of two exponential distributions; and Sultan, Ismail and Al-

Moisheer [16]; who used a mixture of two inverse Weibull distributions. The discrete Poisson distribution was used by Karlis and Xekalaki [17]. Some criteria are used to choose the number of components in finite mixtures models such as McLachlan and Peel [11]. Various authors have suggested the simplest form of testing in LRT for a single component against a two-component model. Here, the test procedure is proposed for k components against $k + 1$ components.

This paper is arranged into six sections. In Section 2, an algorithm is presented to determine the number of finite exponential components by using bootstrapping the LRT in R software packages. Section 3 contains the simulation results from computing the quantiles for an estimated number of finite exponential components using a sequential test. In Section 4, we evaluate the power of the sequential test when determining the number of finite exponential components. Criteria based on the likelihood are applied to determine (estimate) the number of components in finite mixtures models such as the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the Hannan–Quinn information criterion (HQIC), and the consistent Akaike information criterion (CAIC). In Section 5, the sequential test procedures are presented. Finally, in Section 6, the conclusion results are shown for the sequential testing number of finite exponential components.

2. Determining the Number of Components in the Finite Exponential Mixture

In this section, we use a sequential test to specify the number of components in the finite exponential mixture by using a resampling procedure called bootstrap. McLachlan [12] and Feng and McCulloch [13] discussed the idea of bootstrapping the LRT. The general method for determining the number of components is based on the LRT. The LRT statistic is used as appropriate test statistic for testing hypotheses. The test statistic is defined as $-2\ln\lambda$, where λ represents the ratio between the maximized likelihood functions under the null hypothesis (H_0) and the alternative hypothesis (H_1), respectively, (L_0) & (L_1). Equivalently, the test statistic can be written as $2[\ln L_1(\hat{\Theta}) - \ln L_0(\hat{\Theta})]$, where $\hat{\Theta}$ is the maximum likelihood estimator MLE for the parameter Θ .

Consider the hypothesis H_0 : the number of components in the exponential finite mixture is k against H_1 : the number of components in the exponential finite mixture is $k + 1$. The procedure of testing is sequential for $k = 1, 2, \dots$ using the LRT statistic. Bootstrapping the LRT is as follows.

Set $k = 1$.

- (1). Find the MLE of the parameters Θ , of the finite exponential mixture for k and $k + 1$ and calculate the LRT statistic which is referred to as L_{obs} . For the case $k = 1$, the MLE of Θ , $\hat{\Theta} = \bar{x}$, the sample mean.
- (2) Generate a bootstrap B sample of size n (n is the sample size) from the exponential mixture of k components and calculate the value of $-2 \ln \lambda$ after obtaining the MLEs of Θ under $k = k$ and $k = k + 1$.

The EM algorithm for a finite mixture of exponential distribution is used as mentioned in [15], as follows:

$$\hat{p}_i^{(B+1)} = \frac{1}{n} \sum_{j=1}^n \frac{p_i^{(B)} f_i(x_j; \theta_i^{(B)})}{f(x_j; \Theta^{(B)})}, \quad (3)$$

$$\hat{\theta}_i^{(B+1)} = \frac{\sum_{j=1}^n x_j f_i(x_j; \theta_i^{(B)}) / f(x_j; \Theta^{(B)})}{\sum_{j=1}^n f_i(x_j; \theta_i^{(B)}) / f(x_j; \Theta^{(B)})}, \quad (4)$$

where $f(x; \Theta)$ and $f_i(x; \theta_i)$ are represents in equations (1) and (2), respectively.

- (3) The process is repeated independently a number of times B .
- (4) The value of the j^{th} order statistics of the B replications of $-2 \ln \lambda$ can be taken as an estimate of the quantiles of the order $j/B + 1$.
- (5) The bootstrap replications can also be used to provide a test of approximate size α .
- (6) If $L_{\text{obs}} \geq$ the j^{th} order statistic of the B replications of $-2 \ln \lambda$, then the null hypothesis is rejected and we set $k = k + 1$ and go to step 1. Otherwise, the optimal number of components is k and the test is terminated. The bootstrap replications can also be used to provide a test of an approximate size α where $\alpha = 1 - j/B + 1$. For more details, see McLachlan and Peel [11].

3. Simulation Results

In order to find the LRT under the null hypothesis H_0 , we use the simulated data for the sequential test from a mixture of univariate exponential distributions. Accordingly, we use the mix tools package for R, which provides a set of functions to analyze a variety of finite mixture models. The repmix package is used to generate a random sample for a mixture of univariate exponential distributions. Then, we require the MLE of the mixing distribution. To find the best fit of Θ , the function nlm can be used to find the best fit for the model.

