DATA DRIVEN SMOOTH TESTS FOR BIVARIATE NORMALITY *

MAŁGORZATA BOGDAN

Technical University of Wrocław

Summary. Basing upon the idea of construction of data driven smooth tests for composite hypotheses presented in Inglot, Kallenberg and Ledwina (1995, 1996) and Kallenberg and Ledwina (1995c) two versions of data driven smooth test for bivariate normality are proposed. Asymptotic null distributions are derived and consistency of the newly introduced tests against every bivariate alternative with marginals having finite variances is proved. Included results of power simulations show that one of the proposed tests performs very well in comparison with other commonly used tests for bivariate normality.

AMS 1991 subject classifications: 62 G 10, 62 G 20, 62 E 25.

Key words: Schwarz's BIC criterion, tests of bivariate normality, goodness-of-fit, score test, smooth test, Neyman's test, Monte Carlo simulations.

*Research supported by Grant KBN 350 044

1. INTRODUCTION

The problem of testing multivariate normality has attracted a lot of attention in past and recent years. The possible cause of such sustained interest is that many multivariate data analysis methods "rest to some extend on multivariate normality. Most of the distribution theory and optimality of standard test procedures derive directly from this assumption" [cf. Cox and Wermuth (1994)]. Moreover, as noted by Cox and Wermuth (1994) , the exponential family structure of the multivariate normal distribution provides a strong justification for the methods of data reduction, which "hinge on the calculation of sample mean vectors and covariance matrices or 'robust' versions of these quantities". The compendium of information on tests of multivariate normality can be found in Gnanadesikan (1977, pages 161-195), Cox and Small (1978), Mardia (1980), Koziol (1986) and D'Agostino (1986). However, as noted by Koziol (1983), despite of the great amount of articles concerned with this problem, there are relatively few formal methods available for assessing multivariate normality. E.g. asymptotic null distributions and consistency are rarely established. Moreover, "very little has been done by way of power studies for multivariate normality tests" [cf. D'Agostino (1986)].

On the other hand in recent years a renewed interest in smooth tests of goodness-of-fit has been observed. These tests, based on the score statistics in specified exponential families of densities, can be applied to testing both simple and composite hypotheses. For details see Neyman (1937), Javitz (1975), Thomas and Pierce (1979) and Rayner and Best (1989) e.g.

In particular, for testing multivariate normality Koziol (1987) proposed some analogues of Neyman's smooth test. Koziol's statistics are related to the system of Hermite polynomials. More specifically, the overall smooth skewness statistic $\hat{U}_3^2$ and the overall smooth kurtosis statistic $\hat{U}_4^2$ are based on the products of Hermite polynomials of degrees summing up to three and four , respectively. Koziol (1987) observed that $\hat{U}_3^2$ is algebraically equivalent to $nb_{1,p}/6$, where $b_{1,p}$ is the well known Mardia's (1970, 1974) measure of multivariate skewness.

Test statistics $\hat{U}_3^2$ and $\hat{U}_4^2$ can be decomposed into orthogonal components, each depending on one particular function from the set of the products of Hermite polynomials. Koziol (1987) emphasized that examination of the individual components of these statistics "is a valuable adjunct to the overall assessments of skewness and kurtosis afforded by the tests, and should be undertaken whenever further information concerning coordinate-wise departures from normality is desired. Indeed, Small (1980) combines certain subsets of the components of $\hat{U}_3^2$ and $\hat{U}_4^2$ for testing multivariate normality."

Note that Koziol's smooth tests, depending only on sample moments up to fourth order, are not consistent for a wide range of alternatives. Moreover, the results of power simulations presented in Rayner and Best (1989) show that in the lack of knowledge on the class of possible alternatives to multivariate normality it is difficult to decide which components of $\hat{U}_3^2$ and $\hat{U}_4^2$ should be included in the test statistic. It can be observed that by making a wrong decision on this matter we can lose much of the test power.

The problem of the choice of the number of components k in Neyman's smooth test of goodness of fit appears also in the simpler case of testing uniformity [see Inglot, Kallenberg and Ledwina (1994) and Kallenberg and Ledwina (1995a), e.g.]. To deal with it some data driven procedures for choosing k in the smooth test for uniformity has been recently proposed by Bickel and Ritov (1992), Eubank and LaRiccia (1992), Eubank, Hart and

LaRiccia (1993), Ledwina (1994) and Fan (1995).

In the data driven smooth test for uniformity proposed by Ledwina (1994) the number of components in the test statistic is determined by an application of Schwarz's (1978) selection rule. Theoretical results concerning this test proved in Kallenberg and Ledwina (1995a, b) yield its consistency against any alternative to uniformity. Moreover extensive simulations presented in Ledwina (1994) and Kallenberg and Ledwina (1995a) show that the data driven Neyman's test exploiting Schwarz's selection rule compares very well to classical tests and other competitors.

The idea of using Schwarz's selection rule for choosing the number of components in the smooth test statistic was further extended and a very general method of construction of data driven smooth tests for composite hypotheses was proposed and investigated in Inglot et al. (1995, 1996) and Kallenberg and Ledwina (1995c). Theoretical results included in these papers imply the consistency of the data driven smooth tests for composite hypotheses against a broad range of alternatives. Moreover from the simulation study reported in Kallenberg and Ledwina (1995c, d) it follows that these tests work very well in comparison with the well known 'special' tests, as Gini's test for exponentiality and Shapiro-Wilk's test for normality.

The papers of Inglot et al. (1995, 1996) and Kallenberg and Ledwina (1995c) concern a problem of testing composite hypotheses for random variables. In this paper we extend the method of construction of data driven smooth tests on a case of testing bivariate normality. We propose two versions of a data driven smooth test for bivariate normality. These tests are based on the statistics $W_{S(1)}$ and $W_{S(5)}$, which are related to the system of Legendre polynomials. We derive asymptotic null distributions of test statistics and prove consistency of the tests against all alternatives with marginals having finite variances. Main ideas of our proofs are similar to those proposed in Inglot et al. (1995, 1996) and Kallenberg, Ledwina and Rafajłowicz (1996).

The results of the simulation study reported in this paper show that the data driven smooth test based on the statistic $W_{S(5)}$ performs very well in comparison with the well known tests for bivariate normality like the generalization of Shapiro-Wilk's test proposed by Malkovich and Afifi (1973) and tests based on Mardia's (1970, 1974) measures of skewness and kurtosis. Also, the presented comparison of the empirical powers of $W_{S(5)}$ with the simulated powers of some new tests for bivariate normality like Csörgő's (1986) test based on the sample characteristic function, an analogue of smooth test for bivariate normality introduced by Koziol (1987) or Bowman and Foster (1993) test based on the kernel density estimators, shows that the test based on $W_{S(5)}$ works very well and can be recommended as an omnibus test for bivariate normality.

Though this paper is concerned solely with testing bivariate normality the method of construction of data driven smooth tests as well as the methods of proofs of their properties can be easily transferred on a case of testing normality in arbitrary dimension.

2. PRELIMINARIES

We want to test the null hypothesis $\mathcal{H}_0$ that i.i.d. two-dimensional random vectors $X_1, \ldots, X_n$ come from a nondegenerate bivariate normal distribution.

Let X_{1i} and X_{2i} denote the components of the random vector X_i; $X_i = (X_{1i}, X_{2i})^T$.

The random vector X_i comes from a nondegenerate bivariate normal distribution if it possesses a density $f(x)$, $x = (x_1, x_2)^T \in \mathbf{R}^2$, such that

$$f(x) \in \{f(x,\beta): \beta \in \Omega\}$$

where

$$\Omega = \{\beta = (\beta_1, \dots, \beta_5)^T : (\beta_1, \beta_2, \beta_5) \in \mathbf{R}^3, (\beta_3, \beta_4) \in \mathbf{R}_+^2, \beta_3\beta_4 - \beta_5^2 > 0\} \tag{2.1}$$

and

$$f(x,\beta) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp\left\{-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right\} \tag{2.2}$$

with

$$\mu = (\beta_1, \beta_2)^T , \tag{2.3}$$

$$\Sigma = \begin{pmatrix} \beta_3 & \beta_5 \\ \beta_5 & \beta_4 \end{pmatrix} \tag{2.4}$$

and $|\Sigma|$ denoting the determinant of Σ.

Observe that when the random vector X_i has a density $f(x,\beta)$ then

$$\beta = (\mathbf{E}X_{1i}, \mathbf{E}X_{2i}, \mathbf{Var}X_{1i}, \mathbf{Var}X_{2i}, \mathbf{Cov}(X_{1i}, X_{2i}))^T . \tag{2.5}$$

First step in construction of smooth test of $\mathcal{H}_0$ is embedding the density $f(x,\beta)$ into a larger exponential family, which shall be defined below.

2.1. Modeling alternatives via increasing exponential families

Let b_i denote the i-th normalized Legendre polynomial on [0,1] and set additionally $b_0 = \mathbf{1}_{[0,1]}$, where $\mathbf{1}_{\mathcal{A}}$ stands for the indicator function of the set $\mathcal{A}$. Define the function $B_{ij} : [0,1]^2 \to \mathbf{R}$ as

$$B_{ij}(u_1, u_2) = b_i(u_1)b_j(u_2), \quad u_1, u_2 \in [0,1] .$$

Let us order the set of functions $\mathcal{B} = \{B_{ij};\ i,\ j \in \mathbf{N} \cup \{0\}\}$ into the sequence $\tilde{\gamma}$ according to the following rule; the function B_{ij} appears in $\tilde{\gamma}$ before the function B_{lk} if one of the following conditions is fulfilled:

i) $i + j < l + k$,
ii) $i + j = l + k$ and $\max(i,j) > \max(l,k)$,
iii) $i + j = l + k$, $\max(i,j) = \max(l,k)$ and $i > l$.

