Hacettepe Journal of Mathematics and Statistics
Volume 38 (3) (2009), 357 – 372

BOOTSTRAP PERCENTILE CONFIDENCE INTERVALS FOR ACTUAL ERROR RATE IN LINEAR DISCRIMINANT ANALYSIS

Cemal Atakan*

Received 01 : 04 : 2009 : Accepted 16 : 06 : 2009

Abstract

In this study, bootstrap percentile confidence intervals for the actual error rate have been obtained with respect to the bootstrap estimated values of the estimators D, DS, L, LS, O, OS and M. Estimated values for these estimators have been obtained by a simulation study for two multivariate normal populations with different mean vectors and a common variance-covariance matrix. Moreover, traditional confidence intervals for the actual error rate were obtained based on the analytic form of the estimator M. In general, bootstrap percentile intervals are narrower than the traditional intervals obtained with respect to the analytic form of the estimator M. Narrower intervals are desired which will be relatively consistent with the parameter. It is shown that some statistical properties of the estimators can be obtained by the bootstrap method.

Keywords: Discriminant analysis, Error rate, Bootstrap, Percentile intervals.

2000 AMS Classification: 62 F 40, 62 H 30, 65 C 60.

1. Introduction

Discriminant analysis is a statistical technique in which an experimenter makes a number of measurements on an individual and wishes to classify this individual into one of several known populations or categories on the basis of these measurements. It is assumed that the individual can come from a finite number of populations. Each population is characterized by the probability distribution of a random vector $\underline{X}$ associated with the measurements. When the probability distributions are completely known, then the problem is reduced to identifying the allocation rule. If the type of the distribution with unknown parameters is given, then the problem is to identify the allocation rule with respect to the estimation of the parameters which are estimated from related samples.

*Ankara University, Faculty of Science, Department of Statistics, Ankara, Turkey. E-mail: atakan@science.ankara.edu.tr

The main goal of discriminant analysis is to obtain an allocation procedure with minimum error. According to this optimization criterion, it is important to know the probability of the misclassification or error rate for the evaluation of the allocation rules. The probability of misclassification is calculated based on whether the parameters of the distribution are known or not, the sample size and the allocation rule. Useful references for these kind of problems are [1, 13, 14, 15, 21, 24, 29, 31], among many others.

Let Π_1 and Π_2 be two different populations, and $\underline{X} = (X_1, X_2, \dots, X_p)'$ a p dimensional random vector with observed value $\underline{x} \in \mathbb{R}^p$. If $\underline{X}$ is drawn from Π_i, then the probability density function of $\underline{X}$ is $f_i(\underline{x}, \underline{\theta}_i)$; where $\underline{\theta}_i$, $i = 1, 2$, is a vector of parameters. The classification problem is to determine a partition of the sample space into two regions R_1 and R_2 with $R_1 \cup R_2 = \mathbb{R}^p$ and $R_1 \cap R_2 = \emptyset$. If an observed value $\underline{x}$ lies in R_1, then the item with measurement $\underline{x}$ is allocated to Π_1, otherwise it is allocated to Π_2.

When an item is allocated, an error will occur if the item is allocated to a population other than its real population. It is expected that the best allocation rule will allocate items with a minimum error. In general three error rates have been discussed in discriminant analysis; Optimal, Actual and Expected Actual error rates.

In the literature there are many parametric and nonparametric error rate estimators for the actual error rate. In general, the analytic form of the distributions of some of these estimators are not available and thus statistical inferences for these parameters are not possible. Therefore we cannot make some statistical inferences for parameters of these distributions. However, an analytical expression is available for the estimator M [22, 23, 24]. Various techniques, such as the Bootstrap, Jackknife and Cross Validation can be used to obtain the values of the parameters [4, 6, 9]. Bootstrap is a resampling method depending on data and in the literature it has been used to get confidence intervals, too [7, 8, 17, 18, 24, 30, 34].

In this study, bootstrap percentile confidence intervals for the actual error rate have been obtained by using the bootstrap method with respect to the error rate estimators D, DS, L, LS, O, OS and M [1, 13, 15, 20, 21, 24, 29, 31]. These are well known parametric error rate estimators in the literature. Moreover, confidence intervals for the actual error rate have been obtained based on the analytic form of the estimator M. The bootstrap estimates and the estimates obtained from the analytical expression based on the estimator M are compared using a simulation.

Section 2 and 3 introduce the concept of allocation rules and error rates. In Section 4, some error rate estimators are discussed and Section 5 gives an equality for confidence intervals of the actual error rate with respect to the estimator M. Finally, in Section 6, the procedure for estimating bootstrap percentile confidence intervals for the actual error rate is described. The paper concludes with a simulation.

2. Allocation (classification) rules

Allocation rules are constructed using a Linear Discriminant Function (LDF) which was obtained by the likelihood ratio criteria given in [33]. Let Π_1 and Π_2 be two multivariate normal populations with known mean vectors $\underline{\mu}_1$, $\underline{\mu}_2$, and a common covariance matrix Σ. The rule, known as the optimal classification rule, which minimizes the total probability of misclassification, is given by

$$\xi : \begin{cases} U(\underline{x}) > k & \text{classify } \underline{x} \text{ into } \Pi_1 \\ \text{otherwise} & \text{classify } \underline{x} \text{ into } \Pi_2 \end{cases} \tag{2.1}$$

where

$$(2.2) \qquad U(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\underline{\mu}_1 + \underline{\mu}_2)\right]' \Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2)$$

is the population LDF obtained from the ratio of two multivariate normal joint probability density functions $f_1(\underline{x}; (\underline{\mu}_1, \Sigma))$, $f_2(\underline{x}; (\underline{\mu}_2, \Sigma))$ and $k = \ln(q_2/q_1)$ where the q_i's are the prior probabilities of whether the i th observation belongs to the population Π_i or not $(i = 1, 2)$.

