THE UNIVERSITY OF TEXAS AT SAN ANTONIO, COLLEGE OF BUSINESS

Working Paper SERIES

July 20, 2008

Wp# 0057MSS-253-2008

An Extension of the Traditional Classification Rules: the Case of Non-Random Samples

Ricardo Leiva
Departamento de Matem´atica
F.C.E., Universidad Nacional de Cuyo
5500 Mendoza, Argentina

Anuradha Roy
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249, USA

Department of Management Science & Statistics,
University of Texas at San Antonio,
San Antonio, TX 78249, U.S.A

An Extension of the Traditional Classification Rules: the Case of Non-Random Samples

Ricardo Leiva
Departamento de Matemática
F.C.E., Universidad Nacional de Cuyo
5500 Mendoza, Argentina

Anuradha Roy
Department of Management Science and Statistics
The University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249, USA

Abstract

The paper deals with an heuristic generalization of the traditional classification rules by incorporating within sample dependencies. The main motivation behind this generalization is to develop a new classification rule when training samples are not random, but, jointly equicorrelated.

MSC: primary 62H30; secondary 62H12

Keywords: Classification rules; Non-random samples; Jointly equicorrelated training vectors

JEL Code: C30

1 Introduction

Theoretical inference in statistics is primarily based on the assumption of independent and identically distributed random samples drawn from a population. However, it is not necessary that we always have access to samples that are truly random in nature. In these cases the standard inference results fail. Consider an example of a digital image where contiguous pixels are correlated. The correlation exists because sensors take a significant energy from these contiguous pixels, and sensors cover a land region much larger than the size of a pixel. For example, if a pixel represents wheat in an agricultural field, then its neighboring pixels also represent wheat with high probability (Richards et al., 1999). A classification method based on the training samples of these neighboring pixels must take into account this correlation, and equicorrelation could be a reasonable assumption. The rational of this article is to generalize the traditional classification rules by incorporating the existing correlation or dependency of the neighboring training samples.

Considerable progress has been made in relaxing the assumption of independence of neighboring training samples through the concept of equicorrelation (also known as intraclass correlation) in the univariate case. That is, instead of the assumption of random samples, the samples $x_1, \ldots, x_n$ are assumed to be equicorrelated, i.e. the covariance matrix $\mathbf{\Sigma}$ of the vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$ is assumed to be $\mathbf{\Sigma} = (\sigma_0 - \sigma_1)\, \mathbf{I}_n + \sigma_1 \mathbf{J}_{n,n}$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{1}_{n_i}$ is the n_i-variate vector of ones and $\mathbf{J}_{n_1,n_2} = \mathbf{1}_{n_1}\mathbf{1}'_{n_2}$. Several authors (Shoukri and Ward, 1984; Viana, 1982, 1994; Zerbe and Goldgar, 1980; Donner and Bull, 1983; Donner and Zou, 2002; Konishi and Gupta, 1989; Khatri, Pukkila and Rao, 1989; Paul and Barnwal, 1990; Young and Bhandary, 1998; Bhandary and Alam, 2000; Smith and Lewis, 1980; Barghava and Srivastava, 1973; Gupta and Nagar, 1987; Khan and Bhatti, 1998) have used this univariate equicorrelation concept for many different purposes in their studies. Nevertheless, the natural multivariate generalization of this equicorrelation concept has not been explored as thoroughly as its univariate counterpart.

The observed unexpected misclassification probabilities while applying the Fisher (1936) linear discriminant function to multivariate remote sensing data was explained by Basu and Odell (1974) with the assumption of equicorrelated training vector samples. Unfortunately, Basu and Odell did not give any logical solution to this problem. In other words, they did not study the appropriate discriminant function for this problem. Recently, Leiva (2007) obtained a linear classification rule for equicorrelated training vector dependence (defined in Section 2.1), and showed that this generalizes the Fisher's linear classification rule.

The present article builds up on Leiva (2007), and provides a quadratic extension of the traditional classification rules for the non-random samples based on equicorrelated training vectors by using multivariate equicorrelation.

2 Basic concepts

2.1 Equally correlated vectors

Let $\boldsymbol{x}_h$ be a $nm-$variate vector of measurements of n neighboring $m-$variate sample measurements from a population ($h = 1, \ldots, N$). We partition this vector $\boldsymbol{x}_h$ as $\boldsymbol{x}_h = (\boldsymbol{x}'_{h,1}, \ldots, \boldsymbol{x}'_{h,n})'$, where $\boldsymbol{x}_{h,j} = (x_{h,j,1}, \ldots, x_{h,j,m})'$ for $j = 1, \ldots, n$. Let $\boldsymbol{x}$ represent the $n_i m-$variate vector of measurements corresponding to one individual in the ith population. We assume $\boldsymbol{x}$ has constant mean vector structure, i.e. $\mathrm{E}\left[\boldsymbol{x}_h\right] = \boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{1}_n \otimes \boldsymbol{\mu}$ with $\boldsymbol{\mu} \in \mathbf{R}^m$, and partitioned covariance structure, i.e. $\mathbf{\Gamma}_{\boldsymbol{x}} = \mathrm{Cov}\left[\boldsymbol{x}_h\right] = \left(\mathbf{\Gamma}_{\boldsymbol{x}_{h,r},\boldsymbol{x}_{h,s}}\right) = \left(\mathbf{\Gamma}_{h,rs}\right)$, where $\mathbf{\Gamma}_{h,rs} = \mathrm{Cov}\left[\boldsymbol{x}_{h,r}, \boldsymbol{x}_{h,s}\right]$ for $r, s = 1, \ldots, n$.

Definition 1 *The partitioned vector $\boldsymbol{x}_h$ or its component vectors $\boldsymbol{x}_{h,1}, \ldots, \boldsymbol{x}_{h,n}$ are said to be equicorrelated iff*

$$\mathbf{\Gamma}_{\boldsymbol{x}} = \mathbf{I}_n \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1) + \mathbf{J}_{n,n} \otimes \mathbf{\Gamma}_1,$$

where $\mathbf{\Gamma}_0$ is a positive definite symmetric $(m \times m)$ matrix and $\mathbf{\Gamma}_1$ is a symmetric $(m \times m)$ matrix. This matrix $\mathbf{\Gamma}_{\boldsymbol{x}}$ is called equicorrelated covariance matrix, and the matrices $\mathbf{\Gamma}_0$ and $\mathbf{\Gamma}_1$ are called equicorrelation parameters.

The $m \times m$ block diagonals $\mathbf{\Gamma}_0$ represent the variance-covariance matrix of the $m-$variate response variable at any given sample (pixel), whereas $m \times m$ block off diagonals $\mathbf{\Gamma}_1$ represent the covariance matrix of the m response variables between any two neighboring samples (pixels). We assume $\mathbf{\Gamma}_0$ is constant for all samples, and $\mathbf{\Gamma}_1$ is same for all neighboring sample pairs.

If $\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_0 + (n-1)\mathbf{\Gamma}_1$ are non singular matrices, then $\mathbf{\Gamma}_{\boldsymbol{x}}$ is invertible (Lemma 4.3, Ritter and Gallegos, 2002; Leiva, 2007), and

$$\begin{aligned}
\mathbf{\Gamma}_{\boldsymbol{x}}^{-1} &= \mathbf{I}_n \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1)^{-1} - \mathbf{J}_{n,n} \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1)^{-1} \mathbf{\Gamma}_1 (\mathbf{\Gamma}_0 + (n-1)\mathbf{\Gamma}_1)^{-1} \\
&= \mathbf{I}_n \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1)^{-1} + \mathbf{J}_{n,n} \otimes \frac{1}{n} \left[(\mathbf{\Gamma}_0 + (n-1)\mathbf{\Gamma}_1)^{-1} - (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1)^{-1} \right] \\
&= \mathbf{I}_n \otimes \mathbf{A} + \mathbf{J}_{n,n} \otimes \mathbf{B}_n. \qquad (1)
\end{aligned}$$

We notice that $\mathbf{\Gamma}_{\boldsymbol{x}}^{-1}$ has the same format as $\mathbf{\Gamma}_{\boldsymbol{x}}$. This result (1) generalizes the one given by Bartlett (1951) for $m = 1$. The determinant of the matrix $\mathbf{\Gamma}_{\boldsymbol{x}}$ is given by

$$|\mathbf{\Gamma}_{\boldsymbol{x}}| = |(\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1)|^{n-1} |\mathbf{\Gamma}_0 + (n-1)\mathbf{\Gamma}_1|. \qquad (2)$$

2.1.1 Maximum Likelihood estimates of the mean vector and the covariance matrix for equicorrelated samples

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be $nm-$variate vectors of N equally correlated random samples form $\mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{x}}, \mathbf{\Gamma}_{\boldsymbol{x}}) = \mathrm{N}(\mathbf{1}_n \otimes \boldsymbol{\mu}, \mathbf{I}_n \otimes (\mathbf{\Gamma}_0 - \mathbf{\Gamma}_1) + \mathbf{J}_{n,n} \otimes \mathbf{\Gamma}_1)$. The following theorem gives the MLEs of $\boldsymbol{\mu}_{\boldsymbol{x}}$ and $\mathbf{\Gamma}_{\boldsymbol{x}}$.

