Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n2p318

Estimating and updating a linear discriminant function from the mixture of two one-parameter Lindley distributions

By Al-Moisheer, Daghestani, Sultan

Published: 20 June 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 02, June 2022, 318-334
DOI: 10.1285/i20705948v15n2p318

Estimating and updating a linear discriminant function from the mixture of two one-parameter Lindley distributions

A.S. Al-Moisheer*[a], A.F. Daghestani[b], and K.S. Sultan[c]

[a]*Department of Mathematics, College of Science, Jouf University, P.O.Box 848, Sakaka 72351, Kingdom of Saudi Arabia*
[b]*College of Science and Humanities - Jubail,, Imam Abdulrahman Bin Faisal University, , P.O. Box 1982, Dammam 31441, Kingdom of Saudi Arabia*
[c]*Department of Statistics and Operations Research, College of Science, King Saud University, P.O.Box 2455, Riyadh 11451, Kingdom of Saudi Arabia*

Published: 20 June 2022

Finite mixture models have several applications in many fields such as statistics, economics, marking, medicine and reliability analysis. In this paper, we obtain the maximum likelihood estimates of the parameters of the mixture of two one-parameter Lindley distributions by using two types of data namely; classified and unclassified samples. Next, we estimate the linear discriminant function of the underlying mixture model and calculate the total probabilities of misclassification as well as the percentage bias through a series of simulation experiments and some real data sets. Consequently, we study the problem of updating the discriminant function on the basis of data of unknown origin. We consider the updating procedure for the linear discriminant function on the basis of two one-parameter Lindley distributions in situations when the additional observations are mixed or classified. Finally, we study the performance of the updating procedures through some simulation experiments by means of the relative efficiencies.

keywords: finite mixtures, discriminant function, classified and unclassified observations, relative efficiency.

*Corresponding author: asalmoisheer@ju.edu.sa

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

The use of finite mixture distributions for modeling phenomena goes back to the early years of statistics. The mixture models have been considered extensively by many authors such as (Everitt and Hand (1981), Titterington et al. (1985), McLachlan and Basford (1988), Lindsay (1995), McLachlan and Peel (2000) and Al-Moisheer (2021)). Lindley distributions was proposed by Lindley (1958) in the context of fiducial and Bayesian statistics to illustrate the difference between fiducial distribution and posterior distribution. Moreover, the statistical properties of Lindley distributions were discussed by Ghitany et al. (2008) and they have shown that this distribution is a better model for some applications than other distributions such as exponential distribution. Therefore, Al-Moisheer et al. (2021) proposed a finite mixture of two one-parameter Lindley distributions and proved the identifiability. They also studied the statistical properties of this mixture model. Further, Daghestani et al. (2021) suggested a new mixture model in which one of itscomponents is the one parameter Lindley distribution namely the mixture of Lindley and Weibull distributions. On the other point of view, Al-Moisheer et al. (2021) introduced a new mixture model based on the one parameter Lindley distribution namely the mixture of Lindley and inverse Weibull distributions. Also, Al-Moisheer (2021) introduced a new mixture of Lindley and Lognormal Distributions. Here, we present the mathematical formula of the probability density function (pdf) of the mixture of two one-parameter Lindley distributions (MLLD) as follows

$$f(x;p,\theta_1,\theta_2) = p\left(\frac{\theta_1^2}{\theta_1+1}\ \ (1+x)\ e^{-\theta_1 x}\right) + (1-p)\ \left(\frac{\theta_2^2}{\theta_2+1}\ \ (1+x)\ e^{-\theta_2 x}\right),$$

$$0 < p < 1; x > 0; \theta_1, \theta_2 > 0 \quad (1)$$

where $\left(\frac{\theta_i^2}{\theta_i+1}\ \ (1+x)\ e^{-\theta_i x}\right)$ is the pdf of Lindley distribution with one parameter $\theta_i,\ i = 1,2$ and p is the mixing proportion. On the other hand, the cumulative distribution function (cdf) of the MLLD is given by

$$F(x;p,\theta_1,\theta_2) = p\left(1 - \frac{(\theta_1 a + 1 + \theta_1 x)\ \ e^{-\theta_1 x}}{\theta_1+1}\right) + (1-p)\ \left(1 - \frac{(\theta_2 + 1 + \theta_2 x)\ \ e^{-\theta_2 x}}{\theta_2+1}\right),$$

$$x > 0, \theta_1, \theta_2 > 0; 0 < p < 1. \quad (2)$$

Discriminant analysis is a technique for analyzing data to find a set of prediction equations based on independent variables that are used to classify individuals into groups (see McLachlan (2004)). The discriminant analysis of the finite mixture models is an important procedure for many applications. Okamoto (1963) has illustrated an asymptotic expansion for the distribution of the linear discriminant function. The probability of

misclassification for a discriminant rule has estimated by Fukunaga and Kessell (1973). (McLachlan (1975), McLachlan (1977)) has considered the use of the unclassified observations for a special case of equal prior probabilities. Besides, (Ganesalingam and McLachlan (1978), Ganesalingam and McLachlan (1979), Ganesalingam and McLachlan (1981)) have evaluated the efficiency of a linear discriminant function estimated from a mixture of normal populations. However, there are some books which discuss linear discriminant analysis as Duda et al. (2001). Murray et al. (1978) have suggested the updating procedures that are appropriate for non-normal situations. Some authors have considered the discriminant analysis, (see, Amoh and Kocherlakota (1991), Sultan et al. (2013), Al-Moisheer (2016), Ahmad et al. (2010) and Al-Moisheer et al. (2017)).
In this paper, we estimate the discriminant function and the updating process of the MLLD and focus on the linear methods for classification. This paper is organized as follows: In Section 2, we estimate the unknown parameters of the mixture model through the maximum likelihood estimation. We derive the optimal linear discriminant function from the MLLD in Section 3. Further, we estimate the discriminant function from the MLLD in Section 4 according to mixed and classified samples. In order to check the performance of the classification technique we calculate the total probabilities of misclassification. Next, in Section 5 we investigate the problem of updating the discriminant function estimated from the MLLD. Moreover, we discuss the error rate of misclassification and we evaluate the relative efficiency of the mixture and classified discrimination procedures in Section 6. Furthermore, we carry out some simulation experiments to evaluate the efficiency of the linear discriminant function in Section 7. Finally, in Section 8, we apply the estimated linear discriminant function of the MLLD on a set of real data and write our conclusions in Section 9.

