African Journal of Science and Technology (AJST)
Science and Engineering Series Vol. 12, No. 1, pp. 72 - 79

RELATIVE EFFICIENCY OF NON-PARAMETRIC ERROR RATE ESTIMATORS IN MULTI-GROUP LINEAR DISCRIMINANT ANALYSIS

[1]R. Glele KaKai, D. Pelz[2] and R. Palm[3]

[1]Faculty of Agronomic Sciences, University of Abomey-Calavi, 04 BP 1525, Cotonou, Benin.
[2]Department of Forest Biometry, University of Freiburg,
Tennenbacherstr. 4, D-79085, Freiburg (Germany)
[3]Gembloux Agricultural University, Avenue de la Faculté d'Agronomie 8, B-5030,

Email: gleleromain@yahoo.fr

ABSTRACT: *A Monte Carlo study was achieved to assess the relative efficiency of ten non-parametric error rate estimators in 2-, 3- and 5-group linear discriminant analysis. The simulation design took into account the number of variables (4, 6, 10, 18) together with the size sample n so that: n/p = 1.5, 2.5 and 5. Three values of the overlap, e of the populations were considered (e = 0.05, e = 0.1, e = 0.15) and their common distribution was Normal, Chi-square with 12, 8, and 4 df; the heteroscedasticity degree, Γ was measured by the value of the power function, 1-β of the homoscedasticity test related to (1-β = 0.05, 1-β = 0.4, 1-β = 0.6, 1-β = 0.8). For each combination of these factors, the actual error rate was empirically computed as well as the ten estimators. The efficiency parameter of the estimators was their relative error, bias and efficiency with regard to the actual error rate, empirically computed. The results showed the overall best performance e632 estimator. On the contrary, e0, epp, eppCV and eA recorded the lowest performance in terms of mean relative error and mean relative bias. The ranks of the estimators were not influenced by the number of groups but for high values of the later, the mean relative bias of the estimators tend to zero.*

Keywords: *Error rate; Estimation; Efficiency; Multi-group; Linear rule; Simulation.*

INTRODUCTION

Discriminant analysis is a statistical method of allocation of an unknown individual to one group, from at least two foreknown groups, by using a classification rule previously established on well-known individuals. A number of classification rules are available and the most used are linear, quadratic and logistic methods.

Many classification rules have been proposed in literature and the most common is the linear classification rule (Fisher, 1936).

Let us suppose g p-variate populations, $G_k\,(k=1,...,g)$, with mean vectors, $\boldsymbol{\mu}_k\,(k=1,...,g)$ and common covariance matrices, $\boldsymbol{\Sigma}$. The linear rule (LR) is a Normal-based classification rule for which $F = \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ (McLachlan, 1992):

$$\begin{aligned}&\mathrm{LR}\big(\boldsymbol{x}_i, \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})\big) = \\ &\ln(p_k / p_l) + \big(\boldsymbol{x}_i - 0.5(\boldsymbol{\mu}_k + \boldsymbol{\mu}_l)\big)' \,\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_l); \\ &(k, l = 1,...g \ ; \ k \neq l)\end{aligned} \tag{1.1}$$

The unknown observation vector $\boldsymbol{x}_i$ is assigned to G_k if:

$$\mathrm{LR}\big(\boldsymbol{x}_i, \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})\big) \leq 0 \qquad \forall\ l = 1,..., g\ ;\ l \neq k$$

In the case of data samples, LR can be established by replacing in (1.1) the parameters, $\boldsymbol{\mu}_k\,(k=1,...,g)$ and

Σ by their estimates, $\hat{\boldsymbol{\mu}}_k$ (k=1,...,g) and $\hat{\Sigma}_k$; $\hat{\Sigma}$ is considered in (1.1) as the estimated pooled covariance matrix of the k populations.