Theorem 1 *Under the above set-up the maximum likelihood estimate (MLE) of* $\boldsymbol{\mu}_{\boldsymbol{x}}$ *is*

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \mathbf{1}_n \otimes \widehat{\boldsymbol{\mu}} = \mathbf{1}_n \otimes \overline{\boldsymbol{x}},$$

where

$$\overline{\boldsymbol{x}} = \frac{1}{Nn} \sum_{h=1}^{N} \sum_{j=1}^{n} \boldsymbol{x}_{h,j},$$

and the maximum likelihood estimate of $\mathbf{\Gamma}_{\boldsymbol{x}}$ *is*

$$\widehat{\mathbf{\Gamma}}_{\boldsymbol{x}} = \mathbf{I}_n \otimes \left(\widehat{\mathbf{\Gamma}}_0 - \widehat{\mathbf{\Gamma}}_1 \right) + \mathbf{J}_{n,n} \otimes \widehat{\mathbf{\Gamma}}_1,$$

where

$$\widehat{\mathbf{\Gamma}}_0 = \frac{1}{Nn} \sum_{h=1}^{N} \sum_{j=1}^{n} (\boldsymbol{x}_{h,j} - \overline{\boldsymbol{x}}) (\boldsymbol{x}_{h,j} - \overline{\boldsymbol{x}})',$$

$$\textit{and} \quad \widehat{\mathbf{\Gamma}}_1 = \frac{1}{Nn(n-1)} \sum_{h=1}^{N} \sum_{j=1}^{n} \sum_{j \neq i=1}^{n} (\boldsymbol{x}_{h,j} - \overline{\boldsymbol{x}}) (\boldsymbol{x}_{h,i} - \overline{\boldsymbol{x}})'.$$

2.2 Jointly equicorrelated vectors

In this section we introduce the concept of jointly equicorrelated vectors. Let $\boldsymbol{x}_h^{(i)}$ be a $n_i m-$variate vector of measurements of n_i neighboring $m-$variate sample measurements from the ith population ($i = 1, 2,\ h = 1, \dots, N$). We partition this vector $\boldsymbol{x}_h^{(i)}$ as $\boldsymbol{x}_h^{(i)} = (\boldsymbol{x}_{h,1}^{(i)\prime}, \dots, \boldsymbol{x}_{h,n_i}^{(i)\prime})'$, where $\boldsymbol{x}_{h,j}^{(i)} = (x_{h,j,1}^{(i)}, \dots, x_{h,j,m}^{(i)})'$ for $j = 1, \dots, n_i$. Let $\boldsymbol{x}^{(i)}$ represent the $n_i m-$variate vector of measurements corresponding to one individual in the ith population. We assume that $\boldsymbol{x}^{(i)}$ has constant mean vector structure $\boldsymbol{\mu}_{\boldsymbol{x}^{(i)}} = \mathrm{E}\left[\boldsymbol{x}_h^{(i)}\right] = \mathbf{1}_{n_i} \otimes \boldsymbol{\mu}^{(i)}, \boldsymbol{\mu}^{(i)} \in \Re^m$, and partitioned covariance structure $\boldsymbol{\Gamma}_{\boldsymbol{x}^{(i)}} = \mathrm{Cov}\left[\boldsymbol{x}_h^{(i)}\right] = \left(\boldsymbol{\Gamma}_{\boldsymbol{x}_{h,r}^{(i)}, \boldsymbol{x}_{h,s}^{(i)}}\right) = \left(\boldsymbol{\Gamma}_{h,rs}^{(i)}\right)$, where $\boldsymbol{\Gamma}_{h,rs}^{(i)} = \mathrm{Cov}\left[\boldsymbol{x}_{h,r}^{(i)}, \boldsymbol{x}_{h,s}^{(i)}\right]$ for $r, s = 1, \dots, n_i$.

Definition 2 *Vectors* $\boldsymbol{x}_h^{(1)} = \left(\boldsymbol{x}_{h,1}^{(1)\prime}, \dots, \boldsymbol{x}_{h,n_1}^{(1)\prime}\right)'$ *and* $\boldsymbol{x}_h^{(2)} = \left(\boldsymbol{x}_{h,1}^{(2)\prime}, \dots, \boldsymbol{x}_{h,n_2}^{(2)\prime}\right)'$ *(or, equivalalently vectors* $\boldsymbol{x}_{h,1}^{(1)}, \dots, \boldsymbol{x}_{h,n_1}^{(1)}$ *and* $\boldsymbol{x}_{h,1}^{(2)}, \dots, \boldsymbol{x}_{h,n_2}^{(2)}$*) are said to be jointly equicorrelated iff* $\boldsymbol{x}_h^{(i)}$, $i = 1, 2$, *is an equicorrelated vector with equicorrelation parameters* $\boldsymbol{\Gamma}_0^{(i)}$ *and* $\boldsymbol{\Gamma}_1^{(i)}$*, and Cov* $\left[\boldsymbol{x}_{h,r}^{(1)}, \boldsymbol{x}_{h,s}^{(2)}\right] = \boldsymbol{\Gamma}$*, where* $\boldsymbol{\Gamma}$ *is a symmetric matrix. That is, vectors* $\boldsymbol{x}_h^{(1)}$ *and* $\boldsymbol{x}_h^{(2)}$ *are jointly equicorrelated if the covariance matrix* $\boldsymbol{\Gamma}_{\boldsymbol{x}}$ *of the partitioned* $(n_1 + n_2)m-$*variate vector* $\boldsymbol{x}_h = (\boldsymbol{x}_h^{(1)\prime}, \boldsymbol{x}_h^{(2)\prime})'$ *is*

$$\begin{pmatrix} \mathbf{I}_{n_1} \otimes \left(\boldsymbol{\Gamma}_0^{(1)} - \boldsymbol{\Gamma}_1^{(1)}\right) + \mathbf{J}_{n_1,n_1} \otimes \boldsymbol{\Gamma}_1^{(1)} & \mathbf{J}_{n_1,n_2} \otimes \boldsymbol{\Gamma} \\ \mathbf{J}_{n_2,n_1} \otimes \boldsymbol{\Gamma} & \mathbf{I}_{n_2} \otimes \left(\boldsymbol{\Gamma}_0^{(2)} - \boldsymbol{\Gamma}_1^{(2)}\right) + \mathbf{J}_{n_2,n_2} \otimes \boldsymbol{\Gamma}_1^{(2)} \end{pmatrix}. \tag{3}$$

This matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$ *is called jointly equicorrelated covariance matrix, and the matrices* $\boldsymbol{\Gamma}_0^{(1)}, \boldsymbol{\Gamma}_0^{(2)}, \boldsymbol{\Gamma}_1^{(1)}, \boldsymbol{\Gamma}_1^{(2)}$, *and* $\boldsymbol{\Gamma}$ *are called jointly equicorrelated parameters.*

