Journal of Statistical and Econometric Methods, vol.5, no.3, 2016, 11-31
ISSN: 1792-6602 (print), 1792-6939 (online)
Scienpress Ltd, 2016

Some zero mean classification functions with unequal prior probabilities and non-normality

Michael Asamoah-Boaheng[1], Atinuke Adebanji[2] and Morire Labeodan[3]

Abstract

In this study, the problem of classifying a new observation vector into one of the known groups ($\pi_i, i = 1,2$) distributed multivariate normal when the mean vectors are equal and the training data contaminated with outliers to be non-normal. Four classification rules are considered for equal and unequal prior probabilities and non-normality based on: Bartlett and Please method (BPM), Bayesian Posterior Probability Approach (BPP), the Quadratic Discriminant Function (QDF) and the Absolute Euclidean Distance Classifier method (AEDC). Female liked sex twins extracted from Stocks (1933) twin data is used for analysis and performance evaluation is based on Cross Validation (CV) and Balanced Error Rate (BER). While all four functions recorded higher error rates, BPM method was very sensitive to outliers. The QDF performed better with the least error rate under non-normality. BPM outperformed all the other classification rules under unequal

[1] Institute of Research, Innovation and Development (IRID), Kumasi Polytechnic, E-mail:asboaheng@yahoo.com
[2] Department of Mathematics, Kwame Nkrumah University of Science and Technology, P.M.B, KNUST, e-mail:tinuadebanji@yahoo.com
[3] Easydat Consulting and Projects. Nelspruit, South Africa.

Article Info: *Received* : July 3, 2016. *Revised* : July 29, 2016.
Published online : September 1, 2016.

prior probabilities. Similar results were obtained from the simulation study.

Keywords: Mean Error Rates; Bartlett and Please Method; Absolute Euclidean Distance Classifier; Outliers.

1 Introduction

We consider the problem of zero mean difference classification of a new *p*-variate observations into two known groups π_1 and π_2 of independently distributed multivariate normal with mean vectors $\mu_i, i = 1,2$ and positive definite variance covariance matrix, $\Sigma_{p\times p}, n \geq p$. The mean vectors are assumed equal with a zero mean difference and $\Sigma_1 \neq \Sigma_2$. The population parameters are replaced by their sample estimates in the functions. Manjunath et al (2012) indicated that classical discriminant analysis focuses on Gaussian and non-parametric models where the unknown densities were replaced by their kernel densities.

Okamoto (1961) one of the earliest contributors to the subject, studied the problem of discrimination with common mean and different covariance structure of two multidimensional normal populations. This was followed by the study of Bartlett and Please (1963) where uniform covariance matrices were used to obtain a linear discriminant function an assumption of equal and unequal correlation coefficient for classification. The optimal discriminant rule has to be based on the difference between the group covariance matrices with some authors assuming uniformity in covariance matrices. (McLachlan, 2004). Geisser and Desu (1968) assumed various structures of uniform covariance matrices with application of several classification methods. Another school of thought is the Bayesian approach pioneered by Geisser and Desu (1968, 1975). Lachenbruch (1975) absolute linear rule performed well for two group classification with contaminated training data while QDF performed poorly. Ganeslingam (2006) compared the performance of QDF and AEDC with zero mean difference vector assumption resulting in AEDC outperforming the QDF.

In evaluating the classification functions, one of the ways of judging the performance of several classification procedures is to calculate their error rates or misclassification probabilities. Some of the error rate estimation methods are the optimal error rates (OER), the apparent error rate (APER), the balanced error rate (BER) and the leaving-one-out method (LOO). (Lachenbruch, 1968). An assessment of error rate estimators was studied by (Lachenbruch, 1968) and (Krzanowski and hand, 1997) paying special attention to the leave-one-out method. The leave-one-out rule seeks to overcome the drawback of re-substitution by process of cross-validation. The estimator was investigated in simulation study, both in absolute terms and in comparison with a popular bootstrap estimator.

Several comparative studies on the QDF and Linear Discriminant Function (LDF) for two group classifications have been considered. Lachenbruch et al (1975) and Lachenbruch et al (1977) studied the robustness of LDF and effects of non-normality on the QDF. Their results indicated that the actual error rates were considerably larger than the optimal rates in the case of zero mean difference. Also, non-normal samples generally under the QDF did not do substantially worse than when applied to normal samples. In our study we consider the zero mean classification problem for non-normal training data and unequal prior probabilities. The sample size ratios were set at 1:1, 1:2, 1:3 and 1:4.