The number of the finite exponential components test is applied to choose the number of components ($k = 2, 3, 4$). To calculate the quantiles of the LRT tests, we simulate the null distribution of $-2 \ln \lambda$ for the sample sizes at $n = 20, 50, 75$, and 100 according to the stopping criteria described in [14] by using $\text{acc} = 10^{-5}$ as the level of accuracy. Each distribution is generated for the parameters in the model with 500 bootstrap samples. Tables 1–3 present the quantiles at the significance levels 10%, 5%, 2.5%, 1%, and 2%. The test always rejected for each $k = 2$ and $k = 3$ the number of components at sample size $n = 20$, and the test was repeated for five times, as shown in Tables 2 and 3. The epitome from the simulation is at $k = 3$, sample sizes are $n = (20, 50, 75, 100)$, and the choices of parameters are (0.8, 0.15, 0.5, 2, 4). Then, the best estimates for the optimal values for L_{obs} are (2.03, 3.6, 6.54, 10.3), respectively. Further, simulation results depend on the following factors.

As shown in Tables 1–3, to calculate the quantiles, when the sample sizes increase, so do the values of the quantiles;

TABLE 1: Quantiles for estimated number of finite exponential components at $k = 2$.

		K=2									
$(p_1, \theta_1, \theta_2)$	α	1	2	3	4	5	1	2	3	4	5
N=20						N=50					
(0.9, 0.35, 5)	0.01	1.63	1.86	1.82	1.55	1.93	2.85	3.45	2.5	2.53	2.62
	0.025	—	—	—	1.99	—	—	—	3.19	3.53	3.28
	0.05	—	—	—	—	—	—	—	—	—	—
	0.1	—	—	—	—	—	—	—	—	—	—
	0.2	—	—	—	—	—	—	—	—	—	—
(0.75, 0.05, 3)	0.01	-0.0681	-0.0186	0.053	0.111	0.0597	0.281	0.136	0.533	0.171	0.115
	0.025	0.0529	0.106	0.386	0.473	0.36	0.462	0.358	0.928	0.902	0.391
	0.05	1.27	—	1.63	—	—	1.36	0.886	1.7	2.12	0.931
	0.1	—	—	—	—	—	2.53	2.23	2.95	3.32	3.11
	0.2	—	—	—	—	—	—	—	—	—	—
(0.8, 0.16, 5)	0.01	0.888	0.694	0.445	0.815	1.74	1.45	0.866	1.44	1.9	1.69
	0.025	—	1.86	—	—	—	2.82	2.78	2.41	2.88	2.8
	0.05	—	—	—	—	—	—	—	3.39	—	—
	0.1	—	—	—	—	—	—	—	—	—	—
N=75						N=100					
(0.9, 0.35, 5)	0.01	5.15	5.28	5.23	5.67	5.42	8.29	8.19	7.99	8.13	7.91
	0.025	6.27	5.97	6.31	6.04	6.36	9.16	8.74	9.03	9.64	8.81
	0.05	—	—	—	—	—	—	—	9.85	—	9.58
	0.1	—	—	—	—	—	—	—	—	—	—
	0.2	—	—	—	—	—	—	—	—	—	—
(0.75, 0.05, 3)	0.01	1.23	1.71	1.38	0.637	0.878	2.84	4.66	5.51	2.08	4.08
	0.025	2.62	2.66	2.69	2.51	2.51	5.15	5.81	6.72	4.13	5.4
	0.05	3.68	4.17	4.55	3.44	3.61	7.12	7.98	8.55	7.16	7.27
	0.1	5.75	6.18	6.13	5.13	6.51	9.77	—	—	9.45	9.57
	0.2	—	—	—	—	—	—	—	—	—	—
(0.8, 0.16, 5)	0.01	4.72	3.56	5.32	4.44	4.65	6.87	7.12	8.46	8.51	7.68
	0.025	5.87	5.65	6.36	5.26	6.21	9.69	9.83	10.1	9.43	9.15
	0.05	—	—	—	—	—	—	—	—	—	—
	0.1	—	—	—	—	—	—	—	—	—	—
	0.2	—	—	—	—	—	—	—	—	—	—

