Estimation of a discriminant function from a mixture of two inverse Weibull distributions

K. S. Sultan & A. S. Al-Moisheer

Department of Statistics and Operations Research, College of Science, King Saud University, PO Box 2455, Riyadh, 11451, Saudi Arabia

Published online: 21 Sep 2011.


To link to this article: http://dx.doi.org/10.1080/00949655.2011.614245

PLEASE SCROLL DOWN FOR ARTICLE
Estimation of a discriminant function from a mixture of two inverse Weibull distributions

K.S. Sultan* and A.S. Al-Moisheer†

Department of Statistics and Operations Research, College of Science, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia

(Received 7 February 2011; final version received 10 August 2011)

The classification of a random variable based on a mixture can be meaningfully discussed only if the class of all finite mixtures is identifiable. In this paper, we find the maximum-likelihood estimates of the parameters of the mixture of two inverse Weibull distributions by using classified and unclassified observations. Next, we estimate the nonlinear discriminant function of the underlying model. Also, we calculate the total probabilities of misclassification as well as the percentage bias. In addition, we investigate the performance of all results through a series of simulation experiments by means of relative efficiencies. Finally, we analyse some simulated and real data sets through the findings of the paper.

Keywords: finite mixtures; maximum-likelihood estimation; EM algorithm; discriminant function; bias; mean-square error; relative efficiency and Monte Carlo simulations

1. Introduction

Mixture models play an important role in many applicable fields, such as medicine, psychology, cluster analysis, life testing and reliability analysis and so on. Mixture models have been considered extensively by many authors, for an excellent survey of estimation techniques, discussion and applications, see [1–6]. AL-Hussaini and Sultan [7] have reviewed some properties and estimation techniques of finite mixtures of some life-time models. Recently, Sultan et al. [8,9] have discussed the properties of the mixture of two inverse Weibull distribution (MTIWD) and they have proved that the MTIWD is identifiable. Also, they have investigated the hypotheses testing for the number of components. In this paper, we derive and estimate the nonlinear discriminant function from MTIWD with a common shape parameter based on different schemes of sampling.

MTIWD with common shape parameter has its pdf as:

\[ f(t; \Theta) = \sum_{i=1}^{2} p_i f_i(t; \Theta_i), \quad p_1 + p_2 = 1, \]  

*Corresponding author. Email: ksultan@ksu.edu.sa
†Present address: Department of Mathematics, College of Science, Al-Jouf University, PO Box 2014, Al-Jouf, Sakaka Kingdom of Saudi Arabia.

© 2013 Taylor & Francis
where $\Theta = (p_1, \alpha_1, \alpha_2, \beta)$, $\Theta_i = (\alpha_i, \beta), i = 1, 2,$ and $f_i(t; \Theta_i)$, is the pdf of the $i$th component

$$f_i(t; \Theta_i) = \beta t^{-\beta} \exp[-(\alpha_i t)^{-\beta} - \beta \ln \alpha_i], \quad t \geq 0, \; \alpha_i, \beta > 0, \; i = 1, 2. \quad (2)$$

The cdf of MTIWD is given by

$$F(t; \Theta) = \sum_{i=1}^{2} p_i F_i(t; \Theta_i), \quad (3)$$

where $F_i(t; \Theta_i)$, the cdf of the $i$th component is given by

$$F_i(t; \Theta_i) = \exp[-(\alpha_i t)^{-\beta}], \quad t \geq 0, \; \alpha_i, \beta > 0, \; i = 1, 2. \quad (4)$$

Unclassified data have been studied in the context of estimating mixtures and discriminant functions; however, there have been few practical applications. Choi [10] has proposed the use of the unclassified observations for the classification problem. Fukunaga and Kessel [11] and Moore et al. [12] have estimated the probability of misclassification for a discriminant rule. McLachlan [13,14] has discussed the use of unclassified observations for a special case of equal prior probabilities. The estimate he suggested asymptotically gives Fisher’s linear discriminant function, but it is not useful for estimating posterior probabilities. This problem can be corrected by a suitable scaling of the estimate [15]. Amoh [16] has estimated the discriminant function from a mixture of two inverse Gaussian distributions. Mahmoud and Moustafa [17] have estimated a discriminant function from a mixture of two gamma distributions when the sample size is small. Ahmad [18] 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. To illustrate the differences that can occur in practice with using either the classification or mixture approaches, Ganesalingam and McLachlan [24] have investigated a case study where the two approaches were employed to classify some medical data. There are some other references for using mixture models in the survival analysis are [25–28].