2 Methods and Materials

2.1 Zero Mean Classification Functions

Here we discuss the classification functions evaluated in this study.

2.1.1 Bartlett and Please approach to Equal mean discrimination

Bartlett and Please (1963) addressed the problem of equal mean discrimination for two populations using the well-known twins data of Stocks (1933). They adopted the general uniform covariance structure

$$\Sigma_i = \sigma_i^2 \{(1-\rho_i) I_p + \rho_i 1_p 1'_p \} \tag{1}$$

Where **1** is a column vector of **1's** and ρ is the population correlation coefficient between any two variables. This pattern of equal covariance and equal variances in Σ is variously referred to as *uniformity*, *compound symmetry*. Bartlett and Please standardised the first covariance matrix corresponding to the monozygotic twins as;

$$\Sigma_1 = \{(1-\rho_1) I + \rho_1 1_p 1'_p \} \tag{2}$$

They further assumed that, Σ_2 could not be simultaneously standardised to unit variances and was given by;

$$\Sigma_2 = \sigma^2 \{(1-\rho_2) I + \rho_2 1_p 1'_p \} \tag{3}$$

σ_1^2 was set equal to 1 with $\rho_1 = \rho_2 = \rho$ where σ^2 is obtained as the ratio of the sum of squares of the Dizygotic twins π_2 to that of the Monozygotic twins π_1. With reference to the two uniform covariance matrices above, the inverses of Σ_1 and Σ_2 are given below.

$$\Sigma_1^{-1} = \frac{I}{(1-\rho_1)} - \frac{\rho_1}{(1-\rho_1)} \frac{E}{(1+(p-1)\rho_1)} \quad \text{and} \tag{4}$$

$$\Sigma_2^{-1} = \frac{1}{\sigma^2} \left[\frac{I}{(1-\rho_2)} - \frac{\rho_2}{(1-\rho_2)} \frac{E}{(1+(p-1)\rho_2)} \right] \tag{5}$$

With $1_p 1'_p = E$, E *i*s the matrix with entries equal to unity. Therefore using the log likelihood (likelihood ratio) $f_1(x)/f_2(x)$ and ignoring the additive constants gives the optimal discriminant function for equal mean vectors and correlations

($\rho_1 = \rho_2 = \rho$) . The ideal discriminant function involves only Z_1 and Z_2 so that these two quantities are plotted and gives a resulting straight line boundary. The classification rule becomes: assign z to π_1 if and only if

$$Z_1 - \frac{\rho}{1+(p-1)\rho} Z_2 \leq \frac{p(1-\rho)}{1-\sigma^{-2}} - 2c \tag{6}$$

Otherwise assign z to π_2 where p=the number of independent variables, $c = q_2/q_1$ with q_2 and q_1 being the prior probabilities for the two respective groups, $Z_1 = tr(zz')$ and $Z_2 = tr(Ezz')$ with z being the observational vector belonging to either π_1 or π_2.

2.1.2 Bayesian Posterior Probability Approach for Classification

Let π_j denote the proportion of units in the total observations/units in population j . We denote π_j as the prior probability of membership in population j . Considering the probability of unit u belonging to group j, given that the unit has a particular observation vector X_u . This probability, denoted by $P(j \mid X_u)$, is the posterior probability of membership in population j. Hence,

$$P(j \mid X_u) = \frac{P_j . f(X_u \mid j)}{\sum_{j^i=1}^{j} P_j . f(X_u \mid j')} \tag{7}$$

By using equation (10), the total number of misclassification errors is minimized. And unit u is assigned to population j if

$P(j \mid X_u) > P(j' \mid X_u)$ (Huberty and Olejnik, 2006).