for example, in Table 1, at $k = 2$ in the choice of parameter $(0.9, 0.35, 5)$ for sample size $n = 20$, we get the result of $\alpha = 0.01$ for the five repetitions and once at the significant level of $\alpha = 0.025$, while for $n = 100$, we get the results for the significant levels of $\alpha = (0.01, 0.025, 0.05)$. The same goes for $k = 3$ and $k = 4$ (see Figure 1; see also $k = 4$, which shows how the sample size affects the value of the quantiles and the level of acceptance for α). With respect to the initial values of the parameters, as the initial vector consists of $(p_1, \dots, p_{(k-1)}, \theta_1, \dots, \theta_{(k)})$ the simulation results reveal that when there is a large difference between the initial of the parameter $(\theta_{(k)})$ the number of rejected ones decreases and get the quantile results for high values of α . This is clear at $k = 3$ with the parameters $(0.12, 2.97, 2.92)$ and the maximum number of accepted quantiles is in $\alpha = 0.05$, while in the parameters choices $(0.52, 2, 4)$, the accepted values of the LRT at the levels of significance are $\alpha = 0.01$ and 0.5 for sample size $n = 75$. For the mixing proportion, when the initial value of p_1 or the sum of $(p_1, \dots, p_{(k-1)})$ is closer to 1, the number of rejected ones decreases. This is clear when we compare the results of the parameters at $k = 3$ ($p_1 = 0.6, p_2 = 0.3$). Then, the accepted values of the level of significance are up to $\alpha = 0.025$ for a large sample size $n = 100$. On the other hand, for the mixing proportions

$(p_1 = 0.8, p_2 = 0.15)$, the accepted values of α are up to 0.5 in the same sample. Finally, according to the increasing number of components k and the increasing levels of significance (α), it is obvious in the tables for the LRT that the maximum level of significance (α) at $k = 4$ is 0.9 and at $k = 3$ it is 0.5 , even though at $k = 2$, it is 0.1 .

4. The Power of the Sequential Test

The empirical power of the sequential test for the k components against $k + 1$ is evaluated for sample sizes $n = 20, 50, 75$, and 100 and $k = 1, 2, 3, 4$. The empirical power is defined as the proportion of times H_0 was rejected when the data were generated under H_1 . The power is simulated for 500 bootstrap samples and the different choices of parameters. Also, the power is calculated for the significance levels at 10%, 5%, 2.5%, and 1%.

For each case, we study the effect of increasing the distance between the parameters on the power of the test. Also, we study the effect of increasing the mixing proportion and the sample size n for the test. The power results are shown in Tables 4–6. For each case, when the sample size increases, the powers improve for every component ($k = 2, 3, 4$). To test the $k = 2$ components against $k = 3$, the

TABLE 2: Quantiles for estimated number of finite exponential components at $k = 3$.

$(p_1, p_2, \theta_1, \theta_2, \theta_3)$	α	$N = 20$					$K = 3$				
		1	2	3	4	5	1	2	3	4	5
(0.85, 0.05, 0.12, 2.97, 2.92)	0.01	—	—	—	—	—	1.87	2.01	1.48	1.41	1.51
	0.025	—	—	—	—	—	2.48	2.36	2.6	2.06	2.16
	0.05	—	—	—	—	—	2.99	3.15	3.16	2.48	2.92
	0.1	—	—	—	—	—	—	—	—	—	—
	0.2	—	—	—	—	—	—	—	—	—	—
	0.5	—	—	—	—	—	—	—	—	—	—
	0.8	—	—	—	—	—	—	—	—	—	—
	0.01	—	—	—	—	—	0.659	0.424	0.757	0.0385	0.297
(0.8, 0.15, 0.52, 2, 4)	0.025	—	—	—	—	—	1.13	0.869	1.1	0.491	0.69
	0.05	—	—	—	—	—	1.57	1.53	1.43	0.951	1.09
	0.1	—	—	—	—	—	2.27	2.13	1.98	1.53	1.88
	0.2	—	—	—	—	—	3.1	2.81	2.68	2.71	2.56
	0.5	—	—	—	—	—	—	—	—	—	—
	0.01	—	—	—	—	—	1.85	1.45	1.63	1.25	1.78
(0.6, 0.3, 0.48, 3, 5)	0.025	—	—	—	—	—	2.3	2.27	2.2	2.06	2.74
	0.05	—	—	—	—	—	3.34	—	3.31	2.96	3.54
	0.1	—	—	—	—	—	—	—	—	—	—
	0.01	—	—	—	—	—	—	—	—	—	—
		$N = 75$					$N = 100$				
		1	2	3	4	5	1	2	3	4	5
(0.85, 0.05, 0.12, 2.97, 2.92)	0.01	3.64	3.1	3.98	4.05	4.01	6.22	6.44	6.6	5.07	6.32
	0.025	4.9	4.89	4.31	4.6	4.65	7.44	7.69	8.21	7.05	7.36
	0.05	5.93	5.76	5.7	5.74	5.8	9.11	8.82	9.34	8.89	8.81
	0.1	—	—	—	—	—	—	9.83	—	—	10.2
	0.2	—	—	—	—	—	—	—	—	—	—
(0.8, 0.15, 0.52, 2, 4)	0.01	1.46	1.82	1.58	1.6	1.76	2.61	2.45	2.36	2.9	2.09
	0.025	2.18	2.07	1.95	2.18	2.36	3.04	3.42	3.96	3.9	2.9
	0.05	3.13	3.04	3.15	2.79	3.31	3.77	4.18	4.85	4.47	4.08
	0.1	3.99	3.74	3.9	3.75	4.13	5.17	5.17	6.01	5.53	5.17
	0.2	5.09	4.81	4.8	4.83	5.11	6.67	6.94	7.12	6.93	6.67
	0.5	—	—	—	—	—	9.96	9.65	—	10.2	9.93
(0.6, 0.3, 0.48, 3, 5)	0.01	4.94	3.91	4.43	4.95	5.02	9.17	8.07	8.31	6.36	7.59
	0.025	—	6.12	6.03	5.42	6.16	10	9.64	9.54	7.6	8.89
	0.05	—	—	—	—	—	—	—	—	10.2	—
	0.1	—	—	—	—	—	—	—	—	—	—

