RT-MAE 2006-19

CORRECTIONS FOR THE SCORE STATISTIC USED IN THE CLASSIFICATION OF PATTERNS

Manuel R. Sena Jr, Gauss M. Cordeiro
and
Lúcia P. Barroso

Palavras-Chave: Bartlett-type correction, pattern recognition, score statistic.
Classificação AMS: 62F04, 62H15, 62H30.

- Dezembro 2006 -

CORRECTIONS FOR THE SCORE STATISTIC USED IN THE CLASSIFICATION OF PATTERNS

Manoel R. Sena Jr.
Departamento de Estatística
Universidade Federal de Pernambuco
Av. Prof. Moraes Rego, 1235, Recife, Brazil, 50670-901
manoel@de.ufpe.br

Gauss M. Cordeiro
Universidade Federal Rural de Pernambuco
gauss@deinfo.ufrpe.br

Lúcia P. Barroso
Universidade de São Paulo
lbarroso@ime.usp.br

Key Words: Bartlett-type correction, pattern recognition, score statistic.

ABSTRACT

The score statistic, used for taking the decision in the classification area, has been subject of various research efforts in the direction to improve matching of an observed distribution to a known distribution. A problem in using this tool is that, in general, it is necessary that the number of observations of the object to be classified, used in estimating the parameters must be sufficiently large to guarantee the convergence of the statistic being tested. In situations where this number is small or moderate, the error of classification can be substantial. In a system of signatures verification, Nelson, Turin and Hastie (1994) have developed a method of classification based on the score statistic which approximates to the chi-squared distribution, obtaining excellent results in terms of rates of errors of classification, depending on the sample sizes. It is possible to obtain exact distributions for the statistics beta and Hotelling.

Cordeiro and Ferrari (1991) have shown that all statistics which have asymptotic chi-squared distribution can be expanded up to order n^{-1}, where n is the sample size, and therefore be corrected to obtain better results in relation to the critical values of this distribution. A simulated study and a real example involving signatures are presented, to compare the performance of the original method and the two methods that define the score statistic and its Bartlett-type corrected statistic.

1. INTRODUCTION

Bartlett corrections have been widely used to improve the chi-squared distribution of the likelihood ratio tests. More recently, extensions of these corrections, called Bartlett-type corrections, have been used to approximate other statistics that have asymptotic chi-squared distribution. In particular, Cordeiro and Ferrari (1991) have shown how the original statistics may be transformed to a new corrected statistic which has a chi-squared distribution to order n^{-1}, where n is the sample size. The main purpose of this paper is to use this approach to the score statistic proposed by Nelson, Turin and Hastie (1994) for signatures verification, and thus to improve the current rates of classifications. A simulated study involving genuine and false signatures of an author is presented to measure the performance of the original and the modified procedure, where errors rates are estimated in terms of false acceptance and false rejection.

2. PATTERN RECOGNITION

In rather general terms, pattern recognition process is based on defining an association rule, capable of classifying any individual (possibly multivariate) observation from a given population, into one of a predefined number of classes. Using this rule, any posterior observation may be associated with the existing classes, with a controlled margin of error. For defining the underlying set of association rules, three main approaches have been used to tackle the pattern recognition problem: *i*) structural pattern recognition, where symbolic architectures are constructed, with subsequent stages determining the decision rule (see e.g. Fu, 1986), *ii*) neural networks, which use observation sample to "learn" (through a training process) the characteristics of the pattern to be recognized (see Khanna, 1990), and *iii*) statistical pattern recognition, where discriminant analysis is used to define a function that minimizes probability of misclassification (see e.g. Schalkoff, 1992). In this work, we concentrate on the last of the three approaches.

From the statistical point of view, pattern recognition may be formulated for two distinct situations: *i*) when we are able to collect information about all of the patterns, and based on these observations we can find functions to discriminate theses groups and associate a new object to one of theses patterns, *ii*) when we can collect information about only one of the patterns and we have to associate a new object to this pattern or not.

In the first case, studies made by Press and Wilson (1978), suggest the use of proportional function of Mahalanobis distance as discriminant, when the observations are supposed to follow a multivariate normal distribution, or the use of the logistic regression when the normality assumption is violated. In the second case, Nelson, Turin and Hastie (1994) have recently proposed a method for classifying genuine signatures, assuming normality for the vector of observations. In this paper we shall deal only with the second situation and we present methods that evaluate the classification of patterns, considered for signatures verification. We use only the genuine observations to construct the statistic of the test for the decision rule.

3. CONSTRUCTION OF THE SCORE STATISTIC

Consider the conditional distribution of a given random vector $\mathbf{x}$ which belongs to class C_i $(i = 1, 2)$, and has a multivariate normal distribution given by

$$f(\mathbf{x}|C_i) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{x}-\mu_i)\right\}, \; i = 1, 2,$$

where μ_i and $\mathbf{\Sigma}_i$ are, respectively, the vector of averages and the matrix of covariances of the class C_i, and p is the number of features of vector $\mathbf{x}$. The goal is to find R_i $(i = 1, 2)$ that belongs to a vectorial space R, satisfying $R_1 \cup R_2 = R$ and $R_1 \cap R_2 = \phi$, in such a way that we can decide to associate a new observation, say $\mathbf{x}_0$, to the C_i class, if $\mathbf{x}_0$ belongs to R_i. In other words, we want to test $H_0 : \mathbf{x}_0 \in C_1$ against $H_0 : \mathbf{x}_0 \in C_2$, defining the error rates as

$$\begin{aligned}
\alpha &= \Pr(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \\
&= \int_{R_2} f(\mathbf{x}|C_1)\mathrm{d}(\mathbf{x}) \\
\beta &= \Pr(\text{accept } H_0 \text{ when } H_0 \text{ is false}) \\
&= \int_{R_1} f(\mathbf{x}|C_2)\mathrm{d}(\mathbf{x}).
\end{aligned}$$

