Available online at http://scik.org
J. Math. Comput. Sci. 6 (2016), No. 5, 907-921
ISSN: 1927-5307

ESTIMATION OF A NONLINEAR DISCRIMINANT FUNCTION FROM A MIXTURE OF TWO EXPONENTIATED–WEIBULL DISTRIBUTIONS

A. M. ABD-ELRAHMAN AND M. A. MOHAMMED*

Department of Mathematics, Assiut University, Assiut 71516, Egypt

Abstract. In this paper, a procedure for finding maximum likelihood estimates (MLEs) of the parameters of a finite mixture of two exponentiated – Weibull distributions (MEW) is presented, using classified and unclassified observations. Estimation of a nonlinear discriminant function on the basis of a small sample size is considered. Its performance is investigated by a series of simulation experiments. The simulations conducted for estimating a nonlinear discriminant function by the maximum likelihood method, on the basis of unclassified data drawn from a mixture of the underlying populations suggest that the error rate can be reduced by a substantial percentage for widely separated populations. Generally, the performance of the mixture discrimination procedure relative to the completely classified procedure, measured by total probabilities is good.

Keywords: mixture of exponentiated – weibull distributions; maximum-likelihood estimation; classification rules; probability of misclassification; Monte Carlo simulation.

2010 AMS Subject Classification: 62E20; 62F10; 62N05.

1. Introduction

The exponentiated – Weibull family was first introduced by Mudholkar and Srivastava [1]. This family is an extension of the Weibull family, which is obtained by adding an additional

*Corresponding author

Received June 11, 2016

shape parameter. The importance of this family lies in its ability to model monotone as well as non-monotone failure rates, which are quite common in reliability and biological studies. This family can be widely and effectively used in reliability applications because it has a wide variety of shapes in its density and failure rate functions making it useful for fitting many types of data.

The cumulative distribution (cdf), probability density (pdf), and failure rate functions of the exponentiated – Weibull distribution are given, respectively, by

$$F(x;\lambda,\alpha)=\left(1-e^{-x^{\lambda}}\right)^{\alpha},\quad x>0,\,(\alpha,\lambda>0),\tag{1.1}$$

$$f(x;\lambda,\alpha)=\alpha\lambda\, x^{\lambda-1}e^{-x^{\lambda}}\left(1-e^{-x^{\lambda}}\right)^{\alpha-1},\tag{1.2}$$

$$R(x;\lambda,\alpha)=\frac{\alpha\lambda x^{\lambda-1}e^{-x^{\lambda}}\left(1-e^{-x^{\lambda}}\right)^{\alpha-1}}{1-\left(1-e^{-x^{\lambda}}\right)^{\alpha}}.\tag{1.3}$$

Mixtures of life distributions occur when two different causes of failure are present each with the same parametric form of life distribution. Finite mixture of distributions have been used as models throughout the history of modern statistics. There are several areas of applications of finite mixture models. For example, in biology it is often required to measure certain characteristics in natural populations of particular species. Samples of individuals are taken from the natural habitat of the species and the characteristics under investigation is recorded for each individual in sample. The distribution of many such characteristics may very greatly with the age of the individuals and age is frequently difficult to ascertain in samples from wide populations. Consequently the biologist observing the population as a whole is dealing with a mixture of distributions, where mixing is over a parameter depending on the unobservable variate age. For examples, see Titterington et al. [2].

A random variable X is said to follow a finite mixture distribution with k components, if the pdf of X can be written as

$$f(x)=\sum_{j=1}^{k}p_j f_j(x),\tag{1.4}$$

where p_j is a non-negative real number known as the j^{th} mixing proportion such that $\sum_{j=1}^{k}p_j=1$ and $f_j(x)$ is the density function known as the j^{th} component, $j=1,2,\cdots,k$.

The property of identifiability is an important consideration on estimating the parameters in a mixture of distributions, testing hypotheses, classification of random variables, etc., can be meaningfully discussed only if the class of all finite mixtures is identifiable. Discussion of identifiability of finite mixtures may be found in several papers, among others, by Teicher [3], [4], Al-Hussaini and Ahmad [5], Ahmad [6] and Ahmad [7]. Recently, new finite mixtures for different distributions have been constructed and studied by many authors. Among others, see Sultan et al. [8], Ahmad et al. [9].