Now, for $i = 1, 2$ and $k = 3 - i$, if

$$\mathbf{H}_{(i)1} = \boldsymbol{\Gamma}_0^{(i)} - \boldsymbol{\Gamma}_1^{(i)}, \tag{4}$$

$$\begin{aligned} \mathbf{H}_{(i)2} &= \boldsymbol{\Gamma}_0^{(i)} + (n_i - 1)\boldsymbol{\Gamma}_1^{(i)}, \\ &= \mathbf{H}_{(i)1} + n_i \boldsymbol{\Gamma}_1^{(i)}, \end{aligned}$$

$$\text{and} \quad \begin{aligned} \mathbf{H}_{(i,k)} &= \left(\boldsymbol{\Gamma}_0^{(i)} + (n_i - 1)\boldsymbol{\Gamma}_1^{(i)}\right) - n_i n_k \boldsymbol{\Gamma} \left(\boldsymbol{\Gamma}_0^{(k)} + (n_k - 1)\boldsymbol{\Gamma}_1^{(k)}\right)^{-1} \boldsymbol{\Gamma}, \\ &= \mathbf{H}_{(i)2} - n_i n_k \boldsymbol{\Gamma} \mathbf{H}_{(k)2}^{-1} \boldsymbol{\Gamma}, \end{aligned}$$

are non singular matrices, then $\boldsymbol{\Gamma}_{\boldsymbol{x}}$ is also non singular and its inverse is given by

$$\boldsymbol{\Gamma}_{\boldsymbol{x}}^{-1} = \begin{pmatrix} \mathbf{I}_{n_1} \otimes \mathbf{A}^{(1)} + \mathbf{J}_{n_1,n_1} \otimes \mathbf{D}^{(1)} & \mathbf{J}_{n_1,n_2} \otimes \mathbf{T} \\ \mathbf{J}_{n_2,n_1} \otimes \mathbf{V} & \mathbf{I}_{n_2} \otimes \mathbf{A}^{(2)} + \mathbf{J}_{n_2,n_2} \otimes \mathbf{D}^{(2)} \end{pmatrix},$$

where

$$\mathbf{A}^{(i)} = \left(\boldsymbol{\Gamma}_0^{(i)} - \boldsymbol{\Gamma}_1^{(i)}\right)^{-1} = \mathbf{H}_{(i)1}^{-1}, \tag{5}$$

$$\begin{aligned}\mathbf{D}^{(i)} &= -\left(\mathbf{\Gamma}_0^{(i)}-\mathbf{\Gamma}_1^{(i)}\right)^{-1}\left[\mathbf{\Gamma}_1^{(i)}-n_k\mathbf{\Gamma}\left(\mathbf{\Gamma}_0^{(k)}+(n_k-1)\mathbf{\Gamma}_1^{(k)}\right)^{-1}\mathbf{\Gamma}\right] \\ &\qquad \left\{\left(\mathbf{\Gamma}_0^{(i)}-\mathbf{\Gamma}_1^{(i)}\right)+n_i\left[\mathbf{\Gamma}_1^{(i)}-n_k\mathbf{\Gamma}\left(\mathbf{\Gamma}_0^{(k)}+(n_k-1)\mathbf{\Gamma}_1^{(k)}\right)^{-1}\mathbf{\Gamma}\right]\right\}^{-1},\end{aligned}$$

$$\mathbf{T}=-\left(\mathbf{A}^{(1)}+n_1\mathbf{B}_{n_1}^{(1)}\right)\mathbf{\Gamma}\left(\mathbf{A}^{(2)}+n_2\mathbf{D}^{(2)}\right),$$

$$\text{and}\quad \mathbf{V}=-\left(\mathbf{A}^{(2)}+n_2\mathbf{B}_{n_2}^{(2)}\right)\mathbf{\Gamma}\left(\mathbf{A}^{(1)}+n_1\mathbf{D}^{(1)}\right),$$

with

$$\begin{aligned}\mathbf{B}_{n_i}^{(i)} &= -\mathbf{A}^{(i)}\mathbf{\Gamma}_1^{(i)}\left(\mathbf{\Gamma}_0^{(i)}+(n_i-1)\mathbf{\Gamma}_1^{(i)}\right)^{-1}, \\ &= \frac{1}{n_i}\left(\mathbf{H}_{(i)2}^{-1}-\mathbf{H}_{(i)1}^{-1}\right).\end{aligned}$$

Note that if $\mathbf{\Gamma}=\mathbf{0}$, then $\mathbf{T}=\mathbf{0}$, $\mathbf{V}=\mathbf{0}$, and

$$\mathbf{D}^{(i)}=\mathbf{D}_{n_i}^{(i)}=-\left(\mathbf{\Gamma}_0^{(i)}-\mathbf{\Gamma}_1^{(i)}\right)^{-1}\cdot\mathbf{\Gamma}_1^{(i)}\cdot\left\{\mathbf{\Gamma}_0^{(i)}+(n_i-1)\,\mathbf{\Gamma}_1^{(i)}\right\}^{-1}=\mathbf{B}_{n_i}^{(i)}. \tag{6}$$

Therefore,

$$\mathbf{D}_{n_i+1}^{(i)} = -\left(\mathbf{\Gamma}_0^{(i)}-\mathbf{\Gamma}_1^{(i)}\right)^{-1}\mathbf{\Gamma}_1^{(i)}\left[\mathbf{\Gamma}_0^{(i)}+n_i\mathbf{\Gamma}_1^{(i)}\right]^{-1}. \tag{7}$$

Thus, if $\mathbf{\Gamma}=\mathbf{0}$, the inverse of $\mathbf{\Gamma}_{\boldsymbol{x}}$ is given by

$$\begin{aligned}\mathbf{\Gamma}_{\boldsymbol{x}}^{-1} &= \begin{pmatrix}\mathbf{\Gamma}_{\boldsymbol{x}^{(1)}}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Gamma}_{\boldsymbol{x}^{(2)}}^{-1}\end{pmatrix}=\begin{pmatrix}\mathbf{\Gamma}_{n_1}^{(1)-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Gamma}_{n_2}^{(2)-1}\end{pmatrix}, \\ &= \begin{pmatrix}\mathbf{I}_{n_1}\otimes\mathbf{A}^{(1)}+\mathbf{J}_{n_1,n_1}\otimes\mathbf{D}_{n_1}^{(1)} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_2}\otimes\mathbf{A}^{(2)}+\mathbf{J}_{n_2,n_2}\otimes\mathbf{D}_{n_2}^{(2)}\end{pmatrix},\end{aligned}$$

and the determinant of $\mathbf{\Gamma}_{\boldsymbol{x}}$ is given by

$$\begin{aligned}|\mathbf{\Gamma}_{\boldsymbol{x}}| &= \left|\mathbf{I}_{n_1}\otimes\left(\mathbf{\Gamma}_0^{(1)}-\mathbf{\Gamma}_1^{(1)}\right)+\mathbf{J}_{n_1,n_1}\otimes\mathbf{\Gamma}_1^{(1)}\right|\cdot\left|\left[\mathbf{I}_{n_2}\otimes\left(\mathbf{\Gamma}_0^{(2)}-\mathbf{\Gamma}_1^{(2)}\right)+\mathbf{J}_{n_2,n_2}\otimes\mathbf{\Gamma}_1^{(2)}\right]\right. \\ &\qquad \left.-\left(\mathbf{J}_{n_2,n_1}\otimes\mathbf{\Gamma}\right)\left(\mathbf{I}_{n_1}\otimes\mathbf{A}^{(1)}+\mathbf{J}_{n_1,n_1}\otimes\mathbf{B}^{(1)}\right)\left(\mathbf{J}_{n_1,n_2}\otimes\mathbf{\Gamma}\right)\right|, \\ &= \left|\mathbf{I}_{n_1}\otimes\left(\mathbf{\Gamma}_0^{(1)}-\mathbf{\Gamma}_1^{(1)}\right)+\mathbf{J}_{n_1,n_1}\otimes\mathbf{\Gamma}_1^{(1)}\right|\cdot\left|\mathbf{I}_{n_2}\otimes\mathbf{H}_{(2)1}+\mathbf{J}_{n_2,n_2}\otimes\mathbf{\Delta}_1^{(2)}\right|,\end{aligned}$$

where $\mathbf{H}_{(2)1}$ and $\mathbf{A}^{(2)}$ are given in (4) and (5) respectively, and

$$\mathbf{\Delta}_1^{(2)}=\mathbf{\Gamma}_1^{(2)}-n_1\mathbf{\Gamma}\left(\mathbf{A}^{(1)}+n_1\mathbf{B}^{(1)}\right)\mathbf{\Gamma}.$$

That is, $|\mathbf{\Gamma}_{\boldsymbol{x}}|$ is a product of two determinants obtained by formula (2).

