Brazilian Journal of Probability and Statistics
2013, Vol. 27, No. 2, 210–226
DOI: 10.1214/11-BJPS168

Score-type statistics in pattern classification

Manoel R. Sena Jr.[a,1], **Abraão D. C. Nascimento**[a], **Gauss M. Cordeiro**[a] and **Lúcia P. Barroso**[b]

[a]*Federal University of Pernambuco*
[b]*University of São Paulo*

Abstract. Statistical classification methods based on score statistics have received a considerable attention in recent years. The use of these methodologies requires that asymptotic properties relative to such measures are satisfied. In this context, the classification error rates generally present biased values to the nominal level when submitted to small or moderate sample sizes. However, Nelson, Turin and Hastie [*Journal of Pattern Recognition and Artificial Intelligence* **8** (1994) 749–770] proposed a successful classification method based on score statistic described asymptotically by a chi-square law. That proposal presented good results for several sample sizes. On the other hand, stochastic measures with exact distributions described by beta and Hotelling's $\mathcal{T}^2$ laws have also been employed in such situations. This paper presents two Bartlett-type corrections for score statistics considering the method proposed by Cordeiro and Ferrari [*J. Statist. Plann. Inference* **71** (1998) 261–269]. Moreover, Monte Carlo experiments are performed in order to compare the corrected statistics to their respective noncorrected versions and to a classic classifier defined on the nonmodified score statistic. In a confirmatory sense, the proposed methodology is applied to actual signature data, obtained by the Electrical Engineering and Computer Department from the State University of Campinas (UNICAMP, Brazil).

1 Introduction

The term *biometrics* is defined as all the different forms of individual recognition, which have the work-variables based on the distinguishing characteristics. Many researchers have addressed the suggestion of novel techniques in this field (Impedovo and Pirlo, 2008). An important part of biometrics is the signature verification whose attention is focused on the finding by stochastic patterns in the statistical structure of database of signatures. One of the signature verification lines, termed by *off-line*, works with the attributes of signatures instead of using coordinates or function signal, as in the *online* field (Impedovo and Pirlo, 2008). In this context, the utilization of a statistical hypothesis test theory and its corrections has a pivotal importance (Fukunaga, 1990).

[1]First supporter of the project.
Key words and phrases. Bartlett-type correction, score statistic, pattern recognition.
Received June 2011; accepted August 2011.

Assuming a fixed nominal level, the problem of *statistical pattern recognition* aims to define a stochastic decision rule in order to quantify the probability of a new observation to belong to a particular class. Press and Wilson (1978) suggested the use of Mahalanobis distance and logistic regression as discriminant methods when the observations follow multivariate normal and non-normal distributions, respectively. Nelson, Turin and Hastie (1994) considered test statistics for classifying genuine signatures supposing normality for the vector of observations under study.

In order to obtain corrections which yield modified statistics whose first k moments are equal to those of the reference chi-squared distribution to order $O(N^{-1})$ (where N indicates the sample size), the Bartlett-type correction has been successfully used in several applications. Cordeiro and Ferrari (1998) proposed an improvement of this correction under the score statistic, aiming that such measure assumes asymptotically a chi-square distribution to N^{-1} order.

This paper proposes two Bartlett-type corrections based on the first three statistical moments: one of them is based on the distributed $\mathcal{T}^2$ statistic derived by Sena Jr. (1997), while the other correction considers the distributed beta measure proposed by Gnanadesikan and Kettenring (1972). Additionally, such measures are applied to the signature verification context, as studied by Nelson, Turin and Hastie (1994). These corrections for the score statistic take into account the methodology proposed by Cordeiro and Ferrari (1998). In order to compare the corrected statistics to their noncorrected versions and to the one classical classifier defined by the nonmodified score statistic, we consider Monte Carlo experiments on which the estimates for test size are quantified and utilized as comparison criterion. Finally, the discussed methodologies are applied to a database of signatures obtained in the Electrical Engineering and Computer Department from the State University of Campinas (UNICAMP, Brazil).

The remainder of this paper is organized as follows. Section 2 discusses the employed classification method based on the score statistic. In Section 3, such methodology is specified in four particular contexts. Section 4 presents a simulation study by means of Monte Carlo experiments. Moreover, an application to actual data is performed. Finally, the main conclusions are organized in Section 5.

2 Corrected score statistic based on the moment method

The following discussion presents two statistical methods which we use in remainder of this paper: the score statistic and the Bartlett-type correction under the moment method.

2.1 Score statistic (Mahalanobis distance)

Let $\mathbf{x}$ be a p-dimensional random vector with normal distribution whose parametric space (termed as *class* in statistical classification) is represented by $\mathbf{C}_i =$

$\{(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) : \boldsymbol{\mu}_i \in \mathbb{R}^p$ and $\boldsymbol{\Sigma}_i$ is a positive definite matrix with order $p\}$ for $i = 1, 2, \ldots, \mathcal{M}$, where $\mathcal{M}$ represents the number of classes under study. In this approach, it is additionally common to assume two conditions: (i) $\mathbf{C}_i \cap \mathbf{C}_j = \varnothing$, $i \neq j = 1, 2, \ldots, \mathcal{M}$, and (ii) $\bigcup_{k=1}^{\mathcal{M}} \mathbf{C}_k = \mathbb{R}^p \times \mathcal{D}$, where $\mathcal{D}$ is the set of positive definite matrices. The density of $\mathbf{x}$ is given by

$$f_{\mathbf{x}}(\underline{\mathbf{x}}|\mathbf{C}_i) = \frac{1}{\sqrt{(2\pi)^p|\boldsymbol{\Sigma}_i|}} \exp\left\{-\frac{1}{2}(\underline{\mathbf{x}} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\underline{\mathbf{x}} - \boldsymbol{\mu}_i)\right\}, \tag{2.1}$$