The pdf of the MEW distribution is given by

$$f(x|p,\alpha_1,\alpha_2,\lambda) = pf_1(x|\alpha_1,\lambda)+(1-p)f_2(x|\alpha_2,\lambda) \qquad (1.5),$$

where $0 \leqslant p \leqslant 1$, for $j = 1,2$, $f_j(\cdot)$ are as given by (1.2) after indexing α by j and λ is a common shape parameter. The finite MEW distribution has been discussed by Abd-Elrahman and Mohammed [10], they studied the identifiability problem related to this model and obtained MLEs for the parameters based on complete samples. Their results are applied for fitting two different real data sets. They showed that the failure rate function of the MEW model has bathtub shaped, bathtub-constant shaped, increasing or decreasing.

Unclassified data have been studied in the context of estimating mixtures and discriminant functions, there have been few practical applications. Unclassified observations have been used by Fukunga and Kessel [11], Moore *et al.* [12] for estimating the probability of misclassification for a discriminant rule. McLachlan [13], [14], discussed the use of unclassified observations for the special case of equal prior probabilities. For unclassified observations to be of practical use in estimating discriminant function, there must exist good estimates of the parameters of the mixture, this problem has been discussed by Day [15] and O'Neill [16]. The performance of a discriminant function estimated from mixture of two Inverse Gaussian distributions based on small sample size is studied by Amoh [17]. Mahmoud and Moustafa [18] considered the estimation of a discriminant function on the basis of small sample size from a mixture of two gamma distributions, and investigated its performance by a series of simulation experiments.

Ahmad [7] has studied small-sample results for a nonlinear discriminant function estimated from a mixture of two Burr type-XII distributions. Also, Ahmad and Abd-Elrahman [19] have studied a nonlinear discriminant function estimated from a mixture of two Weibull distributions. Mahmoud and Moustafa [20] have studied the errors of misclassification associated with the gamma distribution. Ahmad [21] has studied the efficiency of a nonlinear discriminant function based on unclassified initial samples from a mixture of two Burr type-XII distributions. Moustafa and Ramadan [22] have estimated a discriminant function from a mixture of two Gompertz distributions when the sample size is small. Recently, Ahmad et al. [23] have estimated a discriminant function from a mixture of two Gumbel distributions when the sample size is small. Sultan and Al-Moisheer [24] discussed the estimation of a discriminant function from a mixture of two inverse Weibull distributions based on classified and unclassified observations.

In this article, a mixture of two exponentiated – Weibull distributions is considered and estimation of parameters and nonlinear discriminant functions are studied. Also, three classification procedures, mixture, completely classified and optimal are compared.

Furthermore, we calculate the total probabilities of misclassification as well as the percentage biases. Moreover, we investigate the performance of all results through a series of simulation experiments by means of relative efficiencies. Finally, we give some conclude remarks.

2. Maximum Likelihood Estimation

Let $X_1, X2, \cdots, X_n$ be a random sample of size n drawn from a population whose pdf is given by (1.5). Following the procedure used by Day [15], the likelihood function can be written in the form

$$L(x;\Psi) = \lambda^n \left[\prod_{j=1}^{n} x_j^{\lambda-1} Q(x_j)\right] e^{-\sum_{j=1}^{n} x_j^{\lambda}}, \tag{2.1}$$

where $\Psi = (p,\ \alpha_1,\ \alpha_2,\ \lambda)$ and for $j = 1, 2, \cdots, n$,

$$Q(x_j) = p\,\alpha_1 w_j^{\alpha_1-1} + q\,\alpha_2 w_j^{\alpha_2-1},\ q = 1-p,\ w_j = 1-e^{-x_j^{\lambda}},$$