As the proposal of this work is one particular case where only one of the classes can be observed, (we choose C_1), we cannot evaluate the value of β, therefore supposedly we do not know the conditional distribution of vector $\mathbf{x}$ in the C_2 class. The solution for this problem passes through the following considerations:

1) Define the hypotheses to be tested by

$$H_0 : \mathbf{x}_0 \in C_1$$
$$H_a : \mathbf{x}_0 \notin C_1$$

2) Suppose that $\mathbf{x} \sim N_p(\mu_1, \Sigma_1)$, and define α by

$$\alpha = \int_{R-R_1} f(\mathbf{x}|C_1)\mathrm{d}(\mathbf{x}).$$

The alternative form established to define the subspace $R - R_1$ is to relate the rejection of H_0 with a small probability α, and this is done by increasing $(\mathbf{x} - \mu_1)^T \Sigma_1^{-1}(\mathbf{x} - \mu_1)$. It should be noted that this expression is known as the Mahalanobis distance of vector $\mathbf{x}$ to the center of the distribution (μ_1). If vector $\mathbf{x}$ has multivariate normal distribution, it is easy to verify that this expression follows a chi-squared distribution with p degrees of freedom. From now on, for simplicity, when we suppose that $\mathbf{x}$ belongs to class C_1, we shall use the notation $\mathbf{x} \sim N_p(\mu, \Sigma)$.

In practice, both μ and Σ are unknown. There are many ways of defining the dissimilarity function $d(\mathbf{x})$. Nelson, Turin and Hastie (1994) considered to substitute these parameters for their respective unbiased estimators, assuming that $\mathbf{x} \sim N_p(\mu, \Sigma)$, thus defining the statistic

$d(\mathbf{x})$ given by

$$d(\mathbf{x}) = (\mathbf{x} - \overline{\mathbf{x}})^T \mathbf{S}^{-1}(\mathbf{x} - \overline{\mathbf{x}}), \tag{1}$$

where $\mathbf{x}$ is a new vector, $\overline{\mathbf{x}}$ and $\mathbf{S}$ are, respectively, the vector of averages and the matrix of covariance of the observations that do not include the vector $\mathbf{x}$. Note that $d(\mathbf{x})$ has an asymptotic chi-square distribution, since when n tends to the infinity, $\overline{\mathbf{x}}$ tends to μ and $\mathbf{S}$ tends to $\mathbf{\Sigma}$.

4. CORRECTION OF THE SCORE STATISTIC

Let d be a statistic that has an asymptotic chi-square distribution. Therefore, under some regularity conditions, the null distribution (up to order of n^{-1}) of d, as shown by Chandra (1985), can be written as

$$\Pr(d \leq t) = F_p(t) + \sum_{i=0}^{k} a_i F_{q+2i}(t), \tag{2}$$

where the a_i's are functions of order n^{-1} of the unknown parameters, satisfying $\sum_{i=0}^{k} a_i = 0$, and $F_h(t)$ denotes the distribution function of a random variable with chi-square distribution and h degrees of freedom. Cordeiro and Ferrari (1991) defined the modified statistic (d^*) by

$$d^* = d(1 - \sum_{i=1}^{k} c_i d^{i-1}), \tag{3}$$

where $c_i = 2(\sum_{l=0}^{i-1} a_l)(\mu_i')^{-1}$, with $\mu_i' = E(\chi_p^2)^i = \prod_{l=0}^{i-1}(p + 2l)$, and showed that up to order n^{-1} we have

$$\Pr(d^* \leq t) = \Pr(\chi_p^2 \leq t).$$

The correction $(1 - \sum_{i=1}^{k} c_i d^{i-1})$ in (3) is known as Bartlett-type correction. The usual form of this correction is defined as $d^* = d(1 - B)$, where B is a polynomial of order n^{-1} in the proper statistic d. The development of the corrected statistic d^* is based on the first k moments, as it is seen bellow.

4.1. METHOD OF THE MOMENTS

Cordeiro and Ferrari (1991) showed that the classical Bartlett correction for the likelihood rate test is obtained with $k = 1$ and $a_0 = -a_1 = -b/2$, where b is the term of order n^{-1}

in the calculation of the expected value of the statistic. The term b can be obtained from Lawley (1956). Also, in case of the score statistic, the Bartlett-type correction is a special case of the expression (3) with $k = 3$, that is,

$$d^* = d[1 - (c_1 + c_2 d + c_3 d^2)].$$

Cordeiro and Ferrari (1998) have also used the fact that the c_i's are of order n^{-1} to write the modified statistics $(d^*)^j$ as

$$(d^*)^j = d^j - j d^{j-1}(c_1 d + c_2 d^2 + c_3 d^3), \tag{4}$$

where terms of order inferior to n^{-1} are ignored and, on the basis of the calculation of the expected value for each j, on both sides of the expression (4), we can construct a system of linear equations for the unknown quantities c_1, c_2 and c_3, as function of the first three moments of the statistic d. They concluded, assuming that the c_i's are chosen in such a way that d^* has a chi-square distribution up the order n^{-1}, that

$$\mu_j' = m_j' - j \sum_{i=1}^{3} c_i m_{i+j-1}',$$

where μ_j' and m_j' are the moments of order j of a chi-square distribution and of the statistic d, respectively. This equation can be written as

$$\frac{(m_j' - \mu_j')}{j} = \sum_{i=1}^{3} c_i m_{i+j-1}'.$$

From the expansion (2), we have to order one $m_{i+j-1}' = \mu_{i+j-1}'$, and since the c_i's are of order n^{-1}, we can write

$$\sum_{i=1}^{3} c_i \mu_{i+j-1}' = \frac{b_j}{j} \quad j = 1, 2, \ldots, \tag{5}$$

where b_j is the term of order n^{-1} in the expansion of the jth null moment (m_j') of d.