2.1.3 Quadratic Discriminant Function: Normal Populations with $\Sigma_1 \neq \Sigma_2$

From equation (2), (3) the classification rule for assigning x_0 to π_1 (otherwise to π_2) is given as, classify x_0 as π_i if

$$-\frac{1}{2}x_0^T(\Sigma_1^{-1}-\Sigma_2^{-1})x_0+(\mu_1^T\Sigma_1^{-1}-\mu_2^T\Sigma_2^{-1})x_0-K\geq\left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{P_2}{P_1}\right)\right] \quad (8)$$

Where the constant k is given by $K=\frac{1}{2}\ln\left[\frac{|\Sigma_1|}{|\Sigma_2|}\right]+\frac{1}{2}(\mu_1^T\Sigma_1^{-1}\mu_1-\mu_2^T\Sigma_2^{-1}\mu_2)$. The classification regions are defined by quadratic functions of x. When $\sum_1\neq\Sigma_2$, equation (11) reduces to

$$(\mu_1^T\Sigma_1^{-1}-\mu_2^T\Sigma_2^{-1})x_0-k\geq\left[\ln\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right] \quad (9)$$

The awkward nature of the Quadratic discriminant function occurs in more than two dimensions and can lead to some strange results. (Johnson and Wichern, 2007).

2.1.4 The Absolute Euclidean Distance Classifier (AEDC)}

The Euclidean distance classifier cannot be used when $\mu_1=\mu_2$. The AEDC is used when the absolute values of the components of the observations in Euclidean Distance Classifier (EDC) is considered. It is expected that, this approach does well in a high dimensional data set. AEDC is mostly used and applicable in situations when $\Sigma_1\neq\Sigma_2$. AEDC and QDF are always used as an alternative to LDF. The Euclidean distance classifier (EDC) will allocate an observed vector X to population π_1 if

$$\left\{x-\frac{1}{2}(\mu_1+\mu_2)\right\}^T.(\mu_1+\mu_2)>0 \quad (10)$$

Otherwise to π_2

For $\mu_1=\mu_2$ the absolute values of X, $|X|=Y$, for a three dimensional vector becomes: allocate Y to π_1 if

$$y_1(\mu_1^1-\mu_1^2)-\frac{1}{2}((\mu_1^1)^2-(\mu_1^2)^2)+y_2(\mu_2^1-\mu_2^2)-\frac{1}{2}((\mu_2^1)^2-(\mu_2^2)^2)+y_3(\mu_3^1-\mu_3^2)$$
$$-\frac{1}{2}((\mu_3^1)^2-(\mu_3^2)^2)>0 \qquad (11)$$

Where $\mu_i^{(k)}$ is the mean of the i^{th} component of Y in the k^{th} population. In general, we allocate observation vector X to π_1 if

$$\sum_{i=1}^{p}\left[y_i\left(\sqrt{\sigma_{ii}^{(1)}}-\sqrt{\sigma_{ii}^{(2)}}\right)-\frac{1}{2}\sqrt{\frac{2}{\pi}}(\sigma_{ii}^{(1)}-\sigma_{ii}^{(2)})\right]>0 \qquad (12)$$

2.2 Evaluating Classification functions for two groups

To judge the performance of a sample classification procedure, we calculate its misclassification probability or error rate. Some of the measures of performance that can be calculated for any classification procedure are the Cross Validation (CV) method and the Balanced Error Rate (BER).

2.2.1 Cross Validation Procedure

The leave-one-out method was introduced by Lachenbruch and Mickey (1968) and it stipulates as follows. Let n_{1M}^{CM} and n_{2M}^{CV} be the number of left-out observations misclassified in groups 1 and 2 respectively. A good estimate of the actual error rate is given by:

$$CV=\frac{n_{1M}^{CM}+n_{2M}^{CV}}{n_1+n_2} \qquad (13)$$

2.2.2 The Balanced Error Rate Procedure (BER)

The Balanced Error Rate (BER) statistics is the average of the misclassification can be derived as:

$$BER = \frac{1}{2}\left[\frac{b}{a+b} + \frac{c}{c+d}\right] \tag{14}$$

Where *a* ,*b*, *c* and *d* are entries in the confusion matrix.