2 MLEs

The method of maximum likelihood is used in a wide range of statistical analyses. In this section, we derive the maximum likelihood estimates of the unknown parameters of the MLLD. Let $X_1, X_2, ..., X_n$ be a random sample from MLLD, the log-likelihood function can be written from (1) as follows

$$L^* = \log L = \sum_{j=1}^{n} \log \left(p \frac{\theta_1^2}{\theta_1 + 1} (1 + x_j) e^{-\theta_1 x_j} + (1 - p) \frac{\theta_2^2}{\theta_2 + 1} (1 + x_j) e^{-\theta_2 x_j} \right), \quad (3)$$

Differentiating with respect to the parameters p, θ_1, θ_2 and equating these derivatives to zero, we get

$$L^* = \frac{\partial L^*}{\partial p} = \sum_{j=1}^{n} \frac{f_1(x_j; \theta_1) - f_2(x_j; \theta_2)}{p f_1(x_j; \theta_1) + (1 - p) f_2(x_j; \theta_2)} = 0, \quad (4)$$

$$\frac{\partial L^*}{\partial \theta_1} = \sum_{j=1}^{n} \frac{p(1+x_j)e^{-\theta_1 x_j}\left(\frac{\theta_1(2+\theta_1)}{(\theta_1+1)^2} - \frac{x_j\theta_1^2}{(\theta_1+1)}\right)}{pf_1(x_j;\theta_1)+(1-p)f_2(x_j;\theta_2)} = 0, \tag{5}$$

$$\frac{\partial L^*}{\partial \theta_2} = \sum_{j=1}^{n} \frac{(1-p)(1+x_j)e^{-\theta_2 x_j}\left(\frac{\theta_2(2+\theta_2)}{(\theta_2+1)^2} - \frac{x_j\theta_2^2}{(\theta_2+1)}\right)}{pf_1(x_j;\theta_1)+(1-p)f_2(x_j;\theta_2)} = 0, \tag{6}$$

where

$$f_i(x_j) = \frac{\theta_i^2}{\theta_i+1}\ (1+x_j)\ e^{-\theta_i x_j}, i = 1, 2, ..., n \ \text{ and}, j = 1, 2. \tag{7}$$

The non linear equations (4-6) have to be solved numerically using one of the numerical techniques such as Newton-Raphson method. This has been done by applying the (rootSolve) package in R (see; Kerns(2010)). Section 8 displays the numerical results of the ML estimates. The variance-covariance matrix of the estimates has been obtained by first calculating the second order partial derivatives of the loglikelihood function using equation (4-6) in order to compute Fisher information matrix and then inverting this matrix.

3 Linear Discriminant Analysis

The statistical approach of the discrimination problem or classification considered in this section, deals with a set of observations which come from two populations. Also, the discriminant functions can be either linear in the components of x or nonlinear. A linear discriminant function is created as a linear combination of independent variables. Now if all population parameters are known, we can construct an optimal discriminant function denoted by $LD_o(x)$ based on these parameters. In most cases, the population parameters are unknown then the MLEs will be used in place of the unknown parameters to give the estimated discriminant function. The discriminant analysis method is called linear if the associated Bayes decision rule is linear with respect to the observation x. To define a linear discriminant function as $LD(x) = a + bx$, we use the Bayesian decision theory that is a fundamental statistical approach to the problem of pattern classification. Consequently, we show the formulation of the linear discriminant function of MLLD and how we use it to classify a new observation according to this discriminant function. Let Π_1 and Π_2 be two populations of the one-parameter Lindley distribution with densities $f_i(x)$, $i = 1, 2$, as given in (7). Also, we define W_{1j} as the probability that the observation x_j arises from the i^{th} population. Therefore, we can compute the posterior probabilities that the observation x_j has been generated by each of the two populations. Since the underlying model is mixture then we can calculate the probabilities W_{1j} and $W_{2j} = 1 - W_{1j}$ by using Bayes' formula expressed as follows [see Duda et al. (2001)]

$$W_{1j} = \frac{pf_1(x_j;\Theta_1)}{pf_1(x_i;\theta_1)+(1-p)f_2(x_i;\theta_2)} \quad \text{and} \quad W_{2j} = 1 - W_{1j}, \tag{8}$$

that is

$$W_{1j} = \{1 + \exp[a + bx]\}^{-1}, \tag{9}$$

where

$$a = \ln \frac{(1-p)}{p} + \ln \frac{\theta_2^2(1+\theta_1)}{\theta_1^2(1+\theta_2)} \quad \text{and} \quad b = \theta_1 - \theta_2, \tag{10}$$

and p is the prior probability of the observation coming from Π_1.
Consequently, the probability that an individual x of unknown origin has come from Π_1 is given as

$$Pr(x \in \Pi_1) = \{1 + \exp[LD(x)]\}^{-1}, \tag{11}$$

where $LD(x) = a + bx$. Therefore, we may classify x in Π_1, according to whether the value of the linear discriminant function $LD(x)$ is less or greater than zero. So, if $x \in \Pi_1$ then $(1-p) = 0$ and $a = -\infty$ since $\ln 0 = -\infty$. Thus, the classification rule is to assign a new observation to Π_1 or Π_2 according to the linear discriminant function based on the following result

$$\begin{cases} x \in \pi_1, & if \quad LD(x) < 0, \\ x \in \pi_2, & otherwise. \end{cases} \tag{12}$$