where $\underline{\mathbf{x}}$ is the outcome of the random vector $\mathbf{x}$, $(\cdot)^t$ indicates the vector transposition, $|\cdot|$ represents the determinant of a matrix, and the quantities $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the ith mean vector and covariance matrix, respectively. In statistical classification, subregions $\mathbf{R}_i \subset \mathbb{R}^p$ ($\bigcup_{i=1}^{\mathcal{M}} \mathbf{R}_i = \mathbb{R}^p$) are usually sought for allocating a new observation $\mathbf{x}_0$ to a particular class $\mathbf{C}_i$ when $\mathbf{x}_0 \in \mathbf{R}_i$; or, in otherwise, $\mathbf{x}_0$ belongs to the complement of the set $\mathbf{C}_i$. Considering the simplest case (when $\mathcal{M} = 2$), one aims to test $H_0 : \mathbf{x}_0 \in \mathbf{C}_1$ vs. $H_0 : \mathbf{x}_0 \in \mathbf{C}_2$, adopting $\mathbf{R}_1$ as the critical region on which the null hypothesis H_0 is rejected. To that end, the theoretical errors are given by

$$\alpha = \Pr(\text{Reject } H_0|H_0 \text{ is true}) = \int_{\mathbf{R}_1} f_{\mathbf{x}}(\underline{\mathbf{x}}|\mathbf{C}_1)\, \mathrm{d}\underline{\mathbf{x}}$$

and

$$\beta = \Pr(\text{Not reject } H_0|H_0 \text{ is false}) = \int_{\mathbf{R}_2} f_{\mathbf{x}}(\underline{\mathbf{x}}|\mathbf{C}_2)\, \mathrm{d}\underline{\mathbf{x}},$$

where the differential element $\mathrm{d}\underline{\mathbf{x}}$ is given by $\mathrm{d}\underline{\mathbf{x}} = \prod_{i=1}^{p} \mathrm{d}x_i$ and x_i is the ith entry of vector $\underline{\mathbf{x}}$. In practice, the parameter β is not estimated since the distribution relative to class $\mathbf{C}_2$ is not known. This paper will keep the focus of its investigation on the estimation for the test size, α.

Specifying a nominal level α and the parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$, the rejection region $\mathbf{R}_1$ is easily obtained from the equation

$$\begin{aligned} \alpha &= (2\pi)^{-p/2} \\ &\quad \times \int_{R_1} |\boldsymbol{\Sigma}_1|^{-1/2} \exp\Bigg\{-\frac{1}{2} \underbrace{(\underline{\mathbf{x}} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_1^{-1}(\underline{\mathbf{x}} - \boldsymbol{\mu}_1)}_{d^{\text{pop}}(\underline{\mathbf{x}})}\Bigg\}\, \mathrm{d}\underline{\mathbf{x}}. \end{aligned} \tag{2.2}$$

An important factor of Equation (2.2) is the term $d^{\text{pop}}(\underline{\mathbf{x}})$, known as the *squared Mahalanobis distance of vector* $\underline{\mathbf{x}}$ to the population center. This distance presents a proportional directly relationship to the volume of the $\mathbf{R}_1$ region. Thus, we have that the null hypothesis tends to be rejected for high values of this measure. Moreover, it is easy to verify that this distance follows asymptotically a chi-square distribution with p degrees of freedom (Mardia et al., 1979).

In practice, both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. Assuming that a random vector $\mathbf{x}$ follows a p-variate normal distribution, Nelson, Turin and Hastie (1994) proposed replacing these parameters by their maximum likelihood (ML) estimators. Thus, let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ be a random sample with size N under the random vector $\mathbf{x}$. According to (Mardia et al., 1979), it is known that the ML estimators of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are given by the sample mean $\overline{\mathbf{x}} = [\overline{x}_1 \ \overline{x}_2 \ \cdots \ \overline{x}_p]^t$ ($\overline{x}_i = \sum_{j=1}^{N} x_{ji}/N$, where x_{ji} is the single random variable relative to the jth sample element and to the ith element of vector $\mathbf{x}_j$) and by sample covariance matrix

$$\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^t,$$

respectively. In this context, the sample squared Mahalanobis distance between a p-dimensional new observation $\mathbf{x}_0$ and the center of a random sample $\mathbf{X}$ is represented by

$$d(\mathbf{x}_0) = (\mathbf{x}_0 - \overline{\mathbf{x}})^t \mathbf{S}^{-1}(\mathbf{x}_0 - \overline{\mathbf{x}}). \tag{2.3}$$

The statistic of Equation (2.3) presents asymptotically chi-squared distribution.

Using coefficients in a function of the factors N and p, this paper presents two statistics whose distributions are Hotelling's $\mathcal{T}^2$ and beta. Moreover, two corrections for the score statistic are proposed considering the moment method. Subsequently, this methodology is described.

2.2 The correction by moments

Let d be a test statistic following asymptotically a chi-squared distribution with r degrees of freedom. Adopting mild regularity conditions, Chandra (1985) proved that the cumulative distribution function (c.d.f.) of d under null hypothesis and to order N^{-1} is given by

$$\Pr(d \leq t) = F_p(t) + \sum_{i=0}^{k} a_i F_{q+2i}(t), \tag{2.4}$$

where k is an arbitrary positive integer (in practice, small values for k are required for correcting the first k moments of a studied statistic and its respective c.d.f.), a_i is a function defined on the unknown parameters to order N^{-1} (satisfying $\sum_{i=0}^{k} a_i = 0$), and $F_h(t)$ represents the c.d.f. of a random variable described by a chi-squared distribution with h degrees of freedom. Cordeiro and Ferrari (1998) defined the modified statistic by

$$d^* = d\left(1 - \sum_{i=1}^{k} c_i d^{i-1}\right), \tag{2.5}$$

where $c_i = 2(\sum_{\ell=i}^{k} a_\ell)/(\mu_i')$, $\mu_i' = \{\mathrm{E}(\chi_p^2)^i\} = \prod_{l=0}^{i-1}(p+2l)$, and $\mathrm{E}\{\cdot\}$ represents the statistical expectation operator. Thus, one has that

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

The correction in Equation (2.5) is known as the 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} defined at the statistic d. As it will be seen later, the development of the corrected statistic d^* is based on the first k moments.