When the parameters $\underline{\mu}_1$, $\underline{\mu}_2$ and Σ are unknown, the optimal classification rule evaluated by using the sample LDF $\widehat{\xi}$ given by

$$(2.3) \qquad \widehat{\xi} : \begin{cases} W(\underline{x}) > k & \text{classify } \underline{x} \text{ into } \Pi_1 \\ \text{otherwise} & \text{classify } \underline{x} \text{ into } \Pi_2 \end{cases}$$

where

$$(2.4) \qquad W(\underline{x}) = \left[\underline{x} - \frac{1}{2}(\overline{\underline{x}}_1 + \overline{\underline{x}}_2)\right]' s^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)$$

is the sample LDF which is obtained by the plugging the estimated values $\overline{\underline{x}}_1$, $\overline{\underline{x}}_2$ and s into the population LDF. They are calculated based on the sizes n_1 and n_2 of the training samples from the populations Π_1 and Π_2, respectively [1, 20]. Note that $\overline{\underline{x}}_1$, $\overline{\underline{x}}_2$ and s refer to realizations of the random vectors $\overline{\underline{X}}_1$, $\overline{\underline{X}}_2$ and the random matrix S. Here, the prior probabilities are taken to be equal, which will trivially give $k = 0$.

3. Error rates

In the allocation procedure, if an individual is misclassified then an error is made. The purpose of discriminant analysis is to allocate individuals with a minimum error.

Error rates have been obtained with respect to the distribution of the discriminant function [28]. There are several different types of error rate associated with the allocation rules given in Section 2. These are the optimal, actual (or conditional) and expected actual (or unconditional) error rates.

The optimal error rate is obtained for the rule given in equation (2.1). This is the error rate that would occur when the parameters are known. The actual error rate is given by the rule in (2.3) which is conditional on the estimated parameters. If the expected value operator is defined with respect to all possible training samples, then the expected actual error rate is the expectation of the actual error rate. Useful references are [1, 15, 20, 21, 22, 23, 24, 26, 29].

When the parameters of the multivariate normal distributions are known, then the distribution of $U(\underline{X})$ is, for $(i = 1, 2)$, univariate normal with means $(-1)^i(-\Delta^2/2)$ and common variance Δ^2, where Δ^2 is the Mahalanobis square distance between the two populations, and is given by

$$(3.1) \qquad \Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)'\Sigma^{-1}(\underline{\mu}_1 - \underline{\mu}_2).$$

The distribution of $W(\underline{X})$ is derived under some conditions given in [1, 15, 23, 25, 27, 32], but is very complicated to implement. Consequently, most of the work associated with the error rates has assumed that the sample estimates $\overline{\underline{x}}_1$, $\overline{\underline{x}}_2$ and s are fixed. The conditional distribution of $W(\underline{X})$ with respect to the these values can be obtained by plugging the estimated values in place of the estimators in the expression for $W(\underline{X})$.

When $\underline{X} \sim N_p(\underline{\mu}_i, \Sigma)$, the conditional distributions of $W(\underline{X})$ are univariate normals with means

$$\left[\underline{\mu}_i - \frac{1}{2}(\overline{\underline{x}}_1 + \overline{\underline{x}}_2)\right]' s^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)$$

and the common variance

$$(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)' s^{-1} \Sigma\, s^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)$$

[20]. If a random observation from Π_1 is misclassified to Π_2, the optimal error rate according to the ξ rule given in (2.1) is

$$\begin{aligned} \tau_1(\xi) &= P(U(\underline{X}) \le 0/\underline{X} \in \Pi_1) \\ &= \Phi(-\Delta/2). \end{aligned} \tag{3.2}$$

Similarly, the actual error rate according to the rule $\widehat{\xi}$ given in (2.3) is

$$\begin{aligned} \tau_1(\widehat{\xi}) &= P(W(\underline{X}) \le 0/\underline{X} \in \Pi_1, \overline{\underline{x}}_1, \overline{\underline{x}}_2, s) \\ &= \Phi\left(- \frac{\left[\underline{\mu}_1 - \frac{1}{2}(\overline{\underline{x}}_1 + \overline{\underline{x}}_2)\right]' s^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_2}{\left[(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)' s^{-1}\Sigma s^{-1}(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)\right]^{1/2}} \right), \end{aligned} \tag{3.3}$$

and the expected actual error rate which is the expectation of the actual error rate with respect to all possible samples taken from the populations is given by

$$E(\tau_1(\widehat{\xi})) = E\left[P(W(\underline{X}) \le 0/\underline{X} \in \Pi_1)\right], \tag{3.4}$$

where $\Phi(\,.\,)$ is the cumulative distribution function of the standard normal distribution [16, 22, 24, 26]. If the observation are taken from Π_2 and allocated to Π_1, the expressions will be similar. When the prior probabilities are assumed to be equal for the two populations, the total probability of misclassification for optimal and actual error rates are given by

$$\tau(\xi) = \frac{1}{2}(\tau_1(\xi) + \tau_2(\xi)) \tag{3.5}$$

and

$$\tau(\widehat{\xi}) = \frac{1}{2}(\tau_1(\widehat{\xi}) + \tau_2(\widehat{\xi})), \tag{3.6}$$

respectively.

Note the hierarchy associated with these error rates: the optimal error rate is a function only of the distributions of $\underline{X}$ for the two populations, the actual error rate is a function of the distributions of $\underline{X}$ and the particular training samples selected and the expected actual error rate is a function of the distributions of $\underline{X}$ and the training sample sizes [29].

4. Some error rate estimators

Error rates are calculated based on the distributions of the discriminant functions, but the distribution of the sample LDF based on unknown parameters is quite complicated and thus an analytical expressions for the error rates becomes difficult. Under some conditions, the distribution of the sample LDF is available and it is not very practical to use [1, 15, 23, 25, 27, 29, 32]. Therefore the values of the error rates have been estimated by different error rate estimators, which will be explained below. Each of the methods of error rate estimation described in this section is given a symbol to identify it. The estimators are referred to by $\widehat{\tau}$ with that symbol as a superscript.

Here the widely used error rate estimators D, DS, L, LS, O, OS and M, defined in literature, will be considered. These are known as the parametric error rate estimators and mainly depend on the assumption of normality for a given data set. Under normality, the actual error rate is a known function of the unknown population parameters. If the random observation is taken from Π_1 and allocated to Π_2, analytical expressions for these error rate estimators are given as follows. Similar expressions can also be given when the observation is taken from Π_2 and allocated to Π_1. For more details about error rate estimators, one can look at in [1, 21, 24].