4 Estimation of the Discriminant Function

Estimating discriminant functions is important in the field of discriminant analysis. Usually, the parameters of the populations are unknown. The available data can be used to estimate the unknown parameters of the density function and then the discriminant function can be estimated. Therefore, an important factor in the estimation of the discriminant function is the available data. We will consider the following data types in order to estimate the discriminant function for these data types.
(i) Classified sample: In this data type, we obtain the data by sampling from each population and the origin of each observation is known while the population parameters are unknown. Consequently, we can obtain a classified discriminant function based on the parameter estimates from these classified samples denoted by (c).
(ii) Mixed sample: In this case, we have data obtained by sampling from a mixture population and the origin of each observation is unknown. Our concern here is to construct a mixture discriminant function based on the mixed samples denoted by (m).
(iii) Classified and mixed sample: This data type is a combined sample from classified, Type (i), and mixed, Type (ii), samples. This kind of samples is denoted by cm sample and we use it to update the discriminant function that estimated from MLLD.

4.1 Classified sample (c)

In this type of data, we have initial observations or input vector as $(x_{i1}, x_{i2}, ..., x_{in_i})$ available from Π_i with sample size $n_i, i = 1, 2$ and $n = n_1 + n_2$. In order to obtain the classified linear discriminant function $LD_c(x)$, we will replace the unknown parameters of the following equation by their MLEs calculated from the classified samples as given below.

$$LD_c(x) = \tilde{a} - \tilde{b}x, \tag{13}$$

where $\tilde{a} = \ln \frac{(1-\tilde{p})}{\tilde{p}} + \ln \frac{\tilde{\theta}_2^2(1+\tilde{\theta}_1)}{\tilde{\theta}_1^2(1+\tilde{\theta}_2)}$ and $\tilde{b} = \tilde{\theta}_1 - \tilde{\theta}_2$ and $(\tilde{\theta}_1, \tilde{\theta}_2)$ are the MLEs in the classified samples case are given by [see Ghitany et al. (2008)]

$$\tilde{\theta}_i = \frac{-(\bar{x}_i - 1) + \sqrt{(\bar{x}_i - 1)^2 + 8\bar{x}_i}}{2\bar{x}_i}, \quad i = 1, 2, \tag{14}$$

$$\tilde{p} = \frac{n_1}{n}. \tag{15}$$

4.2 Mixed sample (m)

In the case of this type of data, all initial observations come from the mixture of Π_1 and Π_2. The linear discriminant function for this mixed sample is given by

$$LD_m(x) = \hat{a} + \hat{b}x, \tag{16}$$

where $\hat{a} = \ln \frac{(1-\hat{p})}{\hat{p}} + \ln \frac{\hat{\theta}_2^2(1+\hat{\theta}_1)}{\hat{\theta}_1^2(1+\hat{\theta}_2)}$ and $\hat{b} = \hat{\theta}_1 - \hat{\theta}_2$ and $(\hat{p}, \hat{\theta}_1, \hat{\theta}_2)$ are MLEs of the parameters that given in Section 2 equations (4-6) via using a numerical method such as the Newton-Raphson. We use the R packages software to get the MLEs of the three unknown parameters, (see; Kerns (2010)).

5 Updating Procedure

In this section, we define the updating procedure for the discriminant function. Updating the linear discriminant function shows how the additional observations affect the performance of the discriminant function. To update the linear discriminant function that is estimated from the MLLD we use the third data type. Commonly, we separate the (cm) sample into the following two models.

5.1 Model (I) (classified)

Suppose that a discriminant function has been estimated based on a limited number of classified observations from one-parameter Lindley populations. Also, we have a reasonably large sample of unclassified observations from the MLLD. Therefore, we use these additional observations to improve the estimates of the unknown parameters for estimating the discriminant function. Consequently, we will examine the performance of the updated discriminant function by comparing it with the initial one that was estimated from the classified data.

5.2 Model (II) (unclassified)

In this model, we assume that a linear discriminant function has been estimated by using a sample from MLLD. We then obtain a relatively small sample of classified observations from the one-parameter Lindley populations. We wish to use these classified observations to update the estimates of the parameters for estimating the linear discriminant function. The performance of the updated linear discriminant function will be compared to that of the linear discriminant function estimated from the mixture sample alone. McLachlan and Ganesalingam (1982) have studied the problem of updating a discriminant function on the basis of unclassified data which is a data of unknown origin. Consequently, we can obtain the estimates of the parameters p, θ_1 and θ_2 as a by-product. Let the classified samples be represented by $X_i = (x_{i1}, x_{i2}, ..., x_{in_i})$, for $i = 1, 2$, with $n = n_1 + n_2$ and the mixed sample represented by $Y = (y_1, y_2, , y_m)$.
Let the classified samples be represented by $\underline{X_i} \equiv (x_{i1}, x_{i2}, ..., x_{in_i})$, for i= 1, 2, with $n = n_1 + n_2$ and the mixed sample represented by $\underline{Y} \equiv (y_1, y_2, ..., y_m)$.
Under both models and using the cm sample, the likelihood function is given by

$$L_2(p_1, \alpha_1, \alpha_2, \beta | \underline{x_1}, \underline{x_2}, \underline{y}) = \binom{n}{n_1} p^{n_1}(1-p)^{n_2} \{ \prod_{i=1}^{2} \prod_{j=1}^{n_i} f_i(x_{ij}) \} \prod_{k=1}^{m} f(y_k), \tag{17}$$