Cordeiro and Ferrari (1998) showed that the classical Bartlett correction for the likelihood ratio test statistic is obtained using $k = 1$ and $a_0 = -a_1 = -b/2$, where b is the term of order N^{-1} in the calculation of the statistic expected value. The term b was firstly derived by Lawley (1956). With respect to the proposal of the corrected score statistic, the Bartlett-type correction is a special case of (2.5) assuming $k = 3$, that is,

$$d^* = d\{1 - (c_1 + c_2 d + c_3 d^2)\}.$$

Using the fact that the c_i's are $O(N^{-1})$, Cordeiro and Ferrari (1998) showed that

$$(d^*)^j = d^j - j d^{j-1}(c_1 d + c_2 d^2 + c_3 d^3), \tag{2.6}$$

where the terms of order less than N^{-1} are ignored. Taking the expected value of both sides of Equation (2.6) for each j, a system of linear equations is obtained. Such a system has as unknown variables the quantities c_1, c_2 and c_3, being the complementary elements defined by the first three moments of the statistic d. Thus, Cordeiro and Ferrari (1998) derived the following expression:

$$\mu'_j = m'_j - j \sum_{i=1}^{3} c_i m'_{i+j-1},$$

where μ'_j and m'_j are the moments of order j of a chi-squared distribution and of the statistic d, respectively. This equation can also be written as

$$\frac{(m'_j - \mu'_j)}{j} = \sum_{i=1}^{3} c_i m'_{i+j-1}.$$

From expansion (2.4), one has that $m'_{i+j-1} = \mu'_{i+j-1}$ to $O(1)$ and, since the c_i's are $O(N^{-1})$, it is possible to obtain that

$$\sum_{i=1}^{3} c_i \mu'_{i+j-1} = \frac{b_j}{j} \tag{2.7}$$

for $j \geq 1$, where b_j is the term of order N^{-1} in the expansion of the jth moment (m'_j) of the statistic d.

3 The statistics d and their contexts

This section considers four alternative forms for balancing the statistic d in the presence of a new observation. In this approach, two are based on modified score statistics with known exact distributions, while the others are obtained by the use of the Bartlett-type correction, being asymptotically distributed as chi-squared. In order to classify a new observation (or to identify the false alarm ratio), we consider the following methodologies. Assuming that N observations are independent and follow a normal distribution, the estimators $\overline{\mathbf{x}}$ and $\mathbf{S}$ are calculated and the Mahalanobis distance between such sample estimates and a new observation $\mathbf{x}_0$ is obtained according to Equation (2.3), $d(\mathbf{x}_0)$. Additionally, we also adopt the following method: (i) include the new observation in the sample, (ii) obtain the estimators, and (iii) calculate the distance d. In the last procedure, the influence of a new observation under the estimators is captured. Finally, we obtain two closed-form expressions for Bartlett-type corrections based on two distributed $\mathcal{T}^2$ and beta modified score statistics.

3.1 Inference concerning the $(N+1)$th observation: $d^{\mathcal{T}}$

Let $\overline{\mathbf{x}}$ and $\mathbf{S}$ be the ML estimators of the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ obtained from a random sample with size N from a p-variate normal law equipped with such parameters. Suppose also that we have a new observation of the same population. Thus, we aim to make statistical inference for classifying such observation. The following results represent the relationships among the Wishart, Hotelling's $\mathcal{T}^2$ and Snedecor's $\mathcal{F}$ distributions (Mardia et al., 1979).

Let $\mathbf{y}$ be a random vector with distribution $\mathcal{N}_p(\mathbf{0}, \mathbf{I})$ and $\mathbf{W}$ be an independent random matrix of $\mathbf{y}$ following a Wishart law with the parameters $\mathbf{I}$ and L, termed by $\mathbf{W} \sim \mathcal{W}_p(\mathbf{I}, L)$. Here, p is the dimension of the vector $\mathbf{y}$, $\mathbf{I}$ is the identity matrix of order p, and L is the degrees of freedom. Considering the expression

$$T = L\mathbf{y}^t\mathbf{W}^{-1}\mathbf{y},$$

Sena Jr. (1997) showed that T has a Hotelling's $\mathcal{T}^2$ distribution with parameters p and L, termed by $T \sim \mathcal{T}^2(p, L)$. By the definition of the above concepts, the following theorems are deduced in Mardia et al. (1979).

Theorem 3.1. *Let* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ *be a set of random vectors based on* $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *distribution and* $\mathbf{S} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^t$, *then*

$$(N-1)\mathbf{S} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, N-1).$$

Theorem 3.2. *If* $\mathbf{W} \sim \mathcal{W}_p(\boldsymbol{\Sigma}, L)$, *then*

$$\boldsymbol{\Sigma}^{-1/2}\mathbf{W}\boldsymbol{\Sigma}^{-1/2} \sim \mathcal{W}_p(\mathbf{I}, L).$$

Theorem 3.3. *If* $T \sim \mathcal{T}^2(p, L)$, *then*

$$\left(\frac{T}{L}\right)\left(\frac{L-p+1}{p}\right) \sim \mathcal{F}(p, L-p+1).$$

Considering $\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then

$$\mathbf{y} = \sqrt{\frac{N}{N+1}}\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \overline{\mathbf{x}}) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}).$$

Directly from Theorem 3.1, minor manipulations lead to the following result:

$$\mathbf{W} = (N-1)\boldsymbol{\Sigma}^{-1/2}\mathbf{S}\boldsymbol{\Sigma}^{-1/2} \sim \mathcal{W}_p(\mathbf{I}, N-1).$$

From the last two results, it is possible to show that

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

Writing d as a function of T and using the relation between the $\mathcal{F}$ and $\mathcal{T}^2$ distributions, one has then that

$$d = \frac{N+1}{N}T = \frac{N+1}{N}\frac{(N-1)p}{(N-p)}F,$$

where F is a random variable with $\mathcal{F}$ law whose degrees of freedom are p and $(N-p)$, nominated by $F \sim \mathcal{F}(p, N-p)$.

From the last equation, it is possible to modify the statistic d and so establish a critical point, such that one can assure that at most $\alpha\%$ of the observations do not belong to a prefixed confidence region. Writing the statistic F in function of d, Sena Jr. (1997) obtained an exact distribution for the distance measure given by

$$F = \frac{N-p}{p(N-1)}\frac{N}{N+1}d, \tag{3.1}$$

where $F \sim \mathcal{F}(p, N-p)$. In the remainder of this article, such measure will be called $d^{\mathcal{T}}$.