2.3 Simulation design and stocks data

A multivariate normal distribution data, $N(\mu_i, \Sigma_i)$, $i = 1,2$ from the groups π_1, π_2 and sample size $n_i = 50, i = 1,2$, using the estimates of ρ_1 and ρ_2 from the female like sex twin data to be 0.0478 and 0.2194 respectively with ten variables depicting the behaviour of female like sex twins from Stock twin data was simulated. Fifteen (15) pairs of observations were sampled from each group after 10 replications using simple random sampling without replacement. Mean estimates of σ^2, ρ_1, ρ_2 and ρ were computed after the 10 replications and was used to generate the classification functions and rules. The simulated 10 variate normally distributed data was further contaminated to assume non normality by introducing outliers into both the monozygotic and dizygotic twin groups. The process was therefore repeated and was applied to the already outlined classification methods and their performances were evaluated.

2.3.1 Stocks Female liked sex Data

Female like sex twins comprising thirty (30) pairs of monozygotic twins and 25 pairs of dizygotic twins was used for the data analyses. A sample size of fifteen (15) from each group were selected after 10 replications based on simple random sampling and the estimates of σ^2, ρ_1, ρ_2 were estimated from the mean estimates from the 10 replicated samples. The final sample size $n_1 = n_2 = 15$ was selected based on the closeness of the estimated parameters to that of the mean values of the 10 replicated estimates. The 10 variables selected included: Height (Ht), Weight (Wt), Head length (HL), Head breadth (HB), Head circumference (HC), Interpupillary distance (ID), Systolic blood pressure (SBP),Pulse interval (PI), Strength of left(SGL) grip, Strength of right grip (SGR). Stocks twin data was contaminated to assume non normality after the introduction of outliers into the two twin groups.

3 Main Results

Under the assumption of unequal prior probabilities and equal misclassification cost and non-normality contamination, four classification functions were derived under several varying sampling degrees. The varying degrees of sample selections used based on the groups ratio in the order of Monozygotic: Dizygotic were; 1:2, 1:3, 1:4 and the ratio order of Dizygotic: Monozygotic were 1:2, 1:3, 1:4.

3.1 Bartlett and Please Classification Method for $n_1 \neq n_2, (n_1 : n_2)$

The summary of classification rules under each of the various sampling ratios (1:2, 1:3, 1:4) are presented in the table below.

Table 1: Classification rules under the various sampling ratios

Sample Ratios	Classification rules	Cut-off
$n_1 : n_2$		
1:2	$Z_1 - 0.0655Z_2 \leq 14.96$	14.96
1.3	$Z_1 - 0.0660Z_2 \leq 12.90$	12.90
1:4	$Z_1 - 0.0709Z_2 \leq 17.70$	17.70
$n_2 : n_1$		
1:2	$Z_1 - 0.0702Z_2 \leq 15.90$	15.90
1:3	$Z_1 - 0.0230Z_2 \leq 18.91$	18.91
1:4	$Z_1 - 0.0158Z_2 \leq 17.99$	17.99

Based on the above derived classification rules, the following discriminant scores were obtained as summarised in Table 3.

Table 3: Discriminant scores

1:2		**1:3**		**1:4**	
$D_M < 14.96$	$D_D \geq 14.96$	$D_M < 12.90$	$D_D \geq 12.90$	$D_M < 17.70$	$D_M \geq 17.70$
1.53	15.79	1.53	17.79	1.52	**15.69**
5.01	28.74	5.01	28.74	5.01	28.72
1.57	16.52	1.57	16.52	1.56	**16.38**
2.81	50.15	2.81	50.08	2.79	49.48
4.95	60.19	4.94	60.17	4.84	59.96
	14.36		14.36		14.36
	63.99		63.98		63.84
	38.40		38.38		38.18
	15.58		15.58		15.58
	15.51		15.51		15.43
			132.50		131.49
			44.85		44.19
			80.76		79.83
			43.42		42.97
			18.04		17.90
					20.76
					43.98
					19.67
					59.67
					34.88

3.2 Bayesian Posterior Probability approach

The posterior probability approach for classification using Bayes rule was applied when the prior probabilities were assumed to be unequal. The sample ratios in the order of $n_1 : n_2$ and $n_2 : n_1$ were used as already spelt out in the above section.

Taking the sample ratio, $n_1 : n_2 = 1:2$

The main idea of these varying samples from each respective twin group was to find out the effect of unequal prior probabilities on the classification of the observations into their respective groups based on their scores. Table 4 below summarises the Bayes rule of posterior probabilities. Two (2) out of the ten selected observations from the dizygotic group were misclassified. In all, 80 percent correct classification was observed.