Let us denote the i-th element of $\tilde{\gamma}$ by $\tilde{\gamma}_i$. So we have

$$\tilde{\gamma}_0 = B_{00},\ \tilde{\gamma}_1 = B_{10},\ \tilde{\gamma}_2 = B_{01},\ \tilde{\gamma}_3 = B_{20},\ \tilde{\gamma}_4 = B_{02},\ \tilde{\gamma}_5 = B_{11},\ \dots . \tag{2.6}$$

For each $\beta \in \Omega$ let L be the lower triangular matrix with positive diagonal such that

$$LL^T = \Sigma^{-1} , \tag{2.7}$$

where Σ is as in (2.4). Thus

$$L = L(\beta) = \begin{pmatrix} l_1 & 0 \\ l_2 & l_3 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\beta_4}{|\Sigma|}} & 0 \\ \frac{-\beta_5}{\sqrt{\beta_4|\Sigma|}} & \frac{1}{\sqrt{\beta_4}} \end{pmatrix} . \tag{2.8}$$

For every $i \in \mathbf{N}$ define the function $\gamma_i : \mathbf{R}^2 \times \Omega \to \mathbf{R}$ as

$$\gamma_i(x,\beta) = \tilde{\gamma}_i(\Phi(y_1), \Phi(y_2)) \quad ,$$

where Φ denotes the standard normal distribution function and y_1, y_2 are the coordinates of the vector $y(x,\beta)$ given by

$$y(x,\beta) = (y_1, y_2)^T = L^T(x-\mu). \tag{2.9}$$

By orthonormality of the Legendre polynomials in $L_2([0,1])$ it easily follows that for every $\beta \in \Omega$ the functions $\gamma_1(x,\beta), \gamma_2(x,\beta), \ldots$ satisfy

$$\forall\ j \in \mathbf{N} \quad \int_{\mathbf{R}^2} \gamma_j(x,\beta) f(x,\beta)\, dx = 0, \tag{2.10}$$

$$\forall\ i,j \in \mathbf{N} \quad \int_{\mathbf{R}^2} \gamma_j(x,\beta)\gamma_i(x,\beta) f(x,\beta)\, dx = \delta_{ij}\ , \tag{2.11}$$

where δ_{ij} is the Kronecker delta.

For $k = 1, 2, \ldots$ define the exponential family of densities

$$g_k(x,\theta,\beta) = \exp\{\theta \circ \gamma(x,\beta) - \psi_k(\theta)\} f(x,\beta)\ , \tag{2.12}$$

where

$$\theta = (\theta_1, \ldots, \theta_k) \in \mathbf{R}^k\ ,$$
$$\gamma(x,\beta) = (\gamma_1(x,\beta), \ldots, \gamma_k(x,\beta))\ ,$$
$$\psi_k(\theta) = \log \int_{[0,1]^2} \exp\{\sum_{j=1}^{k} \theta_j \tilde{\gamma}_j(x)\}\, dx$$

and $\circ$ stands for the inner product in $\mathbf{R}^k$.

Observe that testing bivariate normality within the exponential family (2.12) is equivalent to testing $\theta = 0$ against $\theta \neq 0$. Asymptotically optimal test statistic for this testing problem is the score statistic. Its description is given below.

2.2. Smooth test for bivariate normality

The score test statistic for testing $\theta = 0$ within the family (2.12) is defined as follows

$$W_k = n(T_{nk}(\hat{\beta}))^T \{\mathbf{I}_{k\times k} + R_k(\hat{\beta})\} T_{nk}(\hat{\beta})\ , \tag{2.13}$$

where

$$T_{nk}(\beta) = (\overline{\gamma_1}(\beta), \ldots, \overline{\gamma_k}(\beta))^T = \frac{1}{n}\sum_{i=1}^{n} (\gamma_1(X_i,\beta), \ldots, \gamma_k(X_i,\beta))^T\ , \tag{2.14}$$

$$R_k(\beta) = (I_k(\beta))^T \left(I_{\beta\beta} - I_k(\beta)(I_k(\beta))^T\right)^{-1} I_k(\beta)\ , \tag{2.15}$$

with

$$I_k(\beta) = \mathbf{Cov}_{0,\beta}\left\{\frac{\partial \log f(X,\beta)}{\partial \beta_t}, \gamma_j(X,\beta)\right\}_{t=1,\ldots,5,\ j=1,\ldots,k} \tag{2.16}$$

and

$$I_{\beta\beta} = \mathbf{Cov}_{0,\beta}\left\{\frac{\partial \log f(X,\beta)}{\partial \beta_t}, \frac{\partial \log f(X,\beta)}{\partial \beta_u}\right\}_{t=1,\dots,5,\, u=1,\dots,5} ,$$

$\mathrm{I}_{k\times k}$ is the $k \times k$ identity matrix and $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_5)$ denotes the maximum likelihood estimator of β under $\mathcal{H}_0$.

Test rejecting $\mathcal{H}_0$ for large values of W_k is called a smooth test. For a justification see Javitz (1975), Rayner and Best (1989) and Thomas and Pierce (1979), e.g.

Observe that

$$\hat{\beta} = (\bar{X}_1,\, \bar{X}_2, \frac{1}{n}\sum_{i=1}^{n} X_{1i}^2 - \bar{X}_1^2,\, \frac{1}{n}\sum_{i=1}^{n} X_{2i}^2 - \bar{X}_2^2,\, \frac{1}{n}\sum_{i=1}^{n} X_{1i}X_{2i} - \bar{X}_1\bar{X}_2)^T , \tag{2.17}$$

where

$$(\bar{X}_1,\, \bar{X}_2) = (\frac{1}{n}\sum_{i=1}^{n} X_{1i},\, \frac{1}{n}\sum_{i=1}^{n} X_{2i}) .$$

Below, the vector $(\bar{X}_1,\, \bar{X}_2)^T$ shall be denoted by $\hat{\mu}$.

Remark 2.1 Lemma 6.1 (see Appendix) shows that though the matrices $I_k(\beta)$ and $I_{\beta\beta}$ depend on β the matrix $R_k(\beta)$ depends only on k. Moreover in Section 6.2 of Appendix a straightforward computational formula for the matrix $R_k(\beta)$ is given.

Using W_k we must at first decide on the number k of components in the vector $T_{nk}(\beta)$. It turns out that a wrong choice of k may give a considerable loss of power. For an evidence see Kallenberg and Ledwina (1995c), e.g. Therefore, following Kallenberg and Ledwina (1995c) and Inglot et al. (1996), we use a certain modification of Schwarz's selection rule to chose the "right" dimension for the smooth test.

2.3. Schwarz Selection Rule

When the vector of parameters β is known, Schwarz Bayesian Information Criterion for choosing exponential families (2.12), corresponding to successive dimensions up to $d(n)$ yields

$$\mathrm{S}_\beta = \min\{k:\ 1 \le k \le d(n)\ , L_k(\beta) \ge L_j(\beta),\ j = 1, \dots, d(n)\} ,$$

where

$$L_k(\beta) = n \sup_{\theta \in \mathbf{R}^k} \{\theta \circ T_{nk}(\beta) - \psi(\theta)\} - \frac{1}{2} k \log n .$$

An extension of the Schwarz's rule to the situation where β is an unknown nuisance parameter is obtained by inserting the maximum likelihood estimator $\hat{\beta}$ of β (under $\theta = 0$) into the formula S_β. However, a calculation of $\mathrm{S}_{\hat{\beta}}$ is numerically involved. Therefore a simplification would be welcome.

By an inspection of proofs of Theorems 7.3 and 7.4 of Inglot and Ledwina (1995) it is easy to check that these theorems generalize to the case of an orthonormal system of functions in $\mathbf{R}^2$. Thus we have that the maximized loglikelihood

$$n \sup_{\theta \in \mathbf{R}^k} \{\theta \circ T_{nk}(\beta) - \psi_k(\theta)\}$$

is locally equivalent to $\frac{1}{2}n \parallel T_{nk}(\beta) \parallel^2$. Therefore following Kallenberg and Ledwina (1995c) and Inglot, Kallenberg and Ledwina (1996) we use in this paper a modification of the Schwarz rule given by

$$S(1) = \min \{k : 1 \leq k \leq d(n),\ n \parallel T_{nk}(\hat{\beta}) \parallel^2 - k \log n \geq n \parallel T_{nj}(\hat{\beta}) \parallel^2 - j \log n,\ j = 1, \ldots, d(n)\} . \tag{2.18}$$

During the simulation study we observed that in a result of an estimation of the vector of parameters β the first five components of $T_{nk}(\hat{\beta})$ are very small in comparison to the other ones. In a consequence $S(1)$ usually does not chose $k \in \{2,3,4,5\}$. Therefore we considered also its modification

$$S(5) = \min \{k : 5 \leq k \leq d(n),\ n \parallel T_{nk}(\hat{\beta}) \parallel^2 - k \log n \geq n \parallel T_{nj}(\hat{\beta}) \parallel^2 - j \log n,\ j = 5, \ldots, d(n)\} . \tag{2.19}$$

2.4. Test statistics of the data driven smooth tests for bivariate normality

Observe that the test statistic W_k (see (2.13)) and selection rules $S(1)$ and $S(5)$ (see (2.18) and (2.19) are defined only when the sample covariance matrix $\hat{\Sigma}$ is nonsingular. To cover the cases when $\hat{\Sigma}$ is singular we additionally set $W_{S(1)} = W_{S(5)} = \infty$ when $|\hat{\Sigma}| = 0$. Since under the assumption (2.1) the probability of singularity of $\hat{\Sigma}$ is equal to zero this additional definition does not influence the null distributions of $W_{S(1)}$ and $W_{S(5)}$.

The two versions of the data driven smooth test for bivariate normality, considered in this paper, reject the null hypothesis for large values of $W_{S(1)}$ and $W_{S(5)}$, respectively.

Lemma 2.2 Test statistics $W_{S(1)}$ and $W_{S(5)}$ are invariant upon the following transformation of the sequence of the random vectors $X_1, \ldots, X_n$;

$$\forall i \in \{1, \ldots, n\} \quad Y_i = AX_i + B \ , \tag{2.20}$$

where A is any upper triangular matrix with positive diagonal and B is a column vector in $\mathbf{R}^2$.

Proof. By Theorem 4 of Szkutnik (1987) we get that the following standardization of the sequence of the random vectors $X_1, \ldots, X_n$

$$\forall i \in \{1, \ldots, n\} \quad Z_i = (L(\hat{\beta}))^T (X_i - \hat{\mu}) \ , \tag{2.21}$$

where $L(\beta)$ is given by (2.8) and $\hat{\beta}$ is as in (2.17), is invariant upon (2.20).

Since the test statistic W_k as well as the selection rules $S(1)$ and $S(5)$ depend on the sample only via its standardized version given by (2.21) thus the proof of Lemma 2.2 is concluded. □

Corollary 2.3 Under $\mathcal{H}_0$ the distributions of $W_{S(1)}$ and $W_{S(5)}$ do not depend on the vector of parameters β.