3.2 Inference under the Nth observation: d^{B}

Another methodology consists in calculating the N distances between each vector $\mathbf{x}_i$ $(i = 1, 2, \ldots, N)$ and the estimator $\overline{\mathbf{x}}$. Assuming that $\mathbf{x}_i$ follows the $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ law, Gnanadesikan and Kettenring (1972) studied the robustness of statistic u_i expressed by

$$u_i = \frac{N}{(N-1)^2}d_i. \tag{3.2}$$

Such a measure follows the beta distribution with parameters $p/2$ and $(N - p - 1)/2$, where d_i is the sample Mahalanobis distance. According to the above discussion, the following rule is defined: after a new observation $\mathbf{x}_0$ is introduced in the sample, the estimates for $\bar{\mathbf{x}}$ and $\mathbf{S}$ are obtained and the distance between their values and $\mathbf{x}_0$ is then quantified obeying Equation (2.3). In subsequent discussions, the statistic u_i will be represented by d^{B}.

3.3 The correction via the moment method based on beta distribution: d_1^*

Rewriting the Expression (3.2), the following relation of jth moments is obtained by

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

where $u \sim \mathrm{beta}(p/2, (N - p - 1)/2)$. According to Mardia et al. (1979), it is known that the jth moment of the statistic u is given by

$$\mathrm{E}\{u^j\} = \mu_u{}^j = \frac{B(j+a, b)}{B(a, b)},$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ and $\Gamma(\cdot)$ indicates the gamma function. Using the fact that $\Gamma(a) = (a - 1)\Gamma(a - 1)$, the three first moments of the statistic d are expressed by

$$\begin{aligned}
m_1'(d) &= \mathrm{E}\{d\} = \frac{(N-1)^2}{N}\mathrm{E}\{u\} = \frac{N-1}{N}p = p\left(1 - \frac{1}{N}\right),\\
m_2'(d) &= \mathrm{E}\{d^2\} = \left(\frac{(N-1)^2}{N}\right)^2 \mathrm{E}\{u^2\} = p(p+2)\frac{(N-1)^3}{N^2(N+1)}\\
&= p(p+2)\left(1 - \frac{4}{N} + O(N^{-2})\right),\\
m_3'(d) &= \mathrm{E}\{d^3\} = \left(\frac{(N-1)^2}{N}\right)^3 \mathrm{E}\{u^3\} = p(p+2)(p+4)\frac{(N-1)^5}{N^3(N+3)(N+1)}\\
&= p(p+2)(p+4)\left(1 - \frac{9}{N} + O(N^{-2})\right).
\end{aligned}$$

Considering the system formulated in Equation (2.7), the first five moments of the chi-squared distribution are given by $\mu_1' = p$, $\mu_2' = p(p + 2), \ldots,$ and $\mu_5' = p(p + 2)(p + 4)(p + 6)(p + 8)$. Thus, the first three terms b_j's of the expansion for the jth null moment m_j' of the statistic d are given by $b_1 = -p/N$, $b_2 = -4p(p + 2)/N$, and $b_3 = -9p(p + 2)(p + 4)/N$. Therefore, Equation (2.7) is degenerated in the following system of equations:

$$\begin{cases} c_1\mu_1' + c_2\mu_2' + c_3\mu_3' = b_1, \\ c_1\mu_2' + c_2\mu_3' + c_3\mu_4' = \dfrac{b_2}{2}, \\ c_1\mu_3' + c_2\mu_4' + c_3\mu_5' = \dfrac{b_3}{3}. \end{cases} \tag{3.3}$$

From this system, we derive the quantities $c_1 = p/2N$, $c_2 = -1/2N$ and $c_3 = 0$. Based on the terms c_1, c_2 and c_3, we have that the Bartlett-type correction for the statistic u is given by

$$d_1^* = d(1 - (c_1 + c_2 d + c_3 d^2)) = d\left\{1 + \left(\frac{d-p}{2N}\right)\right\},$$

where $\Pr(d \le t) = \Pr(\chi_p^2 \le t) + O(1)$ and $\Pr(d_1^* \le t) = \Pr({\chi_p}^2 \le t) + O(N^{-1})$.

3.4 The correction via the moment method based on $\mathcal{F}$ distribution: d_2^*

From Equation (3.1), one has that

$$\mathrm{E}\{d^j\} = \left(\frac{p(N^2-1)^2}{N(N-p)}\right)^j \mathrm{E}\{F^j\},$$

where $F \sim \mathcal{F}(p, N-p)$. The jth moment of F is expressed by Mardia et al. (1979),

$$\mathrm{E}\{F^j\} = {\mu_F}^j = \left(\frac{N-p}{p}\right)^j \Gamma\left(\frac{p}{2} + j\right)\Gamma\left(\frac{N-p}{2} - j\right)\Big/\left(\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{N-p}{2}\right)\right),$$

where $N > 2j + p$. Through minor algebraic manipulations, one can obtain that

$$\begin{aligned} m_1'(d) &= \mathrm{E}\{d\} = \frac{p(N^2-1)^2}{N(N-p)}\mathrm{E}\{F\} = \frac{N^2-1}{N^2 - N(p+2)}p \\ &= p\left(1 + \frac{p+2}{N} + O(N^{-2})\right), \\ m_2'(d) &= \mathrm{E}\{d^2\} = \left(\frac{p(N^2-1)^2}{N(N-p)}\right)^2 \mathrm{E}\{F^2\} \\ &= p(p+2)\frac{N^4 - 2N^2 + 1}{N^4 - 2N^3(p+3) + N^2(p+2)(p+4)} \\ &= p(p+2)\left(1 + \frac{2(p+3)}{N} + O(N^{-2})\right), \\ m_3'(d) &= \mathrm{E}\{d^3\} = \left(\frac{p(N^2-1)^2}{N(N-p)}\right)^3 \mathrm{E}\{F^3\} \end{aligned}$$

$$= p(p+2)(p+4)\frac{N^6 - 3N^4 + N^2 + 1}{N^6 - A_1(p)N^5 + A_2(p)N^4 - A_3(p)N^3}$$
$$= p(p+2)(p+4)\left(1 + \frac{A_1(p)}{N} + O(N^{-2})\right),$$

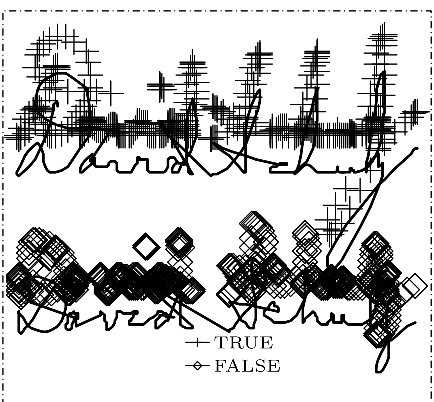
where $A_1(p) = 3p + 12$, $A_2(p) = 3p^2 + 24p + 44$ and $A_3(p) = p^3 + 12p^2 + 44p + 48$.