where $f_i(.)$ and $f(.)$ are given, respectively, by (7) and (1) for $i = 1, 2$. By differentiating $logL_2$ with respect to the parameters $(p, \theta_1,\ \theta_2)$ and setting the partial derivatives to zero, we can obtain the required estimates as illustrated below

$$\frac{\partial \log L_2}{\partial p} = \frac{n_1}{p} - \frac{n_2}{(1-p)} + \sum_{k=1}^{m} \frac{f_1(y_k;\theta_1) - f_2(y_k;\theta_2)}{p f_1(y_k;\theta_1) + (1-p) f_2(y_k;\theta_2)} = 0, \tag{18}$$

$$\frac{\partial \log L_2}{\partial \theta_1} = \sum_{j=1}^{n_1} \frac{(1+x_{1j}) e^{-\theta_1 x_{1j}} \left(\frac{\theta_1(2+\theta_1)}{(\theta_1+1)^2} - \frac{x_{1j}\theta_1^2}{(\theta_1+1)} \right)}{f_1(x_{1j};\theta_1)}$$

$$+ p \sum_{k=1}^{m} \frac{(1+y_k) e^{-\theta_1 y_k} \left(\frac{\theta_1(2+\theta_1)}{(\theta_1+1)^2} - \frac{y_k\theta_1^2}{(\theta_1+1)} \right)}{p f_1(y_k;\theta_1) + (1-p) f_2(y_k;\theta_2)} = 0, \tag{19}$$

$$\frac{\partial \log L_2}{\partial \theta_2} = \sum_{j=1}^{n_2} \frac{(1+x_{2j})e^{-\theta_2 x_{2j}}\left(\frac{\theta_2(2+\theta_2)}{(\theta_2+1)^2} - \frac{x_{2j}\theta_2^2}{(\theta_2+1)}\right)}{f_2(x_{2j};\theta_2)}$$

$$+ (1-p)\sum_{k=1}^{m} \frac{(1+y_k)e^{-\theta_2 y_k}\left(\frac{\theta_2(2+\theta_2)}{(\theta_2+1)^2} - \frac{y_k\theta_2^2}{(\theta_2+1)}\right)}{pf_1(y_k;\theta_1)+(1-p)f_2(y_k;\theta_2)} = 0. \quad (20)$$

Furthermore, we use R packages software to solve the system of non-linear equations (18-20) by applying some numerical methods such as Newton Raphson to get the parameter estimates, (see; Kerns (2010)). All the numerical results are displayed later in Section 7.

6 Error Rates and Relative Efficiencies

Since there is always a possibility of making the wrong classification, we may need to define the probability of an observation misclassifying. Therefore, we can compute this probability as follows:
Suppose $e_{ik}, i = 1,2, k = o,c,m$ denote the conditional probability that an individual from Π_1 is misallocated by the k^{th} discriminant function, where o denotes optimum, c denotes classified and m denotes mixed. We also have e_k that denote the overall error rates which can be obtained by weighting the conditional error rates (total probabilities of misclassification) by the true mixing proportions. So, e_k is given by

$$e_k = p_1 e_{1k} + p_2 e_{2k}. \quad (21)$$

We classify $x \in \Pi_1$, if $LD_k(x) < 0$, for $k = o,c,m$, otherwise $x \in \Pi_2$. The probabilities of misclassifying an observation from Π_i, $i = 1,2$ by the linear discriminant function $LD_k(x)$ is

$$e_{1k} = Pr(a_k + b_k > 0 \mid \Pi_1), \quad k = o,c,m,. \quad (22)$$

Putting $\gamma_k = [-a_k/b_k]$, we have

$$e_{1k} = \begin{cases} 1 - F_1(\gamma_k,\theta_1), & \theta_1 > \theta_2, \\ \\ F_1(\gamma_k,\theta_1), & \theta_1 < \theta_2. \end{cases} \quad (23)$$

where $F_i(.,\theta_i), i = 1,2$ is the cdf of the one parameter Lindley distribution. Similarly, e_{2k} is given by

$$e_{2k} = \begin{cases} F_2(\gamma_k, \theta_2), & \theta_1 > \theta_2, \\ \\ 1 - F_2(\gamma_k, \theta_2), & \theta_1 < \theta_2. \end{cases} \tag{24}$$

We can examine the performance of the updated discriminant functions when the additional observations are classified (Model I) and unclassified (Model II) according to the linear discriminant functions estimated from classified and mixed samples by Monte Carlo simulation experiments. The relative efficiency is a performance measure for the developed updating procedures that obtained based on the total error rates. The relative efficiencies of the updated linear discriminant function relative to the initial classified (c) and mixed (m) linear discriminant functions can be denoted, respectively, by , ε_c and ε_m, where

$$\varepsilon_c = \frac{\bar{e}_c - \bar{e}_{MII}}{\bar{e}_c - \bar{e}_{MI}}, \tag{25}$$

and

$$\varepsilon_m = \frac{\bar{e}_m - \bar{e}_{MII}}{\bar{e}_m - \bar{e}_{MI}}. \tag{26}$$

Also, we measure the asymptotic relative efficiencies of the updated linear discriminant function by

$$\varepsilon_\infty = \frac{\bar{e}_{MI} - e_o}{\bar{e}_{MII} - e_o}. \tag{27}$$

where MI denotes the updating according to completely classified data (Model I), MII denotes the updating according to mixture (unclassified) data (Model II) and o denotes optimum.