power is increased for large sample sizes, starting from $n = 50$ and over. For the $k = 3$ components against $k = 4$ for small sample sizes of $n = 20$, the power is always decreased. Performances of empirical power are affected by the sample sizes and not the reverse (see Figures 2–4).

5. Application

The sequential test procedure is applied in two real data as follows.

5.1. Application (1). The data considered in this application are given by Maswadah [18]. It represents the maximum flood levels (in millions of cubic feet per second) of the Susquehanna River in Harrisburg, Pennsylvania, over 20 four-year periods (1890–1969). The observations are as follows: 0.654, 0.613, 0.315, 0.449, 0.297, 0.402, 0.379, 0.423, 0.379, 0.324, 0.269, 0.740, 0.418, 0.412, 0.494, 0.416, 0.338, 0.392, 0.484, 0.265.

5.2. Application (2). It is an application of the sequential mityres of an exponential test for fitting exponential mixtures. According to Smith and Naylor [19], the following data represent the strength of 1.5 cm glass fibers.

Data $n = 63$: 0.55, 0.93, 1.25, 1.36, 1.49, 1.52, 1.58, 1.61, 1.64, 1.68, 1.73, 1.81, 2.00, 0.74, 1.04, 1.27, 1.39, 1.49, 1.53, 1.59, 1.61, 1.66, 1.68, 1.76, 1.82, 2.01, 0.77, 1.11, 1.28, 1.42, 1.50, 1.54, 1.60, 1.62, 1.66, 1.69, 1.76, 1.84, 2.24, 0.81, 1.13, 1.29, 1.48, 1.50, 1.55, 1.61, 1.62, 1.66, 1.70, 1.77, 1.84, 0.84, 1.24, 1.48, 1.30, 1.51, 1.55, 1.61, 1.63, 1.67, 1.70, 1.78, 1.89.

We apply the sequential testing of the number of components in an exponential finite mixture for the above two real datasets for sample sizes $n = 20$ and 63, respectively. The sequential results for the two applications are given in Tables 7 and 8. Column 1 in Tables 7 and 8 contains the number of components k in finite mixtures of exponentials. Column 2 contains the values of the LRT statistics for testing k versus $k + 1$. The test's p values are obtained between 0 and 1 by using 500 bootstrap samples, as described previously. The last four columns contain some information criteria that

TABLE 3: Quantiles for estimated number of finite exponential components at $k = 4$.