Taking the sample ratio $n_1 : n_2 = 1:3$

From Table 4, none of the monozygotic observations were misclassified. However six (6) observations were misclassified from the dizygotic twin group with 30 percent error rate of misclassifications.

Table 4: Posterior probability for the sample ratios $n_1 = 1 : n_2 = 2$, $n_1 = 1 : n_2 = 3$ and $n_1 = 1 : n_2 = 4$

1:2				**1:3**			
$P(j \mid x_M)$	$P(j' \mid x_M)$	$P(j' \mid x_D)$	$P(j \mid x_D)$	$P(j \mid x_M)$	$P(j' \mid x_M)$	$P(j' \mid x_D)$	$P(j \mid x_D)$
8.7E-01	1.3E-01	6.8E-01	3.2E-01	5.8E-01	4.2E-01	9.1e-01	9.2E-02
6.7E-01	3.3E-01	7.4E-01	2.6E-01	9.1E-01	9.0E-02	7.3E-01	2.7E-01
6.1E-01	3.9E-01	9.3E-01	7.4E-02	7.4E-01	2.6E-01	6.8E-01	3.2E-01
8.3E-01	1.7E-01	**3.2E-01**	**6.8E-01**	6.2E-01	3.8E-01	9.9E-01	1.4E-02
		6.1E-01	3.9E-01	7.9E-01	2.1E-01	**2.7E-01**	7.3E-01
		9.9E-01	9.9E-03			**4.8E-01**	5.2E-01
		3.7E-01	**6.3E-01**			9.1E-01	8.9E-02
		8.8E-01	1.2E-01			6.9E-01	3.0E-01
		6.1E-01	3.9E-01			5.5E-01	4.5E-01
						1.6E-01	**8.4E-01**
						2.5E-01	**7.5E-01**
						8.8E-01	1.2E-01
						7.3E-01	2.7E-01
						3.9E-01	**6.1E-01**
						2.9E-01	**7.1E-01**

Table 4: continued

1:4			
$P(j \mid x_M)$	$P(j' \mid x_M)$	$P(j' \mid x_D)$	$P(j \mid x_D)$
6.9E-01	3.1E-01	9.2E-01	8.2E-02
9.2E-01	8.3E-02	6.7E-01	3.4.2E-02E-01
7.7E-01	2.4E-01	6.5E-01	3.5E-01
6.4E-01	3.6E-01	9.9E-01	1.5E-02
8.6E-01	1.4E-01	2.5E-01	**7.5E-01**
		3.3E-01	**6.7E-01**
		8.8E-01	1.3E-01
		7.9E-01	2.1E-01
		5.1E-01	4.9E-01
		1.1E-01	8.9E-01
		2.1E-01	7.9E-01
		9.1E-01	8.7E-02
		7.3E-01	2.7E-01
		3.3E-01	6.7E-01
		2.7E-01	7.3E-01
		2.7E-01	7.3E-01
		2.8E-01	7.2E-01
		1.0E+00	1.5E-04
		9.7E01	3.2E-02
		7.4E-01	2.6E-01

Taking the sample ratio $n_1 : n_2 = 1:4$

From Table 4, the posterior probabilities used as a classification rule was hugely affected by the unequal prior probabilities as the sample size of the dizygotic group increased to 20. As a result of this, we observed eight (8) misclassified observations from the dizygotic group. No observation was misclassified from the monozygotic twin group. Hence the classification rule affects the population with larger sample size.

3.3 The QDF approach for $n_1 \neq n_2$

Assuming equal misclassification cost and unequal prior probabilities for $n_1 \neq n_2$, the QDF can be written in this particular case as

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < 1.658 \tag{15}$$

For sampling ratios, 1:3 and 1:4, the functions derived were

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < 1.204 \tag{16}$$

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < 1.585 \tag{17}$$

and resulted in 70 percent and 84 percent correct classification respectively. (*see the discriminant scores in Table 5*). The Table below summarises the discriminant scores for the QDF obtained from the three sampling ratios explained above. From Table 5, the discriminant scores for the sample ratio 1:2 recorded two misclassified observations from the dizygotic group. One (1) and five (5) observations from both the Monozygotic and Dizygotic twin groups were found to be misclassified from their groups respectively for the sample ratio of 1:3. For the sample ratio of 1:4, we observed 6 twin pair observations being misclassified from the dizygotic twin group. The QDF's performance was similar to that of the Bayesian Classifier with the number of misclassified observations increasing, as the sample size selection for the dizygotic group increased.