The remainder of this paper has the following organization. In Section 2, we calculate the maximum-likelihood estimates (MLEs) of the unknown parameters of MTIWD (with common shape parameter). In Section 3, we derive the optimal nonlinear discriminant function from MTIWD. Similarly, in Section 4, we derive the nonlinear discriminant function from MTIWD based on the mixed and classified samples. In Section 5, we discuss the error rate of the misclassification. In Section 6, we carry out some simulation studies to illustrate the performance of the proposed discriminant techniques. In Section 7, we analyse a set of real data. Finally, in Section 8, we draw our concluding remarks.

2. MLEs

Suppose that $t_1, t_2, \ldots, t_n$ is a random sample of size $n$ drawn from MTIWD, given by Equation (1).

The likelihood function generated by the random sample is

$$L(\Theta) = \beta^n \prod_{j=1}^{n} t_j^{-(\beta+1)} Q_j, \quad (5)$$
where

\[ Q_j = \sum_{i=1}^{2} p_i g_1(t_j; \Theta_i) \]  \hspace{1cm} (6)

and

\[ g_1(t_j; \Theta_i) = \exp\{-[\beta \ln \alpha_i + (\alpha_i t_j)^{-\beta}]\}. \]  \hspace{1cm} (7)

By differentiating the log-likelihood function with respect to the unknown parameters, we get

\[ \sum_{j=1}^{n} g_1(t_j; \Theta_1) - g_2(t_j; \Theta_2) \]

\[ Q_j = 0, \]

and

\[ \sum_{j=1}^{n} p_i \beta \alpha_i^{-1} g_1(t_j; \Theta_i)\left[\alpha_i^{-\beta} t_j^{-\beta} - 1\right] \]

\[ Q_j = 0, \quad i = 1, 2, \]  \hspace{1cm} (8)

Now, we define \( W_{ij} \) as the probability that the observation \( t_j \) arises from the \( i \)th component. Since the components are assumed to exist in fixed proportions, \( W_{ij} \) can be expressed as follows:

\[ W_{1j} = \frac{p_1 g_1(t_j; \Theta_1)}{Q_j} \quad \text{and} \quad W_{2j} = 1 - W_{1j}, \]  \hspace{1cm} (9)

that is

\[ W_{ij} = \{1 + \exp[\alpha - \beta z]^{-1}, \quad z = t^{-\beta}, \]  \hspace{1cm} (10)

where

\[ a = \ln \frac{p_2}{p_1} - \beta \ln \frac{\alpha_2}{\alpha_1}, \quad \text{and} \quad b = \alpha_2^{-\beta} - \alpha_1^{-\beta}. \]  \hspace{1cm} (11)

Substitution of the MLEs \( \hat{\alpha} \) and \( \hat{\beta} \) of \( a \) and \( b \) in Equation (9), we get the MLE of \( W_{ij} \). The MLEs of \( p_1, \alpha_1, \alpha_2 \) and \( \beta \) can be obtained by solving the system of equations in Equation (8) as follows:

\[ \hat{p}_1 = \frac{1}{n} \sum_{j=1}^{n} \hat{W}_{1j}, \]

\[ \hat{\alpha}_i = \left( \frac{\sum_{j=1}^{n} \hat{W}_{ij}}{\sum_{j=1}^{n} \hat{W}_{ij} t_j^{-\beta}} \right)^{-1/\beta}, \quad i = 1, 2, \]  \hspace{1cm} (12)

\[ \hat{\beta}^{-1} = \frac{1}{n} \left[ \sum_{j=1}^{n} \ln t_j - \sum_{j=1}^{n} \sum_{i=1}^{2} \hat{W}_{ij} (\alpha_i t_j)^{-\beta} \ln(\alpha_i t_j) - \log(\hat{\alpha}_i) \right]. \]

One can solve the system of Equation (12) iteratively by choosing the initial values \( p_1(0), \alpha_1(0), \alpha_2(0) \) and \( \beta(0) \). Dempster et al. [29] have shown that the expectation maximization (EM) algorithm ensures the convergence of this procedure to a local maximum, irrespective of the starting point. In this paper, we solve the nonlinear system of Equation (8) directly by using the subroutine DNEQNJ from the International Statistical and Mathematical Libraries (IMSL).
The logarithm of the likelihood function (5) can be used to construct the Fisher information matrix \( I(\Theta) \) with the elements

\[
I_{ij}(\Theta) = E \left( -\frac{\partial^2 \ln L(\Theta)}{\partial \theta_i \partial \theta_j} \right), \quad i, j = 1, 2, 3, 4,
\]

where \( \Theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (p_1, \alpha_1, \alpha_2, \beta) \). As we can see from Table 1, the MLEs of the component of \( \Theta \) are consistent estimators. Also, \( \sqrt{n}(\hat{\Theta} - \Theta) \) is asymptotically normal with mean vector 0 and the variance covariance matrix \( I_{ij}^{-1}(\hat{\Theta}) \).

