Commun.Fac.Sci.Univ.Ank.Series A1 Volume 63, Number 2, Pages 109–118 (2014) ISSN 1303-5991

# A STUDY ON MODELING OF PHENOMENA AIR CONDITIONING DATA

## MEHMET YILMAZ AND BUSE BÜYÜM

ABSTRACT. This study is based on modeling of phenomena air-conditioning data set, which has been trying to model in many times by different authors, with mixed exponential distribution with two-component (2MED). Since 1963, the whole data or some part of the data has been taken as real data in modeling study. These studies have been taken into account to detect the best model. Inspiring from studies led by Proschan F. (1963), the obtained results by modeling this data set with 2MED are of our interest. For this purpose, brief summary of the studies in literature has been given. After that, the results have been compared with results of 2MED. We claim that 2MED will be located among the purposed models.

### 1. INTRODUCTION

One can generally point out that data set has only one law in modeling and estimation problems. However, it's possible that obtained data appear as mixture of either same distribution or different distribution family which is called contaminated. For example, first failure time of products that have same function in different qualities and patients' length of stay in hospital can be modelled by mixture of exponential distribution.

Air-conditioning data set that has been obtained for the first time by [12] is probably most analyzed in statistical literature. In [12], the data set has been modelled with exponential distribution. Later in many different studies, while researches have been trying to model this data set, some of them have used nonmixture distributions and some of them have used mixture distributions. For example, the data set has been modelled in [3] by Exponential distribution, in [7] by Gamma distribution, in [9] by Weibull distribution, in [11] by Gamma-Dagum distribution and in [14] by Inverse Rayleigh distribution. Except from these, researches have used Exponential-Poisson (EP), Exponential-Gamma (EG),

©2014 Ankara University

109

Received by the editors Octob. 31, 2014; Accepted: Nov. 30, 2014.

<sup>2000</sup> Mathematics Subject Classification. Primary 05C38, 15A15; Secondary 05A15, 15A18.

Key words and phrases. Air-conditioning data, expectation-maximization algorithm (EM), least square method (LSE), life data modeling, maximum likelihood method (MLE), mixed exponential distribution.

Exponential-Logarithmic (EL) and Exponential-Binomial (EB) as a mixture distributions in [8], [15], [2] and [16].

In addition to studies above, whether the question of that this data set can be modelled or not by 2MED has created the source of the current study. From this point, 2MED and the methods of parameter estimation for this distribution will be introduced firstly. Then, the results that is obtained from other studies and from this study will be compared. In the comparison, Kolmogrov-Smirnov test statistics (KS) will be considered. Parameter estimations, KS values and p values (p) have been obtained by MATLAB.

## 2. MATERIAL AND METHODS

In the literature, [1] and [5] have mentioned finite mixture distributions and methods of parameter estimation for them. In this section, 2MED will be considered.

2.1. Mixed exponential distribution with two-component (2MED). Probability density function (pdf) of 2MED is given below.

 $f(x; \alpha, \theta_1, \theta_2) = \alpha f_1(x; \theta_1) + (1-\alpha) f_2(x; \theta_2) = \alpha \frac{1}{\theta_1} \exp(-x/\theta_1) + (1-\alpha) \frac{1}{\theta_2} \exp(-x/\theta_2)$ where  $\alpha \in (0, 1), \ \theta_i > 0 \ (i = 1, 2), x > 0$ . Similarly the cumulative distribution function (cdf) is as follows.

$$F(x; \alpha, \theta_1, \theta_2) = \alpha F_1(x; \theta_1) + (1 - \alpha) F_2(x; \theta_2) = \alpha (1 - \exp(-x/\theta_1)) + (1 - \alpha) (1 - \exp(-x/\theta_2))$$