7 Simulation Experiments

This section examines the performance of estimating and updating the discriminant functions of the MLLD through different types of data classified (c), mixed (m), classified and mixed (cm) which are separated according to Model (I) and Model (II). The main goal of this section is to show the usefulness of the linear discriminant function and study the behavior of the MLEs of the mixed and classified samples that used to estimate the linear discriminant function from MLLD. In addition, we investigate the performance of $LD_m(x)$ compared with $LD_c(x)$ and $LD_o(x)$, via the error of misclassification criterion. Besides, we update the discriminant function when the additional observations are mixed or classified. Thus, the simulation algorithm is described as follows:

Generate random samples of sizes 50 and 100 from MLLD using different combinations of parameter values such as $\theta_1 = 0.55, 0.70$ with $\theta_2 = 0.5, 0.85$ and different values of the mixing proportion $p = 0.45, 0.75$. Further, we use R package (rlindley) to generate the required random samples from one parameter Lindley distribution via the uniform generator (runif).

Repeat the samples generation according to 1000 replications gives n_1 observations identifiable from the first component, n_2 from the second component yielding a mixed sample of size $n = n_1 + n_2$. The numerical results for the simulation process of the MLLD according to classified and mixed samples are displayed in Tables 1, 2, 3 and 4. In Table 1, we calculate the bias and mean squared error (MSE) of the parameter estimates from the mixed samples and compare them with those from the classified samples based on 1000 repetitions and different choices of MLLD parameters. By using the classified samples we have n_1 is drawn from the first components with parameter θ_1 and n_2 is drawn from the second components with parameter θ_2.
Also, in Tables 1, we observe that, the estimates are consistent since the MSE of the MLEs decreases when the sample size increases for both samples classified and mixed. Moreover, we perform a series of simulation experiments in order to study the performance of $LD_m(x)$ relative to $LD_c(x)$ and $LD_o(x)$ as displayed in Table 2 that show the conditional probabilities of misclassification for the classified and mixed samples with the optimal probabilities of misclassification when $n = 50, 100$.
Therefore, we estimate the unknown parameters of MLLD by MLE and evaluate the total conditional probability of misclassifications, as defined in (22), for the classified, mixture and updated linear discriminant procedures for each generated sample. There were averaged over 1000 repetitions for each combination of MLLD parameters considered and listed in Table 2 where the standard deviation of the probabilities of misclassification is shown in parentheses.
In general, we notice from Table 2 that the standard deviation of $e_{im}, i = 1, 2$ are smaller than that of $e_{ic}, i = 1, 2$ since the mixture samples gives less error in estimating the model unknown parameters. Further, the values of $\bar{e}_m$ are closer to the corresponding optimal values than those of $\bar{e}_c$. In addition, when p increase, then the optimal probabilities of misclassification decrease. Also, when θ_1 or θ_2 increase the optimal values decrease.
Besides, in Table 2, we calculate the relative bias to optimal for both mixed and classified. The first entry in each cell under $|B(\bar{e}_k)|$ is the value of the absolute bias from e_o, standardized by the standard deviations of $e_k, k = m, c$ denoted by sd_kwhich is given by $|\bar{e}_k - e_o|/sd_k, k = m, c$. The second is the value of the ratio of the absolute bias to e_o which is given by $|\bar{e}_k - e_o|/e_o, k = m, c$. So, we notice that the mixed samples have better results than classified samples since the values of the relative bias according to e_o are less in the mixed samples and close to zero.
However, there are some cases where the performance of $\bar{e}_c$ getting better and be more close to the optimality especially when the value of the mixing proportion p increase from 0.45 to 0.75. Conversely, when we increase the value of θ_1 from 0.55 to 0.70 we get bad performance for $\bar{e}_c$. Moreover, the performance of the mixture discrimination procedure is quite well compared with the classified procedure in terms of total probability.
From Table 3, we can see that the total probabilities of misclassification of the updated procedure are close to the corresponding optimal case. However, the value of the total probabilities of misclassification $\bar{e}_m$ decreases as m increase. On the other hand, $\bar{e}_m$ and $\bar{e}_c$ are not good estimates of e_o compared with the total probabilities of misclassification in the updated procedure.
The relative efficiencies and asymptotic relative efficiencies of the updated discriminant

functions, as defined, respectively, in (25), (26) and (27) were calculated and presented in Table 4. In all cases of Table 4 the sample size has a significant effect on the relative efficiency since they are decrease when m increases from 50 to 100 in Model I (classified) and Model II (unclassified). Moreover, when we increasing the mixing proportion p, the asymptotic relative efficiency will decreases at $\theta_1 = 0.55$ and increases at $\theta_1 = 0.70$. In most cases, we notice that the asymptotic value of the relative efficiency decreases when m increases from 50 to 100.

8 Real Data Analysis

In this section, we apply the estimated discriminant function of the MLLD on a real data set to illustrate the importance of the classification process.

Application 1

The first data considered here represents the ordered lifetimes of 20 electronic components and these data are shown as follows
0.03, 012, 0.22, 0.35, 0.73, 0.79, 1.25, 1.41, 1.52, 1.79, 1.8, 1.94, 2.38, 2.4, 2.87, 2.99, 3.14, 3.17, 4.72, 5.09; (see; Razali and Salih (2009)).
The unknown parameters of the MLLD are estimated using the maximum likelihood method. In order to solve the system of non linear likelihood equations, the R package (nleqslv) is used (see; Kerns (2010)). In addition, we evaluate the Kolmogorov-Smirnove (ks) test statistic and its corresponding p-value using the function ks.test() in R (see; Kerns (2010)) to check the goodness of fit of the real data to MLLD. The results of the MLEs of the MLLD parameters and the ks test statistic are displayed below.