$(p_1, p_2, p_3, \theta_1, \theta_2, \theta_3, \theta_4)$	α	$K = 4$									
		1	2	3	4	5	1	2	3	4	5
(0.8, 0.05, 0.11, 0.99, 0.17, 2, 4)		$N = 20$					$N = 50$				
	0.01	—	—	—	—	—	1.25	1.1	1.32	1.33	1.59
	0.025	—	—	—	—	—	1.71	1.63	1.51	1.84	2.02
	0.05	—	—	—	—	—	2.03	2.14	2.2	2.07	2.28
	0.1	—	—	—	—	—	2.53	2.69	2.64	2.48	2.64
	0.2	—	—	—	—	—	3.15	3.14	3.13	3.01	3.19
	0.01	—	—	—	—	—	2.61	2.47	2.38	2.46	2.3
	0.025	—	—	—	—	—	2.87	2.78	2.66	2.72	2.55
	0.05	—	—	—	—	—	3.04	3.1	3.01	3.01	2.84
	0.1	—	—	—	—	—	3.39	3.44	3.47	3.36	3.28
(0.7, 0.2, 0.05, 0.77, 0.95, 2, 3.5)	0.01	—	—	—	—	—	2.37	2.39	2.31	2.25	2.41
	0.025	—	—	—	—	—	2.5	2.57	2.55	2.44	2.53
	0.05	—	—	—	—	—	2.7	2.73	2.71	2.66	2.69
	0.1	—	—	—	—	—	2.93	2.95	2.91	2.94	2.86
	0.2	—	—	—	—	—	3.17	3.11	3.15	3.14	3.12
	0.5	—	—	—	—	—	3.59	3.54	3.54	3.54	3.55
		$N = 75$					$N = 100$				
	0.01	3.55	3.5	3.32	3.62	3.56	5	5.32	5.04	5.2	5.43
	0.025	4.06	3.91	3.94	3.99	3.8	5.57	5.76	5.48	5.69	5.73
	0.05	4.64	4.41	4.34	4.31	4.35	6.08	6.18	6.36	6.37	6.32
(0.8, 0.15, 0.01, 1, 2.5, 3.15, 4)	0.1	5.23	5.18	4.86	5.04	4.95	6.8	7.05	6.94	7.03	7.07
	0.2	5.94	5.82	5.84	5.84	5.75	7.69	7.81	7.87	7.79	7.91
	0.5	—	—	—	—	—	10	9.92	9.88	9.72	9.87
	0.01	4.02	4.47	4.49	4.33	4.29	6.22	6.05	6.15	6.26	6.25
	0.025	4.53	5.03	4.78	4.72	4.79	6.86	6.52	6.85	6.56	6.67
	0.05	5.05	5.3	5.08	5.23	5.08	7.46	7.23	7.52	7.15	7.36
	0.1	5.56	5.66	5.66	5.64	5.61	7.94	7.75	8.13	7.74	7.94
	0.2	6.13	6.28	6.31	6.23	6.21	8.58	8.52	8.66	8.43	8.62
	0.5	—	—	—	—	—	9.88	9.96	10.1	9.88	10.3
	0.01	4.56	4.29	4.46	4.58	4.43	6.21	6.3	6.33	6.36	6.09
(0.8, 0.15, 0.01, 1, 2.5, 3.15, 4)	0.025	4.82	4.54	4.71	4.73	4.68	6.43	6.52	6.59	6.7	6.65
	0.05	5.07	4.8	4.99	5.02	5	6.89	6.87	6.94	6.93	6.89
	0.1	5.31	5.17	5.3	5.31	5.26	7.27	7.41	7.24	7.24	7.28
	0.2	5.67	5.58	5.72	5.64	5.65	7.65	7.83	7.71	7.64	7.74
	0.5	6.44	6.36	6.4	6.37	6.33	8.58	8.61	8.61	8.52	8.66
	0.8	—	—	—	—	—	9.4	9.52	9.48	9.52	9.57
	0.9	—	—	—	—	—	10	9.98	9.99	9.99	9.92

FIGURE 1: Simulation plot quantile at $k = 4$ with parameters (0.8, 0.05, 0.11, 0.99, 0.17, 2.0, 4.0).

TABLE 5: Continued.

$(p_1, p_2, \theta_1, \theta_2, \theta_3)$	α	$K=3$									
		1	2	3	4	5	1	2	3	4	5
$(0.85, 0.05, 0.12, 2.97, 2.92)$		$N=75$					$N=100$				
	0.01	1	1	1	1	1	1	1	1	1	1
	0.025	1	1	1	1	1	1	1	1	1	1
	0.05	1	1	1	1	1	1	1	1	1	1
	0.1	—	—	—	—	—	—	—	—	—	1
	0.2	—	—	—	—	—	—	—	—	—	—
$(0.8, 0.15, 0.52, 2, 4)$	0.01	1	0.992	0.994	1	1	1	1	1	1	1
	0.025	0.992	0.992	0.99	1	0.998	0.996	1	0.99	0.998	0.996
	0.05	0.976	0.982	0.984	0.998	0.986	0.996	0.996	0.984	0.998	0.996
	0.1	0.95	0.958	0.966	0.98	0.968	0.984	0.986	0.968	0.976	0.982
	0.2	0.918	0.916	0.94	0.952	0.922	0.942	0.94	0.938	0.938	0.958
	0.5	—	—	—	—	—	0.764	0.774	—	0.738	0.784
$(0.6, 0.3, 0.48, 3, 5)$	0.01	1	1	1	1	0.998	1	0.996	1	1	1
	0.025	—	0.996	1	1	0.998	1	0.994	1	1	1
	0.05	—	—	—	—	—	—	—	—	0.998	—
	0.1	—	—	—	—	—	—	—	—	—	—

TABLE 6: Power results for estimated number of finite exponential components at $k=4$.