Differentiating the log-Likelihood function with respect to p, $\alpha_1\,\alpha_2$ and λ respectively, and equating to zero, the normal equations are then given by

$$0=\sum_{j=1}^{n}\frac{\alpha_1 w_j^{\alpha_1}-\alpha_2 w_j^{\alpha_2}}{A_j+B_j}, \tag{2.2}$$

$$0=\sum_{j=1}^{n}\frac{p w_j^{\alpha_1}[1+\alpha_1\ln w_j]}{A_j+B_j}, \tag{2.3}$$

$$0=\sum_{j=1}^{n}\frac{q w_j^{\alpha_2}[1+\alpha_2\ln w_j]}{A_j+B_j}, \tag{2.4}$$

$$0=\frac{n}{\lambda}+\sum_{j=1}^{n}\ln x_j\left[1-\frac{x_j{}^{\lambda}}{w_j}-\frac{x_j^{\lambda}\left(\alpha_1 A_j+\alpha_2 B_j\right)}{w_j\left(A_j+B_j\right)}\right], \tag{2.5}$$

where

$$A_j=p\alpha_1 w_j^{\alpha_1}, \qquad B_j=(1-p)\alpha_2 w_j^{\alpha_2}, \tag{2.6}$$

and w_j is as given in (2.1).

Define

$$W_{1j}=\frac{1}{1+\exp(a+b\,r_j(x_j))}=\frac{A_j}{A_j+B_j}, \quad W_{2j}=1-W_{1j}, \quad j=1,2,\cdots,n,$$

where

$$a=\ln\left(\frac{q\,\alpha_2}{p\,\alpha_1}\right), \quad b=\alpha_2-\alpha_1, \quad r(x_j)=\ln\left(1-\exp(-x_j^{\lambda})\right).$$

Hence, in view of (2.2)-(2.5) and using some initial values for the unknown vector of parameters, say $\Psi^{(0)}=(p^{(0)},\alpha_1^{(0)},\alpha_2^{(0)},\lambda^{(0)})'$, an estimate value for Ψ may be then iteratively obtained as follows: for $s=0,1,2,\cdots,1000$,

$$\hat{p}^{(s+1)}=\frac{1}{n}\sum_{j=1}^{n}W_{1j}^{(s)}, \qquad W_{1j}^{(s)}=W_{1j}|_{\Psi=\hat{\Psi}^{(s)}}, \tag{2.7}$$

$$\hat{\alpha}_1^{(s+1)}=-\frac{\sum_{j=1}^{n}W_{1j}^{(s)}}{\sum_{j=1}^{n}W_{1j}^{(s)}\ln w_j^{(s)}}, \qquad w_j^{(s)}=w_j|_{\Psi=\hat{\Psi}^{(s)}}, \tag{2.8}$$

$$\hat{\alpha}_2^{(s+1)}=-\frac{\sum_{j=1}^{n}W_{2j}^{(s)}}{\sum_{j=1}^{n}W_{2j}^{(s)}\ln w_j^{(s)}}, \qquad W_{2j}^{(s)}=W_{2j}|_{\Psi=\hat{\Psi}^{(s)}}, \tag{2.9}$$

where A_j, B_j and w_j are as given in (2.6), and $\hat{\Psi}^{(0)}=\Psi_0$.

In the $(s+1)^{\text{th}}$ iteration, $s = 0, 1, 2, \cdots, 1000$, $\hat{\lambda}^{(s+1)}$ may be obtained as the numerical solution of (2.5), after replacing Ψ by $(\hat{p}^{(s+1)}, \hat{\alpha}_1^{(s+1)}, \hat{\alpha}_2^{(s+1)}, \lambda)'$. This iterative procedure continues until reaching some accuracy. For the iterative procedure the following criterion is used to terminate the iterations.

For $s < 1000$, define

$$\delta = \sum_{i=1}^{4} \left| \frac{\hat{\psi}_i^{(s+1)} - \hat{\psi}_i^{(s)}}{\hat{\psi}_i^{(s)}} \right|,$$

where $\hat{\Psi}^{(s)} = (\hat{\psi}_1^{(s)}, \hat{\psi}_2^{(s)}, \hat{\psi}_3^{(s)}, \hat{\psi}_4^{(s)})' = (\hat{p}^{(s)}, \hat{\alpha}_1^{(s)}, \hat{\alpha}_2^{(s)}, \hat{\lambda}^{(s)})'$. If $\delta \leq 10^{-4}$, the iterative procedure will be then terminated, and $\hat{\Psi}^{(s+1)}$ will be accepted as an estimated value for the unknown vector of parameters Ψ. Otherwise, $\hat{\Psi}^{(1000)}$ will be accepted instead.