3. Optimal discriminant function

Consider the two populations \( \Pi_1 \) and \( \Pi_2 \) corresponding to the densities \( f_i(t) \), \( i = 1, 2 \), as given by Equation (2) with known scale parameters \( \alpha_1 \) and \( \alpha_2 \), respectively, and a known common shape parameter \( \beta \). Then, the nonlinear discriminant function in this case is given by

\[
NLD_o(t) = a - b z, \quad z = t^{-\beta}.
\] (13)

The probability that an individual \( t \) of unknown origin has come from \( \Pi_1 \) is given by

\[
\Pr(t \in \Pi_1) = \left[ 1 + \exp[NLD_o(t)] \right]^{-1},
\]

which is called the posterior probability [30]. Then, we may classify \( t \) in \( \Pi_1 \), if \( NLD_o(t) < 0 \), and in \( \Pi_2 \), if \( NLD_o(t) \geq 0 \). If all parameters of the populations \( \Pi_1 \) and \( \Pi_2 \) are known, then we have an optimal nonlinear discriminant function \( NLD_o(t) \) in Equation (13), where \( a \) and \( b \) are given in Equation (11).

4. Estimated discriminant function

Usually, the parameters of the populations are unknown. 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 two types of data.

1. **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).
2. **Mixed sample.** The case where each observation in the mixture is unclassified, this called mixed (m).

4.1. **Classified sample (c)**

If there are \( n_i \) initial observations available from \( \Pi_i \), with \( n = n_1 + n_2 \), then the nonlinear discriminant function in this case is given by

\[
NLD_c(t) = \tilde{a} - \tilde{b} \tilde{z}, \quad \tilde{z} = t^{-\hat{\beta}},
\] (14)

where \( \tilde{a} = \ln(\tilde{p}_2/\tilde{p}_1) - \hat{\beta} \ln(\tilde{\alpha}_2/\tilde{\alpha}_1) \) and \( \tilde{b} = \tilde{\alpha}_2^{-\hat{\beta}} - \tilde{\alpha}_1^{-\hat{\beta}} \).
Table 1. Estimated biases, mean-square errors and traces at \( p = 0.25 \) and \( \alpha_2 = 3.0 \).