Whatever the rule established is, it is subject to a probability of misclassifications. Then, an actual error rate is associated with any classification rule established on data samples in order to evaluate its efficiency. In practice, it is impossible to precisely determine the actual error rate, because it is only computed on the actual parameters of the populations, which are usually unknown. To solve this problem, some parametric and non parametric estimators of the actual error rate were established (McLachlan, 1992). Parametric estimators were established for two normal homoscedastic groups and the actual error rates estimated, using some parameters related to the considered samples such as the estimated Mahalanobis distance between the two groups. On the contrary, non-parametric error rate estimators do not depend on any hypothesis of use and are based on resampling methods. For two-group discriminant analysis, many comparative studies of error rate estimators have been done in linear discriminant analysis, in order to deduce the ones that have the lowest errors compared with the theoretical actual error rate. A thorough review of these studies is provided by Schiavo and Hand (2000). However, in real world problems, more than two groups are often considered in discriminant analysis. This paper evaluates and compares, by simulation technique, the efficiency of ten non parametric error rate estimators for 2-, 3- and 5-groups submitted to linear discriminant analysis.

ACTUAL ERROR RATE

The actual error rate can be defined as the theoretical proportion of misclassified observations, obtained by validating a classification rule established on data samples to any other observation taken from the same populations. This error rate is useful in practice because it gives the expected misclassification rate when a previously established rule is used.

Let us assume two samples, E_1 and E_2 with p variables and common size n. The mean vectors and the pooled covariance matrix are $\bar{\boldsymbol{x}}_1$, $\bar{\boldsymbol{x}}_2$ and $\boldsymbol{S}$, respectively. Let us also suppose also that these samples are taken from normal populations, P_1 and P_2, with mean vector ($\boldsymbol{\mu}_k$ =1,2). The actual error rate specific to the group k, ec_k ($k=1,2$) and the overall actual error rate are given by McLachlan (1975):

$$ec_k = \Phi\left\{(-1)^k \frac{\left[\boldsymbol{\mu}_k - \frac{1}{2}(\bar{\boldsymbol{x}}_1 + \bar{\boldsymbol{x}}_2)\right]' \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}{\sqrt{(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \boldsymbol{S}^{-1} \Sigma \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)}}\right\} \quad (2.1)$$

and $ec = \sum_{k=1}^{2} p_k ec_k$

where p_k and Φ are, respectively, the prior probability related to the group and the cumulative function of the Normal distribution.

The relations (2.1) can only be used in two-group discriminant analysis when the linear rule is established on two normal homoscedastic populations. In the other cases, the actual error rate associated with a classification rule can be empirically computed, for two groups, by determining the proportion of misclassified observations when the rule is established on the samples E_1 and E_2 and validated on a couple of large samples, of size 10,000 for example.

ESTIMATION OF THE ACTUAL ERROR RATE

For more than two groups submitted to discriminant analysis, only non-parametric estimators can be used to assess the actual error rate associated with an established rule; parametric estimators were only conceived for two-group discriminant analysis. Ten non-parametric error rate estimators were considered in the study and presented below.

Resubstitution estimator, eA (Smith, 1947): i.e., proportion of misclassified observations when the rule was established and validated on the same samples.

Cross validation estimator, eCV (Lachenbruch, 1967): i.e., proportion of misclassified observations when gn discriminant analyses are done on gn -1 observations by removing, at each step, one observation and by allocating the removed observation to one of the considered groups on the basis of the rule established on the gn-1 observations.

eS_1 and eS_2 *Estimators* (Hand, 1986):

$$eS_1 = \frac{2n}{2n+1}eCV \text{ and } eS_2 = \frac{2n}{2n+3}eCV \quad (3.1)$$

epp Estimator (Fukunaga and Kessell, 1972):

$$epp = 1 - \frac{1}{gn}\sum_{i=1}^{gn}\max\left(\hat{\tau}_1(\boldsymbol{x}_i),\ldots,\tau_g(\boldsymbol{x}_i)\right) \quad (3.2)$$

The symbols $\hat{\tau}_k(\boldsymbol{x}_i)(k=1,\ldots,g)$ represent the posterior probability that an individual i of observations vector $\boldsymbol{x}_i$ belongs to population G_k and is defined as:

$$\hat{\tau}_k(\boldsymbol{x}_i) = \hat{f}_k(\boldsymbol{x}_i)/\sum_{l=1}^{g}\hat{f}_l(\boldsymbol{x}_i)$$

where $\hat{f}_k(\boldsymbol{x}_i)$ is the value of the estimated density function at $\boldsymbol{x}_i$ for population G_k.