Proof. Consider a random vector X_i coming from a bivariate normal distribution (2.2) with the vector of parameters β. Observe that if in (2.20) we replace A with $(L(\beta))^T$

and B with $-(L(\beta))^T\mu$, where μ is related to β via (2.3), then Y_i is distributed according to a bivariate normal distribution with the vector of parameters $\beta_0 = (0,0,1,1,0)^T$. Thus Corollary 2.3 is an immediate consequence of Lemma 2.2. □

Remark 2.4 Ordering the functions B_{ij} into the sequence $\tilde{\gamma}$, being important for the definition of $W_{S(1)}$ and $W_{S(5)}$, can be done in many different ways. In our opinion the order imposed by i)–iii) in Section 2.1 is a natural one in case of the lack of knowledge on the class of possible alternatives to bivariate normality. However, for testing bivariate normality against some particular types of alternatives another orders could be more appropriate and could result in a better finite sample performance of corresponding tests. Anyway, asymptotic results on distributions of $W_{S(1)}$ and $W_{S(5)}$, which shall be presented below, hold for each ordering of the functions B_{ij} into the sequence $\tilde{\gamma}$ if the bounds (6.9) [2 can be replaced by any positive constant], (6.13) and (6.14) of Appendix are fulfilled for sufficiently large j. Of course some subsets of $\mathcal{B}$ can be considered as well. However, then the resulting tests miss consistency against some particular alternatives.

3. ASYMPTOTICS OF $\mathbf{W_{S(1)}}$ AND $\mathbf{W_{S(5)}}$ UNDER $\mathcal{H}_0$

Since under the null hypothesis the probability of singularity of the sample covariance matrix $\hat{\Sigma}$ is equal to zero therefore in this section we shall restrict attention to the situation when $\hat{\Sigma}$ is invertible.

By Corollary 2.3 we have that the distributions of $S(1)$, $S(5)$, $W_{S(1)}$ and $W_{S(5)}$ under $\mathcal{H}_0$ do not depend on the vector of parameters β. Therefore we shall derive their asymptotic null distributions assuming that $\beta = \beta_0 = (0,0,1,1,0)^T$.

3.1. Asymptotic behavior of the modified Schwarz rule under $\mathcal{H}_0$

For fixed $k_0 \in \{1,\ldots,d(n)\}$ let $S(k_0)$ denote the following modified version of the Schwarz rule,

$$\begin{aligned} S(k_0) = \min \; &\{k:\ k_0 \le k \le d(n),\ n \parallel T_{nk}(\hat{\beta}) \parallel^2 - k\log n \\ &\ge n \parallel T_{nj}(\hat{\beta}) \parallel^2 - j\log n,\ j = k_0,\ldots,d(n)\}\ . \end{aligned}$$

The following theorem states that under $\mathcal{H}_0$ $S(k_0)$ asymptotically concentrates on the dimension k_0.

Theorem 3.1 If $d(n) = o(n^c)$ for some $c < \frac{1}{16}$ then

$$P_{\beta_0}(S(k_0) > k_0) \to 0 \ \text{ as } \ n \to \infty\ . \tag{3.1}$$

The proof of Theorem 3.1 shall be based on two following lemmas which state how close is $\hat{\beta}$ to β_0 and $T_{nk}(\hat{\beta})$ to $T_{nk}(\beta_0)$.

Lemma 3.2 For each $c_2 > 0$ and $c_3 \in (0, 1/4)$ there exists a positive constant c_4 such that it holds

$$P_{\beta_0}\left(\parallel \hat{\beta} - \beta_0 \parallel > c_2\sqrt{\frac{\log n}{n}} \right) \le c_4 n^{-c_3 c_2^2}\ . \tag{3.2}$$

Proof. Denote by V_i, $1 \leq i \leq n$, the vector of random variables

$$V_i = (X_{1i}, X_{2i}, X_{1i}^2 - 1, X_{2i}^2 - 1, X_{1i}X_{2i}) \ .$$

If for each $z = (z_1, \ldots, z_5) \in \mathbf{R}^5$ we define the function g(z) as

$$g(z) = (z_1^2 + z_2^2 + (z_3 - z_1^2)^2 + (z_4 - z_2^2)^2 + (z_5 - z_1 z_2)^2)$$

then we get

$$\| \hat{\beta} - \beta_0 \|^2 = g(\bar{V}) \ ,$$

where $\bar{V} = \frac{1}{n}\sum_{i=1}^n V_i$.

Since $\lim_{\|z\| \to 0} g(z) \| z \|^{-2} = 1$, thus for each $\delta \in (0,1)$ we have that for sufficiently large n

$$\begin{aligned} P_{\beta_0}\left(\| \hat{\beta} - \beta_0 \| > c_2 \sqrt{\frac{\log n}{n}} \right) &= P_{\beta_0}\left(g(\bar{V}) > c_2^2 \frac{\log n}{n} \right) \\ &\leq P_{\beta_0}\left(\| \bar{V} \|^2 > c_2^2 \frac{\log n}{n}(1-\delta) \right) . \end{aligned} \tag{3.3}$$

By an application of the inequality (101) of Rubin and Sethuraman (1965) to the right-hand side of (3.3) we obtain

$$P_{\beta_0}\left(\| \hat{\beta} - \beta_0 \| > c_2 \sqrt{\frac{\log n}{n}} \right) \leq c_4 n^{-c_2^2(1-\delta)/4} \tag{3.4}$$

and the proof of (3.2) is completed. □

Before we present Lemma 3.3 we introduce the following notation

$$a_k = \frac{(k-k_0)\log n}{n} \ , \quad k > k_0 \ , \tag{3.5}$$

X – a random variable distributed as X_1, (3.6)

$$U_{tj} = \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{\partial \gamma_j(X_i, \beta)}{\partial \beta_t} |_{\beta=\beta_0} - \mathbf{E}_{\beta_0}\left(\frac{\partial \gamma_j(X, \beta)}{\partial \beta_t} |_{\beta=\beta_0} \right) \right\} \ , \tag{3.7}$$

$$U_j = (U_{1j}, \ldots, U_{5j})^T \ , \tag{3.8}$$

$$R_{1j} = (\hat{\beta} - \beta_0)^T U_j \ , \tag{3.9}$$

$$R_{2j} = \frac{1}{2n}\sum_{i=1}^{n} (\hat{\beta} - \beta_0)^T \left(\frac{\partial^2 \gamma_j(X_i, \beta)}{\partial \beta \partial \beta^T} |_{\beta=\vartheta_i} \right) (\hat{\beta} - \beta_0) \ , \tag{3.10}$$

where ϑ_i is a point between $\hat{\beta}$ and β_0 ,

$$Z_j = (\hat{\beta} - \beta_0)^T \mathbf{E}_{\beta_0}\left(\frac{\partial \gamma_j(X, \beta)}{\partial \beta} |_{\beta=\beta_0} \right) . \tag{3.11}$$

Lemma 3.3 If $d(n) = o(n^c)$ for some $c < \frac{1}{16}$ then there exists a constant $\xi \in (0,1)$ such that

$$\lim_{n\to\infty} \sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\| T_{nk}(\hat{\beta}) - T_{nk}(\beta_0) \| \geq \xi \sqrt{\frac{(k-k_0)\log n}{n}} \right) = 0 \ . \tag{3.12}$$

Proof. Notice that by Taylor expansion for some ϑ_i between $\hat{\beta}$ and β_0 it holds

$$\begin{aligned}\gamma_j(x,\hat{\beta}) &= \gamma_j(x,\beta_0) + (\hat{\beta}-\beta_0)^T \frac{\partial \gamma_j(x,\beta)}{\partial \beta}|_{\beta=\beta_0} \\ &+\frac{1}{2}(\hat{\beta}-\beta_0)^T \left(\frac{\partial^2 \gamma_j(x,\beta)}{\partial\beta\partial\beta^T}|_{\beta=\vartheta_i}\right)(\hat{\beta}-\beta_0) \ . \end{aligned} \tag{3.13}$$

Hence

$$\overline{\gamma_j}(\hat{\beta}) - \overline{\gamma_j}(\beta_0) = R_{1j} + R_{2j} + Z_j \ . \tag{3.14}$$

Therefore $\forall\ \epsilon \in (0,1)$ we have

$$\begin{aligned} P_{\beta_0}\left(\| T_{nk}(\hat{\beta}) - T_{nk}(\beta_0) \| \geq \xi\sqrt{a_k}\right) &\leq P_{\beta_0}\left(\sqrt{\sum_{j=1}^k Z_j^2} \geq (1-\epsilon)\xi\sqrt{a_k}\right) \\ +P_{\beta_0}\left(\sqrt{\sum_{j=1}^k R_{1j}^2} \geq \frac{\epsilon}{2}\xi\sqrt{a_k}\right) &+ P_{\beta_0}\left(\sqrt{\sum_{j=1}^k R_{2j}^2} \geq \frac{\epsilon}{2}\xi\sqrt{a_k}\right) \ . \end{aligned} \tag{3.15}$$

We shall show at first that for some $\epsilon_1 \in (0,1)$ and $\xi \in (0,1)$ it holds

$$\lim_{n\to\infty} \sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^k Z_j^2} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right) = 0 \ . \tag{3.16}$$

By the orthonormality of the functions $\gamma_j(x,\beta)$ with respect to $f(x,\beta)$ (see (2.10) and (2.11)) we have that

$$\frac{\partial}{\partial \beta_t}\left(\int_{\mathbf{R}^2} \gamma_j(x,\beta) f(x,\beta) dx\right)|_{\beta=\beta_0} = 0\ , \ \ t = 1,\ldots,5 \ . \tag{3.17}$$

By straightforward calculation it is easy to check that there exists a neighborhood $\tilde{\mathcal{V}}(\beta_0)$ of β_0, such that for each $x \in \mathbf{R}^2$ and $\beta \in \tilde{\mathcal{V}}(\beta_0)$ it holds

$$f(x,\beta) \leq c_5 \exp(-\frac{1}{3} \| x \|^2) \tag{3.18}$$

and

$$\left|\frac{\partial f(x,\beta)}{\partial \beta_t}\right| \leq c_5(1+ \| x \|^2) \exp(-\frac{\| x \|^2}{3}) \ , \tag{3.19}$$

where c_5 is a positive constant.
By the inequalities (6.13), (6.14), (3.18) and (3.19) we may interchange the order of differentiation and integration in (3.17) and we easily obtain

$$\mathbf{Cov}_{\beta_0}\left(\gamma_j(X,\beta_0), \frac{\partial \log f(X,\beta)}{\partial \beta_t}|_{\beta=\beta_0}\right) = -\mathbf{E}_{\beta_0}\left(\frac{\partial \gamma_j(X,\beta)}{\partial \beta_t}|_{\beta=\beta_0}\right) \ . \tag{3.20}$$