$(p_1, p_2, p_3, \theta_1, \theta_2, \theta_3, \theta_4)$	α	$K=4$									
		1	2	3	4	5	1	2	3	4	5
$(0.8, 0.05, 0.11, 0.99, 0.17, 2, 4)$		$N=20$					$N=50$				
	0.01	—	—	—	—	—	0.992	0.996	0.99	0.994	0.99
	0.025	—	—	—	—	—	0.986	0.988	0.99	0.98	0.968
	0.05	—	—	—	—	—	0.976	0.978	0.95	0.964	0.956
	0.1	—	—	—	—	—	0.93	0.912	0.894	0.94	0.92
	0.2	—	—	—	—	—	0.85	0.82	0.85	0.87	0.826
$(0.7, 0.2, 0.05, 0.77, 0.95, 2, 3.5)$	0.01	—	—	—	—	—	0.996	0.998	0.99	0.992	0.996
	0.025	—	—	—	—	—	0.984	0.992	0.99	0.99	0.988
	0.05	—	—	—	—	—	0.978	0.982	0.98	0.972	0.982
	0.1	—	—	—	—	—	0.95	0.952	0.918	0.952	0.956
	0.01	—	—	—	—	—	0.992	0.998	0.992	0.986	0.984
	0.025	—	—	—	—	—	0.976	0.992	0.964	0.964	0.972
$(0.8, 0.15, 0.01, 1, 2.5, 3.15, 4)$	0.05	—	—	—	—	—	0.962	0.972	0.932	0.934	0.938
	0.1	—	—	—	—	—	0.944	0.928	0.884	0.898	0.914
	0.2	—	—	—	—	—	0.85	0.832	0.792	0.812	0.824
	0.5	—	—	—	—	—	0.606	0.568	—	0.544	0.54
		$N=75$					$N=100$				
$(0.8, 0.05, 0.11, 0.99, 0.17, 2, 4)$	0.01	0.994	0.996	1	0.992	0.998	0.996	0.994	1	0.994	0.996
	0.025	0.992	0.992	0.992	0.986	0.992	0.992	0.992	0.996	0.992	0.986
	0.05	0.97	0.978	0.988	0.98	0.978	0.982	0.986	0.98	0.982	0.97
	0.1	0.936	0.934	0.97	0.942	0.942	0.962	0.952	0.97	0.964	0.944
	0.2	0.878	0.86	0.892	0.874	0.886	0.906	0.906	0.922	0.902	0.868
	0.5	—	—	—	—	—	0.606	0.636	0.674	0.664	0.598
$(0.7, 0.2, 0.05, 0.77, 0.95, 2, 3.5)$	0.01	0.998	0.994	0.994	1	1	0.998	1	0.998	1	0.998
	0.025	0.998	0.984	0.99	0.998	0.992	0.994	0.996	0.998	1	0.998
	0.05	0.992	0.978	0.984	0.98	0.988	0.984	0.986	0.978	0.99	0.984
	0.1	0.976	0.96	0.958	0.95	0.97	0.97	0.972	0.95	0.972	0.966
	0.2	0.926	0.904	0.906	0.89	0.912	0.934	0.942	0.91	0.924	0.912
	0.5	—	—	—	—	—	0.76	0.738	0.66	0.74	0.638
$(0.8, 0.15, 0.01, 1, 2.5, 3.15, 4)$	0.01	0.992	0.996	0.992	0.994	0.996	1	0.998	1	0.998	1
	0.025	0.99	0.982	0.984	0.988	0.986	1	0.998	1	0.994	0.994
	0.05	0.984	0.974	0.968	0.982	0.974	0.988	0.988	0.99	0.988	0.98
	0.1	0.946	0.938	0.942	0.952	0.956	0.956	0.956	0.966	0.98	0.954
	0.2	0.88	0.862	0.884	0.888	0.898	0.898	0.872	0.902	0.924	0.876
	0.5	0.648	0.576	0.616	0.616	0.56	0.652	0.634	0.632	0.65	0.642
	0.8	—	—	—	—	—	0.382	0.312	0.362	0.336	0.318
	0.9	—	—	—	—	—	0.208	0.184	0.226	0.192	0.218