Therefore, the first three terms b_j's in the expansion of the jth null moment m'_j concerning to the statistic d is expressed by $b_1 = p(p+2)/N$, $b_2 = 2p(p+2)(p+3)/N$ and $b_3 = p(p+2)(p+4)(3p+12)/N$. Including theses expressions for b_i in the system (3.3), the quantities $c_1 = -(2+p)/2N$, $c_2 = 1/2N$ and $c_3 = 0$ are derived. Thus, the Bartlett-type correction for the statistic F is expressed by

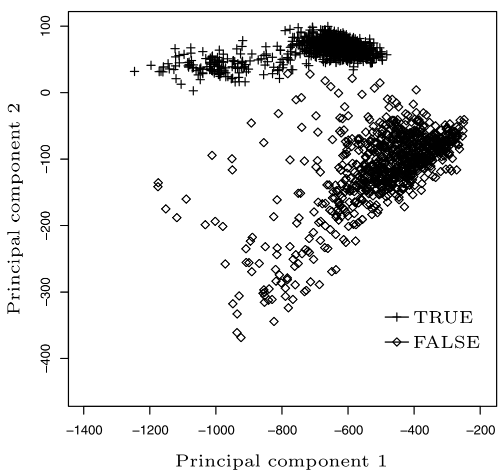
$$d_2^* = d\left\{1 - \left(\frac{p+d+2}{2N}\right)\right\} = d_1^* - d\left(\frac{1+d}{N}\right).$$

4 Applications and simulations

The five discussed score statistics are included in the signature verification context and their performances are investigated. To this end, we use a database of signatures provided by the Electrical Engineering and Computer Department from the State University of Campinas (UNICAMP, Brazil) relative to the author David Schulz. This data set consists of 1000 genuine and 825 false signatures. Figure 1 presents a plot of two signatures and a dispersion graphic between two principal



(a) Real signature



(b) Statistical properties

Figure 1 *Signatures and dispersion graphic between two principal components under the larger eigenvalues.*

components indexed by the two larger eigenvalues based on the correlation matrix of the database. This figure introduces the subsequent analysis based on both simulated and real data with an illustration of the current database. Additionally, Figure 1(b) shows evidence that characteristics relative to the signatures are able for classifying new signatures.

In this section, the empirical study aims to construct a structure with similar characteristics to actual signatures and to quantify the convergence rate relative to the estimated test size under different scenarios. Additionally, such estimates are obtained and assessed from real signatures.

4.1 Synthetic signature

In order to incorporate the signature database characteristics to this simulation study, 100 genuine signatures with 42 basic features of an author were sampled among 1000, using a simple random sample without replacement. Based on such a sample, the pair of sample average vectors and covariance matrices, termed by $(\overline{\mathbf{x}}, \mathbf{S})$, was obtained. Subsequently, we generate vectors according to the multivariate normal and t_3 distributions for different sample sizes (N) and characteristic numbers (p). Adopting as nominal level $\alpha \in \{1\%, 5\%, 10\%\}$, this methodology addresses the analysis of non-normality impact under sizes of tests based on the modified score statistics: (i) the Mahalanobis distance (Mardia et al., 1979), (ii) their Barllet-type corrections (Cordeiro and Ferrari, 1998), (iii) the classical Hotelling's $\mathcal{T}^2$ (Hotelling, 1931), and (iv) one proposed by Gnanadesikan and Kettenring (1972). For notation purposes, the remainder of the paper will refer to such statistics by terms d, d_i^* $(i = 1, 2)$, $d^{\mathcal{T}}$ and d^{B}, respectively.

This simulation considered $N > p$ such that $p \in \{2, 4\}$ and $N \in \{p + 10, p + 11, \dots, 100\}$. The selection process of p features among the 42 available in the signature database was based on the coefficient of variation (CV). Once we specified the value of p, we select the p variables with CV lower values because these situations are more difficult to be imitated in practice.

Establishing the pair value, (N, p), we generate N outcomes of the p-dimensional vector according to the multivariate normal [with density given by Equation (2.1)] or t_3 distribution, whose density function is expressed by McNeil (2006),

$$f_{\mathbf{x}}(\underline{\mathbf{x}}|v, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\Gamma((v+p)/2)}{\Gamma(v/2)(v\pi)^{p/2}} |\boldsymbol{\Sigma}|^{-1/2} \times \left[1 + \frac{1}{v}(\underline{\mathbf{x}} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\underline{\mathbf{x}} - \boldsymbol{\mu})\right]^{-(v+p)/2}, \tag{4.1}$$

where $v > 0$ is the degrees of freedom. In this case, the distribution is said to be 'central' when $\boldsymbol{\mu} = \mathbf{0}$. This simulation considered its noncentral version, replacing the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by outcomes $\overline{\mathbf{x}}$ and $\mathbf{S}$ obtained by signature actual data.