eppcv Estimator (Fukunaga and Kessell, 1972): i.e., computed by using the relation (3.2) in which the posterior probabilities, $\hat{\tau}_k(\boldsymbol{x}_i)\,(k=1,\ldots,g)$ of the observations vector $\boldsymbol{x}_i$ was determined, using the classification rule established on gn-1 observations, the vector $\boldsymbol{x}_i$, being removed.

Jackknife estimator, (Quenouille, 1949): i.e., computed by realising discriminant analyses on gn-1 observations. For each sample of gn-1 observations, the observation being removed, the resubstitution estimator, $eA_{k(i)}$, specific to G_k $(k=1,\ldots,g)$, was computed. By assuming, $e\overline{A}_k$

$$e\overline{A}_k = \frac{1}{n}\sum_{i=1}^{n}eA_{k(i)} \quad (3.3)$$

the Jackknife estimator is computed as:

$$eJc = \frac{1}{g}\left(\sum_{k=1}^{g}\left(eA_k + (n-1)(eA_k - e\overline{A}_k)\right)\right) \quad (3.4)$$

where eA_k is the resubstitution estimator specific to and computed from the overall sample.

Ordinary bootstrap estimator, (Efron, 1983): i.e., computed on 100 bootstrap samples, a sample of size *n* being taken with replacement in each initial sample of size *n*. For each bootstrap sample, the classification rule is established and the resubstitution estimator, eA^*_{kj} $k=1,\ldots,g;\ j=1,\ldots,100$ specific to G_k was computed. The same rule is also used to compute the proportions, r^*_k of misclassified observations, the rule being validated on the initial sample. The bias, $b_k\,(k=1,\ldots,g)$ of eA^*_{kj} is computed as follows:

$$b_k = \frac{1}{100}\sum_{j=1}^{100}(eA^*_{kj} - r^*_k) \quad (3.5)$$

The overall bootstrap estimator is computed as:

$$eboot = \frac{1}{g}\sum_{k=1}^{g}\left(eA_k - b_k\right) \quad (3.6)$$

eA_k being the resubstitution estimator specific to G_k when the rule is established on the gn initial observations.

$e0$ *Estimator* (Chatterjee and Chatterjee, 1983): i.e., computed on 100 bootstrap samples t^*_i $(i=1,\ldots,100)$, taken from the initial sample . For each bootstrap sample, a classification rule is established and the proportion of misclassified observations of t, which do not belong to t^*_i, was computed. The $e0$ estimator is the mean of the 100 proportions.

$e632$ *Estimator* (Efron, 1983): i.e., computed as follows:

$$e632 = 0.368eA + 0.632e0 \quad (3.7)$$

SIMULATION DESIGN

Discriminant model

We consider the case of 2-, 3- and 5-groups submitted to linear discriminant analysis and characterized by their means and covariances matrices. In the case of 2 groups, the mean vector, $\mathbf{m}_k \left(k=1,2\right)$ is:

$$\mathbf{m}_1 = 0;\ \mathbf{m}_2 = (m,0,...,0)';\quad m \in IR^+.$$

The covariance matrix, $\mathbf{\Sigma}_k \left(k=1,\ 2\right)$, is a diagonal matrix with $\boldsymbol{v}_k \left(k=1,\ 2\right)$, the vector of diagonal elements given by

$\boldsymbol{v}_1 = \boldsymbol{v}(1)$; $\boldsymbol{v}_2 = \boldsymbol{v}(\lambda)$ where $\lambda \in \mathrm{IR}^+$ and $\boldsymbol{v}(\lambda) = (\lambda, 1,...,1)'$.