Thus applying the Schwarz inequality to the left-hand side of (3.20), by (2.10) and (2.11), we get

$$\sup_j \left|\mathbf{E}_{\beta_0} \frac{\partial \gamma_j(X,\beta)}{\partial \beta_t}|_{\beta=\beta_0}\right|^2 \leq \mathbf{Var}_{\beta_0}\left(\frac{\partial \log f(X,\beta)}{\partial \beta_t}|_{\beta=\beta_0}\right) \ . \tag{3.21}$$

From (6.7) of Appendix we have

$$\sum_{t=1}^{5} \mathbf{Var}_{\beta_0}\left(\frac{\partial \log f(X,\beta)}{\partial \beta_t}|_{\beta=\beta_0}\right) = 4 \ . \tag{3.22}$$

By (3.11), (3.21) and (3.22) we get

$$P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} Z_j^2} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right) \leq P_{\beta_0}\left(\| \hat{\beta} - \beta_0 \| 2\sqrt{k} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right). \tag{3.23}$$

By (3.2), (3.5) and (3.23) it follows that for each $c_3 \in (0, 1/4)$ there exists $c_4 > 0$ such that

$$P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} Z_j^2} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right) \leq c_4 \exp(-c_3 \frac{1}{4}(1-\epsilon_1)^2 \xi^2 \frac{k-k_0}{k} \log n) \ . \tag{3.24}$$

Observe that for each constant $c < 1/16$ we can choose $c_3 \in (0,1/4)$, $\epsilon_1 \in (0,1)$, $\xi \in (0,1)$ and a natural number K such that $\forall\, k \geq K$ it holds

$$\frac{1}{4} c_3 (1-\epsilon_1)^2 \xi^2 \frac{k-k_0}{k} > c \ .$$

Hence by (3.24) and the assumption on $d(n)$ we get

$$\sum_{k=K}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} Z_j^2} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right) \leq c_4 \frac{d(n)}{n^c} \quad \to 0 \text{ as } n \to \infty \ .$$

Since from (3.24) it follows also that $\forall\, k \in \{k_0+1, \ldots, K-1\}$

$$\lim_{n\to\infty} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} Z_j^2} \geq (1-\epsilon_1)\xi\sqrt{a_k}\right) = 0$$

the proof of (3.16) is completed.

Let us show now that

$$\lim_{n\to\infty} \sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{1j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) = 0 \ . \tag{3.25}$$

For each $n \in \mathrm{N}$ we shall denote by $\mathcal{B}_n$ the event

$$\mathcal{B}_n = \left\{ \| \hat{\beta} - \beta_0 \| \leq \sqrt{\frac{\log n}{n}} \right\}$$

and by $\mathcal{B}_n^c$ its complement.
Observe that $\forall\, k \in \mathrm{N}$ we get

$$P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{1j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) \leq P_{\beta_0}(\mathcal{B}_n^c) + P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{1j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k},\ \mathcal{B}_n\right). \tag{3.26}$$

When $\mathcal{B}_n$ holds then $\forall\, j \in \mathbf{N}$ we have

$$R_{1j}^2 \le \frac{\log n}{n} \sum_{t=1}^{5} U_{tj}^2 \ .$$

Thus

$$P_{\beta_0}\left(\sum_{j=1}^{k} R_{1j}^2 \ge \frac{\epsilon_1^2}{4}\xi^2 a_k,\ \mathcal{B}_n\right) \le P_{\beta_0}\left(\frac{\log n}{n}\sum_{j=1}^{k}\sum_{t=1}^{5} U_{tj}^2 \ge \frac{\epsilon_1^2}{4}\xi^2 a_k\right) \ .$$

Observe that

$$\mathbf{E}_{\beta_0} U_{tj}^2 = \frac{1}{n}\mathbf{Var}_{\beta_0}\left(\frac{\partial\gamma_j(X,\beta)}{\partial\beta_t}\,|_{\beta=\beta_0}\right) \ . \qquad (3.27)$$

Since by (6.13) $\forall\, t \in \{1,\ldots,5\}$

$$\mathbf{Var}_{\beta_0}\left(\frac{\partial\gamma_j(X,\beta)}{\partial\beta_t}\,|_{\beta=\beta_0}\right) \le c_1^2 j^3 \mathbf{E}_{\beta_0}(1+\parallel X \parallel)^2$$

then (3.27) yields

$$\sum_{t=1}^{5} \mathbf{E}_{\beta_0} U_{tj}^2 \le \frac{5 j^3 c_6}{n} \ ,$$

where $c_6 = c_1^2 \mathbf{E}_{\beta_0}(1+\parallel X \parallel)^2$.
From the above, by Markov's inequality, it follows that there exists a constant c_7 such that

$$P_{\beta_0}\left(\sum_{j=1}^{k} R_{1j}^2 \ge \frac{\epsilon_1^2}{4}\xi^2 a_k,\ \mathcal{B}_n\right) < c_7\frac{k^3}{n} \ . \qquad (3.28)$$

Now by the inequalities (3.26), (3.28) and (3.2) (with $c_2 = 1$ and $c_3 = 1/5$) we get

$$\sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{1j}^2} \ge \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) < \frac{c_7 d^4(n)}{n} + \frac{c_4 d(n)}{n^{1/5}} \quad \to 0 \ \text{as}\ n \to \infty$$

and the proof of (3.25) is completed.

By (3.15), (3.16) and (3.25), to prove (3.12) it is enough to show that

$$\lim_{n\to\infty} \sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{2j}^2} \ge \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) = 0 \ . \qquad (3.29)$$

To this end observe that (3.10) and (6.14) imply that for sufficiently large n on the set $\mathcal{B}_n$ it holds

$$|R_{2j}| \le \frac{5}{2} c_1 j^{5/2} \frac{\log n}{n^2} \sum_{i=1}^{n} (1+\parallel X_i \parallel^2) \ . \qquad (3.30)$$

Thus

$$P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{2j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k},\ \mathcal{B}_n\right)$$
$$< P_{\beta_0}\left(\frac{5}{2}k^3 c_1 \frac{\log n}{n^2}\sum_{i=1}^{n}(1+\| X_i \|^2) \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) \ . \tag{3.31}$$

Applying Markov's inequality to the right-hand side of (3.31) we further obtain that there exists a constant c_8, independent of k, such that

$$P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{2j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k},\ \mathcal{B}_n\right) \leq c_8 k^{5/2}\sqrt{\frac{\log n}{n}} \ . \tag{3.32}$$

Inequalities (3.32) and (3.2) (with $c_2 = 1$ and $c_3 = 1/5$) and the assumption on $d(n)$ imply

$$\sum_{k=k_0+1}^{d(n)} P_{\beta_0}\left(\sqrt{\sum_{j=1}^{k} R_{2j}^2} \geq \frac{\epsilon_1}{2}\xi\sqrt{a_k}\right) \ < \ c_8[d(n)]^{7/2}\sqrt{\frac{\log n}{n}} + \frac{c_4 d(n)}{n^{1/5}} \ \to 0 \text{ as } n \to \infty \ .$$

Thus the proof of Lemma 3.3 is completed. □

Proof of Theorem 3.1. To prove (3.1) first observe that for each $k \in \{k_0+1, \ldots, d(n)\}$ and $\xi \in (0,1)$ it holds

$$\begin{aligned} P_{\beta_0}(S(1) = k) &\leq P_{\beta_0}(n \| T_{nk}(\hat{\beta}) \|^2 \geq (k-k_0)\log n) \\ &\leq P_{\beta_0}(n \| T_{nk}(\beta_0) \|^2 \geq (1-\xi)^2(k-k_0)\log n) \\ &\quad + P_{\beta_0}\left(\| T_{nk}(\hat{\beta}) - T_{nk}(\beta_0) \| \geq \xi\sqrt{\frac{(k-k_0)\log n}{n}}\right) \ . \end{aligned} \tag{3.33}$$

By (3.33) and Lemma 3.3 to prove (3.1) it is enough to show that

$$\lim_{n\to\infty} \sum_{k=k_0+1}^{d(n)} P_{\beta_0}(n \| T_{nk}(\beta_0) \|^2 \geq (1-\xi)^2(k-k_0)\log n) = 0 \ . \tag{3.34}$$

We have that

$$\begin{aligned} &P_{\beta_0}(n \| T_{nk}(\beta_0) \|^2 \geq (1-\xi)^2(k-k_0)\log n) \\ &= P_{\beta_0}\left(\sum_{j=1}^{k}\left\{[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)]^2 - 1\right\} \geq \tilde{a}_k\right) \ , \end{aligned} \tag{3.35}$$

where $\tilde{a}_k = (1-\xi)^2(k-k_0)\log n - k$.
Using the orthonormality of the functions $\gamma_j(x,\beta_0)$ with respect to $f(x,\beta_0)$ we easily get

$$\mathbf{E}_{\beta_0}\left[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)\right]^4 = \frac{1}{n}\mathbf{E}_{\beta_0}\gamma_j^4(X,\beta_0) + \frac{3(n-1)}{n}$$

and hence

$$\mathbf{Var}_{\beta_0}\left\{[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)]^2-1\right\}=\frac{1}{n}\mathbf{E}_{\beta_0}\gamma_j^4(X,\beta_0)+2-\frac{3}{n} \ . \tag{3.36}$$

Moreover, by direct calculation we obtain that $\forall\ j\neq l$

$$\mathbf{E}_{\beta_0}\left[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)\right]^2\left[n^{-1/2}\sum_{i=1}^{n}\gamma_l(X_i,\beta_0)\right]^2$$
$$=\frac{1}{n}\mathbf{E}_{\beta_0}\gamma_j^2(X,\beta_0)\gamma_l^2(X,\beta_0)+1-\frac{1}{n} \ .$$

Hence

$$\mathbf{Cov}_{\beta_0}\left(\left[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)\right]^2-1,\left[n^{-1/2}\sum_{i=1}^{n}\gamma_l(X_i,\beta_0)\right]^2-1\right)$$
$$=\frac{1}{n}\mathbf{E}_{\beta_0}\gamma_j^2(X,\beta_0)\gamma_l^2(X,\beta_0)-\frac{1}{n} \ . \tag{3.37}$$