<table>
<thead>
<tr>
<th>Actual values of the parameters</th>
<th>Estimated bias (MSE) and (Tr)</th>
<th>Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n ) ( \alpha_1 ) ( \beta )</td>
<td>( \hat{\alpha}_1 ) ( \hat{\alpha}_2 ) ( \beta )</td>
<td>( \hat{\alpha}_1 ) ( \hat{\alpha}_2 ) ( \beta )</td>
</tr>
<tr>
<td>50 0.25 \ 3</td>
<td>0.0047 (0.0041) \ [0.1295]</td>
<td>0.0021 (0.0007) \ -0.0255 (0.0297) \ 0.0947 (0.1046)</td>
</tr>
<tr>
<td>50 0.5 \ 3</td>
<td>-0.0039 (0.0041) \ [0.1500]</td>
<td>0.0063 (0.0038) \ 0.0141 (0.0291) \ 0.1022 (0.1237)</td>
</tr>
<tr>
<td>50 0.25 \ 4</td>
<td>0.0061 (0.0042) \ [0.1751]</td>
<td>-0.0021 (0.0003) \ -0.0193 (0.0162) \ 0.1225 (0.1698)</td>
</tr>
<tr>
<td>50 0.5 \ 4</td>
<td>-0.0014 (0.0043) \ [0.3181]</td>
<td>0.0043 (0.0021) \ 0.0129 (0.0167) \ 0.2455 (0.3555)</td>
</tr>
<tr>
<td>100 0.25 \ 3</td>
<td>0.0047 (0.0019) \ [0.0607]</td>
<td>-0.0026 (0.0004) \ -0.0034 (0.0164) \ 0.0153 (0.0422)</td>
</tr>
<tr>
<td>100 0.5 \ 3</td>
<td>-0.0011 (0.0023) \ [0.0758]</td>
<td>0.0012 (0.0018) \ 0.0004 (0.0167) \ 0.0681 (0.0596)</td>
</tr>
<tr>
<td>100 0.25 \ 4</td>
<td>0.0039 (0.0019) \ [0.0800]</td>
<td>-0.0029 (0.0002) \ -0.0052 (0.0090) \ 0.0014 (0.0688)</td>
</tr>
<tr>
<td>100 0.5 \ 4</td>
<td>0.0049 (0.0019) \ [0.1227]</td>
<td>0.0015 (0.0008) \ -0.0007 (0.0083) \ 0.0959 (0.1210)</td>
</tr>
<tr>
<td>50 0.25 \ 3</td>
<td>0.0092 (0.0058) \ [0.2221]</td>
<td>0.0203 (0.0004) \ 0.0201 (0.0423) \ 0.1731 (0.2041)</td>
</tr>
<tr>
<td>50 0.5 \ 3</td>
<td>0.0088 (0.0056) \ [0.2425]</td>
<td>0.0043 (0.0020) \ 0.0188 (0.0437) \ 0.1907 (0.2281)</td>
</tr>
<tr>
<td>50 0.25 \ 4</td>
<td>0.0088 (0.0057) \ [0.2709]</td>
<td>0.0009 (0.0002) \ 0.0086 (0.0217) \ 0.1601 (0.2690)</td>
</tr>
<tr>
<td>50 0.5 \ 4</td>
<td>0.0093 (0.0058) \ [0.3305]</td>
<td>0.0030 (0.0009) \ 0.0127 (0.0230) \ 0.2093 (0.3448)</td>
</tr>
<tr>
<td>100 0.25 \ 3</td>
<td>0.0059 (0.0025) \ [0.0727]</td>
<td>0.0020 (0.0002) \ -0.0102 (0.0182) \ 0.0867 (0.0594)</td>
</tr>
<tr>
<td>100 0.5 \ 3</td>
<td>0.0070 (0.0024) \ [0.0758]</td>
<td>0.0033 (0.0008) \ -0.0093 (0.0168) \ 0.0823 (0.0628)</td>
</tr>
<tr>
<td>100 0.25 \ 4</td>
<td>0.0052 (0.0025) \ [0.0895]</td>
<td>0.0008 (0.0001) \ -0.0097 (0.0098) \ 0.1048 (0.0883)</td>
</tr>
<tr>
<td>100 0.5 \ 4</td>
<td>0.0061 (0.0025) \ [0.1155]</td>
<td>0.0029 (0.0004) \ -0.0087 (0.0099) \ 0.1074 (0.1143)</td>
</tr>
</tbody>
</table>
In this case, \( \tilde{\alpha}_1, \tilde{\alpha}_2 \) and \( \tilde{\beta} \) are the MLEs of the parameters from the classified samples defined by

\[
\tilde{\alpha}_i = \left( \frac{n_i}{\sum_{j=1}^{n_i} (1/t^\tilde{\beta}_{ij})} \right)^{-1/\tilde{\beta}}, \quad i = 1, 2, \tag{15}
\]

\[
\tilde{\beta} = n \left\{ \sum_{i=1}^{2} \left[ n_i \sum_{j=1}^{n_i} \log t_{ij} + n_i \sum_{j=1}^{n_i} \left[ \log (1/t_{ij}) / t^\tilde{\beta}_{ij} \right] \right] \right\}^{-1}, \tag{16}
\]

\[
\tilde{p}_1 = \frac{n_1}{n} \quad \text{and} \quad \tilde{p}_2 = 1 - \tilde{p}_1. \tag{17}
\]

Note that the unknown mixing proportion parameter can be obtained explicitly from Equation (17). For solving Equation (16), we use the routine DZBREN from the IMSL and substitute the estimate of \( \tilde{\beta} \) in Equation (15).

### 4.2. Mixed sample \((m)\)

Suppose that all the initial observations are known to be from the mixture of \( \Pi_1 \) and \( \Pi_2 \). The nonlinear discriminant function for this mixed model is given by

\[
\text{NLD}_m(t) = \hat{a} - \hat{b} \hat{z}, \quad \hat{z} = t^{-\hat{\beta}}, \tag{18}
\]

where \( \hat{a} = \ln(\hat{p}_2 / \hat{p}_1) - \hat{\beta} \ln(\tilde{\alpha}_2 / \tilde{\alpha}_1) \) and \( \hat{b} = \tilde{\alpha}_2^{-\hat{\beta}} - \tilde{\alpha}_1^{-\hat{\beta}} \).

In this case, \( \tilde{p}_1, \tilde{\alpha}_1, \tilde{\alpha}_2 \) and \( \tilde{\beta} \) are the MLEs of the parameters which can be obtained by solving the system of Equation (8).