5. APPROACHES TO STATISTIC d

We now consider three alternative forms of balance of statistics d when we evaluate a new observation. In two cases we use the approach for an accurate distribution, and in the third case we use the Bartlett-type correction. In the accurate forms, therefore, we treat the problem in two ways: *i*) assuming initially n observations to calculate the estimators $\bar{\mathbf{x}}$ and $\mathbf{S}$ and after that we measure the distance d for the new observation $\mathbf{x}$, *ii*) we have initially $(n-1)$ observations, and after we include the new observation in the sample for the computation of the estimators, and measure the distance d for all the observations, including the new one. The difference is that in the first case dependence between the new observation and the estimators does not exist, since it was not used in the calculation, while it does not occur in the second case. In the following we present the two approaches and at the end of the section we will use the method of the moments to correct the statistic d up to order n^{-1}.

5.1. ACCURATE METHOD BY CONSIDERING **n** OBSERVATIONS

Assume that we have a sample of n observations from the p-variate normal population to calculate the statistics $\bar{\mathbf{x}}$ and $\mathbf{S}$. Suppose that we have a new observation from the same population. The three theorems and a definition, presented bellow, are based on Wishart, Hotelling T^2 and Snedecor F distributions, discussed for example in Mardia et al.(1979).

DEFINITION: Let $\mathbf{y}$ be a random vector with distribution $N_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W}$ an independent random matrix of $\mathbf{y}$ with Wishart distribution with parameters $\mathbf{I}$ and m, denoted by $\mathbf{W} \sim W_p(\mathbf{I}, m)$, where p is the dimension of the vector $\mathbf{y}$ and of the square matrix $\mathbf{W}$, $\mathbf{I}$ is the identity matrix of dimension p and m is the number of degrees of freedom of $\mathbf{W}$. Write the statistic $\mathbf{T}$ as

$$\mathbf{T} = m\mathbf{y}^T\mathbf{W}^{-1}\mathbf{y}.$$

We can say that $\mathbf{T}$ has a T^2 Hotelling distribution with parameters p and m and use the notation $\mathbf{T} \sim T^2(p, m)$.

THEOREM 1: Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ be n random vectors where each one has $N_p(\mu, \Sigma)$ distribution and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$. Then, $(n-1)\mathbf{S} \sim W_p(\Sigma, n-1)$.

THEOREM 2: If $\mathbf{W} \sim W_p(\Sigma, m)$, then $\Sigma^{-1/2}\mathbf{W}\Sigma^{-1/2} \sim W_p(\mathbf{I}, m)$.

THEOREM 3: If $\mathbf{T} \sim T^2(p, m)$, then $(\frac{\mathbf{T}}{m})(\frac{m-p+1}{p}) \sim F(p, m-p+1)$.

Considering that $\mathbf{x}$ has $N_p(\mu, \Sigma)$ distribution and defining $\mathbf{y} = \sqrt{\frac{n}{n+1}}\Sigma^{-1/2}(\mathbf{x} - \overline{\mathbf{x}})$, it follows that $\mathbf{y}$ has $N_p(\mathbf{0}, \mathbf{I})$ distribution since $\mathbf{x}$ and $\overline{\mathbf{x}}$ are independents. Now, define $\mathbf{W} = (n-1)\Sigma^{-1/2}\mathbf{S}\Sigma^{-1/2}$, it thus follows that $\mathbf{W}$ has $W_p(\mathbf{I}, n-1)$ distribution and therefore, satisfies the conditions of the definition above, that is,

$$\begin{aligned}
\mathbf{T} &= (n-1)\mathbf{y}^T\mathbf{W}^{-1}\mathbf{y} \\
&= \frac{n}{n+1}(\mathbf{x} - \overline{\mathbf{x}})^T\mathbf{S}^{-1}(\mathbf{x} - \overline{\mathbf{x}}) \\
&= \frac{n}{n+1}d \sim T^2(p, n-1).
\end{aligned}$$

Writing d as a function of T and using the relation between the distributions F and T^2, we have

$$\begin{aligned}
d &= \frac{n+1}{n}\mathbf{T} \\
&= \frac{n+1}{n}\frac{(n-1)p}{(n-p)}\mathbf{F},
\end{aligned}$$

where $\mathbf{F}$ is a random variable following a F distribution with p and $(n-p)$ degrees of freedom ($\mathbf{F} \sim F(p, n-p)$). In other words, by Theorem 3, we can easily verify that a variable with T^2 distribution is proportional to that of F distribution and in our case, that we can balance statistic d and so establish a critical point that assures that only one fixed number of observations, let us say $\alpha\%$, are above this value by means of the modification of the statistic d. Thus, writing the statistic $\mathbf{F}$ as a function of the statistic d, we have an accurate distribution for a measure of distance, given by

$$\mathbf{F} = \frac{n-p}{p(n-1)}\frac{n}{n+1}d,$$

where $\mathbf{F} \sim F(p, n-p)$. In the applications, this measure will be called $d(T^2)$ and using the notation $d(T^2) \sim F(p, n-p)$ we shall say that $d(T^2)$ has a F distribution with parameters p and $(n-p)$.