FIGURE 2: Power results at $k = 2$ with parameters (0.9, 0.1, 0.35, 5.0). FIGURE 3: Power results at $k = 3$ with parameters (0.85, 0.05, 0.1, 0.12, 2.97, 2.92). FIGURE 4: Power results at $k = 4$ with parameters (0.8, 0.05, 0.11, 0.04, 0.99, 0.17, 2.0, 4.0).TABLE 7: Sequential results for application 1 at $n = 20$.

k	LRT	p value	AIC	BIC	HQIC	CAIC
1	56.1	1.000	1.98	2.98	2.2	2.18
2	23.5	0.938	4.00	5.99	4.39	4.71
3	4.25	≤ 0.0001	8.00	12.0	8.78	10.7

TABLE 8: Sequential results for application 2 at $n = 63$.

k	LRT	p value	AIC	BIC	HQIC	CAIC
1	16.1	1.000	4	8.32	5.7	4.2
2	1.11	0.106	6	12.5	8.55	6.4
3	0.55	≤ 0.0001	8	16.6	11.4	8.68

Put $k = 1$

Step 1: find the MLEs of the parameters Θ_i of the finite exponential components for k and $k + 1$, respectively, by using the code in R.

For $k = 1$

$n \loglik \rightarrow \text{sum}(d \exp(x, \Theta_i)), \Theta \rightarrow \text{optimize}(n \loglik, x) \text{minimum},$

Other

$\log l \rightarrow \text{Sum}(d \exp(x, \text{Rate}, \log = \text{TRUE})) \rightarrow \text{nlm}(\log l, \text{Rate}, x = x) \text{estimate},$

and calculate the LRT statistic, say L_{obs} .

Step 2: declare the initial various variables $\theta_i^k, p_i^{(k-1)}$.

Step 3: generate a random sample of size n , for the mixture of univariate exponential distributions $r \exp \text{mix}(n, \text{lambda}, \text{rate})$,

$\text{data}_{H_0} \rightarrow r \exp \text{mix}(n, \text{lambda} = \text{vector of } (p, n = k - 1), \text{rate} = \text{vector of } (\Theta_i, n = k))$

$\text{data}_{H_1} \rightarrow r \exp \text{mix}(n, \text{lambda} = \text{vector of } (p, n = k), \text{rate} = \text{vector of } (\Theta_i, n = k))$

Step 4: declare the function LRT to calculate $p^{(k+1)}, \theta_i^{(k+1)}$, and then calculate the LRT by using the EM algorithm equation for finite exponential components, as mentioned in [15].

Step 5: simulate 500 bootstrap samples of size n with initial vector parameters $p^{(k+1)}, \theta_i^{(k)}$, and for each bootstrap sample, calculate $p^{(k+1)}, \theta_i^{(k)}$ for the LRT using the boot package functions in R.

$\text{result}_{H_0} \rightarrow \text{boot}(\text{data}_{H_0}, \text{lrt}_{\text{function}}, B = 500, \text{s type} = f, \text{sim} = \text{ordinary})$

$\text{result}_{H_1} \rightarrow \text{boot}(\text{data}_{H_1}, \text{lrt}_{\text{function}}, B = 500, \text{s type} = f, \text{sim} = \text{ordinary})$

Step 6: find the control variate estimates from a bootstrap output object. Estimate the quantiles using the linear approximation as a control variate.

$\text{Quantile}_{H_0} \rightarrow \text{control}(\text{result}_{H_0}) \text{quantiles}$

Step 7: estimate the p values in result_{H_1}

if $(L_{\text{obs}} > \text{Quantile}_{H_0}[2])$ Don't Reject and find p - value

$\text{mean } v \rightarrow \text{sum}((\text{result}_{H_1} t[, 1]) > = \text{value}[vni, 2]) / \text{results2R}$

ALGORITHM 1: Algorithm for monte carlo simulation R.

are used to choose the number of components in finite mixtures models, such as AIC, BIC, HQIC, and CAIC. These criteria also agree with the results and the p values. The p values are obtained through simulation under H_0 of the LRT. Our results in Tables 7 and 8 lead to the selection of the mixture model with one component. It can also be seen from Tables 7 and 8 that the LRT increases when the number of components decreases, while the information criteria increase as the number of components increases. Briefly, the best mixture models at $k = 1$ because it has the largest p value and the minimum values for the four criteria that are used (Algorithm 1).