### 5. Error rates

We classify \( t \) in \( \Pi_1 \), if \( \text{NLD}_k(t) < 0 \), for \( k = o, c, m \). The probabilities of misclassifying an observation from \( \Pi_i, i = 1, 2 \) by the nonlinear discriminant function \( \text{NLD}_k(t) \) is \( e_{1k} = \Pr(a_k - b_k t^{-\beta} > 0 \mid \Pi_1), k = o, c, m \). Putting \( \gamma_k = [a_k \setminus b_k]^{-1/\beta} \), we have

\[
e_{1k} = \begin{cases} 
F(\gamma_k, \alpha_1, \beta), & \alpha_1 \geq \alpha_2, \\
1 - F(\gamma_k, \alpha_1, \beta), & \alpha_1 < \alpha_2,
\end{cases} \tag{19}
\]

where \( F(\cdot, \alpha, \beta) \) is the cdf of the inverse Weibull distribution given in Equation (4). Similarly, \( e_{2k} \) is given by

\[
e_{2k} = \begin{cases} 
1 - F(\gamma_k, \alpha_2, \beta), & \alpha_1 \geq \alpha_2, \\
F(\gamma_k, \alpha_2, \beta), & \alpha_1 < \alpha_2.
\end{cases} \tag{20}
\]

The overall error rates weighted by the true mixing proportion, \( e_k \) is given by

\[
e_k = p_1 e_{1k} + p_2 e_{2k}. \tag{21}
\]

In order to estimate the values of \( e_{1k}, e_{2k} \) and \( e_k \), one can just replace the parameters in Equations (19) and (20) by their estimates.
6. Simulation experiments

In order to show the usefulness of the proposed techniques in this paper, we carry out some simulation experiments. Our main dual purpose of this study is

- studying the behaviour of the MLEs of the mixed and classified samples,
- investigating the performance of NLD\(_m(t)\) compared with NLD\(_c(t)\) and NLD\(_o(t)\), via the error of misclassification criterion.

Eight different combinations of the parameters were taken \(\alpha_1 = 0.25, 0.5, \alpha_2 = 3.0\) and \(\beta = 3.0, 4.0\), with different values of the mixing proportion \(p_1 = 0.25, 0.5\). A mixed random sample of sizes \(n = 50\) and \(100\) was generated for each combination of the parameters as follows.

1. Generate two uniform variates \(u_1\) and \(u_2\) from the Fortran numerical library (IMSL) using the routine DRNUN.
2. If \(u_1 < p_1\), then use \(u_2\) to generate a random variate \(t\) from the MTIWD as \(t = F_{1}^{-1}(u_2)\).
3. If \(u_1 \geq p_1\), then use \(u_2\) to generate a random variate \(t\) from the MTIWD as \(t = F_{2}^{-1}(u_2)\).

Repeating the above steps \(n\) times gives \(n_1\) observations identifiable from the first component, \(n_2\) from the second component yielding a mixed sample of size \(n = n_1 + n_2\).

In Table 1, we calculate the bias and mean-squared error (MSE) and trace efficiency (Tr) of the estimates from the mixed samples and compare them with those from the classified samples based on 100 repetitions. From Table 1, we observe that, the estimates are consistent and improve when the sample size increases as well as when the mixing proportion increases. Also, we note that when we compare the two sample schemes, none of the two schemes performs better than the other in all cases.

In order to investigate the performance of NLD\(_m(t)\), relative to NLD\(_c(t)\) and NLD\(_o(t)\), we perform a series of simulation experiments as displayed in Table 2. Table 2 shows the conditional probabilities of misclassification of the two discrimination procedures \(\hat{e}_{ik}\) \((i = 1, 2\) and \(k = m, c)\) as estimates of the optimal probabilities of misclassification \(e_{im}\) when \(n = 50, 100\). The standard deviation of the conditional probabilities of misclassification are shown in parentheses. We find that, generally, at \(p_1 = 0.25\), \(\hat{e}_{2k}\) \((k = m, c)\) are closer to the corresponding optimal values than \(\hat{e}_{1k}\). Conversely, at \(p_1 = 0.5\), \(\hat{e}_{1k}\) \((k = m, c)\) are closer to the corresponding optimal values than \(\hat{e}_{2k}\). Considering these conditional probabilities of misclassification as estimates of the optimal probabilities of misclassification are poor when \(d = |\alpha_1 - \alpha_2|\) is small. This is not surprising since when the components of the mixture population are not well separated, it is difficult to discriminate between them. In addition, when the common shape parameters increase, the optimal probabilities of misclassification also increase. When \(\beta\) increases and \(n\) changes from \(50\) to \(100\), the estimates \(\hat{e}_{ik}\) quite reasonably, especially the estimates \(\hat{e}_{2k}\) improve. As expected for every combination parameters, the standard deviation of \(e_{ic}\) is smaller than that of \(e_{im}\) since more information is known in the former case.