Survival function,

$$\begin{split} S(x; \alpha, \theta_1, \theta_2) &= \alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2) = \alpha \exp(-x/\theta_1) + (1 - \alpha) \exp(-x/\theta_2) \\ \text{and the hazard function,} \\ h(x; \alpha, \theta_1, \theta_2) &= \frac{\alpha h_1(x) S_1(x; \theta_1) + (1 - \alpha) h_2(x) S_2(x; \theta_2)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)} \\ &= h_1(x) \frac{\alpha S_1(x; \theta_1)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)} + h_2(x) \frac{(1 - \alpha) S_2(x; \theta_2)}{\alpha S_1(x; \theta_1) + (1 - \alpha) S_2(x; \theta_2)} \\ &= h_1(x) w_1(x; \alpha, \theta_1, \theta_2) + h_2(x) w_2(x; \alpha, \theta_1, \theta_2), \\ \text{where } h_i(x) &= \frac{1}{\theta_i} \text{ and } w_1(.) + w_2(.) = 1. \end{split}$$

#### **3. PARAMETER ESTIMATION METHODS**

In this section, maximum likelihood estimation (MLE) and least square estimation (LSE) are given for 2MED.

3.1. Maximum likelihood estimations (MLE). In general, the likelihood equation systems emerge as nonlinear equations. Because of this, the numerical methods are preferred for the solution of likelihood equations. In this study, Expectation-Maximization (EM) algorithm which is one of the common numeric method will be mentioned. The EM algorithm was found by [4] and also was studied in [10] and [13].

110

Let  $\underline{X} = \{X_1, X_{2,...,} X_n\}$  be a random sampling with independent and identically distributed as 2MED having a pdf  $f(\underline{x}; \underline{\Phi})$  where  $\underline{\Phi} = (\alpha, \theta_1, \theta_2)$  is a parameter vector. The likelihood function, the logarithmic form of the likelihood function of  $\underline{\Phi}$  are given in below, respectively.

$$L(\underline{\Phi}, \underline{x}) = \prod_{j=1}^{n} \left[\sum_{i=1}^{2} \alpha_{i} \frac{1}{\theta_{i}} \exp(-x_{j}/\theta_{i})\right]$$
$$\log L = \sum_{j=1}^{n} \log\left[\sum_{i=1}^{2} \alpha_{i} \frac{1}{\theta_{i}} \exp(-x_{j}/\theta_{i})\right] - \lambda\left(\sum_{i=1}^{2} \alpha_{i} - 1\right)$$

where  $\sum_{i=1}^{2} \alpha_i = 1$ . If the derivative of this function respect to  $\alpha_i, i = 1, 2$  is equalized to zero,

$$\frac{d\log L}{d\alpha_i} = \sum_{j=1}^n \frac{\frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum\limits_{i=1}^n \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} - \lambda = 0$$
$$\sum_{j=1}^n \frac{\frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum\limits_{i=1}^n \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = \lambda$$

If the both side of (3.1) is multiplied with  $\alpha_i$  and sum over index *i*:

$$\sum_{j=1}^{n}\sum_{i=1}^{2}\frac{\alpha_{i}\frac{1}{\theta_{i}}\exp(-x_{j}/\theta_{i})}{\sum_{i=1}^{2}\alpha_{i}\frac{1}{\theta_{i}}\exp(-x_{j}/\theta_{i})}=\lambda\alpha_{i}$$

then  $n = \lambda$ . Based on Bayes rule, the probability that  $x_j$  belongs to  $i^{th}$  component when  $X_j = x_j$  is observed is as follows:

$$P(i \mid x_j) = \frac{\alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}{\sum_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)}.$$

Thus,

then

$$\widehat{\alpha}_i = \frac{\sum\limits_{j=1}^n P(i|x_j)}{n}, \ i = 1, 2.$$

If the derivative of  $\log L$  with respect to  $\theta_i$  is equalized to zero,

$$\frac{d\log L}{d\theta_i} = \sum_{j=1}^n \frac{\frac{\alpha_i}{\theta_i} (\frac{x_j}{\theta_i^2}) \exp(\frac{-x_j}{\theta_i}) - \frac{\alpha_i}{\theta_i^2} \exp(\frac{-x_j}{\theta_i})}{\sum\limits_{i=1}^2 \alpha_i \frac{1}{\theta_i} \exp(-x_j/\theta_i)} = 0$$

$$\vdots$$

$$\hat{\theta}_i = \frac{\sum\limits_{j=1}^n x_j P(i|x_j)}{\sum\limits_{j=1}^n P(i|x_j)}, i = 1, 2$$

(3.1)

is obtained. Its reminded that  $P(2 \mid x_i) = 1 - P(1 \mid x_i)$ , then the solutions will be

$$\widehat{\theta}_1 = \frac{1}{n\widehat{\alpha}_i} \sum_{j=1}^n x_j P(i \mid x_j)$$
$$\widehat{\theta}_2 = \frac{1}{n(1-\widehat{\alpha}_i)} \sum_{j=1}^n x_j (1 - P(i \mid x_j))$$

These are step solutions obtained by EM which steps are given in below.