The Error Rate Estimator D.

This estimator was first proposed by Fisher [11] and was obtained by plugging the estimators of the unknown parameters in the actual error rate given in (3.3). Therefore, it is also known as a plug-in estimator and is given by

$$(4.1)\qquad \widehat{\tau}_1^D = \Phi\left\{-\frac{\left[\underline{X}_1 - \frac{1}{2}(\underline{X}_1 + \underline{X}_2)\right]' S^{-1}(\underline{X}_1 - \underline{X}_2)}{\left[(\underline{X}_1 - \underline{X}_2)' S^{-1} S S^{-1}(\underline{X}_1 - \underline{X}_2)\right]^{1/2}}\right\} = \Phi\left\{-\frac{D}{2}\right\},$$

where D^2 is the estimator of the Mahalanobis square distance, as mentioned in (3.1).

The Error Rate Estimator DS.

The estimator D^2 is biased for Δ^2 and therefore an unbiased estimator $(DS)^2 = c_1 D^2$ is used [21]. This error rate estimator is obtained from the estimator D by setting

$$(4.2)\qquad \widehat{\tau}_1^{DS} = \Phi\left\{-\sqrt{c_1}\frac{D}{2}\right\},$$

where $c_1 = \dfrac{n_1 + n_2 - p - 3}{n_1 + n_2 - 2}$.

The Error Rate Estimator L.

In [19], Lachenburch gives approximate expressions for the actual error rate. In these expressions, replacing the unknown parameter Δ^2 by D^2 gives the L type error rate estimator as follows:

$$(4.3)\qquad \widehat{\tau}_1^L = \Phi\left\{-\frac{\frac{c_2}{2}(D^2 - \frac{p(n_2-n_1)}{n_1 n_2}}{\sqrt{c_3(D^2 + \frac{p(n_1+n_2)}{n_1 n_2})}}\right\},$$

where $c_2 = \dfrac{n_1 + n_2 - 2}{n_1 + n_2 - p - 3}$ and $c_3 = \dfrac{(n_1 + n_2 - 3)(n_1 + n_2 - 2)^2}{(n_1 + n_2 - p - 2)(n_1 + n_2 - p - 3)(n_1 + n_2 - p - 5)}$.

The Error Rate Estimator LS.

In the expression given for the error rate estimator L, use $(DS)^2$ instead of D^2 to get the estimator LS as follows:

$$(4.4)\qquad \widehat{\tau}_1^{LS} = \Phi\left\{-\frac{\frac{1}{2}(D^2 - c_2\frac{p(n_2-n_1)}{n_1 n_2}}{\sqrt{\frac{c_3}{c_2}(D^2 + \frac{p(n_1+n_2)}{n_1 n_2})}}\right\},$$

where c_2 and c_3 are given in (4.3) [19].

The Error Rate Estimator O.

In [25], Okamoto gives an asymptotic expansion for the actual error rate. Then, Anderson [1] uses the estimator D^2 for the unknown parameter Δ^2 as follows:

$$(4.5)\qquad \widehat{\tau}_1^O = \Phi\left\{-\frac{D}{2}\right\} + \varphi\left\{-\frac{D}{2}\right\}\{E_1 + E_2 + E_3\},$$

where $\varphi\{\cdot\}$ is the probability density function of the standard normal distribution and $E_1 = \frac{\left[\frac{D}{16} + \frac{3(p-1)}{4D}\right]}{n_1}$, $E_2 = \frac{\left[\frac{D}{16} - \frac{(p-1)}{4D}\right]}{n_2}$, and $E_3 = \left[\frac{D(p-1)}{4(n_1+n_2-2)}\right]$.

The Error Rate Estimator OS.

Replacing D^2 by $(DS)^2$ in the O estimator gives the OS estimator [1] as

$$\hat{\tau}_1^{OS} = \Phi\left\{-\sqrt{c_1}\frac{D}{2}\right\} + \varphi\left\{-\sqrt{c_1}\frac{D}{2}\right\}\{F_1 + F_2 + F_3\}, \tag{4.6}$$

where $F_1 = \frac{\left[\frac{\sqrt{c_1}D}{16} + \frac{3(p-1)}{4\sqrt{c_1}D}\right]}{n_1}$, $F_2 = \frac{\left[\frac{\sqrt{c_1}D}{16} - \frac{(p-1)}{4\sqrt{c_1}D}\right]}{n_2}$, and $F_3 = \left[\frac{\sqrt{c_1}D(p-1)}{4(n_1+n_2-2)}\right]$.

The Error Rate Estimator M.

Mclachlan [22] considered an asymptotically unbiased error rate estimator for the actual error rate as follows:

$$\begin{aligned}\hat{\tau}_1^{M} = \Phi\left\{-\frac{D}{2}\right\} + \varphi\left\{\frac{D}{2}\right\}&[(p-1)/Dn_1 \\ &+ D\left\{4(4p-1) - D^2\right\}/(32(n_1+n_2-2)) \\ &+ \{(p-1)(p-2)\}/\left(4Dn_1^2\right) \\ &+ \frac{(p-1)}{(64n_1(n_1+n_2-2))}\left\{-D^3 + 8(2p+1)D + \left(\frac{16}{D}\right)\right\} \\ &+ \frac{(D/12288)}{(n_1+n_2-2)^2}\left\{3D^6 - 4(24p+7)D^4\right. \\ &\left.+ 16\left(48p^2 - 48p - 53\right)D^2 + 192(-8p+15)\right\}].\end{aligned} \tag{4.7}$$