The above methodology is nominated as a 'training step' and it was used to obtain $\boldsymbol{\mu}_1$ and $\boldsymbol{\Sigma}_1$, which are used in the estimation of the test size relative to the discussed hypothesis tests. Subsequently, we generate other 1000 observations, termed a 'test step,' in a similar way to that used to generate samples in the training step. Based on the generated observation in the test step, we quantify the test size, which is defined by

$$
\begin{aligned}
\alpha_d &= \Pr\big(d(\mathbf{x}) > \chi^2_{(p,\alpha)}\big),\\
\alpha_{d_1^*} &= \Pr\big(d_1^*(\mathbf{x}) > \chi^2_{(p,\alpha)}\big),\\
\alpha_{d_2^*} &= \Pr\big(d_2^*(\mathbf{x}) > \chi^2_{(p,\alpha)}\big),\\
\alpha_{d^{\mathcal{T}}} &= \Pr\big(d^{\mathcal{T}}(\mathbf{x}) > F(p, n-p, \alpha)\big),\\
\alpha_{d^{\mathrm{B}}} &= \Pr\left(d^{\mathrm{B}}(\mathbf{x}) > \mathrm{beta}\left(\frac{p}{2}, \frac{n-p-1}{2}, \alpha\right)\right),
\end{aligned}
$$

where $\chi^2_{(p,\alpha)}$, $\mathrm{beta}(\frac{p}{2}, \frac{n-p-1}{2}, \alpha)$ and $\mathcal{F}(p, n-p, \alpha)$ are critical values such that the population elements exceed at $\alpha\%$.

This procedure was performed using 1000 Monte Carlo replications, where the estimates for the theoretical errors α_d, $\alpha_{d_1^*}$, $\alpha_{d_2^*}$, $\alpha_{d^{\mathcal{T}}}$ and $\alpha_{d^{\mathrm{B}}}$ are recorded for each replication. The results for the mean of the estimates for test size are presented in Tables 1 and 2. The estimates closest to the nominal levels are highlighted in boldface type in the tables.

Table 1 presents the estimates for test size in multivariate normal data. The hypothesis tests based on the statistics $d^{\mathcal{T}}$ and d_2^* revealed the better performance.

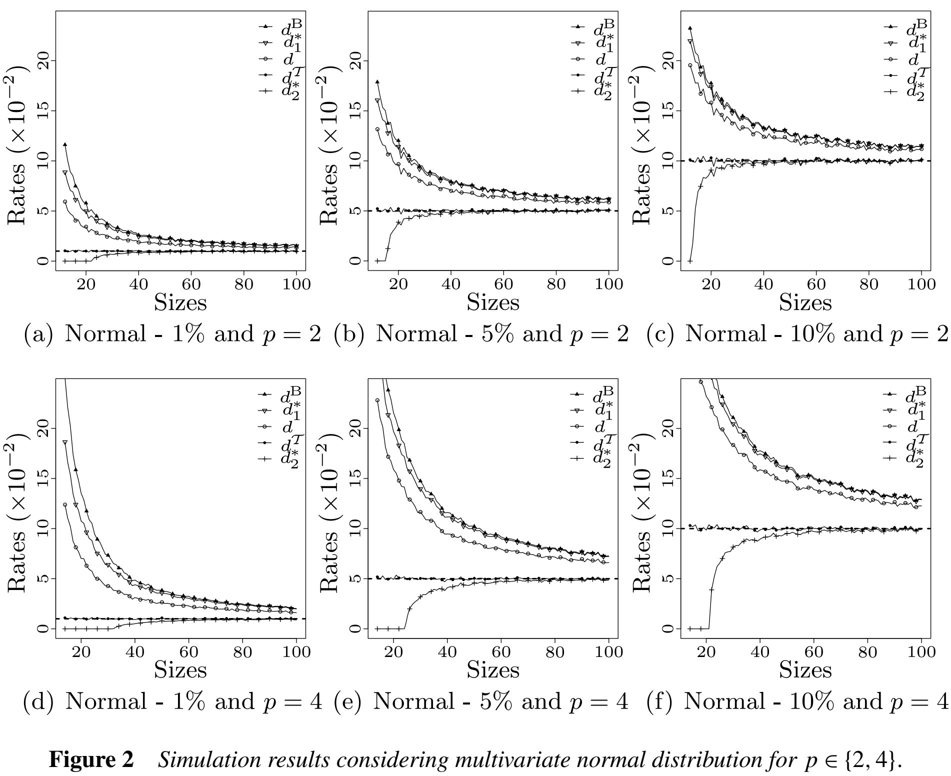
Table 1 *Average percentage of observations that exceed the specified critical values, assuming multivariate normal distribution*

α	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_d$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$
			$p=2$, $N=20$					$p=4$, $N=20$		
1	5.41	**0.96**	3.21	1.17	4.70	13.11	**0.96**	6.87	0.00	10.52
5	11.58	**4.87**	9.28	6.95	11.02	21.30	**4.88**	15.46	9.11	19.29
10	17.32	**9.89**	15.27	13.33	16.89	27.47	**9.67**	22.88	17.72	26.11
			$p=2$, $N=30$					$p=4$, $N=30$		
1	3.58	**1.04**	2.41	1.40	3.28	7.30	**1.05**	4.31	1.71	6.34
5	9.43	**5.13**	8.06	6.72	9.19	14.94	**5.05**	11.74	8.51	14.14
10	15.06	**10.32**	13.89	12.65	14.91	21.26	**10.10**	18.47	15.77	20.65
			$p=2$, $N=50$					$p=4$, $N=50$		
1	2.32	**1.03**	1.77	1.29	2.24	3.84	**1.01**	2.62	1.58	3.57
5	7.27	**4.98**	6.59	5.86	7.18	10.12	**4.97**	8.53	7.04	9.87
10	12.59	**9.83**	11.95	11.31	12.53	16.05	**9.97**	14.57	13.09	15.86

Table 2 *Average percentage of observations that exceed the specified critical values, assuming multivariate t_3 distribution*

α	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_{d}$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_{d}$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$
			$p=2, N=20$					$p=4, N=20$		
1	9.19	**4.56**	7.21	3.34	8.60	18.42	6.87	13.85	**0.00**	16.63
5	13.36	**8.77**	11.90	9.05	13.02	23.35	12.10	19.90	**11.11**	22.21
10	16.83	**12.29**	15.63	13.29	16.57	27.06	**16.02**	24.36	17.25	26.28
			$p=2, N=30$					$p=4, N=30$		
1	7.15	**4.48**	6.10	4.39	6.89	13.30	6.87	10.88	**5.94**	12.59
5	10.96	**8.26**	10.12	8.77	10.81	17.95	**11.54**	16.12	12.37	17.53
10	14.32	**11.48**	13.59	12.44	14.22	21.38	**15.14**	19.94	16.72	21.06
			$p=2, N=50$					$p=4, N=50$		
1	5.70	**4.34**	5.16	4.46	5.61	9.55	**6.51**	8.51	6.72	9.36
5	9.25	**7.77**	8.81	8.17	9.21	13.75	**10.46**	12.86	11.34	13.60
10	12.32	**10.79**	11.93	11.39	12.28	17.04	**13.65**	16.28	14.94	16.94