- (1) Input the initial values.  $(\alpha_i^{(0)}, \theta_i^{(0)}), i = 1, 2.$
- (2) Calculate the  $P(i \mid x_j)$ .
- (3) Calculate  $\alpha_i^{(k)}, \theta_i^{(k)}$
- (4) After calculations of  $\widehat{\alpha}_i$  and  $\widehat{\theta}_i$ , the values replace in logL and get the value of function. For  $\epsilon > 0$  selected small enough  $\log L^{(k)} \log L^{(k-1)} \leq \epsilon$  is provided then the values on the  $k^{th}$  step will be used for parameter estimations. Step 2-4 are repeated until converge is accomplished.

3.2. Least squares estimations (LSE). This method is based on the idea that there is a regression relationship between empirical  $\hat{F}$  and parametric F distributions. Considering ordered observations  $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$  versus empirical distribution  $\hat{F}(x_{(i)}) \equiv \frac{i}{n+1}$ , the vector  $\underline{\Phi}$  which minimizes the following expression is tried to determine. Detailed study was given in [6] for non-mixture Generalized Exponential Distribution. System of equations that is occurred for the solutions for this optimization problem is as follows.

$$\begin{aligned} Q(\underline{\Phi}) &= \sum_{i=1}^{n} (\widehat{F}(x_{(i)}) - F(x_{(i)};\underline{\Phi}))^{2} \\ \frac{dQ}{d\alpha} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}))) \sum_{i=1}^{n} (1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}) \\ -\alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1}))) &= 0 \\ \frac{dQ}{d\theta_{1}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}))) (\frac{\alpha x_{(i)}}{\theta_{1}^{2}} \exp(-\mathbf{x}_{(i)}/\theta_{1})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}))) (\frac{(1 - \alpha)x_{(i)}}{\theta_{2}^{2}} \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}))) (\frac{(1 - \alpha)x_{(i)}}{\theta_{2}^{2}} \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) (\frac{(1 - \alpha)x_{(i)}}{\theta_{2}^{2}} \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2}) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) - (1 - \alpha)(1 - \exp(-\mathbf{x}_{(i)}/\theta_{2})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/\theta_{1})) = 0 \\ \frac{dQ}{d\theta_{2}} &= \sum_{i=1}^{n} (\frac{i}{n+1} - \alpha(1 - \exp(-\mathbf{x}_{(i)}/$$

Since the equations are in a complex structure and are related to parameters, it is difficult to obtain the solutions. Therefore it is necessary to use numerical ways.

## 4. COMPARISON WITH MODELING STUDIES IN LITERATURE

In this section, the results obtained from studies in literature and from current study will be compared. The whole data of failure times in air-conditioning systems which is examined by [12] for the first time is given in Table 1. Some of authors such as [2], [8], [11] and [15] have considered to model whole data set. The others have considered failure times of specific plane identification numbers that are appropriate for their suggested models.

In the current study, both the whole data set and the specific data set have been handled and have been modelled with 2MED. Appealing results about phenomena air-conditioning data have been obtained.