5. Confidence intervals for the actual error rate depending on the estimator M

Analytical expressions for statistical properties of the error rate estimator M are available in the literature [22, 24]. McLachlan [23] considers an approximated $100(1-\alpha)\%$ confidence interval estimation of the actual error rate depending on the estimator M for samples taken from two multivariate normal distributions under the equality of prior probabilities and variance-covariance matrices. This approximated $100(1-\alpha)\%$ confidence interval has the following form

$$\tau_1(\hat{\xi}) : \hat{\tau}_1^M \pm \left\{\sqrt{v_1(DS)}\right\}\Phi^{-1}(1-\alpha/2), \tag{5.1}$$

where DS is defined in (4.2) and $\Phi^{-1}(1-\alpha/2)$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution function. The interval is based on the result that for sufficiently large separate samples sizes n_1 and n_2, the distribution of $\hat{\tau}_1^M - \tau_1(\hat{\xi})$ is approximately normal with zero mean and variance $v_1(\Delta)$ given by

$$\begin{aligned}v_1(\Delta) = \left\{\phi\left(\tfrac{\Delta}{2}\right)\right\}^2&\left[\frac{1}{n_1} + \frac{(\Delta^2/8)}{n_1+n_2} + \left\{\frac{\Delta^2+4(3p-4)\left(p^2-4p+5\right)\left(16/\Delta^2\right)}{(4n_1)^2}\right\}\right. \\ &+ \left\{\frac{\left(\Delta^2-2p\right)/8}{n_1n_2}\right\} + \left\{\frac{\Delta^4+2(11p-16)\Delta^2+8(5p-4)}{64n_1(n_1+n_2-2)}\right\} \\ &\left.+ \left\{\frac{2\Delta^6+16(2p-5)\Delta^4-32(4p-13)\Delta^2}{32(n_1+n_2-2)^2}\right\}\right].\end{aligned} \tag{5.2}$$

If the observation is taken from Π_2 and allocated to Π_1 a similar confidence interval for $\tau_2(\widehat{\xi})$ is obtained just by interchanging n_1 with n_2 in $\tau_1(\widehat{\xi})$.

Here, we take equal sample sizes, $(n_1 = n_2)$. In this case, $v_1(DS) = v_2(DS) = v(DS)$ and then the $100(1-\alpha)\%$ approximated confidence interval estimation of $\tau(\widehat{\xi})$ is given by

$$(5.3) \qquad \tau(\widehat{\xi}) : \widehat{\tau}^M \pm \left\{ \sqrt{v(DS)} \right\} \Phi^{-1} (1 - \alpha/2) \, ,$$

where $\tau(\widehat{\xi})$ is given in (3.6). We will called this type of interval a *traditional confidence interval.*

6. Bootstrap percentile intervals

The bootstrap is a computer intensive method of approximating the unknown sampling distribution on any estimator, and it involves repeated resampling of the observed data [2, 4, 10]. Timmerman and Ter Braak [30] indicate that the method is an alternative to analytically derived results because it does not need complicated derivations, and it usually relies on much weaker assumptions. The bootstrap method can be implemented nonparametrically by using the empirical distribution function constructed from the original data, or parametrically by using the bootstrap cumulative distribution function constructed from the familiar parametric distributions [5,7].

Since the distribution of the error rate estimators is not known, we attempted to use the nonparametric bootstrap procedure. There are many studies on the bootstrap computations such as [2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 18, 24, 30], among many others.

Bootstrapping has been applied widely in statistics when analytical derivations of the distribution of an estimator are intractable, and it can be found to generate distributions close to the underlying true distributions. If analytical expressions for some parameter estimators or statistical properties of these estimators are not available, estimated values of these estimators can be obtained by bootstrapping.

In general, Timmerman and Ter Braak [30] point out that the bootstrap procedure is carried out in three steps: First, resampling from the sample data is used to imitate the sampling process from the population. Secondly, the resampling reveals the bootstrap distribution of the statistic of interest and finally, the bootstrap distribution is used to estimate inferential information, like confidence intervals.

The bootstrap method is also a useful procedure in forming confidence intervals. Different approaches have been proposed to estimate confidence intervals. The first main class of bootstrap confidence intervals is based on the bootstrap standard error which is the standard deviation of the bootstrap distribution and includes the normal, pivotal and the studentized pivotal intervals. On the other hand, the bootstrap confidence intervals are based on a percentile of the bootstrap distribution, which includes the percentile and adjusted percentile intervals. Here, we only consider percentile type confidence intervals because they are both range preserving and transformation respecting. Range preserving would mean that the confidence intervals produced are within the allowable range of the statistics, and transformation respecting would imply that the estimated confidence intervals of a monotonically transformed statistic would match the estimated ones of the untransformed statistic [2, 8, 10, 24, 30 ,34].

The bootstrap technique can be used in discriminant analysis to get the bootstrap type estimated confidence intervals of the actual error rate with respect to different error rate estimators given in Section 4.

We take two samples with sizes n_1 and n_2 from populations Π_1 and Π_2, respectively. For each of these samples, the number B of new replications with the same number n_1 and

n_2 of observations have been taken randomly by replacement. These B observed samples are called the bootstrap samples for each population. From the bootstrap samples, the new observed samples of size B for any error rate estimator of the actual error rate is $\widehat{\tau}^1, \widehat{\tau}^2, \ldots, \widehat{\tau}^B$, where $\widehat{\tau}^b$ is a bth, $(b = 1, 2, \ldots, B)$, estimated value of the error rate estimators given in Section 4. Ordering the error rates as

$$\widehat{\tau}^{(1)} \leq \widehat{\tau}^{(2)} \leq \cdots \leq \widehat{\tau}^{(B)},$$

a $100(1-\alpha)\%$ level bootstrap percentile interval from these ordered observations for the actual error rate is obtained with respect to the error rate estimator of interest. Thus, the observed error rate values are assumed to be $(\alpha/2)$ and $(1-\alpha/2)$ percentiles levels according to the ordered observations for the lower and upper bounds of the percentile interval.