5.2. ACCURATE METHOD BY CONSIDERING $(\mathbf{n-1})$ OBSERVATIONS

Another form to evaluate the distance d is to calculate the n distances of each vector $\mathbf{x}_i$ $(i = 0, 1, 2..., n-1)$ in relation to $\overline{\mathbf{x}}$ weighted by $\mathbf{S}^{-1}$ (the $(n-1)$ observations of the sample plus the new observation, $\mathbf{x}_0$, to be tested). Assuming that $\mathbf{x}_i$ has distribution $N_p(\mu, \Sigma)$, in a study of robustness, Gnanadesikan and Kettenring (1972), showed that the accurate distribution of u_i given by

$$u_i = \frac{n}{(n-1)^2} d_i, \tag{6}$$

is a beta distribution with parameters $\frac{p}{2}$ and $\frac{n-p-1}{2}$, where d_i is given by

$$d_i = d(\mathbf{x}_i) = (\mathbf{x}_i - \overline{\mathbf{x}})^T S^{-1} (\mathbf{x}_i - \overline{\mathbf{x}}). \tag{7}$$

Note that d_i in expression (7) is similar to d of the expression (1), except for the fact that each $\mathbf{x}_i$ is used in the calculations of $\overline{\mathbf{x}}$ and S in (7). Thus, we have an accurate form to evaluate the distance d, therefore when a new observation is introduced, it will be included in the sample for the calculation of $\overline{\mathbf{x}}$ and S, and after that its distance d_0 is evaluated. We call $d(Beta) = u_0$ a weighted form of the distance in (7) and using the notation $d(Beta) \sim Beta(p/2, (n-p-1)/2)$, we say that $d(Beta)$ has a beta distribution with parameters $(p/2)$ and $[(n-p-1)/2]$.

5.3. CORRECTION OF THE STATISTIC d BY THE METHOD OF MOMENTS

We can now calculate the moments of the statistic d through the beta distribution, rewriting the expression (6), raising to the jth power and calculating its expected value, that is,

$$E(d^j) = \left(\frac{(n-1)^2}{n}\right)^j E(u^j),$$

where $u \sim Beta(p/2, (n-p-1)/2)$. Now if $u \sim Beta(a, b)$, the jth moment of u (see Johnson and Wichern (2002)) is given by $E(u^j) = B(j+a, b)/B(a, b)$, where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$, and $\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) dx$. Using the fact that $\Gamma(a) = (a-1)\Gamma(a-1)$, the first three moments of the statistic d are easily obtained

$$\begin{aligned}
m_1'(d) &= E(d) = p(1 - \frac{1}{n}) \\
m_2'(d) &= E(d^2) = p(p+2)(1 - \frac{4}{n} + O(n^{-2})) \\
m_3'(d) &= E(d^3) = p(p+2)(p+4)(1 - \frac{9}{n} + O(n^{-2})).
\end{aligned}$$

Considering the system of equations (5), the five first necessary moments of the chi-squared distribution are:

$$\begin{aligned}
\mu_1' &= p \\
\mu_2' &= p(p+2) \\
&\vdots \\
\mu_5' &= p(p+2)(p+4)(p+6)(p+8).
\end{aligned}$$

Then, the first three terms b_j's in the expansion of jth null moment (m_j') of the statistic d to order n^{-1} are

$$b_1 = -\frac{p}{n}, \quad b_2 = -\frac{4p(p+2)}{n} \quad \text{and} \quad b_3 = -\frac{9p(p+2)(p+4)}{n}.$$

Therefore, our system of equations to determine the values of c_1, c_2 and c_3 based on equations (5) are

$$\begin{aligned}
c_1\mu_1' + c_2\mu_2' + c_3\mu_3' &= b_1 \\
c_1\mu_2' + c_2\mu_3' + c_3\mu_4' &= \frac{b_2}{2} \\
c_1\mu_3' + c_2\mu_4' + c_3\mu_5' &= \frac{b_3}{3}.
\end{aligned}$$

The solution of the system is

$$c_1 = \frac{p}{2n}, \quad c_2 = -\frac{1}{2n} \quad \text{and} \quad c_3 = 0.$$

Therefore, with the values of c_1, c_2 and c_3, we obtain the corrected statistic d^* as

$$d^* = d\left\{1 - (c_1 + c_2 d + c_3 d^2)\right\} = d\left(1 + \frac{d-p}{2n}\right),$$

where

$$\begin{aligned}
\Pr(d \le t) &= \Pr(\chi_p{}^2 \le t) + O(1) \quad \text{and} \\
\Pr(d^* \le t) &= \Pr(\chi_p{}^2 \le t) + O(n^{-1}).
\end{aligned}$$

6. APPLICATIONS

In this section we present applications for on-line signature verification of an author. In the first part we use the data set only to simulate vectors with known distributions. In the second part we use a real set of 1000 genuine signatures and 825 false, obtained in the Departamento de Comunicações of the Faculdade de Engenharia Elétrica e da Computação of the Universidade Estadual de Campinas, Brazil. The objective of the first part is to construct a limitless set of vectors with similar characteristics of real signatures, where we can study the rates of convergence. In the second part, we study a real data set, by assuming that the conditions of normality and independence of the observations are satisfied and we can control the type I error.