Formulas (3.36) and (3.37) lead to

$$\mathbf{Var}_{\beta_0}\left(\sum_{j=1}^{k}\left\{[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)]^2-1\right\}\right)$$
$$=2k+n^{-1}\sum_{j=1}^{k}(\mathbf{E}_{\beta_0}\gamma_j^4(X,\beta_0)-3)+n^{-1}\sum_{j\neq l}\mathbf{E}_{\beta_0}(\gamma_j^2(X,\beta_0)\gamma_l^2(X,\beta_0)-1)$$

and by Chebyshev's inequality we get

$$P_{\beta_0}\left(\sum_{j=1}^{k}\left\{[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)]^2-1\right\}\geq\tilde{a}_k\right) \tag{3.38}$$
$$\leq\tilde{a}_k^{-2}\{2k+n^{-1}\sum_{j=1}^{k}(\mathbf{E}_{\beta_0}\gamma_j^4(X,\beta_0)-3)+n^{-1}\sum_{j\neq l}\mathbf{E}_{\beta_0}(\gamma_j^2(X,\beta_0)\gamma_l^2(X,\beta_0)-1)\} \ .$$

Applying (2.11) and the bound $\gamma_j(x,\beta_0)\leq 2\sqrt{j}$ [see (6.9) of Appendix] we obtain that for each pair of natural numbers $j\leq l$ it holds

$$\mathbf{E}_{\beta_0}\gamma_j^2(X,\beta_0)\gamma_l^2(X,\beta_0)\leq 4j \ .$$

Thus by (3.38) we get that for sufficiently large n there exists a constant c_9, independent of k, such that it holds

$$P_{\beta_0}\left(\sum_{j=1}^{k}\left\{[n^{-1/2}\sum_{i=1}^{n}\gamma_j(X_i,\beta_0)]^2-1\right\}\geq\tilde{a}_k\right)$$
$$<\tilde{a}_k^{-2}\{2k+n^{-1}[k(2k-1)+k(k-1)(4k-5)]\}$$
$$<\tilde{a}_k^{-2}(2k+4k^3n^{-1})<\frac{c_9}{\log^2 n}(2k^{-1}+4kn^{-1}) \ . \tag{3.39}$$

By the well known result on the partial sums of harmonic series we have

$$\sum_{k=1}^{d(n)} \frac{1}{k} = \log d(n) + C + o(1) \ ,$$

where C is the Euler constant and $\lim_{d(n)\to\infty} o(1) = 0$. Hence by (3.35) and (3.39) we get

$$\begin{aligned} &\textstyle\sum_{k=k_0+1}^{d(n)} P_{\beta_0}(n \parallel T_{nk}(\beta_0) \parallel^2 \geq (1-\xi)^2(k-k_0)\log n) \\ &< \tfrac{2c_9}{\log^2 n}(\log d(n) + 2\tfrac{d^2(n)}{n} + C + o(1)) \ . \end{aligned} \tag{3.40}$$

Since $d(n) = o(n^c)$, $c < \frac{1}{16}$, the proof of Theorem 3.1 is completed. □

3.2. Asymptotic distributions of $\mathbf{W_{S(1)}}$ and $\mathbf{W_{S(5)}}$ under $\mathcal{H}_0$

Theorem 3.4 Assume $d(n) = o(n^c)$ for some $c < \frac{1}{16}$. Then under the null hypothesis,

$$W_{S(1)} \xrightarrow{\mathcal{D}} \chi_1^2 \ . \tag{3.41}$$

and

$$W_{S(5)} \xrightarrow{\mathcal{D}} \chi_5^2 \ , \tag{3.42}$$

where χ_k^2 stands for a random variable with a chi-square distribution with k degrees of freedom.

Proof. Note that for $k_0 = 1$ and $k_0 = 5$ it holds

$$\begin{aligned} P_{\beta_0}(W_{S(k_0)} \leq x) &= P_{\beta_0}(W_{k_0} \leq x) - P_{\beta_0}(W_{k_0} \leq x, S(k_0) > k_0) \\ &\quad + P_{\beta_0}(W_{S(k_0)} \leq x, S(k_0) > k_0) \ . \end{aligned}$$

Thus, by Theorem 3.1, to get (3.41) and (3.42) it is enough to show that, under the null hypothesis, for each natural k

$$W_k \xrightarrow{\mathcal{D}} \chi_k^2 \ . \tag{3.43}$$

The statement (3.43) could be inferred from general theory of score statistics [cf. Sen and Singer (1993), Chapter 5]. However, we have collected up to now so many information on ingredients of W_k that we find it easier to show (3.43) in an immediate way. To prove (3.43) we shall argue similarly as in the proof of Theorem 3.1 of Inglot et al. (1996).

By the definition of W_k [see (2.13)], Remark 2.1 and the well known result on the distribution of quadratic forms of normal vectors [cf. e.g. Rao (1973), Chapter 3.2.4] to obtain (3.43) it is enough to show that under the null hypothesis

$$\sqrt{n}T_{nk}(\hat{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (\mathbf{I}_{k\times k} + R_k(\beta_0))^{-1}\right) \ , \tag{3.44}$$

with $R_k(\beta_0)$ as in (2.15).

To prove (3.44) at first recall that $T_{nk}(\beta) = (\bar{\gamma}_1(\beta), \ldots, \bar{\gamma}_k(\beta))^T$ and that from the Taylor expansion (3.13) we get

$$\overline{\gamma_j}(\hat{\beta}) - \overline{\gamma_j}(\beta_0) = R_{1j} + R_{2j} + Z_j \ , \tag{3.45}$$

where R_{1j}, R_{2j} and Z_j are given by (3.9), (3.10) and (3.11) , respectively.
Observe that

$$\hat{\beta} - \beta_0 = \frac{1}{n}\sum_{i=1}^{n} V_i + o_{P_{\beta_0}}(n^{-1/2}) \ , \tag{3.46}$$

where $V_i = (X_{1i}, X_{2i}, X_{1i}^2 - 1, X_{2i}^2 - 1, X_{1i}X_{2i})^T$ and $\| \sqrt{n} o_{P_{\beta_0}}(n^{-1/2}) \| \stackrel{P_{\beta_0}}{\to} 0$.
Thus we easily get

$$\sqrt{n} \| \hat{\beta} - \beta_0 \| = O_{P_{\beta_0}}(1) \ . \tag{3.47}$$

Hence for each natural j

$$\sqrt{n} R_{1j} = \sqrt{n}(\hat{\beta} - \beta_0)^T U_j \stackrel{P_{\beta_0}}{\to} 0 \ .$$

Moreover, by (6.14) we have that for each $\epsilon > 0$

$$\begin{aligned} P_{\beta_0}(\sqrt{n} R_{2j} > \epsilon) &\le P_{\beta_0}\left(\frac{5}{2} c_1 j^{5/2} \sqrt{n} \| \hat{\beta} - \beta_0 \|^2 \frac{1}{n} \sum_{i=1}^{n} (1 + \| X_i \|^2) > \epsilon\right) \\ &\quad + P_{\beta_0}(\hat{\beta} \notin \mathcal{V}(\beta_0)) \quad \to 0 \text{ as } n \to \infty \ . \end{aligned} \tag{3.48}$$

Observe that by (3.20) and (2.16) we have that

$$(Z_1, \ldots, Z_k)^T = -(I_k(\beta_0))^T (\hat{\beta} - \beta_0) \ .$$

Thus from (3.45), (3.46), (3.47) and (3.48) it follows

$$\sqrt{n} T_{nk}(\hat{\beta}) = \sqrt{n} T_{nk}(\beta_0) - \sqrt{n}(I_k(\beta_0))^T \left(\frac{1}{n}\sum_{i=1}^{n} V_i\right) + o_{P_{\beta_0}}(1) \ . \tag{3.49}$$

From (6.1) and (6.7) of Appendix we obtain

$$V_i = I_{\beta_0\beta_0}^{-1} \left(\frac{\partial \log f(X_i, \beta)}{\partial \beta_t} |_{\beta=\beta_0}\right)_{t=1,\ldots,5} \ .$$

Hence by the definition of $I_k(\beta)$ [see(2.16)] the matrix of covariances of the elements of the vector $(I_k(\beta_0))^T V_i$ and the vector $\gamma(X_i, \beta_0) = (\gamma_1(X_i, \beta_0), \ldots, \gamma_k(X_i, \beta_0))^T$ is given by

$$\begin{aligned} \mathbf{E}_{\beta_0}\left[(I_k(\beta_0))^T V_i \gamma^T(X_i, \beta_0)\right] &= (I_k(\beta_0))^T I_{\beta_0\beta_0}^{-1} I_k(\beta_0) \\ &= \mathbf{E}_{\beta_0}\left(\gamma(X_i, \beta_0) V_i^T I_k(\beta_0)\right) \ . \end{aligned} \tag{3.50}$$

Moreover

$$\begin{aligned} \mathbf{E}_{\beta_0}[(I_k(\beta_0))^T V_i V_i^T I_k(\beta_0)] &= (I_k(\beta_0))^T I_{\beta_0\beta_0}^{-1} I_{\beta_0\beta_0} I_{\beta_0\beta_0}^{-1} I_k(\beta_0) \\ &= (I_k(\beta_0))^T I_{\beta_0\beta_0}^{-1} I_k(\beta_0) \ . \end{aligned} \tag{3.51}$$

By the orthonormality of functions $\gamma_1(X_i, \beta_0), \ldots, \gamma_k(X_i, \beta_0)$ with respect to $f(x, \beta_0)$

$$\mathbf{E}_{\beta_0}[\gamma(X_i, \beta_0)\gamma^T(X_i, \beta_0)] = \mathbf{I}_{k\times k} \tag{3.52}$$

and (3.50), (3.51) and (3.52) yield

$$\mathbf{E}_{\beta_0}\left[\left(\gamma(X_i,\beta_0)-(I_k(\beta_0))^T V_i\right)\left(\gamma(X_i,\beta_0)-(I_k(\beta_0))^T V_i\right)^T\right]$$
$$= \mathbf{I}_{k\times k} - (I_k(\beta_0))^T I^{-1}_{\beta_0\beta_0} I_k(\beta_0) \ . \tag{3.53}$$