MLEs for parameters			ks	P-value
$\hat{p}$	$\hat{\theta}_1$	$\hat{\theta}_2$		
0.1006	0.2555	0.7721	0.1291	0.8511

We note that the maximum distance between the data and the fitted MLLD is 0.1291 and that the p-value is 0.851 which indicate that the MLLD is an appropriate model for this data set. Moreover, we apply the linear discriminant function $LD_m(x)$ by (16) for the above data set in order to classify the observations in the mixed data set one by one into either Π_1 or Π_2. We find that 1 out of 20 from the first population Π_1 are classified into Π_1 with probability 0.05 and 19 out of 20 from the second population Π_2 are classified into Π_2 with probability 0.95.

Application 2

The second application is taken from Torabi et al. (2014). The data set consists of 188 observations that represent the number of successive failures for the air conditioning system of each member in a fleet of Boeing 720 jet airplanes (1963). Again, we use the maximum likelihood method to estimate the unknown parameters of the MLLD and we check the data fitting to the MLLD through the ks test statistic and its p-value. From the results shown below, we note that the maximum distance between the data and the fitted model is 0.0497 and that the p-value is 0.7429 which reveal that the model is a good fit for the data set.

MLEs for parameters			ks	P-value
$\hat{p}$	$\hat{\theta}_1$	$\hat{\theta}_2$		
0.4602	0.0121	0.0626	0.0497	0.7429

Also, applying the linear discriminant function $LD_m(x)$ in (16) for this data set, we find that 76 out of 188 from the first population Π_1 are classified into Π_1 with probability 0.4043 whereas 112 out of 188 from the second population Π_2 are classified into Π_2 with probability 0.5957.

9 Conclusion

In this paper, the discriminant analysis of the mixture of two one-parameter Lindley distributions is our main goal. So, we have introduced the mathematical formula of the pdf and cdf of this mixture model. Next, the maximum likelihood method has been used in order to estimate the unknown parameters of the underlying mixture model. Furthermore, we discuss the discriminant function of the MLLD and focus on the linear methods for classification. Generally, the linear discriminant function can be considered as a prototype classifier since it is commonly used and easy for interpretation in many situations. Also, we derive the optimal linear discriminant function from the MLLD. We estimate the discriminant function from the MLLD according to mixed and classified samples and we calculate the total probabilities of misclassification in order to check the performance of the classification techniques. moreover, we investigate the problem of updating procedure for the linear discriminant function on the basis of the MLLD in two different situations when the additional observations are classified or mixed. Besides, we evaluate the performance of all the classification procedures via calculating the error rate of misclassification and the relative efficiency means of the estimated discriminant function from the MLLD through a series of simulation experiments. Therefore, we conclude that the total probabilities of misclassification $\bar{e}_m$ are closer to the corresponding optimal values than those of $\bar{e}_c$. Further, the mixed samples have better results than classified samples since the values of the relative bias according to e_o are less in the mixed samples and close to zero. Thus, the performance of the mixed discriminant procedure is quite well compared with the classified procedure in terms of the total probabilities of misclassification. In general, the total probabilities of misclassification of the updated procedure are close to the corresponding optimal case than $\bar{e}_m$ and $\bar{e}_c$.

References

Ahmad, K. E., Jaheen, Z. F. and Modhesh, A. A. (2010). Estimation of a discriminant function based on small sample size from a mixture of two Gumbel distributions. *Communications in Statistics - Simulation and Computation*, 39(4), 713–725.

Al-Moisheer, A. S., Arfaoui, H. and Maddouri, F. (2017). The asymptotic relative efficiency of a nonlinear discriminant function from a mixture of two inverse Weibull distributions. *Journal of Computational and Theoretical Nanoscience*, 14(2), 1214-1221.

Al-Moisheer, A. S., Daghestani, A. F. and Sultan, K. S. (2021). Mixture of Lindley and Inverse Weibull Distributions: Properties and Estimation. *WSEAS Transactions on Mathematics*, 20(14), 134-143.

Al-Moisheer, A. S., Daghestani, A. F. and Sultan, K. S. (2021). Mixture of Two One-Parameter Lindley Distributions: Properties and Estimation. *Journal of Statistical Theory and Practice*, 15(11), 1-21.

Al-Moisheer, A. S. and Sultan, K. S. (2016) Estimation of a discriminant function from a mixture of two Burr Type III distributions. *Communications in Statistics - Simulation and Computation*, 45, 3760-3775.

Al-Moisheer, A. S. (2021) Sequential test for a mixture of finite exponential distribution. *Journal of Mathematics*, vol. 2021, Article ID 6625853, 10 pages.

Al-Moisheer, A. S. (2021) Mixture of Lindley and Lognormal distributions: properties, estimation and application. *Journal of function spaces*, Volume 2021, Article ID 9358496, 12 pages.

Al-Moisheer, A. S. (2016) Updating a nonlinear discriminant function estimated from a mixture of two Burr Type III distributions. *Journal of Applied Statistics*, 44, 2685-2696.

Amoh, R. K. and Kocherlakota, K. (1991). Updating discriminant functions estimated from inverse Gaussian populations. *Communications in Statistics - Simulation and Computation*, 20, 619-637.

Daghestani, A. F., Sultan, K. S. and Al-Moisheer, A. S. (2021). Mixture of Lindley and Weibull distributions: properties and estimation. *Journal of Statistics Applications & Probability*, 10(2), 301-314.

Duda, R. O., Hart, P. E. and Stork, D. G. (2001). Pattern classification. John Wiley & Sons, New York.

Everitt, B. S. and Hand, D. J. (1981). Finite mixture distribution. Chapman & Hall, London.

Fukunaga, K. and Kessell, D. (1973). Non parametric Bayes error estimation using unclassified samples. *IEEE Transactions on Information Theory*, 19, 434-440.