Furthermore, Redener and Walker [25] show that mixture problems are very often such that the log-likelihood function attains its largest local maxima at several different choices of the parameters of the mixture. That is, for mixture of two exponentiated – Weibull distributions, if the component parameters p, α_1, α_2 and λ are interchanging with the component parameters $(1-p)$, α_2, α_1 and λ the value of the log-likelihood will not change. In our simulation study, it is of interest to estimate the component density parameters. So that, for each sample we calculate the distances $d_1 = |\Psi^* - \Psi|$ and $d_2 = |\Psi^{**} - \Psi|$, where Ψ is the true vector of parameters, $\Psi^* = (\hat{p}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\lambda})'$ and $\Psi^{**} = (1-\hat{p}, \hat{\alpha}_2, \hat{\alpha}_1, \hat{\lambda})'$. If $d_1 \leq d_2$, we accept Ψ^* as an estimate of the vector of parameters. Otherwise, we accept Ψ^{**}.

3. Classification Rules

Consider two populations π_1 and π_2 with corresponding densities $f_i(x), i = 1, 2$, as given by (1.2) with parameters α_1 and α_2, respectively; and common shape parameter λ. Also, consider the nonlinear discriminant function:

$$NLD_o(x) = a + b\, r(x), \qquad r(x) = \ln\left[1 - \exp(-x^{\lambda})\right]. \tag{3.1}$$

The probability that an individual x of unknown origin has come from π_1 is given by

$$P_r(x \in \pi_1) = \frac{1}{1 + \exp\left[NLD_o(x)\right]},$$

which is call the posterior probability, see Afifi and Clark [26]. Then we may classify x in π_1 if $NLD_o(x) < 0$ and in π_2 if $NLD_o(x) \geq 0$.

If all parameters of the populations π_1 and π_2 are known, then we have an optimal nonlinear discriminant function $NLD_o(x)$, which is as given by (3.1), where

$$a = \ln\left(\frac{q\alpha_2}{p\alpha_1}\right), \qquad b = \alpha_2 - \alpha_1. \tag{3.2}$$

3.1 Estimated discriminant functions

Usually, the parameters of the populations are not known. Available data are then used to estimate the parameters in the density functions. The estimated discriminant functions are then constructed. We shall consider the following types of data

i) Classified sample: When data are obtained by sampling from a mixture population and the origin of each observation is determined after sampling, we will call the resulting sample classified "*c*" sample.

ii) Mixed sample: The case where each observation is unclassified will be called mixed "*m*" sample.

3.2 Classified sample case

For i=1,2, let $(x_{i1}, x_{i2}, \ldots, x_{in_i})$ be a classified sample of size n_i from π_i with probability density function given by (1.2); and $n_1 + n_2 = n$.

The solution of the following nonlinear equation ($h(\lambda) = 0$), using, for example, Newton-Raphson iteration scheme, gives an estimated value, $\widetilde{\lambda}$, of the parameter λ, based on the classified sample.

$$g(\lambda) = \frac{n}{\lambda} + \sum_{i=1}^{2}\sum_{j=1}^{n_i} \ln\left(x_{ij}\right)\left\{1 - x_{ij}^{\lambda}\left[1 + \frac{e^{-x_{ij}^{\lambda}}}{w(x_{ij})}\left(1 + \frac{n_i}{\sum_{k=1}^{n_i}\ln w(x_{ik})}\right)\right]\right\}. \tag{3.2.1}$$

$$g(\lambda) = \frac{n}{\lambda} + \sum_{i=1}^{2}\sum_{j=1}^{n_i} \ln\left(x_{ij}\right)\left\{1 - x_{ij}^{\lambda}\left[1 + \frac{e^{-x_{ij}^{\lambda}}}{w(x_{ij})}\left(1 + \frac{n_i}{\sum_{k=1}^{n_i}\ln w(x_{ik})}\right)\right]\right\}. \tag{3.2.2}$$