2.2.1 Maximum Likelihood estimates of the mean vector and the covariance matrix for jointly equicorrelated samples

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ be a $(n_1 + n_2)m-$variate vector of random sample of size N from $\mathrm{N}(\boldsymbol{\mu_x}, \boldsymbol{\Gamma_x})$, with $\boldsymbol{\mu_x} = \left(\mathbf{1}_{n_1}^{'} \otimes \boldsymbol{\mu}^{(1)'}, \mathbf{1}_{n_2}^{'} \otimes \boldsymbol{\mu}^{(2)'}\right)'$ and with $\boldsymbol{\Gamma_x}$ given by (3). The following theorem gives the MLEs of $\boldsymbol{\mu_x}$ and $\boldsymbol{\Gamma_x}$.

Theorem 2 *Under the above set-up the MLE of* $\boldsymbol{\mu_x}$ *is*

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{x}} = \left(\mathbf{1}_{n_1}^{'} \otimes \overline{\boldsymbol{x}}^{(1)\prime}, \mathbf{1}_{n_2}^{'} \otimes \overline{\boldsymbol{x}}^{(2)\prime}\right)',$$

where

$$\overline{\boldsymbol{x}}^{(i)} = \frac{1}{Nn_i} \sum_{h=1}^{N} \sum_{j=1}^{n_i} \boldsymbol{x}_{h,j}^{(i)}.$$

The MLE of $\boldsymbol{\Gamma_x}$ *is*

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{J}_{n_1,n_2} \otimes \widehat{\boldsymbol{\Gamma}} \\ \mathbf{J}_{n_2,n_1} \otimes \widehat{\boldsymbol{\Gamma}} & \mathbf{C}_2 \end{pmatrix},$$

where

$$\mathbf{C}_i = \mathbf{I}_{n_i} \otimes \left(\widehat{\boldsymbol{\Gamma}}_0^{(i)} - \widehat{\boldsymbol{\Gamma}}_1^{(i)}\right) + \mathbf{J}_{n_i n_i} \otimes \widehat{\boldsymbol{\Gamma}}_1^{(i)},\ i = 1, 2, \tag{8}$$

with

$$\widehat{\boldsymbol{\Gamma}}_0^{(i)} = \frac{1}{Nn_i} \sum_{h=1}^{N} \sum_{v=1}^{n_i} \left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right) \left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)', \tag{9}$$

$$\widehat{\boldsymbol{\Gamma}}_1^{(i)} = \frac{1}{Nn_i\left(n_i - 1\right)} \sum_{h=1}^{N} \sum_{v=1}^{n_i} \sum_{v \neq w=1}^{n_i} \left(\boldsymbol{x}_{h,w}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right) \left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)', \tag{10}$$

and

$$\widehat{\boldsymbol{\Gamma}} = \frac{1}{Nn_1 n_2} \sum_{h=1}^{N} \sum_{r=1}^{n_2} \sum_{j=1}^{n_1} \left(\boldsymbol{x}_{h,j}^{(1)} - \overline{\boldsymbol{x}}^{(1)}\right) \left(\boldsymbol{x}_{h,r}^{(2)} - \overline{\boldsymbol{x}}^{(2)}\right)'.$$

When $\boldsymbol{\Gamma} = \mathbf{0}$, the MLE of $\boldsymbol{\Gamma_x}$ reduces to

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{pmatrix}, \tag{11}$$

where $\mathbf{C}_i$, $i = 1, 2$, is given by (8), with $\widehat{\boldsymbol{\Gamma}}_0^{(i)}$ and $\widehat{\boldsymbol{\Gamma}}_1^{(i)}$ given by (9) and (10) respectively.

Moreover, when $\boldsymbol{\Gamma} = \mathbf{0}$ and the equicorrelated parameters of both populations are the same, that is, $\boldsymbol{\Gamma}_0^{(1)} = \boldsymbol{\Gamma}_0^{(2)} \doteq \boldsymbol{\Gamma}_0$ and $\boldsymbol{\Gamma}_1^{(1)} = \boldsymbol{\Gamma}_1^{(2)} \doteq \boldsymbol{\Gamma}_1$, $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}$ is also given by (11) Thus, it is not necessarily true that $\widehat{\boldsymbol{\Gamma}}_0^{(1)} = \widehat{\boldsymbol{\Gamma}}_0^{(2)}$ and $\widehat{\boldsymbol{\Gamma}}_1^{(1)} = \widehat{\boldsymbol{\Gamma}}_1^{(2)}$, and so the structure of $\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}$ is different from the structure of $\boldsymbol{\Gamma_x}$. To avoid this Leiva (2007) suggested to use the following:

$$\widehat{\boldsymbol{\Gamma}}_{\boldsymbol{x}}^{*} = \begin{pmatrix} \mathbf{C}_1^{*} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{*} \end{pmatrix},$$

where

$$\mathbf{C}_i^* = \mathbf{I}_{n_i} \otimes \left(\widehat{\boldsymbol{\Gamma}}_0^* - \widehat{\boldsymbol{\Gamma}}_1^*\right) + \mathbf{J}_{n_i n_i} \otimes \widehat{\boldsymbol{\Gamma}}_1^*, \quad i = 1, 2,$$

and $\widehat{\boldsymbol{\Gamma}}_0^*$ and $\widehat{\boldsymbol{\Gamma}}_1^*$ are given by

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}_0^* &= \frac{n_1 \widehat{\boldsymbol{\Gamma}}_0^{(1)} + n_2 \widehat{\boldsymbol{\Gamma}}_0^{(2)}}{n_1 + n_2}, \\ &= \frac{1}{N(n_1+n_2)} \sum_{h=1}^{N} \sum_{i=1}^{2} \sum_{v=1}^{n_i} \left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)',\end{aligned}$$

and

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}_1^* &= \frac{n_1(n_1-1)\widehat{\boldsymbol{\Gamma}}_1^{(1)} + n_2(n_2-1)\widehat{\boldsymbol{\Gamma}}_1^{(2)}}{n_1(n_1-1) + n_2(n_2-1)}, \\ &= \frac{1}{N[n_1(n_1-1) + n_2(n_2-1)]} \sum_{h=1}^{N} \sum_{i=1}^{2} \sum_{v=1}^{n_i} \sum_{v \neq w=1}^{n_i} \left(\boldsymbol{x}_{h,w}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_{h,v}^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)'.\end{aligned}$$

3 Discriminant analysis with equally correlated vectors

Let $\boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_{n_i}^{(i)}$ be $m-$variate vector training samples of sizes n_i from population Π_i, $i = 1, 2$, where $\boldsymbol{x}_j^{(i)} = (x_{j,1}^{(i)}, \ldots, x_{j,m}^{(i)})'$ for $j = 1, \ldots, n_i$. We define the mn_i-variate vector $\boldsymbol{x}^{(i)}$ as $\boldsymbol{x}^{(i)} = (\boldsymbol{x}_1^{(i)\prime}, \ldots, \boldsymbol{x}_{n_i}^{(i)\prime})'$. The objective is to classify a new individual with measurement vector $\boldsymbol{x}_0 = (x_{0,1}, \ldots, x_{0,m})'$ to one of the populations, using the training samples $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$. The basic assumption in the traditional discriminant analysis is that the vectors $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_{n_1}^{(1)}, \boldsymbol{x}_0, \boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_{n_2}^{(2)}$ are all independent. However, as discussed in the introduction, this assumption may not be appropriate in many cases, as certain type of dependency may possibly exist among these vectors. The main difficulty in these cases is, how to incorporate the dependency in the formulation of the discrimination problem. Even though in this paper we only consider that these vectors have the special kind of dependency, such as jointly equicorrelation, this heuristic idea can also be used with any other type of dependencies that are present in the data.