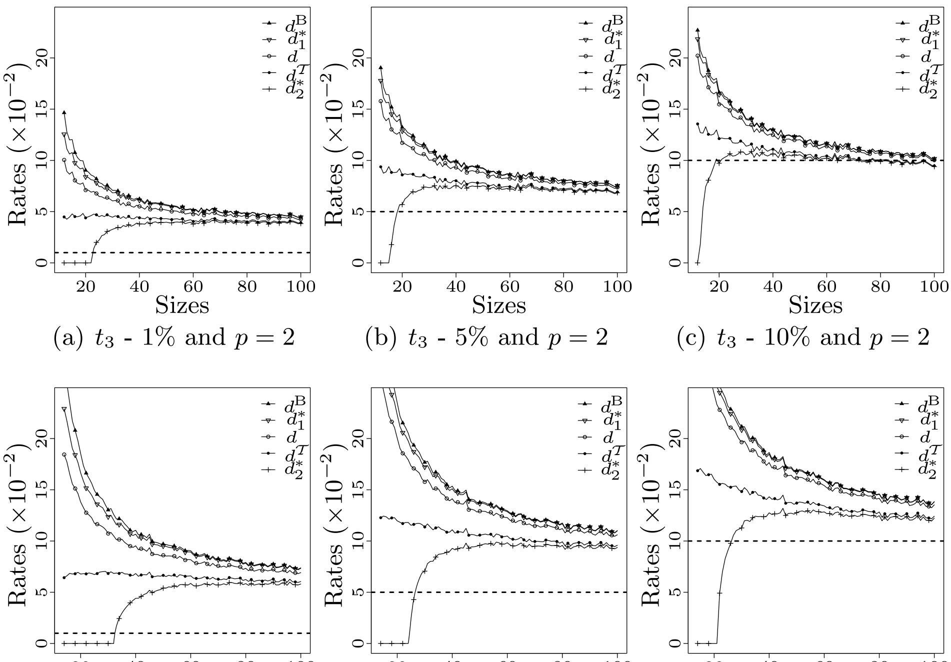
The study for more situations is presented in Figure 2 and indicated that such measures have faster convergence rates with respect to the nominal level.



(a) Normal - 1% and $p=2$ (b) Normal - 5% and $p=2$ (c) Normal - 10% and $p=2$

(d) Normal - 1% and $p=4$ (e) Normal - 5% and $p=4$ (f) Normal - 10% and $p=4$

Figure 2 *Simulation results considering multivariate normal distribution for $p \in \{2, 4\}$.*



(a) t_3 - 1% and $p = 2$ (b) t_3 - 5% and $p = 2$ (c) t_3 - 10% and $p = 2$

(d) t_3 - 1% and $p = 4$ (e) t_3 - 5% and $p = 4$ (f) t_3 - 10% and $p = 4$

Figure 3 *Simulation results considering multivariate t_3 distribution for $p \in \{2, 4\}$.*

Table 2 shows the empirical test sizes submitted to non-normal data. In general terms, the estimates were biased with respect to the adopted nominal levels. The better results were presented by hypothesis tests based on the $d^{\mathcal{T}}$ statistic. The increasing of sample size yielded the estimate values more accurate. Such situations and other cases are illustrated in Figure 3. These plots furnished evidence that the situations less affected by non-normality were ones with $p = 2$ and $\alpha = 10\%$.

4.2 Real signature

This section assesses the performance of the discussed hypothesis tests, adopting as criterion the empirical test sizes. To that end, the following methodology was considered:

1. Specify the values for quantities N and p;
2. Select 1000 sample with size N from the database (nominated by A_i for $i = 1, 2, \ldots, 1000$), using a simple random sample without replacement;
3. For each selected sample A_i, obtain the more significant principal component p whose coefficients are represented by the columns of the matrix C_i. The technique of the principal components was used in order to consider all variables of the database;

Table 3 *Average percentage of observations that exceed the critical values specified, assuming real data*

α	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_{d}$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$	$\widehat{\alpha}_{d^{\mathrm{B}}}$	$\widehat{\alpha}_{d^{\mathcal{T}}}$	$\widehat{\alpha}_{d}$	$\widehat{\alpha}_{d_2^*}$	$\widehat{\alpha}_{d_1^*}$
			$p=2$, $N=20$					$p=4$, $N=20$		
1	7.44	**3.32**	5.72	2.71	6.89	10.71	**2.95**	7.24	0.00	9.32
5	11.16	**7.04**	9.86	7.73	10.84	14.89	**6.02**	11.93	6.79	13.87
10	14.13	**10.19**	13.11	11.42	13.91	18.28	**8.85**	15.75	11.44	17.48
			$p=2$, $N=30$					$p=4$, $N=30$		
1	6.31	**3.47**	5.19	3.85	6.06	8.18	**3.20**	6.16	3.34	7.58
5	10.30	**7.52**	9.45	8.47	10.15	12.38	**6.70**	10.66	8.34	11.97
10	13.36	**10.78**	12.72	11.98	13.26	15.78	**9.79**	14.31	12.31	15.46
			$p=2$, $N=50$					$p=4$, $N=50$		
1	5.21	**3.35**	4.52	3.79	5.11	6.21	**3.22**	5.09	3.89	5.99
5	9.61	**7.90**	9.11	8.59	9.56	10.64	**7.17**	9.65	8.58	10.49
10	12.82	**11.28**	12.45	12.08	12.78	14.08	**10.55**	13.25	12.39	13.97