Table 3 displays the total probabilities of misclassification, with the standard deviations of \(e_{m}\) and \(e_{c}\), in parentheses; and it shows the standardized bias. The first entry in each cell under \(B(\hat{e}_{ik})\) is the value of the absolute bias from \(e_{oi}\), standardized by the standard deviation of \(e_{oi}\) \((SD_{oi})\) which is given by \(|\hat{e}_{ik} - e_{oi}|/SD_{ik}\), \(k = c, m\). The second is the value of the ratio of the absolute bias to \(e_{oi}\), which is given by \(|\hat{e}_{ik} - e_{oi}|/e_{oi}\), \(k = c, m\). \(B\) is the value of the ratio of the bias of \(\hat{e}_{m}\) from \(\hat{e}_{c}\), i.e. \(B = |\hat{e}_{m} - \hat{e}_{c}|/\hat{e}_{c}\). From Table 3, we see that if the distance between the scale and common shape parameters is increased and mixing proportion is decreased, then the standard deviation for \(e_{c}\) appears to decrease. Also, from the trace efficiency, we obtain that the estimates are quite reasonable and \(\hat{e}_{c}\) behaves consistently better than \(\hat{e}_{m}\) and closer to the corresponding optimal one. In addition, the standard deviations for \(e_{c}\) are smaller than the standard deviations for \(e_{m}\).
Table 2. Individual probabilities of misclassification.

<table>
<thead>
<tr>
<th>Actual values of the parameters</th>
<th>Classification procedures</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$ $p_1$ $\alpha_1$ $\alpha_2$ $\beta$</td>
<td>$e_{im}$</td>
<td>$e_{zm}$</td>
<td>$e_{ic}$</td>
<td>$e_{zc}$</td>
<td>$e_{io}$</td>
<td>$e_{zo}$</td>
<td></td>
</tr>
<tr>
<td>50 0.25 0.25 3 3</td>
<td>0.9980 (0.0020)</td>
<td>0.9962 (0.0029)</td>
<td>0.9981 (0.0017)</td>
<td>0.9964 (0.0025)</td>
<td>0.9983</td>
<td>0.9963</td>
<td></td>
</tr>
<tr>
<td>50 0.25 0.5 3 3</td>
<td>0.9825 (0.0187)</td>
<td>0.9806 (0.0114)</td>
<td>0.9862 (0.0099)</td>
<td>0.9816 (0.0095)</td>
<td>0.9864</td>
<td>0.9803</td>
<td></td>
</tr>
<tr>
<td>50 0.25 0.25 3 4</td>
<td>0.9998 (0.0002)</td>
<td>0.9995 (0.0005)</td>
<td>0.9998 (0.0002)</td>
<td>0.9995 (0.0004)</td>
<td>0.9999</td>
<td>0.9996</td>
<td></td>
</tr>
<tr>
<td>50 0.25 0.5 3 4</td>
<td>0.9976 (0.0022)</td>
<td>0.9956 (0.0029)</td>
<td>0.9975 (0.0021)</td>
<td>0.9955 (0.0031)</td>
<td>0.9977</td>
<td>0.9953</td>
<td></td>
</tr>
<tr>
<td>100 0.25 0.25 3 3</td>
<td>0.9981 (0.0014)</td>
<td>0.9961 (0.0020)</td>
<td>0.9981 (0.0012)</td>
<td>0.9961 (0.0019)</td>
<td>0.9983</td>
<td>0.9963</td>
<td></td>
</tr>
<tr>
<td>100 0.25 0.5 3 3</td>
<td>0.9859 (0.0078)</td>
<td>0.9810 (0.0074)</td>
<td>0.9854 (0.0077)</td>
<td>0.9801 (0.0073)</td>
<td>0.9864</td>
<td>0.9803</td>
<td></td>
</tr>
<tr>
<td>100 0.25 0.25 3 4</td>
<td>0.9998 (0.0001)</td>
<td>0.9995 (0.0003)</td>
<td>0.9998 (0.0001)</td>
<td>0.9995 (0.0003)</td>
<td>0.9999</td>
<td>0.9996</td>
<td></td>
</tr>
<tr>
<td>100 0.25 0.5 3 4</td>
<td>0.9976 (0.0017)</td>
<td>0.9953 (0.0025)</td>
<td>0.9974 (0.0016)</td>
<td>0.9951 (0.0023)</td>
<td>0.9977</td>
<td>0.9953</td>
<td></td>
</tr>
<tr>
<td>50 0.5 0.25 3 3</td>
<td>0.9994 (0.0060)</td>
<td>0.9958 (0.0033)</td>
<td>0.9994 (0.0005)</td>
<td>0.9957 (0.0029)</td>
<td>0.9994</td>
<td>0.9957</td>
<td></td>
</tr>
<tr>
<td>50 0.5 0.5 3 3</td>
<td>0.9957 (0.0032)</td>
<td>0.9770 (0.0140)</td>
<td>0.9953 (0.0032)</td>
<td>0.9764 (0.0118)</td>
<td>0.9955</td>
<td>0.9753</td>
<td></td>
</tr>
<tr>
<td>50 0.5 0.25 3 4</td>
<td>0.9999 (0.0001)</td>
<td>0.9995 (0.0005)</td>
<td>0.9999 (0.0001)</td>
<td>0.9995 (0.0005)</td>
<td>1.0000</td>
<td>0.9995</td>
<td></td>
</tr>
<tr>
<td>50 0.5 0.5 3 4</td>
<td>0.9992 (0.0007)</td>
<td>0.9946 (0.0039)</td>
<td>0.9992 (0.0007)</td>
<td>0.9946 (0.0035)</td>
<td>0.9992</td>
<td>0.9945</td>
<td></td>
</tr>
<tr>
<td>100 0.5 0.25 3 3</td>
<td>0.9994 (0.0004)</td>
<td>0.9958 (0.0022)</td>
<td>0.9993 (0.0004)</td>
<td>0.9953 (0.0024)</td>
<td>0.9994</td>
<td>0.9957</td>
<td></td>
</tr>
<tr>
<td>100 0.5 0.5 3 3</td>
<td>0.9957 (0.0020)</td>
<td>0.9760 (0.0088)</td>
<td>0.9951 (0.0023)</td>
<td>0.9744 (0.0098)</td>
<td>0.9955</td>
<td>0.9753</td>
<td></td>
</tr>
<tr>
<td>100 0.5 0.25 3 4</td>
<td>0.9999 (0.0001)</td>
<td>0.9995 (0.0003)</td>
<td>0.9999 (0.0001)</td>
<td>0.9994 (0.0004)</td>
<td>1.0000</td>
<td>0.9995</td>
<td></td>
</tr>
<tr>
<td>100 0.5 0.5 3 4</td>
<td>0.9992 (0.0006)</td>
<td>0.9945 (0.0029)</td>
<td>0.9991 (0.0005)</td>
<td>0.9941 (0.0029)</td>
<td>0.9992</td>
<td>0.9945</td>
<td></td>
</tr>
</tbody>
</table>