Table 5: Discriminant scores for the three sample ratios

1:2		**1:3**		**1:4**	
$D_M < 1.65$	$D_D \geq 1.65$	$D_M < 1.20$	$D_D \geq 1.20$	$D_M < 1.58$	$D_D \geq 1.58$
-0.786	**-0.429**	1.177	6.381	0.855	7.256
-4.780	3.026	-2.828	3.743	-2.392	3.85
-0.425	1.885	-0.257	3.301	0.054	3.622
-1.636	7.692	0.783	10.338	1.226	10.837
-1.744	0.026	-0.858	**-0.194**	-1.163	**0.258**
	2.531		1.644		**0.965**
	4.589		6.457		6.302

	6.765	3.453	5.102
	2.304	2.213	2.482
	-2.845	**-1.515**	**-1.701**
		-0.381	**-0.261**
		5.808	7.11
		3.789	4.371
		0.868	**0.991**
		0.035	**0.402**
			3.456
			4.756
			3.112
			1.876
			4.476

Table 6: Discriminant scores for the sample ratios, $n_2 = 1 : n_1 = 2$

1:2		**1:3**		**1:4**	
$D_M < 1.65$	$D_D \geq 1.65$	$D_M < 1.20$	$D_D \geq 1.20$	$D_M < 1.58$	$D_D \geq 1.58$
-0.536	14.666	-1.078	14.666	-1.175	12.921
-5.851	4.948	-5.769	4.948	-5.316	4.552
-0.691	4.761	-0.621	4.761	-0.537	3.218
-0.726	15.191	-0.984	15.191	-1.364	12.742
-3.811	-3.115	-3.602	-3.115	-3.104	-2.173
-0.161		-1.033		-1.147	
-2.596		-2.41		-2.379	
-3.95		-3.467		-2.924	
-0.524		1.327		2.374	
-2.495		-2.064		-1.987	
		-1.914		-1.684	
		-0.402		-0.487	
		-1.18		-1.205	
		-3.829		-3.475	
		-2.163		-2.219	
				-2.111	
				-3456	
				-3.993	
				-4.980	
				-2.212	

Alternating the sample sizes of the two groups to assume unequal prior probabilities and study the behaviour of the resulting classification rules derived in each case based on effect unequal prior probabilities, the QDF's obtained under the sample ratios 1:2 , 1:3 and 1: 4 were

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < -2.422 \quad (18)$$

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < -2.574 \quad (19)$$

$$\left\{(x-\mu_1)'\Sigma_1^{-1}(x-\mu_1)\right\} - \left\{(x-\mu_2)'\Sigma_2^{-1}(x-\mu_2)\right\} < -2.605 \quad (20)$$

respectively. The proportions of correct classification were 0.60, 0.40 and 0.52 for the three respective functions. The discriminant scores are shown in Table 6.

3.4 Evaluating the performance of the Classification Methods for $n_1 \neq n_2$

Table 7 summarises the error rates of the classification methods and their respective sample ratios under each of the three methods which where applicable to deriving a classification rule under unequal prior probability situation.

The results shows that generally unequal prior probabilities influences the classification rules of the three methods namely Bartlett and Please, QDF and Bayesian Posterior Probability approach. From the Table, it was observed that, the error estimates increased appreciably as the size of one group increases relative to another. The results show the performance of the functions deteriorating as the inequality in sample sizes widened with the QDF recording the highest error rates. Comparatively, all the three methods recorded almost similar error estimates in both the sample selection ratios and their corresponding alternated sampling ratio. Bartlett and Please classification method recorded the least mean error estimates as compared to the QDF and the Bayesian Classifier.