In the case of 3- and 5-groups, the mean vectors, and covariance matrices, are given below:

For 3-groups:

$\mathbf{m}_1 = \mathbf{0}$; $\mathbf{m}_2 = (m,0,...,0)'$ $\mathbf{m}_3 = (0,m,0...,0)'$; $\boldsymbol{v}_1 = \boldsymbol{v}(1)$;
$\boldsymbol{v}_2 = \boldsymbol{v}_3 = \boldsymbol{v}(\lambda)$.

For 5-groups

$\mathbf{m}_1 = \mathbf{0}$; $\mathbf{m}_2 = (m,0,...,0)'$ $\mathbf{m}_3 = (0,m,0...,0)'$; $\mathbf{m}_4 = (-m,0,...0)'$; $\mathbf{m}_5 = (0,-m,0,...,0)'$
$\boldsymbol{v}_1 = \boldsymbol{v}(1)$; $\boldsymbol{v}_2 = \boldsymbol{v}_3 = \boldsymbol{v}_4 = \boldsymbol{v}_5 = \boldsymbol{v}(\lambda)$.

It is known that the linear rule is invariant under a non-singular linear transformation (McLachlan, 1992). So, appropriate linear transformations applied to the simple models proposed above, will help to extend the results of the study to a large variety of real world problems.

To assess the heteroscedasticity degree of the populations, a heteroscedasticity parameter Γ is defined for g populations submitted to discriminant analysis as:

$$\Gamma = -\sum_{k=1}^{g} \ln\left(\left|\mathbf{\Sigma}_k\right| / \left|\mathbf{\Sigma}\right|\right), \qquad (4.1)$$

with $\mathbf{\Sigma}_k$ and $\mathbf{\Sigma}$, being the covariance matrix of G_k and the pooled covariance matrix of the g populations respectively. For data samples, an estimated $\hat{\Gamma}$ can be computed by replacing $\mathbf{\Sigma}_k$ and $\mathbf{\Sigma}$, respectively with $\hat{\mathbf{\Sigma}}_k$ and $\hat{\mathbf{\Sigma}}$.

By considering the discriminant model proposed above, it can analytically be shown that the parameters Γ_g (g = 2, 3 and 5) and λ (defined in Section 4) are linked by the following relations:

$$\Gamma_2(\lambda) = \ln\left[\frac{(1+\lambda)^2}{2^2\lambda}\right]; \Gamma_3(\lambda) = \ln\left[\frac{(1+2\lambda)^3}{3^3\lambda^2}\right];$$
$$\Gamma_5(\lambda) = \ln\left[\frac{(1+4\lambda)^5}{5^5\lambda^4}\right] \qquad (4.2)$$

The inverse of these functions helped in choosing the appropriate values of Γ_g according to λ.

Population features and comparison criteria

The factors considered in the assessment of the efficiency of the non-parametric error rate estimators were the number g of groups (g = 2, 3 and 5), the common distribution of the variables of the p-variate populations that is Normal (named N), Chi-square with 12, 8 and 4 degrees of freedom, named C(12), C(8) and C(4), respectively. The number p of variables was 4, 6, 10, 18; three values of the common size sample, n were considered for each value of $p : n/p$ = 1.5; n/p = 2.5 and n/p = 5. For each number g of groups, four values of the heteroscedasticity degree, $\Gamma_k = (k = 2, 3 \text{ and } 5)$ of the populations were chosen from established empirical power function, 1-β of the homoscedasticity test related to Γ_k under normality case (1-β = 0.05: homoscedasticity; 1-β = 0.4: low heteroscedasticity; 1- =0.6: average heteroscedasticity; 1-β =0.8: high heteroscedasticity. Table 1 presents for each number of groups, the mean values related to each of the four values of 1-β. Three values of the overlap, e of the populations were considered: e = 0.05 (low overlap); e = 0.1 (average overlap) and e = 0.15 (high overlap). The group-prior probabilities were considered equal and the overlap was thus equal to the optimal error rate. For each of the combination of population features described above, the values of the parameter m (defined in section 4) were iteratively computed to obtain each of the three values of the overlap (or optimal error rate) of the populations. However, the expression (2.3) for the computation of the overlap e was difficult to manipulate for $g > 2$ so we used an empirical approach to compute the overlap, e.