7. A simulation study and discussion

In this section, a simulation study have been carried out to evaluate the bootstrap percentile confidence intervals for the actual error rate, $\tau(\widehat{\xi})$, by using the bootstrap estimated values of the estimators based on D, DS, L, LS, O, OS and M. Traditional confidence intervals are also obtained with respect to the analytic expression given based on the estimator M. A simulation study with 10000 replicates was used from two multivariate normal populations with different mean vectors and common variance-covariance matrix. Simulations were performed for the values $p = 2,\ 3,\ 5$, $n_1 = n_2 = 50,\ 100,\ 250$ and $\Delta = 1,\ 2,\ 3$. For each dataset, bootstrap samples were obtained with $B = 1000$ (Efron [8] generally recommends a B of 1000 or more for accurate confidence intervals). In order to obtain the traditional confidence intervals of the actual error rate, $\tau(\widehat{\xi})$, with respect to the estimator M, the confidence level was set as $(1-\alpha) = 0.95$. The Simulation results are given in Tables 1-9.

As can be seen from the tables, when the size of the samples is increased the estimated values of all estimators approximate to a real value of the actual error rate, and thus the estimators are consistent. For all cases, the estimated values of the actual error rate obtained by bootstrapping are both less than the direct estimates of the error rate estimators, and also less than the real value of the actual error rate. Thus we observe that bootstrapping is underestimating the actual error rate. Also, the Bootstrap percentile intervals are narrower than the traditional confidence intervals obtained with respect to the estimator M in all cases. This does not mean that the bootstrap percentile intervals are better than the traditional confidence intervals. However, if traditional confidence intervals cannot be obtained analytically, then the bootstrap percentile intervals can be preferred. Values of the confidence bounds for the traditional intervals are given in the Tables.

When the sample sizes increase, the intervals become narrower for both cases. Thus the bootstrap percentile intervals and the traditional confidence intervals obtained based on the estimator M are consistent with each other. That is, the interval bounds for both cases are close to each other. In other words, confidence bounds obtained based on the bootstrapping technique become close to the traditional bounds when the sample size increases. For small samples, the results seems to be inconsistent and therefore these simulation results are not given in the tables.

If the population means are well separated from each other (Δ large), the value of the error rate decreases and therefore the confidence bound for the error rate becomes narrower. On the other hand, the dimension has no effect on the confidence bounds.

Finally, when statistical properties of the error rate estimators given Section 4 are not available, statistical inferences for the estimators could possibly be obtained by the bootstrap technique.

Table 1. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 1, p = 2$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.3140	D	0.3037	0.2971 (0.0367)	0.2256	0.3683
		DS	0.3064	0.2999 (0.0363)	0.2290	0.3702
		L	0.3131	0.3064 (0.0378)	0.2230	0.3804
		LS	0.3157	0.3091 (0.0374)	0.2364	0.3822
		O	0.3090	0.3025 (0.0375)	0.2299	0.3758
		OS	0.3062	0.2996 (0.0379)	0.2263	0.3738
		M	0.3139	0.3074 (0.0385)	0.2334	0.3832
		Traditional interval based on the estimator M: (0.2126, 0.4153)				
100	0.3112	D	0.3064	0.3031 (0.0261)	0.2522	0.3542
		DS	0.3077	0.3045 (0.0259)	0.2538	0.3553
		L	0.3112	0.3079 (0.0266)	0.2562	0.3601
		LS	0.3125	0.3093 (0.0264)	0.2578	0.3611
		O	0.3091	0.3058 (0.0263)	0.2545	0.3575
		OS	0.3077	0.3044 (0.0265)	0.2528	0.3564
		M	0.3115	0.3082 (0.0266)	0.2564	0.3606
		Traditional interval based on the estimator M: (0.2401, 0.3829)				
250	0.3096	D	0.3077	0.3064 (0.0164)	0.2740	0.3389
		DS	0.3082	0.3069 (0.0164)	0.2746	0.3393
		L	0.3096	0.3083 (0.0165)	0.2757	0.3411
		LS	0.3101	0.3088 (0.0165)	0.2763	0.3415
		O	0.3087	0.3074 (0.0165)	0.2749	0.3401
		OS	0.3082	0.3069 (0.0165)	0.2743	0.3396
		M	0.3097	0.3084 (0.0165)	0.2758	0.3448
		Traditional interval based on the estimator M: (0.2646, 0.3548)				

Table 2. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 2, p = 2$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.1623	D	0.1561	0.1514 (0.0289)	0.0974	0.2083
		DS	0.1589	0.1550 (0.0289)	0.1008	0.2119
		L	0.1622	0.1574 (0.0295)	0.1022	0.2190
		LS	0.1659	0.1610 (0.0295)	0.1057	0.2125
		O	0.1597	0.1549 (0.0292)	0.1002	0.2088
		OS	0.1559	0.1512 (0.0292)	0.0968	0.2157
		M	0.1621	0.1573 (0.0297)	0.1017	0.2372
		Traditional interval based on the estimator M: (0.0870, 0.2372)				

Table 2. Continued

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
100	0.1605	D	0.1572	0.1548 (0.0207)	0.1156	0.1956
		DS	0.1590	0.1566 (0.0207)	0.1174	0.1974
		L	0.1602	0.1578 (0.0209)	0.1182	0.1991
		LS	0.1620	0.1596 (0.0209)	0.1200	0.2008
		O	0.1590	0.1566 (0.0208)	0.1172	0.1975
		OS	0.1571	0.1547 (0.0208)	0.1154	0.1956
		M	0.1602	0.1578 (0.0210)	0.1180	0.1991
		Traditional interval based on the estimator M: (0.1072, 0.2132)				
250	0.1594	D	0.1583	0.1573 (0.0132)	0.1320	0.1833
		DS	0.1590	0.1580 (0.0132)	0.1328	0.1840
		L	0.1595	0.1585 (0.0133)	0.1331	0.1846
		LS	0.1602	0.1593 (0.0133)	0.1339	0.1853
		O	0.1590	0.1580 (0.0133)	0.1327	0.1841
		OS	0.1583	0.1573 (0.0133)	0.1320	0.1833
		M	0.1595	0.1585 (0.0133)	0.1331	0.1846
		Traditional interval based on the estimator M: (0.1259, 0.1931)				