Once $\widetilde{\lambda}$ is obtained, $\widetilde{\alpha}_i$ is then given by

$$\widetilde{\alpha}_i = \frac{n_i}{\sum_{k=1}^{n_i} \ln\left(1 - e^{-x_{ik}^{\widetilde{\lambda}}}\right)}, \qquad i = 1,2.$$

Therefore, the classified nonlinear discriminant function, $NLD_c(x)$, is as given by (3.1) with a, b and λ replaced respectively by $\widetilde{a}$, $\widetilde{b}$ and $\widetilde{\lambda}$, where

$$\left.\begin{array}{rcl} \widetilde{a}_c &=& \widetilde{\alpha}_2 - \widetilde{\alpha}_1, \\ \widetilde{b}_c &=& \ln\left[\frac{(1-\widetilde{p})\widetilde{\alpha}_2}{\widetilde{p}\widetilde{\alpha}_1}\right], \\ \widetilde{\lambda}_c &\equiv& \widetilde{\lambda} \end{array}\right\} \tag{3.2.3}$$

with $\widetilde{p} = \frac{n_1}{n}$.

3.3 Mixed sample case

When all the initial observations are unclassified and they are only known to come from a mixture of π_1 and π_2. For this mixed sample, the nonlinear discriminant function, $NLD_m(\cdot)$, is as given by (3.1) with a, b and λ replaced by

$$\left.\begin{array}{rcl} \widehat{a}_m &=& \widehat{\alpha}_2 - \widehat{\alpha}_1, \\ \widehat{b}_m &=& \ln\left[\frac{(1-\widehat{p})\widehat{\alpha}_2}{\widehat{p}\widehat{\alpha}_1}\right], \\ \widehat{\lambda}_m &\equiv& \widehat{\lambda}, \end{array}\right\} \tag{3.3.1}$$

where

$\hat{p}$, $\hat{\alpha}_1$, $\hat{\alpha}_2$, and $\hat{\lambda}$ are calculated as described in Section (2).

4. Probabilities of misclassification

Let E_{ij}, $(i = 1,2; j = o,c,m)$ denote the conditional probability that an individual from π_i is misclassifying by the jth discriminant function, where

o denotes optimum

c denotes classified

m denotes mixture.

We also denote by E_j the total error rates obtained by weighting the conditional error rates (total probabilities of misclassification) by the true mixing proportions. Consider the nonlinear discriminant function

$$NLD_j(x) = a_j + b_j r\left(x^{\lambda_j}\right).$$

(Where λ_j is the estimated value of the parameter λ using the jth procedure and $r(\cdot)$ is as given in (3.1). We classify x in π_1 if $NLD_j(x) < 0$. Hence

$$E_{1j} = \Pr(a_j + b_j r\left(x^{\lambda_j}\right) > 0|\pi_1).$$

Let

$$\gamma_j = \left\{-\ln\left[1-\left(\frac{(1-p)\,\alpha_2}{p\alpha_1}\right)^{\frac{1}{\alpha_1-\alpha_2}}\right]\right\}^{1/\lambda},$$

then we have

$$E_{1j} = \begin{cases} F(\gamma_j,\,\alpha_1,\,\lambda), & \alpha_1 < \alpha_2, \\ 1-F(\gamma_j,\,\alpha_1,\,\lambda), & \alpha_1 > \alpha_2, \end{cases} \tag{4.1}$$

where $F(\cdot,\,\alpha,\,\lambda)$ is the CDF of the exponentiated – Weibull distribution, which is given by (1.1). Similarly, E_{2j} is given by

$$E_{2j} = \begin{cases} 1-F(\gamma_j,\,\alpha_2,\,\lambda), & \alpha_1 < \alpha_2, \\ F(\gamma_j,\,\alpha_2,\,\lambda), & \alpha_1 > \alpha_2, \end{cases} \tag{4.2}$$