Ganesalingam, S. and McLachlan, G. J. (1978). The efficiency of a linear discriminant function based on unclassified initial samples. *Biometrika*, 65(3), 658-662.

Ganesalingam, S. and McLachlan, G. J. (1979). Small sample results for a linear discriminant function estimated from a mixture of normal populations. *Journal of Statistical Computation and Simulation*, 9, 151-158.

Ganesalingam, S. and McLachlan, G. J. (1981). Some efficiency results for the estimation of the mixing proportion in a mixture of two normal distributions. *Biometrika*, 37, 23-33.

Ghitany, M. E., Atieh, B. and Nadarajah, S. (2008). Lindley distribution and its applications. *Mathematics and Computers in Simulation*, 78, 493-506.

Kerns, G. Jay. (2010). Introduction to probability and statistics using R. ISBN: 978-0-557-24979-4. 386 pages. Publisher: G. Jay Kerns; first edition (January 1, 2010).

Lindley, D. V. (1958). Fiducial distribution and Bayes' theorem. *Journal of the Royal Statistical Society*, 20, 102-107.

Lindsay, B. G. (1995). Mixture Models: Theory, Geometry, and Applications. The Institute of Mathematical Statistics, Hayward, California.

McLachlan, G. J. (1975). Iterative reclassification procedure for constructing an asymptotically optimal rule of allocation. *Journal of the American Statistical Assosiation*, 70, 365-369.

McLachlan, G. J. (1977). Estimating the linear discriminant function from initial samples containing a small number of unclassified observations. *Journal of the American Statistical Assosiation*, 72, 403-406.

McLachlan, G. J. and Basford, K. (1988). Mixture Models: Inference and Applications to Clustering. Marcel Dekker, New York.

McLachlan, G. J. and Ganesalingam, S. (1982). Updating a discriminant function on the basis of unclassified data. *Communications in Statistics - Simulation and Computation*, 11(6), 753-767.

McLachlan, G. J. and Peel, D. (2000). Finite mixture models. John Wiley & Sons, New York.

McLachlan, G. J. (2004). Discriminant analysis and statistical pattern recognition. John Wiley & Sons, New York.

Murray, G. J. and Titterington, D. M. (1978). Estimation problem with data from mixture. *Journal of the Royal Statistical Society. Series C*, 27, 325-334.

Okamoto, M. (1963). An asymptotic expansion for the distribution of the linear discriminant function. *The Annals of Mathematical Statistics*, 34(4), 1286-1301.

Razali, A. M. and Salih, A. A. (2009). Combining two Weibull distributions using a mixing parameter. *European Journal of Scientific Research*, 31(2), 296-305.

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

Sultan, K. S. and Al-Moisheer, A. S. (2013). Updating a nonlinear discriminant function estimated from a mixture of two inverse Weibull distributions. *Statistical Papers*, 54, 163–175.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). Statistical Analysis of Finitee Mixture Distribution. John Wiley & Sons, Chichester.

Torabi, H., Falahati-Naeini, N. and Montazeri, N. H. (2014). An extended generalized Lindley distribution and its applications to lifetime data. *Statistical Research and Training Center*, 11, 203-222.

Table 1: Estimated Biases, MSEs of the MLEs for the MLLD parameters for classified and mixed samples

Actual values of the parameters				Classified Samples						Mixed Samples					
				Bias			MSE			Bias			MSE		
p	θ_1	θ_2	n	$\tilde{p}$	$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\tilde{p}$	$\tilde{\theta}_1$	$\tilde{\theta}_2$	$\hat{p}$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{p}$	$\hat{\theta}_1$	$\hat{\theta}_2$
0.45	0.55	0.50	50	-0.0663	0.7326	0.3173	0.0049	0.5410	0.1030	-0.0289	0.0179	0.0206	0.2126	0.0758	0.0038
			100	-0.0155	0.5699	0.3109	0.0006	0.3273	0.0991	-0.0815	0.0206	0.0164	0.2011	0.0309	0.0020
0.45	0.70	0.50	50	-0.0699	0.9321	0.3158	0.0049	0.8687	0.0997	-0.3856	-0.1106	0.0798	0.1732	0.0167	0.0106
			100	-0.0200	0.7284	0.3002	0.0004	0.5306	0.0901	-0.3839	-0.1097	0.0769	0.1642	0.0144	0.0078
0.45	0.55	0.85	50	-0.0459	0.7089	0.5531	0.0087	0.5750	0.3633	-0.3913	0.1480	-0.1615	0.1707	0.0284	0.0327
			100	-0.0039	0.5467	0.5430	0.0010	0.3021	0.3057	-0.3988	0.1471	-0.1640	0.1667	0.0246	0.0299
0.45	0.70	0.85	50	-0.0096	0.7581	0.7469	0.0069	0.6713	0.8054	0.5434	0.0851	-0.0676	0.2958	0.0151	0.0124
			100	0.0018	0.6997	0.5702	0.0012	0.5166	0.3535	0.5434	0.0818	-0.0711	0.2953	0.0103	0.0086
0.75	0.55	0.50	50	-0.0290	0.2425	0.9791	0.0011	0.0597	0.9963	-0.6994	-0.0183	0.0374	0.4957	0.0043	0.0055
			100	-0.0285	0.2123	0.8165	0.0010	0.0455	0.6840	-0.6868	-0.0227	0.0394	0.4725	0.0029	0.0036
0.75	0.70	0.50	50	0.0002	0.2123	1.3038	0.0036	0.0631	2.0632	-0.6401	-0.0569	0.1438	0.4135	0.0085	0.0259
			100	0.0006	0.2062	1.2413	0.0019	0.0508	1.6955	-0.6232	-0.0591	0.1417	0.3905	0.0059	0.0225