Table 3. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 3, p = 2$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.0693	D	0.0661	0.0635 (0.0180)	0.0322	0.1006
		DS	0.0690	0.0663 (0.0184)	0.0342	0.1040
		L	0.0700	0.0672 (0.0186)	0.0347	0.1055
		LS	0.0730	0.0702 (0.0190)	0.0368	0.1090
		O	0.0684	0.0657 (0.0184)	0.0337	0.1035
		OS	0.0654	0.0629 (0.0180)	0.0316	0.1000
		M	0.0692	0.0654 (0.0188	0.0337	0.1050
		Traditional interval based on the estimator M: (0.0240, 0.1145)				
100	0.0680	D	0.0666	0.0653 (0.0131)	0.0417	0.0919
		DS	0.0681	0.0667 (0.0132)	0.0428	0.0936
		L	0.0685	0.0671 (0.0133)	0.0431	0.0942
		LS	0.0700	0.0686 (0.0134)	0.0443	0.0959
		O	0.0678	0.0686 (0.0132)	0.0426	0.0933
		OS	0.0663	0.0650 (0.0131)	0.0414	0.0917
		M	0.0682	0.0668 (0.0134)	0.0457	0.0940
		Traditional interval based on the estimator M: (0.0363, 0.0873)				

Table 3. Continued

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
250	0.0673	D	0.0666	0.0660 (0.0083)	0.0505	0.0828
		DS	0.0671	0.0666 (0.0084)	0.0510	0.0835
		L	0.0673	0.0668 (0.0084)	0.0511	0.0837
		LS	0.0679	0.0673 (0.0084)	0.0516	0.0843
		O	0.0670	0.0665 (0.0084)	0.0509	0.0834
		OS	0.0664	0.059 (0.0083)	0.0503	0.0827
		M	0.0672	0.0666 (0.0084)	0.0510	0.08536
		Traditional interval based on the estimator M: (0.0471, 0.0873)				

Table 4. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 1, p = 3$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.3182	D	0.2987	0.2879 (0.0355)	0.2168	0.3580
		DS	0.3024	0.2917 (0.0350)	0.2215	0.3608
		L	0.3118	0.3007 (0.0369)	0.2271	0.3741
		LS	0.3153	0.3043 (0.0369)	0.2317	0.3766
		O	0.3083	0.2974 (0.0369)	0.2241	0.3713
		OS	0.3044	0.2933 (0.0374)	0.2193	0.3682
		M	0.3176	0.3064 (0.0385)	0.2307	0.3844
		Traditional interval based on the estimator M: (0.2150, 0.4201)				
100	0.3133	D	0.3042	0.2988 (0.0257)	0.2480	0.3494
		DS	0.3060	0.3006 (0.0256)	0.2502	0.3509
		L	0.3110	0.3055 (0.0264)	0.2537	0.3577
		LS	0.3128	0.3073 (0.0262)	0.2558	0.3591
		O	0.3091	0.3036 (0.0263)	0.2520	0.3555
		OS	0.3072	0.3017 (0.0264)	0.2498	0.3540
		M	0.3137	0.3082 (0.0268)	0.2556	0.3614
		Traditional interval based on the estimator$M$: (0.2418, 0.3855)				
250	0.3105	D	0.3069	0.3047 (0.0166)	0.2724	0.3371
		DS	0.3076	0.3054 (0.0166)	0.2732	0.3377
		L	0.3096	0.3075 (0.0168)	0.2748	0.3403
		LS	0.3103	0.3082 (0.0168)	0.2756	0.3409
		O	0.3088	0.3066 (0.0168)	0.2741	0.3393
		OS	0.3081	0.3059 (0.0168)	0.2733	0.3387
		M	0.3106	0.3084 (0.0169)	0.2757	0.3415
		Traditional interval based on the estimator M: (0.2655, 0.3558)				

Table 5. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 2, p = 3$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.1648	D	0.1538	0.1468 (0.0285)	0.0934	0.2031
		DS	0.1588	0.1517 (0.0286)	0.0979	0.2079
		L	0.1624	0.1551 (0.0294)	0.0999	0.2130
		LS	0.1673	0.1600 (0.0294)	0.1045	0.2177
		O	0.1598	0.1526 (0.0291)	0.0979	0.2101
		OS	0.1548	0.1477 (0.0291)	0.0933	0.2052
		M	0.1648	0.1574 (0.0299)	0.1012	0.2163
		Traditional interval based on the estimator M: (0.0893, 0.2404)				
100	0.1617	D	0.1559	0.1524 (0.0205)	0.1134	0.1930
		DS	0.1584	0.1548 (0.0205)	0.1158	0.1954
		L	0.1602	0.1566 (0.0208)	0.1170	0.1978
		LS	0.1626	0.1590 (0.0208)	0.1194	0.2001
		O	0.1589	0.1553 (0.0207)	0.1160	0.1964
		OS	0.1565	0.1529 (0.0207)	0.1136	0.1939
		M	0.1614	0.1577 (0.0210)	0.1178	0.1993
		Traditional interval based on the estimator M: (0.1082, 0.2146)				
250	0.1599	D	0.1577	0.1563 (0.0133)	0.1311	0.1822
		DS	0.1588	0.1572 (0.0133)	0.1320	0.1832
		L	0.1594	0.1580 (0.0133)	0.1326	0.1841
		LS	0.1604	0.1589 (0.0133)	0.1335	0.1850
		O	0.1589	0.1575 (0.0133)	0.1321	0.1835
		OS	0.1580	0.1565 (0.0133)	0.1312	0.1825
		M	0.1599	0.1584 (0.0134)	0.1330	0.1846
		Traditional interval based on the estimator M: (0.1263, 0.1935)				

Table 6. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 3, p = 3$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.0707	D	0.0648	0.0610 (0.0177)	0.0303	0.0972
		DS	0.0688	0.0647 (0.0182)	0.0330	0.1018
		L	0.0701	0.0660 (0.0186)	0.0336	0.1039
		LS	0.0742	0.0699 (0.0191)	0.0364	0.1086
		O	0.0685	0.0645 (0.0183)	0.0326	0.1019
		OS	0.0645	0.0607 (0.0178)	0.0299	0.0973
		M	0.0708	0.0666 (0.0190)	0.0335	0.1053
		Traditional interval based on the estimator M: (0.0254, 0.1163)				