6. Conclusion

In this paper, the sequential testing of the number of components in exponential finite mixture is discussed. Simultaneously, the optimal number of components for a finite exponential mixture is determined to provide the appropriate fit with the data. A resampling approach to determine the number of components is used via B bootstrap samples. Bootstrap samples are generated from the finite exponential components under H_0 . The value of $-2 \ln \lambda_n$ is evaluated for each bootstrap sample. This process is repeated for 500 bootstraps to obtain the j^{th} order statistic as an estimate of the quantile. The power results for the estimated number of finite exponential components are computed. Two applications of real data are used to illustrate the sequential test. Thus, the innovation in this sequential test is that it permits the testing of the hypothesis of k components in an exponential mixture against $k + 1$ components, along with the determination of the optimal exponential mixture. It thus provides a general method than the one commonly

used in exponential mixtures which focuses on testing one component versus a mixture of two components. The importance of this sequential test lies in cluster analysis and other applications. Finally, it is clear that our sequential test, which was applied to finite exponential mixtures, can be applied to finite mixture models from any family of mixtures.

Data Availability

The data used to support the findings of this study were obtained from Maswadah [18] and Smith and Naylor [19].

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] W. Gui and M. Aslam, "Acceptance sampling plans based on truncated life tests for weighted exponential distribution," *Communications in Statistics - Simulation and Computation*, vol. 46, no. 3, pp. 2138–2151, 2017.
- [2] G. S. Rao, M. Aslam, and O. H. Arif, "Estimation of reliability in multicomponent stress-strength based on two parameter exponential weibull distribution," *Communications in Statistics: Theory and Methods*, vol. 46, no. 15, pp. 7495–7502, 2017.
- [3] N. Khan, M. Aslam, M. S. Aldosari, and C.-H. Jun, "A multivariate control Chart for monitoring several exponential quality characteristics using EWMA," *IEEE Access*, vol. 6, no. 1, pp. 70349–70358, 2018.
- [4] M. Aslam and O. H. Arif, "Testing of grouped product for the weibull distribution using neutrosophic statistics," *Symmetry*, vol. 10, no. 9, p. 403, 2017.

- [5] M. Aslam, "Design of sampling plan for exponential distribution under neutrosophic statistical interval method," *IEEE Access*, vol. 6, no. 1, pp. 64153–64158, 2018.
- [6] B. S. Everitt and D. J. Hand, *Finite Mixture Distribution*, Chapman & Hall, London, UK, 1981.
- [7] D. M. Titterton, A. F. M. Simth, and U. E. Makov, *Statistical Analysis of Finite Mixture Distribution*, Wiley & Sons, Chichester, UK, 1985.
- [8] G. J. McLachlan and K. E. Basford, *Mixture Models: Applications to Clustering*, Marcel Dekker, New York, NY, USA, 1988.
- [9] B. G. Lindsay, *Mixture Models: Theory, Geometry and Applications*, The Institute of Mathematical Statistics, Hayward, CA, USA, 1995.
- [10] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley & Sons, New York, NY, USA, 1997.
- [11] G. J. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, New York, NY, USA, 2000.
- [12] G. J. McLachlan, "On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture," *Applied Statistics*, vol. 36, no. 3, pp. 318–324, 1987.
- [13] Z. D. Feng and C. E. McCulloch, "Using bootstrap likelihood ratios in finite mixture models," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 3, pp. 609–617, 1996.
- [14] W. Seidel, K. Mosler, and M. Alker, "A cautionary note on likelihood ratio tests in mixture models," *Annals of the Institute of Statistical Mathematics*, vol. 52, no. 3, pp. 481–487, 2000a.
- [15] W. Seidel, K. Mosler, and M. Alker, "Likelihood ratio tests based on subglobal optimization: a power comparison in exponential mixture models," *Statistical Papers*, vol. 41, no. 1, pp. 85–98, 2000b.
- [16] K. S. Sultan, M. A. Ismail, and A. S. AL-Moisheer, "Testing the number of components of the mixture of two inverse Weibull distributions," *International Journal of Computer Mathematics*, vol. 86, no. 4, pp. 693–702, 2009.
- [17] D. Karlis and E. Xekalaki, "On testing for the number of components in a mixed Poisson model," *Annals of the Institute of Statistical Mathematics*, vol. 51, no. 1, pp. 149–162, 1999.
- [18] M. Maswadah, "Conditional confidence interval estimation for the Inverse Weibull distribution based on censored generalized order statistics," *Journal of Statistical Computation and Simulation*, vol. 73, no. 12, pp. 887–898, 2003.
- [19] R. L. Smith and J. C. Naylor, "A comparison of maximum likelihood and bayesian estimators for the three-parameter weibull distribution," *Applied Statistics*, vol. 36, no. 3, pp. 358–369, 1987.