3.1 Classification with jointly equicorrelated training vectors.

In this section we derive the Bayesian decision rule to classify a vector of measurements $\boldsymbol{x}_o$ into one of the populations Π_1 and Π_2 using the two sets of training samples $\boldsymbol{x}^{(1)} = (\boldsymbol{x}_1^{(1)\prime}, \ldots, \boldsymbol{x}_{n_i}^{(1)\prime})'$ and $\boldsymbol{x}^{(2)} = (\boldsymbol{x}_1^{(2)\prime}, \ldots, \boldsymbol{x}_{n_i}^{(2)\prime})'$ from the two populations Π_1 and Π_2 respectively. We assume that $\boldsymbol{x}^{(1)}, \boldsymbol{x}_0, \boldsymbol{x}^{(2)}$ are jointly equicorrelated, where the vector $\boldsymbol{x}_o$ has the same parameters as the training vectors of the population it belongs. We also assume that the covariance matrix $\boldsymbol{\Gamma}$ between the vectors of two populations is $\mathbf{0}$, i.e. we assume that the two sets of samples from the two populations are uncorrelated. More precisely, let $\boldsymbol{x} = (\boldsymbol{x}_1^{(1)\prime}, \ldots, \boldsymbol{x}_{n_1}^{(1)\prime}, \boldsymbol{x}_0', \boldsymbol{x}_1^{(2)\prime}, \ldots, \boldsymbol{x}_{n_2}^{(2)\prime})' = (\boldsymbol{x}^{(1)\prime}, \boldsymbol{x}_0', \boldsymbol{x}^{(2)\prime})'$ be the $(n_1 + 1 + n_2)\, m-$variate vector with mean $\boldsymbol{\mu}_{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{x}}$. If the vector $\boldsymbol{x}_0$ belongs to population Π_1 then

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \left(\mathbf{1}'_{n_1+1} \otimes \boldsymbol{\mu}^{(1)'}, \mathbf{1}'_{n_2} \otimes \boldsymbol{\mu}^{(2)\prime}\right)' \doteq \boldsymbol{\mu}_{\boldsymbol{x}(1)},$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1}^{(1)} & \mathbf{1}_{n_1} \otimes \boldsymbol{\Gamma}_1^{(1)} & \mathbf{0} \\ \mathbf{1}'_{n_1} \otimes \boldsymbol{\Gamma}_1^{(1)} & \boldsymbol{\Gamma}_0^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{n_2}^{(2)} \end{bmatrix} \doteq \boldsymbol{\Gamma}_{\boldsymbol{x}(1)}.$$

And, if the vector $\boldsymbol{x}_0$ belongs to population Π_2 then

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \left(\mathbf{1}'_{n_1} \otimes \boldsymbol{\mu}^{(1)'}, \mathbf{1}'_{n_2+1} \otimes \boldsymbol{\mu}^{(2)\prime}\right)' \doteq \boldsymbol{\mu}_{\boldsymbol{x}(2)},$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1}^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_0^{(2)} & \mathbf{1}'_{n_2} \otimes \boldsymbol{\Gamma}_1^{(2)} \\ \mathbf{0} & \mathbf{1}_{n_2} \otimes \boldsymbol{\Gamma}_1^{(2)} & \boldsymbol{\Gamma}_{n_2}^{(2)} \end{bmatrix} \doteq \boldsymbol{\Gamma}_{\boldsymbol{x}(2)}.$$

Therefore, assuming normality,

$$\boldsymbol{x} \sim \begin{cases} \mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{x}(1)}, \boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right) & \text{if} \quad \boldsymbol{x}_0 \in \Pi_1 \\ \mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{x}(2)}, \boldsymbol{\Gamma}_{\boldsymbol{x}(2)}\right) & \text{if} \quad \boldsymbol{x}_0 \in \Pi_2 \end{cases}.$$

3.1.1 Known parameters

We assume $\boldsymbol{\Gamma}_{\boldsymbol{x}(1)} \neq \boldsymbol{\Gamma}_{\boldsymbol{x}(2)}$. Thus, under the assumptions of equal prior probabilities and misclassification costs for both populations, the (theoretical) Bayesian classification rule is given by

$$\begin{aligned} \boldsymbol{x}_0 \in \Pi_1 \quad &\Longleftrightarrow \quad \boldsymbol{x} \sim \mathrm{N}\left(\boldsymbol{\mu}_{\boldsymbol{x}(1)}, \boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right), \\ &\Longleftrightarrow \quad q(\boldsymbol{x}) \doteq -\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} - \boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1}\right)\boldsymbol{x} + \left(\boldsymbol{\mu}'_{\boldsymbol{x}(1)}\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} - \boldsymbol{\mu}_{\boldsymbol{x}(2)}\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1}\right)\boldsymbol{x} \geq k, \end{aligned}$$

where the threshold k is given by

$$k = \frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right|}{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}\right|}\right) + \frac{1}{2}\left(\boldsymbol{\mu}'_{\boldsymbol{x}(1)}\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1}\boldsymbol{\mu}_{\boldsymbol{x}(1)} - \boldsymbol{\mu}_{\boldsymbol{x}(2)}\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1}\boldsymbol{\mu}_{\boldsymbol{x}(2)}\right).$$

Now, since $\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}$ and $\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}$ have the forms

$$\boldsymbol{\Gamma}_{\boldsymbol{x}(1)} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1+1}^{(1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{n_2}^{(2)} \end{bmatrix},$$

and

$$\boldsymbol{\Gamma}_{\boldsymbol{x}(2)} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1}^{(1)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{n_2+1}^{(2)} \end{bmatrix},$$

we have

$$\begin{aligned} \left|\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right| &= \left|\boldsymbol{\Gamma}_{n_1+1}^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_{n_2}^{(2)}\right|, \\ &= \left|\boldsymbol{\Gamma}_0^{(1)} - \boldsymbol{\Gamma}_1^{(1)}\right|^{n_1} \left|\boldsymbol{\Gamma}_0^{(2)} - \boldsymbol{\Gamma}_1^{(2)}\right|^{n_2-1} \cdot \left|\boldsymbol{\Gamma}_0^{(1)} + n_1\boldsymbol{\Gamma}_1^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(2)} + (n_2-1)\boldsymbol{\Gamma}_1^{(2)}\right|, \end{aligned}$$

and

$$
\begin{aligned}
\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}\right| &= \left|\boldsymbol{\Gamma}_{n_1}^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_{n_2+1}^{(2)}\right|, \\
&= \left|\boldsymbol{\Gamma}_0^{(1)} - \boldsymbol{\Gamma}_1^{(1)}\right|^{n_1-1} \left|\boldsymbol{\Gamma}_0^{(2)} - \boldsymbol{\Gamma}_1^{(2)}\right|^{n_2} \cdot \left|\boldsymbol{\Gamma}_0^{(1)} + (n_1-1)\boldsymbol{\Gamma}_1^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(2)} + n_2\boldsymbol{\Gamma}_1^{(2)}\right|.
\end{aligned}
$$

Therefore,

$$
\frac{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right|}{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}\right|} = \frac{\left|\boldsymbol{\Gamma}_0^{(1)} - \boldsymbol{\Gamma}_1^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(1)} + n_1\boldsymbol{\Gamma}_1^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(2)} + (n_2-1)\boldsymbol{\Gamma}_1^{(2)}\right|}{\left|\boldsymbol{\Gamma}_0^{(2)} - \boldsymbol{\Gamma}_1^{(2)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(1)} + (n_1-1)\boldsymbol{\Gamma}_1^{(1)}\right| \cdot \left|\boldsymbol{\Gamma}_0^{(2)} + n_2\boldsymbol{\Gamma}_1^{(2)}\right|}.
$$

It is also clear that

$$
\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1+1}^{(1)-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{n_2}^{(2)-1} \end{bmatrix},
$$

and

$$
\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1} = \begin{bmatrix} \boldsymbol{\Gamma}_{n_1}^{(1)-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}_{n_2+1}^{(2)-1} \end{bmatrix},
$$

and these can be written as

$$
\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} = \begin{bmatrix} \mathbf{I}_{n_1} \otimes \mathbf{A}^{(1)} + \mathbf{J}_{n_1} \otimes \mathbf{D}_{n_1+1}^{(1)} & \mathbf{1}_{n_1} \otimes \mathbf{D}_{n_1+1}^{(1)} & \mathbf{0} \\ \mathbf{1}_{n_1}' \otimes \mathbf{D}_{n_1+1}^{(1)} & \mathbf{A}^{(1)} + \mathbf{D}_{n_1+1}^{(1)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n_2} \otimes \mathbf{A}^{(2)} + \mathbf{J}_{n_2} \otimes \mathbf{D}_{n_2}^{(2)} \end{bmatrix},
$$