We present below (without loss of generality), the computational method of e for three *p*-variate populations, P_1, P_2 and P_3, of theoretical density functions f_1, f_2 and f_3. In the discriminant model considered in section 4 the differences between the means vectors were only carried by the first two variables of the populations. In such cases, the other variables did not influence the overlap, of the populations. So, it can be deduced from equation (2.3) that, for equal group-prior probabilities:

$e = \frac{1}{3}(e_1 + e_2 + e_3)$ with:

Table 1. Values of Γ_k according to the 4 values of 1-β

	$g = 2$	$g = 3$	$g = 5$
$1 - \beta = 0.05$	0	0	0
$1 - \beta = 0.4$	1.2686	1.6331	2.1446
$1 - \beta = 0.6$	1.7009	2.1901	2.8644
$1 - \beta = 0.8$	2.1851	2.7979	3.6571

$$e_1 = \int_{f_2(x) > f_1(x) \text{ and } f_2(x) > f_3(x) \mid x \in P_1} f_1(x)dxdy + \int_{f_3(x) > f_1(x) \text{ and } f_3(x) > f_2(x) \mid x \in P_1} f_1(x)dxdy$$

$$e_2 = \int_{f_1(x) > f_2(x) \text{ and } f_1(x) > f_3(x) \mid x \in P_2} f_2(x)dxdy + \int_{f_3(x) > f_1(x) \text{ and } f_3(x) > f_2(x) \mid x \in P_2} f_2(x)dxdy$$

$$e_3 = \int_{f_1(x) > f_2(x) \text{ and } f_1(x) > f_3(x) \mid x \in P_3} f_3(x)dxdy + \int_{f_2(x) > f_1(x) \text{ and } f_2(x) > f_3(x) \mid x \in P_3} f_3(x)dxdy \quad (4.3)$$

In equation (4.3), e_1, e_2 and e_3 represented the group-conditional error rates of the Bayes rule. The used empirical approach considered these conditional error rates as the volume of solids constituted of successive elementary volumes of width, dx ($dx = x_{i+1} - x_i$), length, dy ($dy = y_{i+1} - y_i$) and height, the value of the bivariate probability density function at $dx(dx,dy)$. The same method was used in the case of 2 and 5 groups.

A total of 1728 combinations of the factors were considered and for each of them, 100 samples of size gn were generated from the *g* populations. For each of them, the 10 non-parametric error rate estimators were computed. The actual error rate ec was also empirically computed for each sample by validating the established linear rule on a large sample of size 10,000*g* and used to calculate the Relative Error (RE), the Relative Bias (RB) and the Relative Efficiency (RE_{ff}) of each estimator:

$$\text{RE} = 100\frac{|\text{estimator} - ec|}{ec}; \quad \text{RB} = 100\frac{(\text{estimator} - ec)}{ec}; \quad \text{RE}_{\text{ff}} = \frac{\text{RE(estimator)}}{\min(\text{RE})} \quad (4.4)$$

In equation (4.4), the symbol min(RE) represents the relative error of the best estimator for the considered sample. The Mean Relative Error (MRE), the Mean Relative Bias (MRB) and the Mean Relative Efficiency (MRE_{ff}) related to each estimator were computed for each of the 1728 combinations of the factors.

RESULTS

The MRE of the non-parametric estimators for each combination of the factors were replaced by ranks. For a given combination of the factors, the ranks of the error rate estimators were computed, the estimator of the lowest relative error having the rank 1. The median ranks of the estimators were calculated for each factor level as well as their median rank for all the 1728 combinations of factors and placed in Table 2. It can be noticed that $e632$ is the overall best estimator; the other estimators of good performance were eS_2 and eS_1. On the contrary, $e0$, epp and eA recorded the lowest relative efficiencies. The ranks of the ten estimators for each level of population features did not globally depend on the number *g* of groups, except eS_2 estimator whose relative performance slightly decreased with increased number of groups. The population features seemed not to have influenced the ranks of the estimators. However, $eboot$ and $eppCV$ improved their ranks for increased values of the ratio n/p whereas an opposite trend was observed, not only in the case of eJc, but also eS_1 and eS_2, especially for 5-groups. Moreover, the relative efficiency of $eppCV$ and $e632$ became low with the increased overlap of the populations. The median rank of the estimators for the levels of population features did not help in analysing the quantitative difference between their performances.