Set $A_k(\beta_0) = (I_k(\beta_0))^T I^{-1}_{\beta_0\beta_0} I_k(\beta_0)$ and observe that

$$\begin{aligned} A_k(\beta_0) &= (I_k(\beta_0))^T I^{-1}_{\beta_0\beta_0}\left(I_{\beta_0\beta_0} - I_k(\beta_0)(I_k(\beta_0))^T\right)\left(I_{\beta_0\beta_0} - I_k(\beta_0)(I_k(\beta_0))^T\right)^{-1} I_k(\beta_0) \\ &= R_k(\beta_0) - A_k(\beta_0)R_k(\beta_0) \ , \end{aligned}$$

with $R_k(\beta_0)$ as in (2.15). Thus we have that

$$(\mathbf{I}_{k\times k} - A_k(\beta_0))(\mathbf{I}_{k\times k} + R_k(\beta_0)) = \mathbf{I}_{k\times k} + R_k(\beta_0) - A_k(\beta_0) - A_k(\beta_0)R_k(\beta_0) = \mathbf{I}_{k\times k} \ . \tag{3.54}$$

Hence by Multivariate Central Limit Theorem (3.49) and (3.53) yield (3.44). □

4. CONSISTENCY OF TESTS BASED ON $\mathbf{W}_{S(1)}$ AND $\mathbf{W}_{S(5)}$

Assume $X_1, \ldots, X_n$ are independent random vectors, where $X_i = (X_{1i}, X_{2i})^T$. Suppose that each X_i is distributed according to $\mathcal{P}$ on $\mathbf{R}^2$, where $\mathcal{P}$ satisfies

$$\mathbf{E}_{\mathcal{P}} X_{11}^2 < \infty \quad \text{and} \quad \mathbf{E}_{\mathcal{P}} X_{21}^2 < \infty \ . \tag{4.1}$$

To prove consistency of $W_{S(1)}$ and $W_{S(5)}$ under alternatives satisfying (4.1) we shall first prove the following theorem.

Theorem 4.1 Assume that $\mathcal{P}$ is a probability measure on $\mathbf{R}^2$ satisfying (4.1) and such that the covariance matrix of X_{11} and X_{21} is positive definite. Assume further that $\mathcal{P}$ is not bivariate normal distribution with the related vector of parameters

$$\beta = (\mathbf{E}_{\mathcal{P}} X_{11},\ \mathbf{E}_{\mathcal{P}} X_{21},\ \mathbf{Var}_{\mathcal{P}} X_{11},\ \mathbf{Var}_{\mathcal{P}} X_{21},\ \mathbf{Cov}_{\mathcal{P}}(X_{11}, X_{21}))^T \ . \tag{4.2}$$

Then, $\lim_{n\to\infty} d(n) = \infty$ implies

$$W_{S(1)} \xrightarrow{\mathcal{P}} \infty \tag{4.3}$$

and

$$W_{S(5)} \xrightarrow{\mathcal{P}} \infty \ . \tag{4.4}$$

Proof. By Lemma 2.2, without loosing generality, we may assume that $\beta = \beta_0$. Let $\mathcal{N}(\beta_0)$ stand for the bivariate normal distribution with the vector of parameters β_0.

From Lemma 6.4 of Appendix we have that if $\mathcal{P} \neq \mathcal{N}(\beta_0)$ then there exists a natural K such that

$$\mathbf{E}_{\mathcal{P}} \gamma_K(X, \beta_0) \neq 0 \ . \tag{4.5}$$

In what follows we shall assume that K is the smallest number possessing the above property.

Consider a fixed $k \in \{1, \ldots, K\}$. By (6.13) we have that if $\hat{\beta} \in \mathcal{V}(\beta_0)$ then $\forall\ j \in \{1, \ldots, k\}$

$$
\begin{aligned}
|\overline{\gamma_j}(\hat{\beta}) - \overline{\gamma_j}(\beta_0)| &= \left| \frac{1}{n} \sum_{i=1}^{n} (\gamma_j(X_i, \hat{\beta}) - \gamma_j(X_i, \beta_0)) \right| \\
&\leq \sqrt{5} j^{3/2} c_1 \parallel \hat{\beta} - \beta_0 \parallel \frac{1}{n} \sum_{i=1}^{n} (1 + \parallel X_i \parallel) \ . \qquad (4.6)
\end{aligned}
$$

Since the second moments of X_{11} and X_{21} exist the law of large numbers implies

$$\hat{\beta} \xrightarrow{\mathcal{P}} \beta_0 \qquad (4.7)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} (1 + \parallel X_i \parallel) \xrightarrow{\mathcal{P}} 1 + \mathbf{E}_{\mathcal{P}} \parallel X \parallel \ . \qquad (4.8)$$

From the law of large numbers we have also that

$$\bar{\gamma}_j(\beta_0) \xrightarrow{\mathcal{P}} \mathbf{E}_{\mathcal{P}} \gamma_j(X, \beta_0) \ . \qquad (4.9)$$

Combination of (4.6), (4.7), (4.8) and (4.9) yields

$$\bar{\gamma}_j(\hat{\beta}) \xrightarrow{\mathcal{P}} \mathbf{E}_{\mathcal{P}} \gamma_j(X, \beta_0) \ . \qquad (4.10)$$

Since for $k < K$ $\mathbf{E}_{\mathcal{P}} \gamma_k(X, \beta_0) = 0$ then by (4.10) we have

$$\forall\, k < K \quad \parallel T_{nk}(\hat{\beta}) \parallel^2 \xrightarrow{\mathcal{P}} 0 \ . \qquad (4.11)$$

Moreover by (4.5) and (4.11)

$$\parallel T_{nK}(\hat{\beta}) \parallel^2 \xrightarrow{\mathcal{P}} (\mathbf{E}_{\mathcal{P}} \gamma_K(X, \beta_0))^2 > 0 \ . \qquad (4.12)$$

Let n_0 be the smallest natural number such that $d(n) \geq K$ for all $n > n_0$. By (4.11) and (4.12) we have that for all $n > n_0$ and $k < K$

$$
\begin{aligned}
&\mathcal{P}(S(1) = k) \\
&\leq \mathcal{P}\left(\parallel T_{nK}(\hat{\beta}) \parallel^2 - \parallel T_{nk}(\hat{\beta}) \parallel^2 \leq \frac{(K-k)\log n}{n} \right) \quad \rightarrow \quad 0 \quad \text{as} \quad n \rightarrow \infty \ .
\end{aligned}
$$

Thus

$$\lim_{n \to \infty} \mathcal{P}(S(1) \geq K) = 1 \qquad (4.13)$$

and

$$\lim_{n \to \infty} \mathcal{P}(S(5) \geq K) = 1 \ . \qquad (4.14)$$

Since for each k $R_k(\beta)$ is nonnegative definite, we have

$$W_k \geq n \parallel T_{nk}(\hat{\beta}) \parallel^2 \ .$$

Thus (4.3) and (4.4) easily follow from (4.12), (4.13) and (4.14). □

Observe that when the covariance matrix Σ of X_{11} and X_{21} is singular then the sample covariance matrix $\hat{\Sigma}$ is singular with probability one and the null hypothesis is almost surely rejected. Thus Theorems 4.1 and 3.4 imply

Corollary 4.2 If $\lim_{n\to\infty} d(n) = \infty$ and $d(n) = o(n^c)$ with $c < 1/16$ then tests based on $W_{S(1)}$ and $W_{S(5)}$ are consistent against each alternative to bivariate normality satisfying (4.1).

5. Simulations

5.1. Introduction

In this section we present the results of simulation study in which we compared the empirical powers of some data driven smooth tests for bivariate normality with simulated powers of other tests.

All the computations were performed with the double-precision arithmetic on supercomputers of WCSS Wrocław and on a SUN10 station in the Institute of Mathematics of the Technical University of Wrocław. Programs and procedures were written in the C programming language by Krzysztof Bogdan under KBN 350 044 Grant. His kind help and useful remarks are gratefully appreciated.

For simulation purposes we used a pseudo random generator of the substract-with-borrow type, described in Marsaglia and Zaman (1991). This generator, with 39 seed values and lags 39 and 25, produces floating point numbers from [0,1] with 24-bit fractions. It has an immense period about 10^{279}.

We simulated the critical values and powers of tests for the sample sizes $n = 25, 50$ and 100, upon 10000 runs in each case. The significance level is $\alpha = 0.05$.

5.2. Test statistics under study

Apart from tests based on $W_{S(1)}$ and $W_{S(5)}$ in the simulation study we considered also two other versions of data driven smooth test for bivariate normality.

First of these tests is obtained by an application of an another modification of the Schwarz Criterion for choosing k in the test statistic W_k. Following Inglot, Kallenberg and Ledwina (1995) we applied the selection rule given by

$$\mathbf{S1} = \min\{k: \ 1 \le k \le d(n),\ W_k - k\log n \ge W_j - j\log n,\ j = 1, \ldots, d(n)\}. \qquad (5.1)$$

As before, to cover the cases when the sample covariance matrix $\hat{\Sigma}$ is singular set $W_{\mathbf{S1}} = \infty$ when $|\hat{\Sigma}| = 0$. The test rejects the null hypothesis for large values of $W_{\mathbf{S1}}$.

The last of the considered data driven smooth tests for bivariate normality results from an application of the Schwarz selection rule of a form (2.18) for the choice of the number of components in the smooth test for bivariate normality based on Hermite polynomials. In this case the test statistic $H_{\tilde{S}}$ is defined as follows.

For each natural i and j let $\mathrm{h}_{ij}(x)$ be the function on $\mathbf{R}^2$ defined by

$$\mathrm{h}_{ij}(x) = \hat{\mathrm{h}}_i(x_1)\hat{\mathrm{h}}_j(x_2) \ , \quad x = (x_1, x_2)^T \in \mathbf{R}^2 \ ,$$

where $\hat{\mathrm{h}}_i$ denotes the i-th normalized Hermite polynomial.
Let us order the set of functions $\mathcal{H} = \{\mathrm{h}_{ij} \ ; \ \ i + j \ge 3\}$ into the sequence $h = (h_1, h_2, \ldots)$

according to the rules i), ii), iii), specified in Section 2.1 . Thus we have

$$h_1 = \mathrm{h}_{30}, \quad h_2 = \mathrm{h}_{03}, \quad h_3 = \mathrm{h}_{21}, \dots \; .$$

The smooth test for bivariate normality related to Hermite polynomials is based on the statistic

$$H_k = n \sum_{j=1}^{k} \left(\frac{1}{n} \sum_{i=1}^{n} h_j(L(\hat{\beta})(X_i - \hat{\mu})) \right)^2 ,$$

with $L(\beta)$ defined by (2.8).