and

$$
\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1} = \begin{bmatrix} \mathbf{I}_{n_1} \otimes \mathbf{A}^{(1)} + \mathbf{J}_{n_1} \otimes \mathbf{D}_{n_1}^{(1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{(2)} + \mathbf{D}_{n_2+1}^{(2)} & \mathbf{1}_{n_2}' \otimes \mathbf{D}_{n_2+1}^{(2)} \\ \mathbf{0} & \mathbf{1}_{n_2} \otimes \mathbf{D}_{n_2+1}^{(2)} & \mathbf{I}_{n_2} \otimes \mathbf{A}^{(2)} + \mathbf{J}_{n_2} \otimes \mathbf{D}_{n_2+1}^{(2)} \end{bmatrix}.
$$

where $\mathbf{A}^{(i)}$, $i = 1, 2$, are given in (5), and $\mathbf{D}_{n_i}^{(i)}$, $\mathbf{D}_{n_i+1}^{(i)}$ are given in (6). Now, writing $\boldsymbol{x} = (\boldsymbol{x}_1^{(1)\prime}, \ldots, \boldsymbol{x}_{n_1}^{(1)\prime}, \boldsymbol{x}_0', \boldsymbol{x}_1^{(2)\prime}, \ldots, \boldsymbol{x}_{n_2}^{(2)\prime})'$, we have

$$
\begin{aligned}
q(\boldsymbol{x}) &\doteq -\frac{1}{2}\boldsymbol{x}'\left(\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} - \boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1}\right)\boldsymbol{x} + \left(\boldsymbol{\mu}_{\boldsymbol{x}(1)}'\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}^{-1} - \boldsymbol{\mu}_{\boldsymbol{x}(2)}'\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}^{-1}\right)\boldsymbol{x} \\
&= -\frac{1}{2}n_1^2\overline{\boldsymbol{x}}^{(1)\prime}\left(\mathbf{D}_{n_1+1}^{(1)} - \mathbf{D}_{n_1}^{(1)}\right)\overline{\boldsymbol{x}}^{(1)} + \frac{1}{2}n_2^2\overline{\boldsymbol{x}}^{(2)\prime}\left(\mathbf{D}_{n_2+1}^{(2)} - \mathbf{D}_{n_2}^{(2)}\right)\overline{\boldsymbol{x}}^{(2)} \\
&\quad + n_1^2\boldsymbol{\mu}^{(1)\prime}\left(\mathbf{D}_{n_1+1}^{(1)} - \mathbf{D}_{n_1}^{(1)}\right)\overline{\boldsymbol{x}}^{(1)} - n_2^2\boldsymbol{\mu}^{(2)\prime}\left(\mathbf{D}_{n_2+1}^{(2)} - \mathbf{D}_{n_2}^{(2)}\right)\overline{\boldsymbol{x}}^{(2)} \\
&\quad + n_1\boldsymbol{\mu}_1^{(1)\prime}\mathbf{D}_{n_1+1}^{(1)}\overline{\boldsymbol{x}}^{(1)} - n_2\boldsymbol{\mu}^{(2)\prime}\mathbf{D}_{n_2+1}^{(2)}\overline{\boldsymbol{x}}^{(2)} \\
&\quad - \frac{1}{2}\boldsymbol{x}_0'\left(\left(\mathbf{A}^{(1)} + \mathbf{D}_{n_1+1}^{(1)}\right) - \left(\mathbf{A}^{(2)} + \mathbf{D}_{n_2+1}^{(2)}\right)\right)\boldsymbol{x}_0 \\
&\quad + n_1\left(\boldsymbol{\mu}^{(1)\prime} - \overline{\boldsymbol{x}}^{(1)\prime}\right)\mathbf{D}_{n_1+1}^{(1)}\boldsymbol{x}_0 - n_2\left(\boldsymbol{\mu}^{(2)\prime} - \overline{\boldsymbol{x}}^{(2)\prime}\right)\mathbf{D}_{n_2+1}^{(2)}\boldsymbol{x}_0 \\
&\quad + \boldsymbol{\mu}^{(1)\prime}\mathbf{A}^{(1)}\boldsymbol{x}_0 - \boldsymbol{\mu}^{(2)\prime}\mathbf{A}^{(2)}\boldsymbol{x}_0 + \boldsymbol{\mu}^{(1)\prime}\mathbf{D}_{n_1+1}^{(1)}\boldsymbol{x}_0 - \boldsymbol{\mu}^{(2)\prime}\mathbf{D}_{n_2+1}^{(2)}\boldsymbol{x}_0,
\end{aligned}
$$

$$
\begin{aligned}
\text{and} \quad k &= \frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(1)}\right|}{\left|\boldsymbol{\Gamma}_{\boldsymbol{x}(2)}\right|}\right)+\frac{1}{2}\left(\boldsymbol{\mu}'_{\boldsymbol{x}(1)}\boldsymbol{\Gamma}^{-1}_{\boldsymbol{x}(1)}\boldsymbol{\mu}_{\boldsymbol{x}(1)}-\boldsymbol{\mu}_{\boldsymbol{x}(2)}\boldsymbol{\Gamma}^{-1}_{\boldsymbol{x}(2)}\boldsymbol{\mu}_{\boldsymbol{x}(2)}\right)\\
&= \frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Gamma}_0^{(1)}-\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(1)}+n_1\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(2)}+(n_2-1)\boldsymbol{\Gamma}_1^{(2)}\right|}{\left|\boldsymbol{\Gamma}_0^{(2)}-\boldsymbol{\Gamma}_1^{(2)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(1)}+(n_1-1)\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(2)}+n_2\boldsymbol{\Gamma}_1^{(2)}\right|}\right)\\
&\quad+\frac{1}{2}\boldsymbol{\mu}^{(1)\prime}\left(n_1^2\mathbf{D}^{(1)}_{n_1+1}+2n_1\mathbf{D}^{(1)}_{n_1+1}-n_1^2\mathbf{D}^{(1)}_{n_1}+\left(\mathbf{A}^{(1)}+\ \mathbf{D}^{(1)}_{n_1+1}\right)\right)\boldsymbol{\mu}^{(1)}\\
&\quad-\frac{1}{2}\boldsymbol{\mu}^{(2)\prime}\left(n_2^2\mathbf{D}^{(2)}_{n_2+1}+2n_2\mathbf{D}^{(2)}_{n_2+1}-n_2^2\mathbf{D}^{(2)}_{n_2}+\left(\mathbf{A}^{(2)}+\ \mathbf{D}^{(2)}_{n_2+1}\right)\right)\boldsymbol{\mu}^{(2)}.
\end{aligned}
$$

Finally, in the inequality $q(\boldsymbol{x})\geq k$, we can have only those terms involving $\boldsymbol{x}_0$ on the left hand side and obtain the inequality $t(\boldsymbol{x})\geq c$, where

$$
\begin{aligned}
t(\boldsymbol{x}) &= -\frac{1}{2}\boldsymbol{x}_0'\left(\mathbf{A}^{(1)}-\mathbf{A}^{(2)}\right)\boldsymbol{x}_0-\frac{1}{2}\boldsymbol{x}_0'\left(\mathbf{D}^{(1)}_{n_1+1}-\ \mathbf{D}^{(2)}_{n_2+1}\right)\boldsymbol{x}_0\\
&\quad -n_1\overline{\boldsymbol{x}}^{(1)\prime}\ \mathbf{D}^{(1)}_{n_1+1}\boldsymbol{x}_0+n_2\overline{\boldsymbol{x}}^{(2)\prime}\ \mathbf{D}^{(2)}_{n_2+1}\boldsymbol{x}_0+\boldsymbol{\mu}^{(1)\prime}\mathbf{A}^{(1)}\boldsymbol{x}_0\\
&\quad -\boldsymbol{\mu}^{(2)\prime}\mathbf{A}^{(2)}\boldsymbol{x}_0+(n_1+1)\,\boldsymbol{\mu}^{(1)\prime}\ \mathbf{D}^{(1)}_{n_1+1}\boldsymbol{x}_0-(n_2+1)\,\boldsymbol{\mu}^{(2)\prime}\ \mathbf{D}^{(2)}_{n_2+1}\boldsymbol{x}_0,
\end{aligned}
$$