PLAN	NE IDI	ENTIF	ICAT	ION N	UMBE	ZR	ing dat	aset				
7907	7908	7909	7910	7911	7912	7913	7914	7915	7916	7917	8044	8045
194	413	90	74	55	23	97	50	359	50	130	487	102
15	14	10	57	320	261	51	44	9	254	493	18	209
41	58	60	48	56	87	11	102	12	5		100	14
29	37	186	29	104	7	4	72	270	283		7	57
33	100	61	502	220	120	141	22	603	35		98	54
181	65	49	12	239	14	18	39	3	12		5	32
	9	14	70	47	62	142	3	104			85	67
	169	24	21	246	47	68	15	2			91	59
	447	56	29	176	225	77	197	438			43	134
	184	20	386	182	71	80	188				230	152
	36	79	59	33	246	1	79				3	27
	201	84	27	15	21	16	88				130	14
	118	44	153	104	42	106	46					230
	34	59	26	35	20	206	5					66
	31	29	326		5	82	5					61
	18	118			12	54	36					34
	18	25			120	31	22					
	67	156			11	216	139					
	57	310			3	46	210					
	62	76			14	111	97					
	7	26			71	39	30					
	22	44			11	63	23					
	34	23			14	18	13					
		62			11	191	14					
		130			16	18						
		208			90	163						
		70			1	24						
		101			16							_
		208			52							
					95							

e data sot Table 1

4.1. The Studies Using Non-Mixture Distribution. In [3], the failure times of 8044 have been handled and have been modelled with exponential distribution like in [12]. As a result of modeling, KS=0.1873 and p=0.7282 are obtained. As the same data set has been modelled by 2MED, the results are given in the table below.

LSE			MLE		
$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$
0.1881	4.3057	135.3318	0.2011	6.5146	133.6494
KS Stat.		p value	KS Stat.		p value
0.1501		0.9132	0.1634		0.8698

Table 2. Parameter estimations, KS and p values for 8044

When the KS values have been compared, one can noticed that the values obtained from 2MED for two methods are less than the values obtained from exponential distribution. Therefore, 2MED is more successful about modeling this data set.

The failure times of 7912 have been taken into account in [7]. In this study, firstly the exponential distribution has been handled and KS=0.2419 and p=0.0497 have been found. So it has been pointed out that exponential distribution is not suitable for 7912. Then the generalized exponential and gamma distributions have been regarded. KS and p values about these two distributions are given in the following table.

Table 3. KS and p values for Gamma and Generalized Exponential Distribution					
DISTRIBUTION	METHOD	KS VALUE	p VALUE		
Gamma	MLE	0.1706	0.3135		
Generalized Exponential	MLE	0.1744	0.2926		

As the same data set has been modelled by 2MED, parameter estimations, KS and p values are shown in the table below.

LSE			MLE		
$\widehat{\alpha}$	$\widehat{\theta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$
0.5534	105.1185	17.6043	0.6514	83.1754	15.555
KS Stat.		p value	KS Stat.		p value
0.1292		0.6517	0.1241		0.6987

Table 4. Parameter estimations, KS and p values for 7912

When the KS values have been compared for these three distributions, it has been seen that the smallest KS value belongs to 2MED. It shows that the best model for 7912 can be made by 2MED.

Both the failure times of 7913 and 7914 have been considered in [9] and Weibull distributions has been suggested for modeling. For 7913 KS=0.08833 p>0.5 and for 7914 KS=0.08953 p>0.5 have been found. Parameter estimations, KS and p values for 2MED are shown in the tables below.

LSE			MLE		
$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$
0.4006	84.6545	85.2081	0.4007	76.7802	76.8380
KS Stat.		p value	KS Stat.		p value
0.0787		0.9913	0.0802		0.9984

Table 5. Parameter estimations, KS and p values for 7913

LSE			MLE		
$\hat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$
0.5067	37.2292	108.8640	0.4705	63.9843	64.2501
KS Stat.		p value	KS Stat.		p value
0.0799		0.9947	0.0835		0.9909

Table 6. Parameter estimations, KS and p values for 7914

As seen from the tables, KS value obtained from 2MED is less than obtained from Weibull distribution for two aircrafts. Therefore one can be said that modeling with 2MED is more appropriate instead of modeling with Weibull distribution.

Considering failure times of 7910, it is indicated in [14] that Inverse Rayleigh distribution is quite suitable for this data set. As a result of modeling, KS=0.21379 and p=0.43879 have been found.

As the same data set has been modelled by 2MED, parameter estimations, KS and p values are shown in the table below.