Table 7: Error rates for the classification method under unequal prior probability case

Classification methods	**Error Rates**		**Mean Error rate**
	CV	**BER**	
Bartlett and Please			
$n_1 : n_2$			
1:2	0.133	0.050	0.092
1:3	0.100	0.000	0.050
1:4	0.120	0.150	0.135
$n_2 : n_1$			
1:2	0.200	0.250	0.225
1:3	0.200	0.200	0.200
1:4	0.160	0.175	0.168
The Bayesian Rule			
$n_1 : n_2$			
1:2	0.266	0.150	0.208
1:3	0.300	0.333	0.316
1:4	0.340	0.234	0.287
$n_2 : n_1$			
1:2	0.133	0.040	0.086
1:3	0.350	0.433	0.391
1:4	0.300	0.398	0.349
The QDF Approach			
$n_1 : n_2$			
1:2	0.440	0.250	0.345
1:3	0.250	0.167	0.209
1:4	0.300	0.342	0.321
$n_2 : n_1$			
1:2	0.333	0.350	0.342
1:3	0.150	0.300	0.225
1:4	0.320	0.425	0.373

The QDF recorded the highest mean error rates. This results shows some partial linkage with the research work of Ganeslingam et al (2006) where the QDF as compared to the AEDC method performed poorly. Bartlett and Please method was observed to perform much better than the other methods for the provision of maximum separation between the two populations. These findings conforms to the results from the research work of Bartlett and Please (1963), Desu and Geisser (1973) where their linear discriminant functions obtained with uniform covariance matrix performed better in the provision of maximum separation under equal prior probabilities.

3.5 Performance evaluation under non-normality

The functions were evaluated and after outliers were introduced to alter the normality state of the data and the classification was done with equal and unequal prior probabilities. The outliers were introduced into the first five observations in each of the twin groups, one at a time and a classification rule was obtained in each case. The errors incurred in the classification of the twin pair observations are presented in Table 8. Generally the performance of all the classification methods with equal and unequal prior probabilities deteriorated after the introduction of outliers into the twin data. However the mean error estimates of AEDC method performed slightly better than the Bayesian Posterior Probability approach with a mean error rate of 0.381 under equal prior probabilities and 0.375 and 0.350 for unequal prior probabilities based on the predetermined choice of sampling ratios. In other words, the AEDC method recorded the least error rate of 0.339 and hence provides better separation than the remaining methods under non normality. This also shows conformation with the study by Ganeslingam et al (2006) in comparing the performance of AEDC and QDF with AEDC outperforming the QDF. It was also observed that, The Bartlett and Please approach performed poorly under non normality assumption with an error estimates for both equal and unequal prior probabilities ranging from 0.466 to 0.667. The QDF performed appreciably better for the equal prior probability case, but the performance under unequal prior probabilities was abysmal, that is with a recorded mean error rate for the two sample ratios as 0.523 and 0.625 for the sample ratios of 1:2 and 1:3 respectively. (*see Table 8*).

Table 8: Evaluation of the classification methods under Non-normality

Classification methods	**Error Rates**		**Mean Error rate**
	CV	**BER**	
Bartlett and Please			
$n_1 = n_2$	0.466	0.466	0.466
$n_1 : n_2$			
1:2	0.800	0.850	0.825
1:3	0.756	0.757	0.762
1:4	0.699	0.861	0.780
$n_2 : n_1$			
1:2	0.614	0.801	0.708
1:3	0.600	0.733	0.667
1:4	0.703	0.788	0.746
The Bayesian Rule			
$n_1 = n_2$			
$n_1 : n_2$			
$n_1 = n_2$	0.352	0.410	0.381
1:2	0.400	0.350	0.375
1:3	0.423	0.422	0.423
1:4	0.478	0.317	0.398
$n_2 : n_1$			
1:2	0.301	0.123	0.212
1:3	0.333	0.166	0.249
1:4	0.314	0.107	0.211
The QDF Approach			
$n_1 = n_2$	0.300	0.300	0.300
$n_1 : n_2$			
1:2	0.446	0.600	0.523
1:3	0.545	0.555	0.550

1:4	0.500	0.689	0.595
$n_2 : n_1$			
1:2	0.800	0.607	0.704
1:3	0.750	0.500	0.625
1:4	0.713	0.578	0.646
The AEDC method			
$n_1 = n_2$	0.333	0.345	0.339

This result conforms to the Lachenbruch et al (1977) on the effect of non-normality on QDF. These results show some conformity with the study by Lachenbruch (1975), where after contaminating the twin data, he discovered that, the performance of the QDF was very poor, but the absolute linear discriminant function performed reasonably well. Lachenbruch et al (1977) work on the effects of non-normality on QDF found that the actual error rate for QDF was considerably larger than the optimal rate in the case of zero mean difference.