The selection rule $\tilde{S}$ for choosing the number of components in H_k is given by

$$\tilde{S} = \min\{k : \; 1 \le k \le d(n), \; H_k - k \log n \ge H_j - j \log n, \; j = 1, \dots, d(n)\}. \tag{5.2}$$

Again, to cover the cases when $\hat{\Sigma}$ is singular set $H_{\tilde{S}} = \infty$ when $|\hat{\Sigma}| = 0$.

The data driven smooth test for bivariate normality based on Hermite polynomials rejects the null hypothesis for large values of $H_{\tilde{S}}$.

Observe that in case of the test based on Hermite polynomials two versions of the Schwarz selection rule, computed according to the formulas (2.18) and (5.1), are equivalent.

Applying the methods of proofs from Kallenberg, Ledwina and Rafajłowicz (1996) it is easy to show that if $d(n) = o(\log n)$ then

$$P_{\beta_0}(\tilde{S} \ge 2) \quad \to \quad 0 \quad \text{as} \quad n \to \infty$$

and, that under the null hypothesis,

$$H_{\tilde{S}} \xrightarrow{\mathcal{D}} \chi_1^2 \; . \tag{5.3}$$

The restriction on the range of $d(n)$ is stronger than in case of tests exploiting the system of Legendre polynomials. A reason is the unboundedness of Hermite polynomials.

While carrying out simulations for data driven smooth tests for bivariate normality we observed that for sample sizes considered in this paper ($n \le 100$) increasing $d(n)$ above the level $d(n) = 15$ has no impact on the powers of these tests. Therefore we fixed $d(n) = 15$ for all the analyzed tests and sample sizes.

In the simulation study we considered also the following test statistics:

$b_{1,2}$	-	Mardia's (1970) measure of skewness,
$b_{2,2}$	-	Mardia's (1970) measure of kurtosis,
W^*	-	the generalization of Shapiro-Wilk's statistic proposed by Malkovich and Afifi (1973),
M	-	Csörgő's (1986) maximal deviation of the sample characteristic function from the hypothetical one,
K	=	$\hat{U}_3^2 + \hat{U}_4^2$, where $\hat{U}_3^2$ and $\hat{U}_4^2$ are Koziol's (1987) smooth skewness and kurtosis statistics , respectively,
ISE	-	Bowman and Foster's (1993) integrated squared error statistic, exploiting the kernel density estimators.

For the sample sizes $n = 25$ and $n = 50$ the coefficients $(a_1, \dots, a_n)$, needed to calculate the generalization of Shapiro-Wilks statistic W^*, were taken from Shapiro and Wilk (1965).

For the sample size $n = 100$ we used the approximated values $(b_1, \ldots, b_{100})$ of these coefficients calculated according to the formula

$$b_i = \frac{m_i}{\sqrt{\sum_{i=1}^{100} m_i^2}} ,$$

where m_i is the expected value of the i-th standard normal order statistic [cf. Shapiro and Francia (1972)]. The values of m_i, $1 \le i \le 100$ were taken from Harter (1961).

5.3. Critical values

The critical values of tests based on $b_{1,2}$ and $b_{2,2}$ were taken from Mardia (1974) and the critical values for ISE from Bowman and Foster (1993). The critical values for all other tests considered in this paper were obtained by simulations and can be found in Tables 1, 2 and 4 below. In Table 3 we present empirical distributions under the null hypothesis of the applied selection rules.

Table 1 Empirical critical values of $W_{S(1)}$, $W_{S(5)}$, $W_{\mathbf{S}1}$ and $H_{\tilde{S}}$, for $n = 25$ and $n = 50$, based on 10000 samples in each case; $\alpha = 0.05$.

	$n = 25$				$n = 50$			
$d(n)$	$W_{S(1)}$	$W_{S(5)}$	$W_{\mathbf{S}1}$	$H_{\tilde{S}}$	$W_{S(1)}$	$W_{S(5)}$	$W_{\mathbf{S}1}$	$H_{\tilde{S}}$
1	3.6661	—	3.6661	3.0817	3.8177	—	3.8177	3.3493
2	3.6661	—	5.3286	4.5834	3.8177	—	5.3202	5.0110
3	3.6661	—	5.9020	5.2100	3.8177	—	5.6511	5.4201
4	3.6661	—	6.1687	5.4534	3.8177	—	5.8605	5.5945
5	3.6661	10.1912	6.3573	5.5026	3.8177	10.8138	5.9177	5.6979
6	3.6661	10.8179	6.3957	5.5046	3.8177	11.3326	5.9570	5.7211
7	3.6661	11.1939	6.4285	5.5654	3.8177	11.6000	5.9714	5.7677
8	3.6661	11.8640	6.4285	5.6148	3.8177	11.7719	5.9739	5.8395
9	3.6661	12.0510	6.4305	5.6632	3.8177	11.8168	5.9739	5.8570
10	3.6661	12.0916	6.4305	5.6632	3.8177	11.8210	5.9848	5.8880
11	3.6661	12.0951	6.4305	5.6632	3.8177	11.8210	5.9848	5.8880
12	3.6661	12.1519	6.4305	5.6632	3.8177	11.8211	5.9848	5.8880
13	3.6661	12.1568	6.5645	5.6632	3.8177	11.8211	6.0169	5.9060
14	3.6661	12.1568	6.5645	5.6632	3.8177	11.8211	6.0169	5.9171
15	3.6661	12.1568	6.5645	5.6632	3.8177	11.8211	6.0169	5.9594

Table 2 Empirical critical values of $W_{S(1)}$, $W_{S(5)}$, $W_{\mathbf{S1}}$ and $H_{\tilde{S}}$, for $n = 100$, based on 10000 samples in each case; $\alpha = 0.05$.

	$n = 100$			
$d(n)$	$W_{S(1)}$	$W_{S(5)}$	$W_{\mathbf{S1}}$	$H_{\tilde{S}}$
1	3.8359	—	3.8359	3.7077
2	3.8359	—	5.2590	5.1698
3	3.8359	—	5.4485	5.3094
4	3.8359	—	5.4744	5.3493
5	3.8359	10.9370	5.4744	5.3762
6	3.8359	11.2064	5.4810	5.4144
7	3.8359	11.3478	5.4810	5.5179
8	3.8359	11.3734	5.4810	5.5624
9	3.8359	11.3763	5.4810	5.5694
10	3.8359	11.3763	5.4810	5.5921
11	3.8359	11.3763	5.4810	5.6245
12	3.8359	11.3763	5.4810	5.6497
13	3.8359	11.3763	5.4904	5.6795
14	3.8359	11.3763	5.4904	5.7010
15	3.8359	11.3763	5.4904	5.7225

Table 3 Counts of $\{S(1) = s\}$, $\{S(5) = s\}$, $\{\mathbf{S1} = s\}$ and $\{\tilde{S} = s\}$ under $\mathcal{H}_0$, $d(n) = 15$, based on 10000 samples in each case.

	$n = 25$				$n = 50$				$n = 100$			
s	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$
1	10000	–	9108	9346	10000	–	9383	9428	10000	–	9619	9532
2	0	–	602	371	0	–	456	340	0	–	322	293
3	0	–	136	97	0	–	78	62	0	–	43	40
4	0	–	59	28	0	–	45	24	0	–	11	10
5	0	8975	35	15	0	9420	14	13	0	9654	2	6
6	0	677	8	15	0	455	6	9	0	290	2	8
7	0	205	4	20	0	86	2	11	0	45	0	2
8	0	93	2	21	0	28	1	3	0	10	0	2
9	0	33	1	32	0	9	0	5	0	1	0	0
10	0	9	0	2	0	1	0	15	0	0	0	14
11	0	2	0	6	0	0	0	15	0	0	0	12
12	0	4	0	9	0	1	0	17	0	0	0	17
13	0	2	41	11	0	0	15	12	0	0	1	17
14	0	0	4	8	0	0	0	12	0	0	0	4
15	0	0	0	19	0	0	0	34	0	0	0	43

From Table 1 it is seen that simulated critical values of $W_{S(1)}$ and $W_{S(5)}$ for $n \geq 50$ and $d(n) = 15$ are close to their asymptotic values [equal to 3.841 and 11.070 , respectively].

Note that the small values of first components of $T_{nk}(\hat{\beta})$ [cf. Section 2.3] do not prevent the selection rule **S1** from choosing values from the set $\{2, 3, 4, 5\}$. This is due to a fact that **S1** depends on the vector $T_{nk}(\hat{\beta})$ via the test statistic W_k.

Comparing $\tilde{S}$ with other selection rules it can be observed that $\tilde{S}$ converges to 1 even slower than **S1**. We believe that such behavior of $\tilde{S}$ is due to unboundedness of Hermite polynomials.

Table 4 Empirical critical values of K, M and W^* based on 10000 samples in each case; $\alpha = 0.05$.

	$n = 25$	$n = 50$	$n = 100$
K	15.8416	17.9338	19.2285
M	0.7084	0.6555	0.6041
W^*	0.8958	0.9443	0.966

5.4. Alternative distributions

The list of alternatives considered in this paper contains among others all the alternatives previously analyzed by Malkovich and Afifi (1973) and Rayner and Best (1989).

The first group of alternatives contains mixtures of bivariate normal distributions with density functions given by

$$h(x, p, \beta) = pf(x, \beta_0) + (1 - p)f(x, \beta) \ , \quad x \in \mathbf{R}^2 \ ,$$

where $p \in (0, 1)$, $f(x, \beta)$ is as in (2.2) and $\beta_0 = (0, 0, 1, 1, 0)^T$. To be specific we consider the alternatives

notation	p	β^T
M(1)	0.5	(3,3,1,1,0)
M(2)	0.25	(3,3,1,1,0)
M(3)	0.5	(0,0,3,3,0)
M(4)	0.5	(0,0,1,1,0.9)
M(5)	0.75	(3,3,3,3,2.7)
M(6)	0.25	(3,3,3,3,2.7)

In the second group of alternatives the random vectors $X_i = (X_{1i}, X_{2i})^T$, $i = 1, \ldots, n$, are such that X_{1i} and X_{2i} are independently and identically distributed. In the Table 5 below we specify the alternatives by describing the distribution of X_{1i}. Here Z denotes a $N(0, 1)$ r.v., R is a uniform r.v. on [0,1] and $\phi_{\mu,\sigma}(x)$ is the density of normal distribution.