From the last column in Table 3, we see that the mixture discrimination procedure relative to the classified performs poorly for small $\beta$ and $p_1 = 0.25$ and $n = 50$. As $\beta$ increases, the performance improves. Also, in the last three columns in Table 3, there are some cases where the relative bias to the optimal and classified is equal, which means that the estimates in the two methods are satisfactory.

Generally, the performance when $p_1 = 0.50$ is superior to the performance when $p_1 = 0.25$. When the sample size increases from $n = 50$ to 100, all the estimates for the combinations of parameters improve. The standard deviations are smaller, as expected, the bias are considerably reduced. The performance of the mixture discrimination procedure is quite well compared with the completely classified procedure in terms of total probability.

6.1. Illustrative example

In this example, we generate a sample of size 50 from MTIWD when $\Theta = (0.5, 0.5, 3.0, 4.0)$ as follows:

0.2830, 1.9498, 0.6168, 0.2898, 0.4701, 2.1599, 0.8493, 0.7084, 0.2719, 1.6059, 2.7528, 0.5613, 3.1907, 3.1489, 3.3101, 3.3418, 2.0476, 2.3872, 1.6750, 3.0929, 0.4822, 0.2475, 2.5571, 0.2865, 1.4391, 1.6022, 0.3826, 0.3001, 1.8647, 0.2822, 3.0974, 0.3035, 0.8712, 0.4030, 1.7564, 0.3767, 0.2610, 1.7605, 0.2615, 0.3550, 0.3763, 0.3863, 1.8954, 1.8184, 0.2733, 0.3283, 0.3085, 2.0152, 0.2713, 2.0070.

We apply the nonlinear discriminant function $NLD_m(t)$ given in Equation (18) for the above mixed simulated data. The MLEs of the parameters are

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$\hat{p}_1$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLEs</td>
<td>0.4586</td>
<td>0.5113</td>
<td>3.1505</td>
<td>4.1232</td>
</tr>
</tbody>
</table>

From the above table, we see that the MLEs perform very well (Figure 1). Next, we designate the $t$ in $\Pi_1$, if $NLD_m(t) < 0$, and in $\Pi_2$, if $NLD_m(t) \geq 0$. By using the discriminant functions for
Table 3. Probabilities of misclassification and percentage biases.