$$
\begin{aligned}
\text{and} \quad c &= \frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Gamma}_0^{(1)}-\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(1)}+n_1\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(2)}+(n_2-1)\boldsymbol{\Gamma}_1^{(2)}\right|}{\left|\boldsymbol{\Gamma}_0^{(2)}-\boldsymbol{\Gamma}_1^{(2)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(1)}+(n_1-1)\boldsymbol{\Gamma}_1^{(1)}\right|\cdot\left|\boldsymbol{\Gamma}_0^{(2)}+n_2\boldsymbol{\Gamma}_1^{(2)}\right|}\right)\\
&\quad+\frac{1}{2}\boldsymbol{\mu}^{(1)\prime}\left(n_1^2\mathbf{D}^{(1)}_{n_1+1}+2n_1\mathbf{D}^{(1)}_{n_1+1}-n_1^2\mathbf{D}^{(1)}_{n_1}+\left(\mathbf{A}^{(1)}+\ \mathbf{D}^{(1)}_{n_1+1}\right)\right)\boldsymbol{\mu}^{(1)}\\
&\quad-\frac{1}{2}\boldsymbol{\mu}^{(2)\prime}\left(n_2^2\mathbf{D}^{(2)}_{n_2+1}+2n_2\mathbf{D}^{(2)}_{n_2+1}-n_2^2\mathbf{D}^{(2)}_{n_2}+\left(\mathbf{A}^{(2)}+\ \mathbf{D}^{(2)}_{n_2+1}\right)\right)\boldsymbol{\mu}^{(2)}\\
&\quad+\frac{1}{2}n_1^2\overline{\boldsymbol{x}}^{(1)\prime}\left(\mathbf{D}^{(1)}_{n_1+1}-\mathbf{D}^{(1)}_{n_1}\right)\overline{\boldsymbol{x}}^{(1)}-\frac{1}{2}n_2^2\overline{\boldsymbol{x}}^{(2)\prime}\left(\mathbf{D}^{(2)}_{n_2+1}-\mathbf{D}^{(2)}_{n_2}\right)\overline{\boldsymbol{x}}^{(2)}\\
&\quad-n_1^2\boldsymbol{\mu}^{(1)\prime}\left(\mathbf{D}^{(1)}_{n_1+1}-\mathbf{D}^{(1)}_{n_1}\right)\overline{\boldsymbol{x}}^{(1)}+n_2^2\boldsymbol{\mu}^{(2)\prime}\left(\mathbf{D}^{(2)}_{n_2+1}-\mathbf{D}^{(2)}_{n_2}\right)\overline{\boldsymbol{x}}^{(2)}\\
&\quad-n_1\boldsymbol{\mu}^{(1)\prime}\mathbf{D}^{(1)}_{n_1+1}\overline{\boldsymbol{x}}^{(1)}+n_2\boldsymbol{\mu}^{(2)\prime}\mathbf{D}^{(2)}_{n_2+1}\overline{\boldsymbol{x}}^{(2)}.
\end{aligned}
$$

Therefore, when the parameters are known, the (theoretical) Bayesian decision rule to classify a new vector $\boldsymbol{x}_0$ is

$$
\boldsymbol{x}_0\in\Pi_1 \iff q(\boldsymbol{x})\geq k \iff t(\boldsymbol{x})\geq c. \qquad (12)
$$

Note that $\overline{\boldsymbol{x}}^{(1)}$ and $\overline{\boldsymbol{x}}^{(2)}$ appear in the classification rule along with $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ when all the parameters are known.

Also note that when $\boldsymbol{\Gamma}_1^{(i)}=\mathbf{0}$, for $i=1,2$, i.e. when there are no correlations between the neighboring samples for both the populations, then $\mathbf{A}^{(i)}=\boldsymbol{\Gamma}_0^{(i)-1}$, and $\mathbf{D}^{(i)}_{n_i}=\mathbf{0}=\mathbf{D}^{(i)}_{n_i+1}$. As a result the theoretical classification rule no more depends on $\overline{\boldsymbol{x}}^{(1)}$ and $\overline{\boldsymbol{x}}^{(2)}$, and $t(\boldsymbol{x})=t(\boldsymbol{x}_0)$ reduces to

$$
t(\boldsymbol{x}_0)=-\frac{1}{2}\boldsymbol{x}_0'\left(\boldsymbol{\Gamma}_0^{(1)-1}-\boldsymbol{\Gamma}_0^{(2)-1}\right)\boldsymbol{x}_0+\left(\boldsymbol{\mu}^{(1)\prime}\boldsymbol{\Gamma}_0^{(1)-1}-\boldsymbol{\mu}^{(2)\prime}\boldsymbol{\Gamma}_0^{(2)-1}\right)'\boldsymbol{x}_0,
$$

and the threshold c reduces to

$$
c=\frac{1}{2}\ln\left(\frac{\left|\boldsymbol{\Gamma}_0^{(1)}\right|}{\left|\boldsymbol{\Gamma}_0^{(2)}\right|}\right)+\frac{1}{2}\left(\boldsymbol{\mu}^{(1)\prime}\boldsymbol{\Gamma}_0^{(1)-1}\boldsymbol{\mu}^{(1)}-\boldsymbol{\mu}^{(2)\prime}\boldsymbol{\Gamma}_0^{(2)-1}\boldsymbol{\mu}^{(2)}\right).
$$

Therefore, we see that the classification rule (12) reduces to the traditional (theoretical) quadratic classification rule under the uncorrelated random samples assumption. In particular when the equicorrelation parameters are same for both the populations, i.e when

$$\begin{aligned} \mathbf{\Gamma}_0^{(1)} &= \mathbf{\Gamma}_0^{(2)} \doteq \mathbf{\Gamma}_0, \\ \text{and} \quad \mathbf{\Gamma}_1^{(1)} &= \mathbf{\Gamma}_1^{(2)} \doteq \mathbf{\Gamma}_1, \end{aligned}$$

and when $\mathbf{\Gamma}_1 = \mathbf{0}$, the classification classification rule (12) becomes Fisher's linear classification rule, that is,

$$\begin{aligned} t\left(\boldsymbol{x}_0\right) &= \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)' \mathbf{\Gamma}_0^{-1} \boldsymbol{x}_0, \\ \text{and} \quad c &= \frac{1}{2}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)' \mathbf{\Gamma}_0^{-1}\left(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2\right). \end{aligned}$$

3.1.2 Unknown parameters

In this case also we assume $\mathbf{\Gamma}_{\boldsymbol{x}(1)} \neq \mathbf{\Gamma}_{\boldsymbol{x}(2)}$. We also assume that all the parameters are unknown. To obtain the sample classification rule we replace $\mathbf{A}^{(i)}$, $\mathbf{D}_{n_i+1}^{(i)}$ and $\mathbf{D}_{n_i}^{(i)}$ by their MLEs $\widehat{\mathbf{A}}^{(i)}$, $\widehat{\mathbf{D}}_{n_i+1}^{(i)}$ and $\widehat{D}_{n_i}^{(i)}$ in the expressions of $t\left(\boldsymbol{x}\right)$ and c. The estimates $\widehat{\mathbf{A}}^{(i)}$, $\widehat{\mathbf{D}}_{n_i+1}^{(i)}$ and $\widehat{D}_{n_i}^{(i)}$ are obtained from (5), (6) and (7) by replacing the parameters $\boldsymbol{\mu}^{(i)}$, $\mathbf{\Gamma}_0^{(i)}, \mathbf{\Gamma}_1^{(i)}$ by their ML estimates

$$\widehat{\boldsymbol{\mu}}^{(i)} = \overline{\boldsymbol{x}}^{(i)} = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{x}_j^{(i)} \quad \text{for} \quad i = 1, 2,$$

$$\begin{aligned} \widehat{\mathbf{\Gamma}}_0^{(i)} &= \frac{1}{n_i}\sum_{v=1}^{n_i}\left(\boldsymbol{x}_v^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_v^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)', \\ \text{and} \quad \widehat{\mathbf{\Gamma}}_1^{(i)} &= \frac{1}{n_i\left(n_i - 1\right)}\sum_{v=1}^{n_i}\sum_{v \neq w=1}^{n_i}\left(\boldsymbol{x}_w^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)\left(\boldsymbol{x}_v^{(i)} - \overline{\boldsymbol{x}}^{(i)}\right)', \end{aligned}$$

respectively. Then, the sample Bayesian decision rule to classify a new measurement vector $\boldsymbol{x}_0$ when parameters are unknown is given by

$$\boldsymbol{x}_0 \in \Pi_1 \iff \widehat{q}\left(\boldsymbol{x}\right) \geq \widehat{k} \iff \widehat{t}\left(\boldsymbol{x}\right) \geq \widehat{c},$$