	·· ·· ·· ·· ·· ·· ·· ·· ··					
LSE			MLE			
$\widehat{\alpha}$	$\widehat{\theta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$	
0.8418	63.8913	1230	0.5934	44.4595	233.3597	
KS Stat.		p value	KS Stat.		p value	
0.1718		0.7059	0.1	.917	0.5749	

Table 7. Parameter estimations, KS and p values for 7910

When a comparison is made according to KS values, it can be seen that 2MED is better about modeling this data set.

Considering whole data set, Weibull and Gamma distributions have been used as non-mixture distribution in [8] and [2]. KS and p values obtained from these distributions are given in the table below.

DISTRIBUTION	KS VALUE	p VALUE
WEIBULL	0.0509	0.6393
GAMMA	0.0634	0.3586

Table 8. KS and p values for Weibull and Gamma Distribution

As [8] and [2], also [15] has proposed Weibull and Gamma distribution for whole data set. According to [15], KS and p values are given in the following table.

Fal	Cable 9. KS and p values for Weibull and Gamma Distribution						
	DISTRIBUTION	KS VALUE	p VALUE				
	WEIBULL	0.0519	0.6025				
	GAMMA	0.0625	0.3665				

Finally, specified type of Gamma Dagum (GD) such as Gamma-Burr III (GB III), Gamma-Fisk or Gamma-Log Logistic (GF of GLLog), Zografos and Balakrishnan-Dagum (ZB-D), ZB-Burr III (ZB-B III), ZB-Fisk of ZB-Log Logistic (ZB-F or ZB-LLog), Burr III(B III) and Fisk or Log Logistic (F or LLog), Gamma Exponentiated Weibull (GEW) have been proposed for whole data set in [11]. Following table includes KS values for these distributions.

DISTRIBUTION	KS VALUE
GD	0.0401
ZB-D	0.0982
ZB-Burr III	0.0782
ZB-Fisk	0.0467
D	0.0421
GEW	0.3863

Table 10. KS values for GD, ZB-D, ZB-Burr III, ZB-Fisk, D and GEW Distribution

The values obtained from 2MED are given in the below.

Table 11	Table 11. Parameter estimations, KS and p values for whole data set					
LSE			MLE			
$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{ heta}_2$	$\widehat{\alpha}$	$\widehat{ heta}_1$	$\widehat{\theta}_2$	
0.7234	60.5029	198.4490	0.4311	46.6148	128.4549	
KS Stat.		p value	KS Stat.		p value	
0.0460		0.7410	0.0482		0.6881	

It is came out that 2MED is more appropriate about modeling according to KS criteria between 8 distributions proposed to model whole data set. Furthermore, 2MED is fourth order between distributions proposed in [11] so it can be said that 2MED have located among the suggested distributions.

4.2. The Studies Using Mixture Distribution. The failure times of 7912 have been modelled by mixed generalized exponential distribution with two component in [16]. As a result of modeling obtained values are as KS=0.1391 and p=0.5598.

In current study, this data set has been considered that can be modelled by 2MED. Parameter estimations, KS and p values are given in Table 4 according to MLE and LS methods.

It has been seen that the data set has been modelled both mixed and non-mixture distributions. As comparing results, it can be seen that 2MED is more successful about modeling this data set.

Considering whole data set, in addition to non-mixture distributions, mixture distributions such as EP and EG have been handled in [8]. KS and p values obtained from these distributions are given in the table below.

Table 12. KS and p values for EP and EG Distribution					
DISTRIBUTION	KS VALUE	p VALUE			
EP	0.0470	0.7351			
EG	0.0494	0.6759			

Besides [8], [15] have proposed EL to model whole data set.

able 13. KS and p values for EL, EP and EG Distribution			
DISTRIBUTION	KS VALUE	p VALUE	
EL	0.0491	0.6731	
EP	0.0470	0.7248	
EG	0.0499	0.6302	

[2] has contributed [8] and [15] proposing EB in their study. The obtained values are as follows.

DISTRIBUTION	KS VALUE	p VALUE
EB	0.0470	0.7356
$\operatorname{EL}$	0.0491	0.6828
EP	0.0470	0.7351
EG	0.0494	0.6759

Table 14. KS and p values for EB, EL, EP and EG Distribution

The values of 2MED for whole data set are given in table 11.