Table 6. Continued

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
100	0.0687	D	0.0657	0.0638 (0.0128)	0.0404	0.0901
		DS	0.0677	0.0657 (0.0130)	0.0420	0.0923
		L	0.0684	0.0663 (0.0131)	0.0424	0.0933
		LS	0.0704	0.0683 (0.0133)	0.0440	0.0995
		O	0.0676	0.0656 (0.0130)	0.0418	0.0924
		OS	0.0656	0.0637 (0.0128)	0.0403	0.0901
		M	0.0687	0.0667 (0.0132)	0.0425	0.0939
		Traditional interval based on the estimator M: (0.0368, 0.1007)				
250	0.0676	D	0.0663	0.0655 (0.0082)	0.0500	0.0823
		DS	0.0671	0.0663 (0.0083)	0.0507	0.0832
		L	0.0674	0.0666 (0.0083)	0.0509	0.0835
		LS	0.0682	0.0673 (0.0084)	0.0516	0.0843
		O	0.0671	0.0663 (0.0083)	0.0506	0.0832
		OS	0.0663	0.0655 (0.0082)	0.0500	0.0823
		M	0.0675	0.0667 (0.0083)	0.0510	0.0837
		Traditional interval based on the estimator M: (0.0474, 0.0876)				

Table 7. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 1, p = 5$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.3263	D	0.2907	0.2790 (0.0335)	0.2012	0.3408
		DS	0.2965	0.2781 (0.0329)	0.2085	0.3455
		L	0.3105	0.290 8(0.0353)	0.2166	0.3637
		LS	0.3159	0.2966 (0.0345)	0.2238	0.3678
		O	0.3083	0.2886 (0.0359)	0.2139	0.3636
		OS	0.3020	0.2820 (0.0364)	0.2062	0.3583
		M	0.3263	0.3055 (0.0384)	0.2265	0.3870
		Traditional interval based on the estimator M: (0.2169, 0.4358)				
100	0.3177	D	0.3000	0.2904 (0.0252)	0.2399	0.3404
		DS	0.3028	0.2932 (0.0249)	0.2432	0.3427
		L	0.3106	0.3007 (0.0261)	0.2486	0.3529
		LS	0.3133	0.3034 (0.0258)	0.2419	0.3550
		O	0.3090	0.2992 (0.0262)	0.2471	0.3515
		OS	0.3061	0.2962 (0.0264)	0.2437	0.3491
		M	0.3180	0.3079 (0.0272)	0.2541	0.3627
		Traditional interval based on the estimator M: (0.2434, 0.3926)				

Table 7. Continued

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
250	0.3122	D	0.3048	0.3009 (0.0163)	0.2687	0.3332
		DS	0.3059	0.3020 (0.0162)	0.2699	0.3341
		L	0.3092	0.3053 (0.0166)	0.2725	0.3382
		LS	0.3103	0.3064 (0.0165)	0.2737	0.3391
		O	0.3085	0.3045 (0.0165)	0.2718	0.3374
		OS	0.3074	0.3034 (0.0166)	0.2706	0.3364
		M	0.3121	0.3081 (0.0168)	0.2749	0.3415
		Traditional interval based on the estimator M: (0.2662, 0.3579)				

Table 8. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 2, p = 5$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.1698	D	0.1491	0.1376 (0.0275)	0.0850	0.1927
		DS	0.1566	0.1449 (0.0277)	0.0917	0.2000
		L	0.1622	0.1501 (0.0289)	0.0947	0.2078
		LS	0.1697	0.1575 (0.0290)	0.1016	0.2150
		O	0.1596	0.1476 (0.0288)	0.0926	0.2051
		OS	0.1518	0.1400 (0.0286)	0.0858	0.1973
		M	0.1698	0.1571 (0.0301)	0.0904	0.1927
		Traditional interval based on the estimator M: (0.0904, 0.2469)				
100	0.1642	D	0.1541	0.1482 (0.0205)	0.1095	0.1884
		DS	0.1578	0.1519 (0.0205)	0.1130	0.1920
		L	0.1607	0.1547 (0.0210)	0.1150	0.1958
		LS	0.1644	0.1583 (0.0210)	0.1186	0.1994
		O	0.1595	0.1534 (0.0209)	0.1139	0.1994
		OS	0.1557	0.1497 (0.0209)	0.1104	0.1907
		M	0.1645	0.1582 (0.0214)	0.1178	0.2002
		Traditional interval based on the estimator M: (0.1106, 0.2183)				
250	0.1608	D	0.1566	0.1542 (0.0130)	0.1291	0.1800
		DS	0.1581	0.1557 (0.0130)	0.1305	0.1815
		L	0.1593	0.1569 (0.0131)	0.1315	0.1829
		LS	0.1607	0.1583 (0.0131)	0.1329	0.1844
		O	0.1588	0.1564 (0.0131)	0.1310	0.1824
		OS	0.1573	0.1549 (0.0131)	0.1296	0.1809
		M	0.1607	0.1583 (0.0132)	0.1327	0.1846
		Traditional interval based on the estimator M: (0.1270, 0.1945)				

Table 9. Estimated values and bootstrap percentile intervals for the actual error rate ($\Delta = 3, p = 5$)