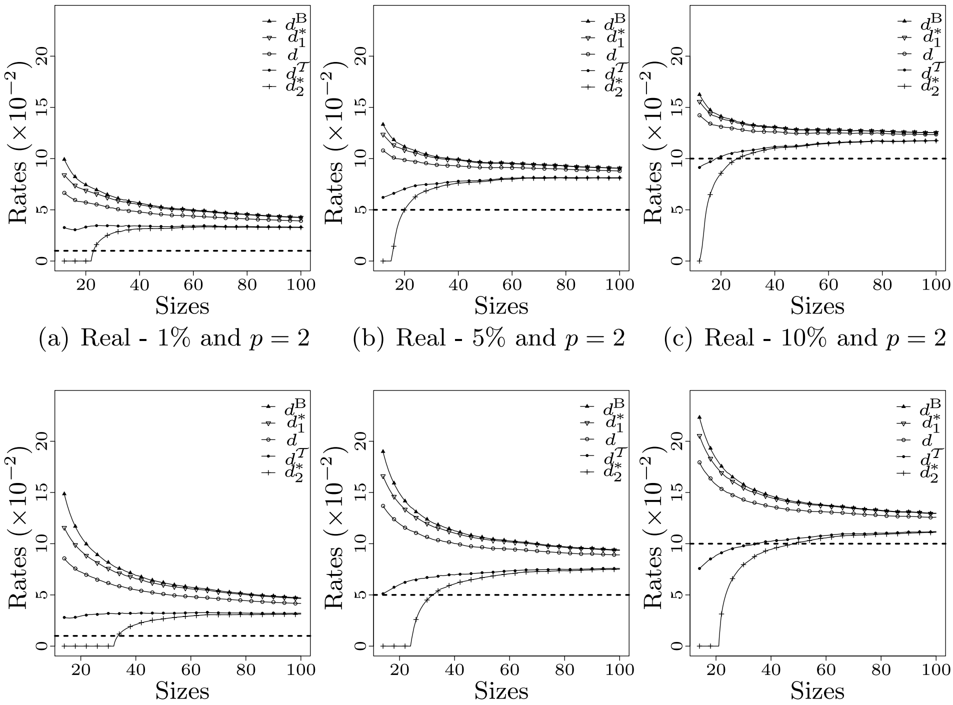
4. In the training step, consider the transformation $A_i^* = A_i C_i$ for $i = 1, 2, \ldots,$ 1000 and calculate $\widehat{\boldsymbol{\mu}} = \overline{A_i^*}$ and $\widehat{\boldsymbol{\Sigma}} = \mathrm{Cov}(A_i^*)$;
5. As a test step, take B_i as complementary of A_i with respect to the database observations and consider $B_i^* = B_i C_i$. Finally, calculate the means of the estimates (in each one selected sample) for the sizes of tests based on distance between the elements of B_i^* and the pair $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$.

The results of the application of the above methodology to the actual data are organized in Table 3 and Figure 4. Table 3 shows the performance concerning any particular situations. Such results furnished evidence that the statistics $d^{\mathcal{T}}$ and d_2^* presented the better rates of false alarm, that is, classify as false a genuine signature. Figure 2 reveals the extension of these results, considering a higher number of sample sizes. In this case, one can note that the nominal levels are overestimated. This fact is justified by the simulation under non-normality situation which presented the same behavior. The empirical test size relative to the measures $d^{\mathcal{T}}$ and d_2^* presented faster convergence degrees to the adopted significance levels.

5 Conclusions

This paper proposed two test statistics based on modified score statistics. Two among them are based on scaling of such measures, while the others were obtained under Bartlett-type corrections. Such methodologies were applied to the classification context.

In order to assess the performance of discussed measures, we employed a Monte Carlo experiment. The parameters of this simulation were obtained from an actual database of signatures. Additionally, such study utilized the estimated test size as



(a) Real - 1% and $p = 2$ (b) Real - 5% and $p = 2$ (c) Real - 10% and $p = 2$

(d) Real - 1% and $p = 4$ (e) Real - 5% and $p = 4$ (f) Real - 10% and $p = 4$

Figure 4 *Results based on real signatures for* $p \in \{2, 4\}$.

comparison criterion. The results showed evidence that the hypothesis tests defined by $d^{\mathcal{T}}$ and d_2^* presented the better estimates for test size. Finally, the proposed methodology was extended to the actual data.

References

Chandra, T. K. (1985). Asymptotic expansions of perturbed chi-square variables. *Sankhyã: The Indian Journal of Statistics, Ser. A* **47**(1), 100–110. MR0813448

Cordeiro, G. M. and Ferrari, S. L. P. (1998). A note on Bartlett-type correction for the first few moments of test statistics. *Journal of Statistical Planning and Inference* **71**(1-2), 261–269. MR1651828

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, 2nd ed. *Computer Science and Scientific Computing*. San Diego: Academic. MR1075415

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

Hotelling, H. (1931). The generalization of Student's ratio. *Annals of Mathematical Statistics* **2**, 360–378.

Impedovo, D. and Pirlo, G. (2008). Automatic signature verification: The state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C—Applications and Reviews* **38**(5), 609–635.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**(3/4), 295–303. MR0082237

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press. MR0560319

McNeil, A. J. M. (2006). Review of *Multivariate t Distributions and Their Applications*, by Samuel Kotz and Saralees Nadarajah. *Journal of the American Statistical Association* **101**, 390–391.

Nelson, W., Turin, W. and Hastie, T. (1994). Statistical methods for on-line signature verification. *Journal of Pattern Recognition and Artificial Inteligence* **8**(3), 749–770.

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

Sena Jr., M. R. (1997). Aplicações Estatísticas em Reconhecimento de Padrões em Ênfase em Verificação de Assinaturas. Ph.D. thesis, State Univ. Campinas.

M. R. Sena Jr.
A. D. C. Nascimento
G. M. Cordeiro
Departamento de Estatística
Universidade Federal de Pernambuco
50740-540, Recife, PE
Brasil
E-mail: manoel@de.ufpe.br
abraao.susej@gmail.com
gauss@de.ufpe.br

L. P. Barroso
Departamento de Estatística
Universidade de São Paulo
05508-090, Rua do Matão, 1010-SP
Brasil
E-mail: lbarroso@ime.usp.br