where

$$\begin{aligned} \widehat{t}\left(\boldsymbol{x}_0\right) &= \overline{\boldsymbol{x}}^{(1)\prime}\left(\widehat{\mathbf{A}}^{(1)} + \widehat{\mathbf{D}}_{n_1+1}^{(1)}\right)\boldsymbol{x}_0 - \overline{\boldsymbol{x}}^{(2)\prime}\left(\widehat{\mathbf{A}}^{(2)} + \widehat{\mathbf{D}}_{n_2+1}^{(2)}\right)\boldsymbol{x}_0 \\ &\quad -\frac{1}{2}\boldsymbol{x}_0'\left(\left(\widehat{\mathbf{A}}^{(1)} + \widehat{\mathbf{D}}_{n_1+1}^{(1)}\right) - \left(\widehat{\mathbf{A}}^{(2)} + \widehat{\mathbf{D}}_{n_2+1}^{(2)}\right)\right)\boldsymbol{x}_0, \\ \text{and} \quad \widehat{c} &= \frac{1}{2}\ln\left(\frac{\left|\widehat{\mathbf{\Gamma}}_0^{(1)} - \widehat{\mathbf{\Gamma}}_1^{(1)}\right| \cdot \left|\widehat{\mathbf{\Gamma}}_0^{(1)} + n_1\widehat{\mathbf{\Gamma}}_1^{(1)}\right| \cdot \left|\widehat{\mathbf{\Gamma}}_0^{(2)} + (n_2 - 1)\widehat{\mathbf{\Gamma}}_1^{(2)}\right|}{\left|\widehat{\mathbf{\Gamma}}_0^{(2)} - \widehat{\mathbf{\Gamma}}_1^{(2)}\right| \cdot \left|\widehat{\mathbf{\Gamma}}_0^{(1)} + (n_1 - 1)\widehat{\mathbf{\Gamma}}_1^{(1)}\right| \cdot \left|\widehat{\mathbf{\Gamma}}_0^{(2)} + n_2\widehat{\mathbf{\Gamma}}_1^{(2)}\right|}\right) \\ &\quad +\frac{1}{2}\overline{\boldsymbol{x}}^{(1)\prime}\left(\widehat{\mathbf{A}}^{(1)} + \widehat{\mathbf{D}}_{n_1+1}^{(1)}\right)\overline{\boldsymbol{x}}^{(1)} - \frac{1}{2}\overline{\boldsymbol{x}}^{(2)\prime}\left(\widehat{\mathbf{A}}^{(2)} + \widehat{\mathbf{D}}_{n_2+1}^{(2)}\right)\overline{\boldsymbol{x}}^{(2)}. \end{aligned}$$

When the vectors $\boldsymbol{x}_1^{(1)}, \ldots, \boldsymbol{x}_{n_1}^{(1)}, \boldsymbol{x}_0, \boldsymbol{x}_1^{(2)}, \ldots, \boldsymbol{x}_{n_2}^{(2)}$ are uncorrelated, that is, when $\boldsymbol{\Gamma}_1^{(i)} = \mathbf{0}$, the corresponding sample classification rule reduces to

$$\begin{aligned} \widehat{t}\,(\boldsymbol{x}_0) &= \left[\overline{\boldsymbol{x}}^{(1)\prime}\widehat{\boldsymbol{\Gamma}}_0^{(1)-1} - \overline{\boldsymbol{x}}^{(2)\prime}\widehat{\boldsymbol{\Gamma}}_0^{(2)-1}\right]\boldsymbol{x}_0 - \frac{1}{2}\boldsymbol{x}_0'\left(\widehat{\boldsymbol{\Gamma}}_0^{(1)-1} - \widehat{\boldsymbol{\Gamma}}_0^{(2)-1}\right)\boldsymbol{x}_0, \\ \text{and} \quad \widehat{c} &= \frac{1}{2}\ln\left(\frac{\left|\widehat{\boldsymbol{\Gamma}}_0^{(1)}\right|}{\left|\widehat{\boldsymbol{\Gamma}}_0^{(2)}\right|}\right) + \frac{1}{2}\overline{\boldsymbol{x}}^{(1)\prime}\widehat{\boldsymbol{\Gamma}}_0^{(1)-1}\overline{\boldsymbol{x}}^{(1)} - \frac{1}{2}\overline{\boldsymbol{x}}^{(2)\prime}\widehat{\boldsymbol{\Gamma}}_0^{(2)-1}\overline{\boldsymbol{x}}^{(2)}. \end{aligned}$$

And, this is the traditional sample quadratic classification rule when all the samples are uncorrelated in both the populations.

4 Conclusions

This study presents a new approach for the generalization of the traditional classification rules. The new classification rule can be used when the assumption of uncorrelated training samples is violated. The generalization of the classification rule for more than two populations is straightforward. The extension of the proposed classification rule when $\boldsymbol{\Gamma} \neq \mathbf{0}$ is under progress, and we will report it in a future correspondence. The heuristic idea of incorporating joint equicorrelation among the neighboring sample vectors can easily be applied to many other types of dependence such as classification of time series.

References

Barghava, R.P., Srivastava, M.S., 1973. On Tukey's confidence intervals for the contrasts in the means of the intraclass correlation model. J.R. Statist. Soc. B 34, 147-152.

Bartlett, M.S., 1951. An inverse matrix adjustment arising in discriminant analysis. Annals of Mathematical Statististics 22(1), 107-111.

Basu, J.P., Odell, P.L., 1974. Effect of intraclass correlation among training samples on the misclassification probabilities of Bayes' procedure. Pattern Recognition 6, 13-16.

Bhandary, M., Alam, M.K., 2000. Test for the equality of intraclass correlation coefficients under unequal family sizes for several populations. Communications in Statistics-Theory and Methods 29, 755-768.

Donner, A., Bull, S., 1983. Inferences concerning a common intraclass correlation coefficient. Biometrics, 39, 771-775.

Donner, A., Zou, G., 2002. Testing the equality of dependent intraclass correlation coefficients. Statistician, 51, 367-379.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179-188.

Gupta, A.K., Nagar, D.K., 1987. Nonnull distribution of the likelihood ratio criterion for testing equality of covariance matrices under intraclass correlation model. Communications in

Statistics-Theory and Methods 16, 3323-3341.
Khan, S., Bhatti, M.I., 1998. Predictive inference on equicorrelated linear regressio models. Applied mathematics and Computation 95, 205-217.
Khatri, C.G., Pukkila, T.M., Rao, C.R., 1989. Testing intraclass correlation coefficient. Communications in Statistics - Simula 18, 755-768.
Konishi, S., Gupta, A.K., 1989. Testing the equality of several intraclass correlation coefficients. J. Statist. Planning and Inference 21, 93-105.
Leiva R., 2007. Linear discrimination with equicorrelated training vectors. Journal of Multivariate Analysis 98(2) 384-409.
Paul, S.R., Barnwal, R.K., 1990. Maximum likelihood estimation and a C(α) test for a common intraclass correlation. Statistician 39, 15-30.
Richards, J.A., Jia, X., 1999. Remote sensing and digital image analysis: an introduction. Springer, Berlin.
Ritter, G., Gallegos, M.T., 2002. Bayesian object identification: variants. Journal of Multivariate Analysis 81(2) 301-334.
Shoukri, M., Ward, R., 1984. On the estimation of the intraclass correlation. Communications in Statistics-Theory and Methods 13, 1239-1255.
Smith, J.R., Lewis, T.O., 1980. Determining the effects of the intraclass correlation on factorial experiments. Communications in Statistics-Theory and Methods 9, 1253-1364.
Viana, M.A.G., 1982. Combined estimators for the equicorelation coefficient. Communications in Statistics-Theory and Methods 11, 1483-1504.
Viana, M.A.G., 1994. Combined maximun likelihood estimates for the equicorelation coefficient. Biometrics 50, 813-820.
Young, D.J., Bhandary, M., 1998. Test for equality of intraclass correlation coefficients under unequal family sizes. Biometrics 54, 1363-1373.
Zerbe, G.O., Goldgar, D.E., 1980. Comparison of intraclass correlation coefficients with the ratio of two independent F statistics. Communications in Statistics-Theory and Methods 9(15), 1641-1655.