6.1. SIMULATED DATA

From the signatures set, we extract 100 genuine and 100 false signatures, with 42 basic features of an author to generate our data set. We first calculate the vectors of averages (μ_1 and μ_2) and the matrices of covariances (Σ_1 and Σ_2). These will only serve as reference, without any other specific purpose, because these values could be freely determined. Thus, we generate vectors according to the multivariate normal distribution, beyond the distributions chi-square and uniform with variances and covariances structures, in accordance with a combination of the values of p and n. The objective is to compare the pre-established percentile values (in our case $\alpha = 1\%, 5\%$ and 10%) and the percentages of observations that exceed the specified critical values for each one of the supposed distributions of d: without correction, called simply d; with accurate distributions, obtained by taking d to have exact T^2 and *beta* distributions, denoted here $d(T^2)$ and $d(Beta)$ respectively, and by means of the Bartlett-type correction denoted d^*.

Table 1 : Average percentage of observations that exceed the specified critical values, assuming normal distribution.

	$p = 2$ and $n = 5$				$p = 4$ and $n = 10$			
α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$
1	0.00	0.48	0.40	0.00	0.00	0.90	0.48	0.00
5	0.00	2.63	7.89	0.00	0.00	2.12	3.14	0.00
10	0.00	13.73	15.82	0.00	0.03	11.60	10.36	4.32
	$p = 2$ and $n = 10$				$p = 4$ and $n = 20$			
1	0.00	0.77	0.67	0.00	0.00	0.96	0.95	0.31
5	0.23	2.61	3.13	1.20	2.28	5.90	5.95	4.43
10	4.59	8.52	7.20	6.81	6.68	10.56	10.55	9.37
	$p = 2$ and $n = 20$				$p = 4$ and $n = 50$			
1	0.21	1.03	1.13	0.66	0.81	1.63	1.68	1.45
5	3.56	5.25	5.13	4.78	3.73	5.07	5.06	4.87
10	6.68	8.61	8.92	8.11	8.85	10.20	10.27	10.03

Keeping the assumption that n is always greater than p, this implies for practical purposes a choice of values of p smaller than the initial 42 entries of the data base. We adopt a process of choice of the main features of the signatures based on the coefficient of variation, CV, in order to select the p features that present smaller CV's. Therefore, we believe that features with high values of CV's are more easily to be imitated by falsifiers. Having established the value of p, we generate n observations of the vector with this dimension and respective distribution, called training sample, that serves for estimating μ_1 and Σ_1, represented by $\bar{\mathbf{x}}$ and S, respectively, and used in the calculation of the statistical tests. After that, we generate 1000 observations, in a similar way to the used procedure to generate the training sample, called test sample, that serves to obtain the rates of type I errors given by

$$\begin{aligned}
\alpha_d &= \Pr\left(d(\mathbf{x}) > \chi^2_{(p,\alpha)}\right), \\
\alpha_{d(Beta)} &= \Pr\left(d_{Beta}(\mathbf{x}) > Beta\left(\frac{p}{2}, \frac{n-p-1}{2}, \alpha\right)\right), \\
\alpha_{d(T^2)} &= \Pr\left(d_{T^2}(\mathbf{x}) > F(p, n-p, \alpha)\right), \\
\alpha_{d^*} &= \Pr\left(d^*(\mathbf{x}) > \chi^2_{(p,\alpha)}\right),
\end{aligned}$$

where $\chi^2_{(p,\alpha)}$, $Beta\left(\frac{p}{2}, \frac{n-p-1}{2}, \alpha\right)$ and $F(p, n-p, \alpha)$ are critical values that corresponds to the elements of the population that exceeds $\alpha\%$.

Table 2 : Average percentage of observations that exceed the specified critical values, assuming uniform distribution.

	$p=2$ and $n=5$				$p=4$ and $n=10$			
α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$
1	0.00	0.00	0.00	0.00	0.78	2.67	0.02	1.42
5	0.00	6.36	3.15	0.00	1.96	3.91	0.19	2.92
10	0.00	2.46	6.44	0.00	2.43	3.72	0.69	3.08
	$p=2$ and $n=10$				$p=4$ and $n=20$			
1	0.00	0.39	0.61	0.00	0.58	1.20	0.14	0.94
5	0.00	0.36	0.58	0.30	1.56	2.11	0.56	1.93
10	2.40	6.18	6.49	4.64	2.23	2.67	0.88	2.51
	$p=2$ and $n=20$				$p=4$ and $n=50$			
1	0.00	0.02	0.01	0.00	0.71	0.22	0.32	0.85
5	0.10	0.42	0.49	0.35	1.98	2.2	1.62	2.18
10	4.19	6.01	5.46	5.65	2.28	2.44	1.74	2.43

This procedure was repeated 100 times for some combinations of the sample size ($n = 5, 10, 20$ and 50) and dimension of the vector ($p = 2$ and 4), in accordance with the specified distribution (normal, uniform and chi-square) to calculate the average of the percentile values. The results of the estimate rates of errors $\widehat{\alpha}_d$, $\widehat{\alpha}_{d(Beta)}$, $\widehat{\alpha}_{d(T^2)}$ and $\widehat{\alpha}_{d^*}$ are given in Tables

1, 2 and 3. It can be noticed that, in accordance with each method, it is expected that only $\alpha\%$ of the observations used for the tests exceed the specified critical values.

From Table 1 we can confirm that, in case of the normal distribution, the rates of type I errors are usually closer to the corresponding nominal values for the accurate statistics, $d(Beta)$ and $d(T^2)$, than the corrected statistic d^*. When n is considered much bigger than p, the correction presents competitive results. When breaking the normality assumption, the Bartlett-type corrected statistic presents better performance, in the sense of that the estimate rates of type I errors follow up the growth of the expected nominal level.