Table 2 : Individual Probabilities of Misclassification from MLLD and relative biases to optimal

Actual values of the parameters				Classification Procedures									Relative Bias to Optimal	
				Mixtures			Classified			Optimal				
p_1	θ_1	θ_2	n	$\bar{e_{1m}}$	$\bar{e_{2m}}$	$\bar{e_m}$	$\bar{e_{1c}}$	$\bar{e_{2c}}$	$\bar{e_c}$	e_{1o}	e_{2o}	e_o	$\lvert B(\bar{e}_m)\rvert$	$\lvert B(\bar{e}_c)\rvert$
					(Std.dev)			(Std.dev)						
0.45	0.55	0.50	50	0.8755	0.1246	0.4625	1.1951	-0.1350	0.4635	1.1149	-0.0965	0.4487	0.0825	0.0196
				(0.3311)	(0.3310)	(0.1672)	(1.6658)	(0.1697)	(0.7554)				0.0308	0.0330
			100	0.9016	0.0984	0.4599	1.1740	-0.0876	0.4801				0.0742	0.2457
				(0.2990)	(0.2990)	(0.1510)	(0.1994)	(0.1654)	(0.1278)				0.0250	0.0700
0.45	0.70	0.50	50	0.8947	0.1053	0.4606	1.7421	-0.2194	0.6633	0.5088	0.3372	0.4144	0.2967	0.5339
				(0.3082)	(0.3082)	(0.1557)	(1.0330)	(0.0653)	(0.4662)				0.1115	0.6006
			100	0.9330	0.0670	0.4567	1.7134	-0.2355	0.6415				0.3323	1135.5
				(0.2522)	(0.2521)	(0.1273)	(0.0004)	(0.0002)	(0.0002)				0.1021	0.5480
0.45	0.55	0.85	50	0.9459	0.0542	0.4554	0.3135	0.6798	0.5150	0.6007	0.1880	0.3737	0.7086	0.4163
				(0.2283)	(0.2283)	(0.1153)	(0.5254)	(0.4427)	(0.3394)				0.2186	0.3781
			100	0.9541	0.0460	0.4546	0.0688	0.8981	0.5249				0.7561	4.4602
				(0.2120)	(0.2120)	(0.1070)	(0.0238)	(0.0585)	(0.0339)				0.2165	0.4046
0.45	0.70	0.85	50	0.9438	0.0562	0.4556	0.7816	0.2930	0.5129	0.7735	0.1460	0.4284	0.2343	0.1948
				(0.2300)	(0.2300)	(0.1161)	(0.7524)	(0.4928)	(0.4337)				0.0635	0.1972
			100	0.9467	0.0533	0.4553	0.6509	0.3596	0.4907				0.2376	0.1728
				(0.2241)	(0.2241)	(0.1132)	(0.5337)	(0.4890)	(0.3606)				0.0628	0.1454
0.75	0.55	0.50	50	0.0503	0.9497	0.2751	0.1376	0.6969	0.2774	0.00001	0.99997	0.2500	0.1843	0.7507
				(0.2180)	(0.2180)	(0.1362)	(0.0424)	(0.0714)	(0.0365)				0.1004	0.1096
			100	0.0006	0.9994	0.2503	0.1203	0.7516	0.2781				1.500	2.0362
				(0.0003)	(0.0003)	(0.0002)	(0.0076)	(0.0504)	(0.0138)				0.0012	0.1124
0.75	0.70	0.50	50	0.0201	0.9799	0.2600	0.1689	0.5701	0.2692	0.0138	0.9389	0.2451	0.1703	0.0377
				(0.1401)	(0.1401)	(0.0875)	(0.8478)	(0.2517)	(0.6389)				0.0608	0.0983
			100	0.0081	0.9919	0.2540	0.1881	0.5932	0.2894				0.1598	0.6110
				(0.0891)	(0.0891)	(0.0557)	(0.0776)	(0.1729)	(0.0725)				0.0363	0.1807

Table 3 : Comparison the updated LD function with the classified and mixed discriminant functions

Actual values of the parameters				m	Classification Procedures				
					Mixed	Classified	updating process		Optimal
n	p	θ_1	θ_2		$\bar{e}_m$	$\bar{e}_c$	$\bar{e}_{MI}$	$\bar{e}_{MII}$	$\bar{e}_o$
40	0.45	0.55	0.50	50	0.4632701	0.581802	0.4505838	0.4469334	0.4486678
				100	0.4632675	0.1085137	0.4507917	0.4480607	
	0.45	0.70	0.50	50	0.4600274	0.6450514	0.4503814	0.4462947	0.4144267
				100	0.4599350	0.6377379	0.4498390	0.4483494	
40	0.75	0.55	0.50	50	0.2856652	0.2523172	0.2605125	0.2490420	0.2499992
				100	0.2827162	0.2522563	0.2574985	0.2489723	
	0.75	0.70	0.50	50	0.2600314	0.2642143	0.2598034	0.2486379	0.2450835
				100	0.2540382	0.2665034	0.2548139	0.2488766	

Table 4 : Efficiency of the updated LD function

Actual values of the parameters				m	ε_c	ε_m	ε
n	p	θ_1	θ_2				
40	0.45	0.55	0.50	50	1.0278	1.2877	-1.1047
				100	0.9920	1.2189	-3.4984
	0.45	0.70	0.50	50	1.0210	1.4237	1.1282
				100	1.0079	1.1475	1.0439
40	0.75	0.55	0.50	50	-0.3996	1.4560	-10.9834
				100	-0.6265	1.3381	-7.3029
	0.75	0.70	0.50	50	3.5313	49.9715	4.1413
				100	1.5079	-6.6541	2.5653