The overall error rates weighted by the true mixing proportion, E_j, is given by

$$E_j = pE_{1j} + (1-p)E_{2j}. \tag{4.3}$$

5. Simulation experiments

A series of simulation experiments were performed to investigate the performance of $NL_m(x)$ relative to $NL_c(x)$ and $NL_o(x)$ for small samples. Three combinations of the parameters were taken in order to cover three different shapes as depicted in Table 1. For each combination of the parameters classified and mixture samples of sizes $n = 40$ and $n = 100$ were generated from the mixture distribution. The following procedure was used for generating a mixture sample:

- Two independent observations U_1 and U_2 are generated from $\mathscr{U}(0,1)$ using DRNUN routine from IMSL [27].
- If $U_1 \leq p$, then $X = \theta\left(-\ln\left(-U_2^{\frac{1}{\alpha_1}} + 1\right)\right)^{\frac{1}{\lambda}}$.
  Otherwise, $X = \theta\left(-\ln\left(-U_2^{\frac{1}{\alpha_2}} + 1\right)\right)^{\frac{1}{\lambda}}$.

X is then an observation from a MEW distribution. This procedure is continue n times. The n resulting observations will be a mixed sample of size n from a MEW.

Using the parameters and the calculated estimates the individual and total conditional probabilities of misclassification, as defined in (4.1)-(4.3), were evaluated for the completely classified and mixture discrimination procedures for each sample generated. These were averaged over 1000 repetitions for each combination of parameters considered. The sample means of the individual and total conditional probabilities of misclassification are denoted by $\bar{E}_{ij}$ and $\bar{E}_j$ $(i = 1,2,\ j = m,c)$ respectively. The corresponding optimal probabilities of misclassification E_{1o}, E_{2o} and E_o were also evaluated for each parameter combination. Table 2 shows the individual probabilities of misclassification for the three discrimination procedures for $n = 40$ and $n = 100$. The standard deviations for the conditional probabilities of misclassification are shown in parentheses. We find that generally $\bar{E}_{1j}$ $(j = c,m)$ are closer to the corresponding optimal values than $\bar{E}_{2j}$, considering these conditional probabilities of misclassification as estimates of the optimal probabilities of misclassification. We observe that for small values of $d = |\alpha_1 - \alpha_2|$, the estimates $\bar{E}_{1j}$ are poor with $\bar{E}_{2j}$, consistently exceeding E_{2o}. This is not surprising since when the parameters are small the components of the mixture population are not well-separated and hence it is very difficult to discriminate between them. The variance associated with $\bar{E}_{ij}$ are quite large and every $\bar{E}_{ij}$ lies within one standard deviation of the corresponding optimal value E_{io}. Also for every parameter combination, the standard deviation of E_{ic} is smaller than that of E_{im} since more information is known in the former case.

Table 3 shows the total probabilities of misclassification with the standard deviation of E_m and E_c shown in parentheses. Also shown are the standardized biases. The first entry in each cell under $B(\bar{E}_j)$ is the value of the absolute bias from E_o standardized by the standard deviation of $\bar{E}_j$ and the second is the value of the ratio of the absolute bias to E_o. On the other hand, B is the value of the ratio of the bias of $\bar{E}_m$ from $\bar{E}_c$.

From Table 3, we see that the total conditional probabilities of misclassification as estimates of the optimal probabilities are poor when $d = |\alpha_1 - \alpha_2|$ is small. The standard deviations are smaller than those for individual probabilities but they are still quite large. When d is large the estimates are quite good and $\bar{E}_c$ dose consistently better than $\bar{E}_m$. From the last column, we see that the mixture discriminant procedure relative to the classified performs poorly for d large and as d decrease the performance improves. When the sample size is increased from $n = 40$ to $n = 100$ all the estimates of the three combinations of parameters considered improve.

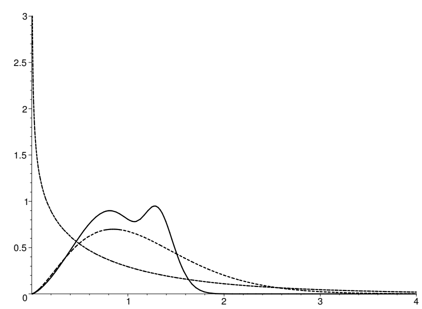

FIGURE 1. Graphs for the pdf with different shapes, Bi-Modal (Solid), Uni-Modal (Dot), and L-Shape (Dash).