Similarly, it has been concluded that the best distribution to model for whole data set is 2MED when KS values obtained from [2], [8], [15] and current study have been compared.

### 5. CONCLUSION

The current study has regarded 10 different studies about modeling aircraft data set in the literature. Some of the studies have modelled the data set with non-mixture distributions and some of them have used mixture of two different distributions or same distributions.

Examining the results given in detail in Section 4, it is observed that the best model is 2MED almost all of the comparisons among 16 distributions. In case 2MED may not be recommended as the best model due to obtained KS values according to the MLE for the generally. Even in this case, 2MED has located in the best of the first three distribution. In addition to this case, 2MED is the fourth order to model the whole data set among proposed 6 distributions in [11]. Therefore suggested mixed distribution emerges as a distribution can be suggested and first come to mind.

Additionally, considering the results of the current study, 2MED is more appropriate to propose as a model instead of exponential distribution for the data set modelled with exponential distribution for the first time by [12].

Acknowledgement 1. This research has been funded by A.U. BAP Office.

## References

- Açıkgöz, I., 2007. Sonlu karma dağılımlarda parametre tahmini, PhD Thesis, Graduate School of Natural and Applied Sciences, Ankara University, Turkey.
- [2] Chahkandi, M. and Ganjali, M., 2009. On some lifetime distributions with decreasing failure rate. Computational Statistics and Data Analysis, Vol. 53, , pp. 4433- 4440.
- [3] Davison, A.C. and Hinkley D.V., 1997. Bootstrap methods and their application, Cambridge Series in Statistical and Probabilistic Mathematics, 1 edition, 594 p.
- [4] Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist.Soc. Ser. B 39, 1-38.
- [5] Everitt, E.S. and Hand, D.J., 1981. Finite Mixture Distributions, London: Chapman and Hall.
- [6] Gupta, R.D. and Kundu, D., 2000. Generalized exponential distribution: different method of estimations, Journal of Statistical Computation Simulation, Vol. 00, pp. 1 - 22.

- [7] Gupta R.D. and Kundu, D., 2003. Closeness of Gamma and Generalized Exponential Distribution. Communications in Statistics, Volume 32, Issue 4, pp. 705-721.
- [8] Kus, C. 2007. A new lifetime distribution, Computational Statistics & Data Analysis, 51 4497 - 4509.
- [9] Lin, C. and Ke, S., 2013. Estimation of P(Y<X) for location-scale distributions under joint progressively type-II right censoring. Quality Technology & Quantitative Management, Vol. 10, No. 3, pp. 339-352.
- [10] G.J. McLachlan, G.J. and Krishnan, T., 2008. The EM algorithm and extensions. John Wiley & Sons, Inc., Hoboken, New Jersey.
- [11] Oluyede, B. O., Huang, S. and Pararai, M., 2014. A New Class of Generalized Dagum Distribution with Applications to Income and Lifetime Data, Journal of Statistical and Econometric Methods, vol.3, no.2, 125-151.
- [12] Proschan, F., 1963. Theoretical explanation of observed decreasing failure rate. Technometrics 5, 375-383.
- [13] Ren, Y., 2011. The methodology of flowgraph models, PhD Thesis, Department of Statistics London School of Economics and Political Science.
- [14] A.I. Shawky, A.I., Badr, M.M., 2012. Estimations and prediction from the inverse rayleigh model based on lower record statistics. Life Science Journal (9;1).
- [15] Tahmasbi, R., Rezaei, S., 2008. A two-parameter lifetime distribution with decreasing failure rate. Computational Statistics and Data Analysis, Vol. 52, pp.3889-3901.
- [16] Tian, Y., Tian, M. and Zhu, Q., 2014. Estimating a Finite Mixed Exponential Distribution under Progressively Type-II Censored Data, Vol. 43(17), pages 3762-3776. *Current address*: Ankara University, Faculty of Sciences, Dept. of Statistics, Ankara, TURKEY

 $E-mail\ address:$  yilmazm@science.ankara.edu.tr, busebuyum@gmail.com URL: http://math.science.ankara.edu.tr