Table 3 : Average percentage of observations that exceed the specified critical values, assuming chi-squared distribution.

	$p=2$ and $n=5$				$p=4$ and $n=10$			
α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$
1	0.00	6.40	0.00	0.30	0.72	3.04	0.01	1.55
5	0.34	5.37	0.02	1.45	1.93	3.71	0.16	2.66
10	1.75	6.20	0.26	3.22	1.96	3.06	0.32	2.48
	$p=2$ and $n=10$				$p=4$ and $n=20$			
1	0.57	2.90	0.20	1.75	0.52	1.07	0.23	0.80
5	1.67	3.03	0.70	2.48	1.82	2.27	0.54	2.13
10	4.04	5.28	2.29	4.74	2.25	2.64	1.23	2.53
	$p=2$ and $n=20$				$p=4$ and $n=50$			
1	0.64	1.21	0.53	0.97	0.72	0.21	0.44	0.81
5	2.10	2.69	2.23	2.58	1.67	1.91	1.21	1.87
10	2.83	3.34	2.20	3.23	2.24	2.42	1.15	2.39

We also observe that the estimate rates of type I errors are in general conservative when the generating distribution is not normal, with rare exception of some values of the accurate statistic $d(Beta)$. This can be understood as a positive fact, because in practice we shall hardly encounter a normality situation. It can be also observed that the estimate rates of type I errors are closer to the nominal level when the rate between n and p grows. This can be justified, in part, by the fact that the weight or the correction in d depends on this relation and tends to be closer to 1 when n grows in absolute terms or in relation to p. The improvement caused by the Bartlett-type correction is clear if it is compared with the statistic without correction, but it can also be seen in relation to the other two statistics, when the normality assumption is violated. This can be observed from Table 3, in cases $(p = 2, n = 5)$ and $(p = 4, n = 10)$. It is noted that there are almost no differences between the estimate rates of type I errors for the different values of α in the accurate statistics $d(Beta)$: (6.40, 5.37 and 6.20) and (3.04, 3.71 and 3.06) while the estimate rates for $d(T^2)$ are much smaller than the expected values: (0.00, 0.02 and 0.26) and (0.01, 0.16 and 0.32), respectively, when it would have to be (0.01, 0.05 and 0.10) in the four cases.

6.2. REAL SIGNATURE DATA

The data used in this section is the real data of signatures of an author, consisting of a set of 1000 observations of genuine signatures, obtained by means of an electronic digitizer, and codified in vectors of dimension 42. Each dimensional unit considered is a feature of the signature, for example the maximum height or the minimum width. Also a sample with 825 false signatures was considered. We assume, from now on, that each signature represents a p-variate observation of the normal distribution $(p = 2, 4$ and $10)$. We choose 100 samples of sizes n $(n = 5, 10, 20, 30$ and $50)$, depending on the value of p, of the genuine signatures to calculate the values of $\bar{\mathbf{x}}$ and S that will be used in the calculation of $d, d(Beta), d(T^2)$ and d^*. The rest will be used to estimate the rates of errors (type I and II), represented here by $\alpha_{(.)}$,

$$
\begin{aligned}
\alpha_d &= \Pr(d(\mathbf{x}) \geq \chi^2_{(p,\alpha)}), \\
\alpha_{d(Beta)} &= \Pr\left(d_{Beta}(\mathbf{x}) \geq Beta(\frac{p}{2}, \frac{n-p-1}{2}, \alpha)\right), \\
\alpha_{d(T^2)} &= \Pr\left(d_{T^2}(\mathbf{x}) \geq F(p, n-p, \alpha)\right), \\
\alpha_{d^*} &= \Pr(d^*(\mathbf{x}) \geq \chi^2_{(p,\alpha)}),
\end{aligned}
$$

where $\mathbf{x} \sim N_p(\mu_g, \Sigma_g)$, and $\beta_{(.)}$

$$\begin{aligned}
\beta_d &= \Pr(d(\mathbf{x}) \leq \chi^2_{(p,\alpha)}), \\
\beta_{d(Beta)} &= \Pr\left(d_{Beta}(\mathbf{x}) \leq Beta(\frac{p}{2}, \frac{n-p-1}{2}, \alpha)\right), \\
\beta_{d(T^2)} &= \Pr\left(d_{T^2}(\mathbf{x}) \leq F(p, n-p, \alpha)\right), \\
\beta_{d^*} &= \Pr(d^*(\mathbf{x}) \leq \chi^2_{(p,\alpha)}),
\end{aligned}$$

where $\mathbf{x} \sim N_p(\mu_f, \Sigma_f)$.

The subscript g and f that appear in the representation of the normal distribution, indicate, respectively, a sample of genuine and false signatures. As in the simulated case, based in the set of 100 samples, each with n observations and dimension p, we calculate the average number of genuine observations that exceed the respective critical values, as well as the average number of false observations that are smaller than this respective critical value. Results in the rates of types I and II errors are given in Tables 4, 5 and 6.

Table 4 : Estimated rates of type I and II errors (in %), for the real observations of signatures, fixed at $p = 2$.