TABLE 1. Three Different Mixed Populations.

Populations		Parameters			
No.	Type	p	α_1	α_2	λ
1	Bi-Modal	0.4	13.5	0.7	3.5
2	Uni-Modal	0.4	1.9	1.6	1.5
3	L-Shape	0.4	0.9	0.6	0.9

TABLE 2. Individual probabilities of misclassification.

PT	n	Classification Procedures					
		Mixtures		Completely Classified		Optimal	
		$\bar{E}_{1m}$	$\bar{E}_{2m}$	$\bar{E}_{1c}$	$\bar{E}_{2c}$	E_{1o}	E_{2o}
Bi-Modal	40	0.82929 (0.05469)	0.96950 (0.01607)	0.82797 (0.04854)	0.96987 (0.01357)	0.83192	0.97124
	100	0.83259 (0.03460)	0.96999 (0.01077)	0.83127 (0.03002)	0.97039 (0.00888)	0.83192	0.97124
Uni-Modal	40	0.15180 (0.24109)	0.94800 (0.10338)	0.22905 (0.22097)	0.85770 (0.18756)	0.04600	0.97417
	100	0.14808 (0.23124)	0.93456 (0.13218)	0.14114 (0.15909)	0.91515 (0.13133)	0.04600	0.97417
L-Shape	40	0.18684 (0.27692)	0.92444 (0.15714)	0.25897 (0.18962)	0.87259 (0.14325)	0.19753	0.91221
	100	0.18856 (0.25316)	0.91310 (0.15841)	0.21293 (0.12980)	0.90417 (0.07432)	0.19753	0.91221

TABLE 3. Total probabilities of misclassification and percentage biases.

PT	n	Classification Procedures			Relative Bias		
		Mixtures	Completely Classified	Optimal	Optimal		Completely Classified
		$\bar{E}_m$	$\bar{E}_c$	E_o	$B(\bar{E}_m)$	$B(\bar{E}_c)$	B
Bi-Modal	40	0.91754	0.91583	0.91551	0.08055	0.01321	0.00187
		(0.02518)	(0.02388)		0.00222	0.00034	
	100	0.91598	0.91543	0.91551	0.02576	0.00505	0.00059
		(0.01792)	(0.01603)		0.00050	0.00009	
Uni-Modal	40	0.64095	0.62494	0.60291	0.50220	0.34976	0.02561
		(0.07575)	(0.06300)		0.06309	0.03655	
	100	0.63039	0.61338	0.60291	0.49734	0.23916	0.02773
		(0.05526)	(0.04380)		0.04559	0.01738	
L-Shape	40	0.65167	0.64400	0.62634	0.28516	0.28192	0.01190
		(0.08882)	(0.06265)		0.04044	0.02820	
	100	0.63968	0.63446	0.62634	0.20094	0.19341	0.00822
		(0.06638)	(0.04201)		0.02130	0.01297	

Conflict of Interests

The authors declare that there is no conflict of interests.

REFERENCES

[1] G. S. Mudholkar and D. K. Srivastava, Exponentiated Weibull family for analyzing bathtub failure rate data, *IEEE Transactions on Reliability* **42**,(1993), 299-302.

[2] D. M. Titterington, A. F. M. Smith and U. E. Makov, *Statistical Analysis of Finite Mixture Distributions.* John Wiley and Sons, New York,(1985).

[3] H. Teicher, Identifiability of finite mixtures, *Ann. Math. Statist.* **33**,(1961), 244-248.

[4] H. Teicher, Identifiability of finite mixtures, *Ann. Math. Statist.* **34**,(1963), 1265-1269.

[5] E. K. Al-Hussaini and K. E. Ahmad, On the identifiabilty of finite mixtures of distributions. *IEEE Trans. Inform. Theory* **27** (5),(1981), 664-668.

[6] K. E. Ahamad, Identifiability of finite mixtures using a new transform. *Ann. Inst. Statist. Math.* **40** (2),(1988), 261-265.