n	$\tau(\widehat{\xi})$	Error rate estimators	Estimates	Bootstrap estimates (St. Dev.)	Bootstrap lower bounds	Bootstrap upper bounds
50	0.0737	D	0.0662	0.0559 (0.0171)	0.0266	0.0908
		DS	0.0681	0.0615 (0.0178)	0.0304	0.0976
		L	0.0703	0.0634 (0.0184)	0.0313	0.1008
		LS	0.0765	0.0693 (0.0192)	0.0355	0.1079
		O	0.0685	0.0617 (0.0182)	0.0302	0.0987
		OS	0.0624	0.0561 (0.0174)	0.0263	0.0917
		M	0.0739	0.0665 (0.0194)	0.0327	0.1058
		Traditional interval based on the estimator M: (0.0281, 0.1198)				
100	0.0702	D	0.0643	0.0610 (0.0127)	0.0381	0.0868
		DS	0.0672	0.0638 (0.0130)	0.0404	0.0901
		L	0.0683	0.0648 (0.0132)	0.0410	0.0916
		LS	0.0713	0.0677 (0.0135)	0.0434	0.0949
		O	0.0675	0.0641 (0.0131)	0.0404	0.0906
		OS	0.0645	0.0612 (0.0128)	0.0381	0.0873
		M	0.0701	0.0665 (0.0135)	0.0421	0.0939
		Traditional interval based on the estimator M: (0.0380, 0.1021)				
250	0.0681	D	0.0658	0.0644 (0.0082)	0.0496	0.0811
		DS	0.0669	0.0656 (0.0082)	0.0500	0.0823
		L	0.0674	0.0660 (0.0083)	0.0503	0.0829
		LS	0.0685	0.0671 (0.0084)	0.0514	0.0842
		O	0.0671	0.0657 (0.0083)	0.0501	0.0826
		OS	0.0659	0.0645 (0.0082	0.0491	0.0813
		M	0.0681	0.0667 (0.0084)	0.0509	0.0837
		Traditional interval based on the estimator M: (0.0479, 0.0882)				

Acknowledgement Many thanks to the referee, whose helpful comments and suggestions led to a drastic reorganization and improvement in this manuscript.

References

[1] Anderson, T. W. *An introduction to multivariate statistical analysis, Second edition* (John Wiley& Sons. Inc., New York, 1984).

[2] Davison, A. C. and Hinkley, D. V. *Bootstrap methods and their application* (Cambridge University Press, Cambridge, 1997).

[3] DiCicio, T. J. and Romano, J. P. *A Review of bootstrap confidence intervals* (with discussion), J. R. Statist. Soc. B. **50**, 338–370, 1988.

[4] Efron, B. *Bootstrap methods: Another look at the Jackknife*, Ann. Math. Statist. **7**, 1–26, 1979.

[5] Efron, B. *Nonparametric standard errors and confidence intervals* (with discussion), Canad. J. Statist. **9**, 139–172, 1981.

[6] Efron, B. *The Jackknife, the Bootstrap and other resampling plans* (J. W. Arrowsmith Ltd., Bristol, England, 1982).

[7] Efron, B. *Bootstrap confidence intervals for a class of parametric problems*, Biometrica **72**, 45–58, 1985.
[8] Efron, B. *Better Bootstrap confidence intervals* (with discussion), J. Amer. Statist. Assoc. **82**, 171–200, 1987.
[9] Efron, B. and Gong, G. *A Leisurely look at the Bootstrap, the Jackknife and Cross Validation*, The American Statistician **37**, 36–48, 1983.
[10] Efron, B. and Tibshirani, R. *Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy* (with discussion), Statist. Science **1**, 54–77, 1986.
[11] Fisher, R. A. *The Use of the multiple measurements in taxonomic problems*, Annals of Eugenics **7**, 179–188, 1936.
[12] Hall, P. *Theoretical comparison of Bootstrap confidence intervals* (with discussion), Ann. Statist. **16**, 927–985, 1988.
[13] Hand, D. *Recent advances in error rate estimation*, Pattern Recognition **10**, 211–222, 1986.
[14] Hand, D. *Construction and assessment of classification rules* (John Wiley & Sons, 1997).
[15] Hills, M. *Allocation rules and their error rates*, Journal of the Royal Statistical Society, Ser. B. **28**, 1–20, 1966.
[16] John, S. *Errors in discrimination*, Ann. Math. Statist. **32**, 1125–1144, 1961.
[17] Krzanowski, W. J. *Data-based interval estimation of classification error rates*, Journal of Applied Statistics **28** (5), 585–595, 2001.
[18] Kiers, H. A. L. *Bootstrap confidence intervals for three way methods*, J. Chemometrics **18**, 22–36, 2004.
[19] Lachenbruch, P. A. *An Almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis*, Biometrics **23**, 639–645, 1967.
[20] Lachenbruch, P. A. *Discriminant analysis* (Hafner Press., New York, 1975).
[21] Lachenbruch, P. A. and Mickey, M. R. *Estimation of error rates in discriminant analysis*, Technometrics **10**, 1–11, 1968.
[22] McLachlan, G. J. *An Asimptotic unbiased technique for estimation the error rates in discriminant analysis*, Biometrics **30**, 239-249, 1974.
[23] McLachlan, G. J. *Confidence intervals for the conditional probability of misallocation in discriminant analysis*, Biometrics **32**, 161–167, 1975.
[24] McLachlan, G. J. *Discriminant analysis and statistical pattern recognition* (John Wiley & Sonsc Inc., New York, 1992).
[25] Okamoto, M. *An asymptotic expansion for the distribution of the linear discriminant function*, Annals of Math. Statistics **34**, 1286–1301, 1963.
[26] Sayre, J. W. *The Distribution of the actual error rates in discriminant analysis*, J. Amer. Statist. Assoc. **75**, 201–205, 1980.
[27] Sitgreaves, R. *Some results on the distribution of the W-classification statistics* in *Studies in Item Analysis and Prediction*, H. Solomon ed., Stanford, Calif. Stanford University Press., 241–251, 1961.
[28] Smith, C. A. B. *Some examples of discrimination*, Annals of Eugenics **18**, 272–282, 1947.
[29] Snapinn, S. M. *An Evaluation of smoothed error rate estimators in discriminant analysis* (Institute of Statistics Mimeo Series No. 1438, Dept. of Biostatistics, Univ. of N.C. at Chapel Hill, 1983).
[30] Timmerman, M. E. and Ter Braak, C. J. F. *Bootstrap confidence intervals for principal response curves*, Comput. Statis. and Data Anal. **52** (4), 1837–1849, 2007.
[31] Toussaint, G. T. *Bibliography on estimation of misclassification*, IEEE Trans. Inform. Theory IT-20, 472–479, 1974.
[32] Wald, A. *On Statistical problems arising in the classification of an individual into one of two groups*, Ann. Math. Statist. **15**, 145–162, 1944.
[33] Welch, B. L. *Note on the discriminant functions*, Biometrika **31**, 218–220, 1939.
[34] Wasserman, L. *All of nonparametric statistics* (Springer, New York, USA, 2006).