4 Conclusions

This paper studied the performance of four classification methods after their evaluation under unequal prior probabilities and non-normality. The BPM outperformed the QDF and BPP under unequal prior probabilities with the QDF performing poorly in the classification of the twin pair observations. Generally the performance of all the classification methods with equal and unequal prior probabilities deteriorated after the introduction of outliers into the data, while the BPM was found to be very sensitive to outliers since it performed poorly under non-normality. The mean error estimate of AEDC method performed slightly better than the BPP with mean error rate of 0.381. The QDF performed appreciably better under equal prior probability case and its performance was abysmal under unequal prior probability situation. Similar results were obtained after the application of the simulated data in evaluating the performance of the classification methods under unequal prior probabilities and under non-normally distributed training samples.

References

[1] Ariyo, O.S., and Adebanji, A.O., Effect of Misclassification Costs on the Performance Functions. Paper Presented at the 45th Annual Conference of the Science Association of Nigeria (SAN) held Niger Delta University, Wilberforce, Island , Bayelsa State Functions, (2010).

[2] Bartlett, M.S., and Please, N.W., Discrimination in the case of zero mean differences, *Biometrika*, 47, (1963), 185-189.

[3] Desu, M. M., and Geisser, S., Methods and applications of equal-mean discrimination, *Discriminant Analysis and Applications*, (1973), 139-161.

[4] Ganeslingam, S., Nanthakumar, A. S., Ganesh, S, A comparison of quadratic discriminant function with discriminant function based on absolute deviation from the mean, *Journal of Statistics and Management Studies*, 9, (2006), 441-457.

[5] Geisser, S. and Desu, M.M., Predictive zero-mean uniform discrimination. *Biometrika* 55, (1968), 519-524.

[6] Huberty, C. J., and Olejnik, S, Applied MANOVA and Discriminant Analysis, Second edition, A John Wiley and Sons, Inc., Publication, (2006), 289-384.

[7] Hyodo, M., and Kubokawa.,A variable selection criterion for linear discriminant rule and its optimality in high dimensional and large sample data, *Journal of multivariate Analysis*, 123, (2014), 364-379.

[8] Johnson, R. A., and Wichern., D.W., *Applied Multivariate Statistical Analysis*, Pearson Education Inc, New Jersey NJ, 2007, 575-585.

[9] Krzanowski, W.J., and Hand., Assessing error rate estimators: The leave-one-out reconsidered, *Australian J. Statist*., 39, (1997), 35-46.

[10] Lachenbruch, P. A., and Mickey, R., Estimation of Error Rates in Discriminant Analysis, *Technometrics*, 10(1), (1968),1-11.

[11] Lachenbruch, P. A., Sneeringer, C., L. T., Revo, Robustness of the linear and quadratic discriminant functions to certain types of non-normality, *Computational Statistics*, 1, (1973), 39-56.

[12] Lachenbruch, P. A., Zero-mean difference discrimination and the absolute linear discriminant function, *Biometrika*, 62, (1975), 397-401.

[13] Lachenbruch, P. A., Clarke, W., Broffit, B., Lin, L.,. *The Effect of Non-Normality on Quadratic Discriminant Function*, MEDINFO 77 Shives/Wolf. IFEP, North Holland Publishing Co: (1977), 101-104.

[14] Manjunath, B. G., Frick, M., Reiss, R. D., Some notes on external discriminant analysis, *Journal of Multivariate analysis*, 103(1), (2012), 107-115.

[15] Marco, V. R., Young, D. M., Tubne, D. W., Asymptotic expansions and estimation of the expected error rate for equal-mean discrimination with uniform covariance structure, *Biometrika*, 29, (1987), 103-111.

[16] McLachlan, G.J., *Discriminant Analysis and Statistical Pattern Recognition*, A John Wiley and Sons Inc. Publication, Hoboken, New Jersey, (2004) 57-58.

[17] Okamoto, M., *Discrimination for variance matrices*, Osaka Math 13, (1961), 1-39.

[18] Stocks, P. A., Biometric investigation of twins: Part II, *Annals of Eugenics*, 5, (1933), 1-55.