<table>
<thead>
<tr>
<th>Actual values of the parameters</th>
<th>Classification procedures</th>
<th>Relative bias to</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mixtures</td>
<td>Classified</td>
<td>Optimal</td>
<td>Optimal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{e}_m$</td>
<td>$\bar{e}_c$</td>
<td>$e_o$</td>
<td>$B(\bar{e}_m)$</td>
</tr>
<tr>
<td>$n$</td>
<td>$p_1$</td>
<td>$\alpha_1$</td>
<td>$\alpha_2$</td>
<td>$\beta$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.25</td>
<td>0.25</td>
<td>3</td>
<td>3</td>
<td>0.9967 (0.0026)</td>
</tr>
<tr>
<td>50</td>
<td>0.25</td>
<td>0.5</td>
<td>3</td>
<td>3</td>
<td>0.9816 (0.0116)</td>
</tr>
<tr>
<td>50</td>
<td>0.25</td>
<td>0.25</td>
<td>3</td>
<td>4</td>
<td>0.9996 (0.0004)</td>
</tr>
<tr>
<td>50</td>
<td>0.25</td>
<td>0.5</td>
<td>3</td>
<td>4</td>
<td>0.9962 (0.0026)</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.25</td>
<td>3</td>
<td>3</td>
<td>0.9966 (0.0018)</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.5</td>
<td>3</td>
<td>3</td>
<td>0.9824 (0.0071)</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.25</td>
<td>3</td>
<td>4</td>
<td>0.9996 (0.0003)</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.5</td>
<td>3</td>
<td>4</td>
<td>0.9959 (0.0022)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.25</td>
<td>3</td>
<td>3</td>
<td>0.9977 (0.0019)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>3</td>
<td>0.9868 (0.0081)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.25</td>
<td>3</td>
<td>4</td>
<td>0.9997 (0.0003)</td>
</tr>
<tr>
<td>50</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>4</td>
<td>0.9970 (0.0023)</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.25</td>
<td>3</td>
<td>3</td>
<td>0.9976 (0.0013)</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>3</td>
<td>0.9861 (0.0051)</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.25</td>
<td>3</td>
<td>4</td>
<td>0.9997 (0.0002)</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.5</td>
<td>3</td>
<td>4</td>
<td>0.9969 (0.0017)</td>
</tr>
</tbody>
</table>

Figure 1. Histogram of the simulated data.
mixed sample in Equation (18) and classifying the observations in the mixed samples one by one into either $\Pi_1$ or $\Pi_2$. The 20 out of 50 are classified into $\Pi_1$ with the probability of 0.4 and 30 out of 50 are classified into $\Pi_2$ with the probability of 0.6.

7. Data analysis

The following maintenance data were reported on active repair times (hours) for an airborne communication transceiver [31, p. 156]:

0.2, 0.3, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5.

We show a rough indication of the goodness of fit for our model by plotting the superimposed for the data shows that the MTIWD is a good fit in Figure 2. In addition, to confirm our results, we use the Kolmogorov–Smirnov test and found that the maximum distance between the data and the fitted of the MTIWD is 0.0853 which gives a good fit at 10% level of significance. Next, we use the method of maximum likelihood to estimate the unknown parameters as given below:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{p}_1$</th>
<th>$\hat{\alpha}_1$</th>
<th>$\hat{\alpha}_2$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE</td>
<td>0.8512</td>
<td>0.8069</td>
<td>1.4618</td>
<td>1.0512</td>
</tr>
</tbody>
</table>

Then, we apply the nonlinear discriminant function $\text{NLD}_{m}(t)$ by Equation (18) for the above data set and we classify the observations in the mixed data set one by one into either $\Pi_1$ or $\Pi_2$, we have 34 out of 46 from $\Pi_1$ are classified into $\Pi_1$ with the probability of 0.7391 and 12 out of 46 from $\Pi_2$ are classified into $\Pi_2$ with the probability of 0.2609.

8. Concluding remarks

In this paper, the maximum-likelihood method is used to obtain the MLEs of the parameters of MTIWD based on the mixed and classified samples. Then, the MLEs are utilized to estimate the corresponding nonlinear discriminant functions of both the mixed and the classified cases. The efficiencies of the estimated nonlinear discriminant functions are investigated and compared with
the corresponding optimal one through a series of Monte Carlo simulations in terms of the MSEs and trace. Illustrative example and a set of real data are analysed by using the proposed discriminant procedures. Generally, by using the total probability, we conclude that the performance of the mixture discriminant procedure behaves well compared with the completely classified procedure.

Acknowledgements

The authors thank the referees for their helpful comments, which improved the presentation of the paper. This project was supported by King Saud University, Deanship of Scientific Research, College of Science Research Center.

References