Table 2: Median ranks of the estimators according to the populations features

	$e632$			eS_2			eS_1			eJc			$eboot$			eCV			$eppCV$			$e0$			epp			eA		
G	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**	**2**	**3**	**5**
Global	1	1	1	2	3	4	4	4	4	5	5	5	5	5.5	6	6	6	5	7	7	6	7	7	8	8	9	8	10	10	10
N	1	1	1	2	2	3	4	3.5	4	4.5	5	5	5	6	6	5	5	5	7.5	8	7	7	7	8	8	9	9	10	10	10
C(12)	1	1	1	2	3	4	4	4	4	4	5	5	5	5	6	6	5.5	5	7	7	6	7	7	8	8	9	9	10	10	10
C(8)	1	1	1	2	3	4	4	4	4	5	5	5	5	5	6	6	6	5	7	7	6	7	7	8	8	9	8	10	10	10
C(4)	1	1	1	2	3	4	4	4	4	5	5	5	5	5	6	6	6	5	7	7	7	7	8	8	8	9	8	10	10	10
e=0.05	1	1	1	2	3	3	4	5	4	5	5	5	5	4	6	6	6	6	6	6	5	8	8	8	8	9	9	10	10	10
e=0.10	1	1	1	2	3	3.5	4	4	4	4	5	5	5	6	6	5	5	5	7	7	6	7	7	8	9	9	8	10	10	10
e=0.15	4.5	1	1	2	3	5	4	3	3	5	5	4	6	6	6	6	5	5	8	8	7	6	7	7	8	9	8.5	10	10	10
1-β=0.05	1	1	1	2.5	3	3	4	4	4	5	5	5	6	5	6	6	6	6	7	7	6	7	7.5	8	9	9	9	10	10	10
1-β=0.4	1	1	1	2	3	3.5	4	4	4	5	5	5	5	6	6	6	6	6	7	7	6	7	7	8	8	9	8	10	10	10
1-β=0.6	1	1	1	2	3	4	4	4	4	4	5	5	5	6	6	6	5	5	7	7	6	8	7	8	8	9	9	10	10	10
1-β=0.8	2	1	1	2	3	4	4	4	3	4.5	5	5	5	5	6	5	6	5	7	7	7	7	7	8	8	9	8	10	10	10
n/p=1.5	1	1	1	2	2	3	4	4	3	4	4	4	7	6	6	6	5	5	8	7	7	7	7	7	9	9	9	10	10	10
n/p=2.5	1	1	1	2	3	3	4	4	4	5	5	5	5	4	5	6	6	5	8	8	7	7	7	8	9	9	9	10	10	10
n/p=5	1	1	1	2	4	5	4	4	5	5	6	6	3	4	5	5	6	6	6	4	2	8	8	9	7	9	7	10	10	10

Boxplots of the mean relative efficiencies (MRE_{ff}) of the error rate estimators were presented in Figure 1.

This figure confirms the best performance of $e632$, as well as eS_2, eS_1, eJc, $eboot$ and eCV with however, a loss of efficiency of about 28 % of the latter compared to $e632$, which is equivalent to a mean relative error of 12.8 % for these estimators for 10 % of relative error for $e632$. Except the resubstitution estimator, eA, that presented a loss of efficiency of more than 100 % compared to $e632$, the other estimators presented losses of efficiency that vary from 28 % to 70 % compared to $e632$. As far as the dispersion of the MRE_{ff} of the estimators was concerned, Figure 1 shows the very low variability of $e632$, which maintains its best performance over the various populations features considered in this study. Estimators eCV, $e0$, epp and eA that present the lowest performance are also the less stable.