α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\beta}_d$	$\widehat{\beta}_{d(Beta)}$	$\widehat{\beta}_{d(T^2)}$	$\widehat{\beta}_{d^*}$
				$n = 5$				
1	0.00	7.70	5.53	0.00	45.78	36.32	35.04	45.78
5	0.00	8.55	7.69	0.00	61.27	47.97	43.49	61.27
10	0.00	11.48	9.27	0.00	50.95	32.31	30.98	50.95
				$n = 10$				
1	0.00	6.71	5.58	4.51	75.04	63.35	62.15	69.83
5	6.06	8.84	7.98	7.48	64.68	52.07	51.49	57.56
10	9.14	11.67	11.27	10.67	53.26	46.08	44.45	48.72
				$n = 20$				
1	2.98	4.32	3.97	3.82	85.10	79.09	78.93	81.16
5	7.76	9.21	8.67	8.85	69.88	65.27	64.97	66.45
10	11.59	12.65	12.27	12.44	57.96	55.21	54.99	55.82

Table 5: Estimated rates of type I and II errors (in %), for the real observations of signatures, fixed at $p = 4$.

α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\beta}_d$	$\widehat{\beta}_{d(Beta)}$	$\widehat{\beta}_{d(T^2)}$	$\widehat{\beta}_{d^*}$
				$n = 10$				
1	0.00	6.09	4.94	0.00	78.48	65.94	63.42	78.48
5	0.00	11.02	9.21	4.53	74.18	44.45	43.20	68.90
10	4.80	13.40	11.98	8.98	69.67	39.48	38.05	54.56
				$n = 20$				
1	2.94	5.73	5.12	4.50	82.45	64.34	64.13	71.66
5	5.98	8.92	8.76	7.87	60.12	47.43	46.59	51.66
10	10.81	13.55	12.79	12.70	49.19	42.49	42.17	44.35
				$n = 50$				
1	3.49	4.47	4.45	4.26	75.81	70.23	70.63	71.28
5	7.95	8.98	8.93	8.83	52.63	49.15	49.21	49.68
10	11.29	12.15	12.05	12.05	42.48	40.55	40.66	40.78

The great challenge to find the best classification method is the estimated rate of type II error, for which the estimate value is widely large. It is noticed that in the case ($p = 2$ and $n = 5$) the estimate rates for type I error, for the two accurate methods, already present good results; however, it would be inadmissible a rate of approximately 31% of false acceptance, that was the lesser estimate rate ($\widehat{\beta}_{d(T^2)}$ with $\alpha = 10\%$). This indicates that the value of p must be a little bigger, therefore the same with $p = 4$ (see Table 5) this picture of high in the estimate rate of type II error continues, still that the estimates of the type I error are acceptable.

Table 6 : Estimated rates of type I and II errors (in %), for the real observations of signatures, fixed at $p = 10$.

α	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d(Beta)}$	$\widehat{\alpha}_{d(T^2)}$	$\widehat{\alpha}_{d^*}$	$\widehat{\beta}_d$	$\widehat{\beta}_{d(Beta)}$	$\widehat{\beta}_{d(T^2)}$	$\widehat{\beta}_{d^*}$
				$n = 20$				
1	0.00	4.51	3.89	0.00	93.12	34.97	31.23	93.12
5	0.00	10.09	8.85	4.20	89.67	16.47	15.17	42.92
10	3.04	12.28	11.34	6.89	47.92	10.86	9.84	2.15
				$n = 30$				
1	1.21	4.61	4.36	2.85	66.97	22.38	21.80	35.52
5	4.03	8.85	8.69	6.82	25.62	12.29	12.10	15.94
10	8.51	13.76	13.27	11.84	12.17	7.06	6.88	8.49
				$n = 50$				
1	2.35	4.68	4.65	3.88	2.61	16.25	16.17	18.52
5	7.02	9.74	9.77	9.05	12.72	9.25	9.13	9.98
10	11.47	13.97	13.87	13.39	7.56	5.95	5.89	6.28

The estimates only decay for acceptable values with $p = 10$ and $\alpha = 10\%$ (see Table 6). In these conditions, all the methods presented acceptable estimate rates, except the case without correction for $n = 20$. In this case, the corrected statistic presents better performance, since it estimates smaller rates type I error, with equivalents esteem rates of type II error.

7. COMMENTS AND CONCLUSIONS

In this work, we introduce two new techniques for signature classification, one accurate, using the T^2 of Hotelling distribution and another using Bartlett-type correction. The first seems to be more reliable in terms of application, since when we take a new observation, it is enough to directly calculate the distance without requiring to include it in the sample for the calculation of $\overline{\mathbf{x}}$ and $\mathbf{S}$ as it makes the one that uses the beta distribution. The Bartlett-type correction seems more appropriate when the normality condition breaks down.

The corrected statistic presents superior performance, in relation to the estimated rates of type I errors with low type II error, when compared to the statistics without correction. The case of the application involving signatures, we believe that it is impracticable a system that demands that the customer registers 20, 30 or 50 times signatures himself (herself), to compose the sample of training. However, others techniques of normalization and election of the ideal dimension of the vector can be used to construct a competitive system, as well as detailed studies more in relation to the number of features, p, and alternative forms of election. When we consider ($n = 20$ and $p = 10$) and establish $\alpha = 10\%$, the corrected statistic can be considered satisfactory for a system of verification of signatures, therefore, literature considers a good system of verification those that obtain rates around 20% for the addition of the errors (Plamondon and Lorette (1989)). The advantage of this case is that the rates of false rejections can previously be informed to the customer, that is, he (she) would be knowing that about 10% of the times it would have its proper rejected signature.

BIBLIOGRAPHY

Chandra, T.K. (1985). Asymptotic expansion of perturbed chi-squared variables. *Sankhya A*, **47**, 100–110.