[7] K. E. Ahmad, Small sample results for a nonlinear discriminant function estimated from a mixture of two Burr type XII distributions. *Comput. Math. Appl.* **28** (5), (1994), 13-20.

[8] K. S. Sultan, M. A. Ismail and A. S. A-Moisheer, On mixture of two inverse Weibull distributions: Properties and estimation. *Computional Statistics and Data Analysis* 51, (2007), 5377-5387.

[9] K. E. Ahmad, Z. F. Jaheen, H. S. and Mohamed, Finite mixture of Burr typy XII distribution and its reciprocal: properties and applications. *Statistical Paper* **52**, (2011), 835-845.

[10] A. M. Abd-Elrahman, and M. A. Mohammed, Etimation of mixed exponentiated Weibull parameters in life testing. *The* 4^{th} *International Conference for Young Researchers.* (Faculty of science, Assiut University, Assiut, Egypt). (2014).

[11] K. Fukunga, and D. Kessel, Non-parametric Bayes error estimation using unclassified samples *IEEE Transactions on Information Theory. IT.* **19**, (1973), 434-440.

[12] D. S. Moore, S. J. Whitsitt, and D. A. Landgrebe, Variance comparisons for unbiased estimators of probabilities of correct classification *IEEE Transactions on Information Theory, IT.* **21**, (1976), 102-105.

[13] G. J. McLachlan, Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation *J. Amer. Statist. Ass.* **70**, (1975), 365-369.

[14] G. J. McLachlan,Estimating the linear discriminant functions from initial samples containing a small number of unclassified observations *J. Amer. Statist. Ass.* **72**, (1977), 403-406.

[15] N. E. Day, Estimating the components of a mixture of two normal distributions, *Biometrika* **56**, (1969), 463-474.

[16] T. J. O'Neill, Normal distribution with unclassified data *J. Amer. Statist. Ass.* **73**, (1978), 821-826.

[17] R. K. Amoh, Estimation of a discriminant function from a mixture of two inverse Gaussian distributions when the sample size is small *J. Statist. Comp. Simul.* **20**, (1985), 275-286.

[18] M. A. Mahmoud, and H. M. Moustafa, Estimation of a discriminant function from a mixture of two gamma distributions when the sample size is small. *Mathl. Comput. Modelling* **18**, (1993), 87-95.

[19] K. E. Ahmad, and A. M. Abd-Elrahman, Updating a nonlinear discriminant function estimated from a mixture of two Weibull distributions. *Math. Comput. Modelling* **19** (11), (1994), 41-51.

[20] M. A. W. Mahmoud, and H. M. Moustafa, Errors of misclassification associated with gamma distribution *Math. Comput. Modelling* **22** (3), (1995), 105-119.

[21] K. E. Ahmad, The efficiency of a nonlinear discriminant function based on unclassified initial samples from a mixture of two Burr type XII populations. *Comput. Math. Appl.* **30** (2), (1995), 1-7.

[22] H. M. Moustafa, and S. G. Ramadan, On MLE of a nonlinear discriminant function from a mixture of two Gompertz distributions based on small sample sizes. *J. Stat. Comput. Simul.* **73**, (2003), 867-885.

[23] K. E. Ahmad, Z. F. Jaheen, and A. A. Modhesh, Estimation of a discriminant function based on small sample size from a mixture of two Gumbel distributions. *Comm. Statist. Simulation Comput.* **39** (4), (2010), 713-725.

[24] K. S. Sultan, and A. S. Al-Moisheer, Estimation of a discriminant function from a mixture of two inverseWeibull distributions. *Journal of Statistical Computation and Simulation* **83**(3), (2013), 405-416

[25] R. A. Redener, H. F. Walker, Mixture densities, maximum likelihood and the EM algorithm. *Siam Review* **26** (2), (1984), 195-239.

[26] A. A. Afifi, and V. Clark, *Computer Aided Multivariate Analysis*. Lifetime Learning Publications, Belmont, Calif, (1984).

[27] IMSL *IMSL Reference Manual*, IMSL, Inc., Houston, TX, 1995.