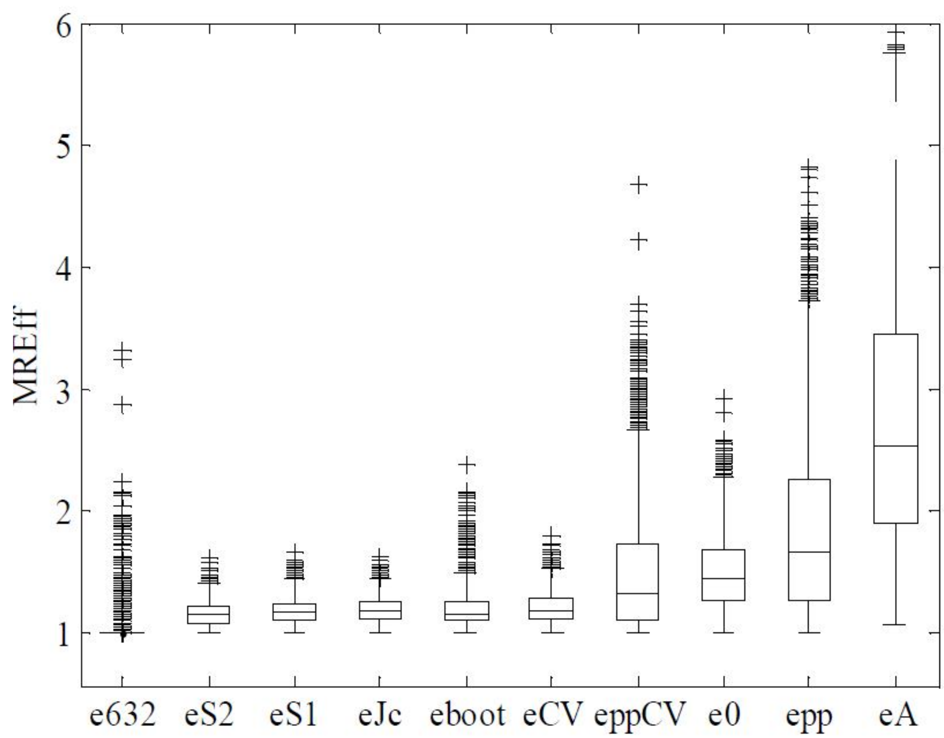

Figure 1: Boxplots of the MRE_{ff} of the estimators

The Mean Relative Bias (MRB) helps in appreciating the direction of the deviation of the estimators' performance for 2-, 3- and 5-groups. Table 3 shows that almost all the non parametric estimators performed well when the number of groups became more important. For 2- and 3-group discriminant analyses, eS_1 and eJc present the lowest absolute MRB (2.5 % for 2-groups and 0.1 for 3-groups) whereas for 5-groups, $e632$ became the best with 0.2 % of absolute MRB.

The resubstitution estimator, eA, presents the most optimistic bias whereas $e0$ presented the most pessimistic one.

Table 3. Mean and standard deviation of the MRB of the error rate estimators

Estimation	$g=2$		$g=3$		$g=5$	
	m	σ	m	σ	m	σ
$e632$	-14.9	9	-0.5	5.1	-0.2	3.7
eS_2	-3.1	5	-5.6	4.9	-6.7	4.7
eS_1	2.4	5.6	-0.2	4	-1.4	2.8
eJc	2.5	5.3	0.1	3.9	-0.6	2.8
$eboot$	31.8	15.1	23.4	8.4	16.3	5.8
eCV	5.6	7	2.8	4.8	1.5	3.3
$eppCV$	39.6	14.8	-26.3	14.1	-14.6	11.8
$e0$	26.7	14.9	24.2	13	20.7	10.3
epp	-38.4	21.1	-36.9	17.4	-26.4	16.1
eA	-86.3	11.2	-43.1	16.6	-35.9	14.1