Cordeiro, G.M. and Ferrari, S.L.P. (1991). A modified score test statistic having chi-squared distribution to order n^{-1}. *Biometrika*, **78**, 573–578.

Cordeiro, G.M. and Ferrari, S.L.P. (1998). A note on Bartlett-type correction for the first few moments. *Journal of Statistical Planning and Inference*, **71**, 261–269.

Fu, K.S. (1986). A Step towards unification of syntactic and statistical pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, **8**, 398–404.

Gnanadesikan, R. and Kettenring, J.R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, **28**, 81–124.

Johnson, R. A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis*, 5th

Ed. New Jersey: Prentice-Hall.

Khanna, T. (1990). *Foundation of Neural Networks.* Reading: Addison-Wesley.

Lawley, D.N. (1956). A general method for approximating to the distribution of the likelihood ration criteria. *Biometrika*, **43**, 295–305.

Mardia, .V., Kent, J. T. and Bibby, J.M. (1979). *Multivariate Analysis.* London: Academic Press.

Nelson, W., Turin, W. and Hastie, T. (1994). Statistical methods for on-line signature verification. *Journal of Pattern Recognition and Artificial Inteligence*, **8**, 749–769.

Plamondon, R. and Lorette, G. (1989). Automatic signature verification and writer identification: the state of the art. *Pattern Recognition*, **22**, 107–131.

Press, S. J. and Wilson, S. L. (1978). Choosing between regression and discriminant analysis. *Journal of the American Statistical Association*, **73**, 699–705.

Schalkoff, R. (1992). *Pattern Recognition: Statistical, Structural and Neural Approaches.* New York: John Wiley & Sons.

ÚLTIMOS RELATÓRIOS TÉCNICOS PUBLICADOS

2006-01 - SILVA, M.F., FERRARI, S.L.P., CRIBARI-NETO, F. Improved likelihood inference for the shape parameter in Weibull regression. 2006. 24p. (RT-MAE 2006-01)

2006-02 - LACHOS, V.H., BOLFARINE, H., ARELLANO-VALLE, R. B., MONTENEGRO, L. C. Inference in multivariate skew-normal regression models. 2006. 21p. (RT-MAE 2006-02)

2006-03 - BUENO, V.C. Redundancy allocation for a coherent system of dependent components. 2006. 16p. (RT-MAE 2006-03)

2006-04 - ESPINHEIRA, P.L., FERRARI, S. L. P., CRIBARI-NETO, F. On Beta regression residuals. 2006. 17p. (RT-MAE 2006-04)

2006-05 - BAZÁN, J. L., BOLFARINE, H., BRANCO, M. D. A generalized skew probit class link for binary regression. 2006. 22p. (RT-MAE 2006-05)

2006-06 - MIRANDA, J. C. S. DE; MORETTIN, P. A. On the estimation of the intensity of point processes via wavelets. 2006. 45p. (RT-MAE 2006-06)

2006-07 - DUARTE, D.; GALVES, J.; GARCIA, N. L. Markov approximation and consistent estimation of unbounded probabilistic suffix trees. 2006. 12p. (RT-MAE 2006-06)

2006-08 - RIFO, L. R.; WECHSLER, S. Regularly varying densities for St. Petersburg envelopes. 2006. 8p. (RT-MAE 2006-08)

2006-09 - FERRARI, S. L. P.; CYSNEIROS, A. H. M. A. Skovgaad's adjustment to likelihood ratio tests in exponential family nonlinear models. 2006. 15p. (RT-MAE 2006-09)

2006-10 - POLPO, A., PEREIRA, C. A. B., Parallel and series systems: rates, distributions and sub-distributions. 2006. 13p. (RT-MAE 2006-10)

2006-11 - LACHOS, V. H., BOLFARINE, H., GENTON, M.G. Bayesian inference for measurement error models with flexible skew-symmetric distributions. 2006. 29p. (RT-MAE 2006-11)

2006-12 - BUENO, V.C. A Martingale Estimator for a Structural Importance Measure of a Coherent System's Component Through its Critical Level.

2006-13 - BUENO, V.C. Estimating system´s component reliability importance under dependence conditions. 2006. 09p. (RT-MAE 2006-13)

2006-14 - PANZIERI, A., BELITSKY, V., ROCHA, P. R. M., ALMEIDA PRADO, F. P. An implicit measure of tail dependence and its application for identification of the charge of dependence between assets in market booms and crashes. 2006. 29p. (RT-MAE-2006-14)

2006-15 - ALMEIDA, A. E. ELIAN, S. N., NOBRE, S. J. Modificações e Alternativas aos Teste de Levene e de Brown e forsythe para Igualdade de variâncias e Médias. 2006. 26p. (RT-MAE-2006-15)

2006-16 - ROMEO, J.S., TANAKA, N.T., LIMA, A.C.P. Functional Central Limit Theorem for Copulas in Survival Analysis. 2006. 24p. (RT-MAE-2006-16)

200-17 - ESPINHEIRA, P.L., FERRARI, S.L.P., CRIBARI-NETO, F. Influence Diagnostics ina a Class of Econometric Regression Models for Rates and Proportions. 2006. 19p. (RT-MAE-2006-17)

2006-18 - ACHIC,B.G.B., SANDOVAL M.C., YOSHIDA, O.S. Homoscedastioc Controlled calibration model. 2006. 21p. (RT-MAE-2006-18)

The complete list of "Relatórios do Departamento de Estatística", IME-USP, will be sent upon request.

Departamento de Estatística
IME-USP
Caixa Postal 66.281
05311-970 - São Paulo, Brasil