DISCUSSION AND CONCLUSION

The estimation of the actual error rate for practical use is one of the relevant topics in discriminant studies and a synthesis of the various estimators of the actual error rate was provided in McLachlan (1992). Most studies have been done to compare in two group-discriminant analyses the performance of the error rate estimators, especially associated with the linear classification rule, and a synthesis of them was done by Schiavo and Hand (2000). The originality of our study is that the relative efficiency of non-parametric error rate estimators can be analysed in multi-group discriminant analysis. The obtained results help to point out the overall best efficiency of $e632$ irrespective of the number of the considered groups. For two-group linear discriminant analysis, many studies come to almost the same conclusions (Wehberg and Schumacher, 2004; Glèlè Kakaï et al., 2003). Other studies pointed out the efficiency of this estimator for non-linear classification rules. Jain et al. (1987), using multivariate normal distributions in nearest neighbour discriminant analysis, found that $e632$ out-performed all the other estimators (eCV, $eboot$ and $e0$). However, we noticed from the present study that for high overlap of the populations in the case of two groups, the performance of this estimator decreased. Fitzmaurice et al. (1991), using two-group discriminant analysis concluded that $e632$ became less reliable as the true actual error rate increased above 0.35, but more reliable as the true error rate decreased. Other estimators that performed well in the present study were eS_1, eS_2 and eJc. On the contrary, $e0$, epp, $eppCV$ and eA recorded the lowest performance, in most of the cases considered in the study.

The ranks of estimators were less influenced by the populations' features, probably due to the fact that they were all based on resampling methods that do not replicate conditions of use. However, the number of groups had a high impact on the performance of the estimators. The latter became more efficient as the number of groups increased.

The highest positive relative bias was obtained by eA whereas $e0$ had the highest and negative relative bias. These results have already been obtained by Wehberg and Schumacher (2004), Chatterjee and Chatterjee (1983) and Chernick and Murthy (1985) who qualified eA and $e0$ as the optimist and pessimist estimators, respectively.

REFERENCES

Chatterjee S., Chatterjee S. (1983). Estimation of misclassification probabilities by bootstrap methods. *Communication in Statistics – Simulation and Computation*, 12, 645-656.

Chernick M. R., Murthy V. K. (1985). Properties of bootstrap samples. *American Journal of Mathematical and Management Sciences*, 5, 161-170.

Efron B. (1983). Estimating the error of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, 78, 316-331.

Fitzmaurice G. M., Krzanowski W. J., Hand D. J. (1991). A Monte Carlo study of the 632 bootstrap estimator of error rate. *Journal of Classification*, 8, 239-250.

Fukunaga K., Kessell D. L. (1972). Application of optimum error reject function. *IEEE Transactions on Information Theory*, 18, 814-817.

Glèlè Kakaï R., Palm. R. (2006). Methodological contribution to control heteroscedasticity in discriminant analysis studies. *Global Journal of Pure and Applied Sciences*, 12 (1), 107-110.

Glèlè Kakaï R., Piraux F., Fonton, N. Palm R. (2003). Comparaison empirique des estimateurs de taux d'erreur en analyse discriminante. *Revue de Statistique Appliquée*, 51(2), 91-104.

Hand D. J. (1986). Recent advances in error rate estimations, *Pattern Recognition Letters*, 4, 335-346.

Jain A. K., Dubes R. C., Chen C. (1987). Bootstrap techniques for error rate estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9, 628-633.

Lachenbruch P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23, 639-645.

McLachlan G. J. (1975). Confidence intervals for the conditional probability of misallocation in discriminant analysis. *Biometrics*, 31, 161-167.

McLachlan G. J. (1992). *Discriminant analysis and statistical pattern recognition*, Wiley, New York.

Quenouille M. H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society* (*Serie B*), 11, 18-84.

Schiavo R. A., Hand D. J. (2000). Ten more years of error rate research. *International statistical review*, 68, 295-310.

Smith C.A.B. (1947). Some examples of discrimination. *Annals of Eugenics*, 13, 272-282.

Wehberg S., Schumacher M. (2004). A comparison of Nonparametric Error Rate Estimation Methods in Classification Problems. *Biometrical Journal*, 46, 35-47.

ACKNOWLEDGEMENTS

This work was supported by the Alexander-Von-Humboldt Foundation (AvH) and the Third World Academy of Sciences (TWAS).