Table 5 Marginal distributions of the alternatives from the second group

notation	density or definition of X_{1i}
exponential(1)	$e^{-x}, \quad x > 0$
$LN(0,1)$	$(\sqrt{2\pi}x)^{-1}\exp\{\frac{-\ln^2 x}{2}\}, \quad x > 0$
uniform(0,1)	$1, \quad 0 < x < 1$
logistic	$e^x(1+e^x)^{-2}$
χ_k^2	$\sqrt{2}\Gamma(k/2)^{-1}x^{k/2-1}e^{-\frac{1}{2}x}, \quad x > 0$
Student(k)	$\frac{\Gamma[(k+1)/2]}{\sqrt{k\pi}\Gamma(k/2)}(1+\frac{x^2}{k})^{-(k+1)/2}$
$S(\alpha)$	symmetric stable with an index of stability α
Gamma(p;q)	$q^{-p}\{\Gamma(p)\}^{-1}x^{p-1}e^{-x/q}, \quad x > 0$
Beta(p;q)	$x^{p-1}(1-x)^{q-1}\{B(p,q)\}^{-1}, \quad 0 \le x \le 1$
SU(d)	$X_{1i} = \sinh\frac{Z}{d}$
TU(l)	$X_{1i} = R^l - (1-R)^l$
SC(p;d)	$p\phi_{0,d}(x) + (1-p)\phi_{0,1}(x)$
LC(p;m)	$p\phi_{m,1}(x) + (1-p)\phi_{0,1}(x)$
$LSC(p_1, p_2; m_1, m_2)$	$p_1\phi_{m_1,1}(x) + p_2\phi_{m_2,1}(x) + (1-p_1-p_2)\phi_{0,1}(x)$

In the third group of alternatives the random vectors $X_i = (X_{1i}, X_{2i})^T$, $i = 1, \ldots, n$, are such that X_{1i} and X_{2i} are independently distributed and X_{2i} is a standard normal r.v. To denote these alternatives we used the symbol A$\times N$, where A defines the distribution of X_{1i} according to Table 5.

Moreover we analyzed the alternative denoted by $\mathcal{X}$. $X_i = (X_{1i}, X_{2i})^T$ is distributed according to $\mathcal{X}$ if $R_i^2 = X_{1i}^2 + X_{2i}^2$ has a chi-square distribution with 2 degrees of freedom and the distribution of the angle α between the vector $(X_{1i}, X_{2i}) \in \mathbf{R}^2$ and the axis $Y = 0$ is given by

$$P(\alpha = 0) = P(\alpha = \pi/2) = P(\alpha = \pi) = P(\alpha = 3/2\pi) = 1/4 \ .$$

Observe that the alternative $\mathcal{X}$ is not absolutely continuous with respect to Lebesgue measure in $\mathbf{R}^2$ and it is very different from bivariate normal distribution. We have included this alternative to demonstrate shortcomings of tests based on some measures of skewness, like the test based on $b_{1,2}$, and also of those whose test statistics depend on the sample $X_1, \ldots, X_n$ via $S_1, \ldots, S_n$, where $S_i = (X_i - \hat{\mu})^T\hat{\Sigma}^{-1}(X_i - \hat{\mu})$, like the test based on $b_{2,2}$.

5.5. Power simulations

Table 6 Estimated powers based on 10000 samples in each case; $\alpha = 0.05$, n=25.

alternatives	$b_{1,2}$	$b_{2,2}$	K	W^*	M	ISE	$W_{S(1)}$	$W_{S(5)}$	W_{S1}	$H_{\tilde{S}}$
M(1)	2	18	3	3	2	35	5	12	7	2
M(2)	18	8	10	19	15	55	6	32	20	15
M(3)	18	14	18	15	16	12	9	12	13	16
M(4)	18	14	21	20	27	17	10	16	13	16
M(5)	61	17	41	56	34	63	10	61	49	48
M(6)	18	9	15	15	14	13	11	15	11	15
exponential(1)	90	55	78	85	58	94	85	94	96	91
$LN(0,1)$	99	85	95	97	86	99	96	99	100	99
uniform(0,1)	0	47	0	1	0	26	3	18	23	0
logistic	19	15	19	17	15	11	11	15	17	19
χ^2_{10}	33	16	26	29	17	30	30	32	37	34
Student(4)	40	39	43	37	36	30	22	39	41	41
Student(6)	25	22	27	23	22	15	14	22	24	26
$S(1.8)$	40	37	42	39	36	30	25	37	39	41
Gamma(2;1)	66	33	53	59	36	67	59	70	76	67
Beta(2.5;1.5)	3	13	2	5	2	17	11	13	12	3
SU(1.5)	33	31	36	31	28	23	18	32	35	35
TU(0.7)	0	34	0	1	0	16	3	11	12	0
SC(0.1;3)	48	46	53	47	43	34	25	46	49	51
LC(0.2;3)	23	8	14	20	9	41	32	41	39	25
LC(0.5;3)	1	30	0	2	1	21	4	16	17	1
$LSC(0.5, 0.5; -6, 6)$	39	38	41	40	38	28	21	37	39	40
Student(6) $\times N$	16	12	17	15	13	11	5	14	15	16
$LSC(0.04, 0.01; 3, 6) \times N$	31	23	30	30	27	20	30	27	27	31
$LSC(0.08, 0.02; -2, 4) \times N$	18	14	20	18	16	11	16	15	17	18
$\mathcal{X}$	11	6	20	62	7	66	55	99	93	16

Table 7 Estimated powers based on 10000 samples in each case; $\alpha = 0.05$, n=50.

alternatives	$b_{1,2}$	$b_{2,2}$	K	W^*	M	ISE	$W_{S(1)}$	$W_{S(5)}$	W_{S1}	$H_{\tilde{S}}$
M(1)	2	36	3	4	2	80	5	36	14	2
M(2)	39	8	13	26	18	92	7	64	49	28
M(3)	24	25	27	19	28	16	9	19	17	23
M(4)	22	23	35	31	51	31	10	26	17	22
M(5)	96	22	79	75	59	94	24	92	87	88
M(6)	35	12	28	25	23	21	14	26	17	27
exponential(1)	100	82	99	99	87	100	99	100	100	100
$LN(0,1)$	100	99	100	100	99	100	100	100	100	100
uniform(0,1)	0	92	0	3	0	63	3	75	66	0
logistic	25	26	31	22	26	14	11	25	23	29
χ^2_{10}	68	26	50	56	30	52	58	66	74	68
Student(4)	56	65	69	56	62	48	28	65	62	64
Student(6)	35	39	44	33	39	22	15	38	36	41
$S(1.8)$	60	62	66	59	61	44	36	59	58	64
Gamma(2;1)	96	55	85	91	62	94	92	97	99	96
Beta(2.5;1.5)	5	27	1	14	1	38	22	37	33	6
SU(1.5)	47	55	59	45	52	37	20	55	52	55
TU(0.7)	0	80	0	3	0	43	2	48	38	0
SC(0.1;3)	67	75	79	69	74	52	32	72	71	77
LC(0.2;3)	54	7	24	40	11	75	65	79	80	55
LC(0.5;3)	0	67	0	5	0	54	4	55	43	0
$LSC(0.5, 0.5; -6, 6)$	60	63	64	63	61	39	30	59	59	64
Student(6) $\times N$	22	20	28	22	24	13	5	23	21	26
$LSC(0.04, 0.01; 3, 6) \times N$	53	42	53	51	48	30	49	48	47	54
$LSC(0.08, 0.02; -2, 4) \times N$	26	24	33	27	29	13	17	26	26	30
$\mathcal{X}$	9	5	58	92	15	100	100	100	100	17

Table 8 Estimated powers based on 10000 samples in each case; $\alpha = 0.05$, n=100.

alternatives	$b_{1,2}$	$b_{2,2}$	K	W^*	M	ISE	$W_{S(1)}$	$W_{S(5)}$	W_{S1}	$H_{\tilde{S}}$
M(1)	1	60	4	2	79	100	6	89	44	2
M(2)	77	8	33	22	25	100	27	96	89	65
M(3)	24	45	40	33	50	27	10	32	22	29
M(4)	22	40	60	59	83	61	10	47	25	33
M(5)	100	32	100	81	87	100	82	100	100	100
M(6)	62	17	54	42	39	42	19	52	28	48
exponential(1)	100	98	100	100	99	100	100	100	100	100
$LN(0,1)$	100	100	100	100	100	100	100	100	100	100
uniform(0,1)	0	100	0	0	2	98	4	100	99	0
logistic	29	46	48	41	43	22	12	43	32	40
χ^2_{10}	96	43	83	80	49	84	89	96	98	96
Student(4)	68	90	90	85	88	74	37	90	84	85
Student(6)	44	65	67	61	63	36	17	62	51	60
$\mathcal{S}(1.8)$	77	86	88	85	85	64	51	82	78	86
Gamma(2;1)	100	81	100	99	99	100	100	100	100	100
Beta(2.5;1.5)	17	54	3	10	1	76	46	90	73	23
SU(1.5)	57	83	82	76	80	62	25	83	74	76
TU(0.7)	0	99	0	0	1	86	2	97	87	0
SC(0.1;3)	78	95	96	93	95	75	42	92	89	94
LC(0.2;3)	91	6	57	53	15	98	95	99	99	92
LC(0.5;3)	0	94	0	2	1	93	7	97	87	0
$LSC(0.5, 0.5; -6, 6)$	78	85	86	86	84	53	38	79	77	86
Student(6) $\times N$	26	34	42	40	40	19	6	38	28	37
$LSC(0.04, 0.01; 3, 6) \times N$	77	67	78	77	73	48	73	74	72	80
$LSC(0.08, 0.02; -2, 4) \times N$	33	43	54	53	52	21	20	45	39	48
$\mathcal{X}$	6	5	100	100	85	100	100	100	100	42

Table 9 Counts of $\{S(1)=s\}$, $\{S(5)=s\}$, $\{\mathbf{S1}=s\}$ and $\{\tilde{S}=s\}$ under the alternative L$N(0,1)$, $d(n)=15$, based on 10000 samples in each case.

	$n=25$				$n=50$				$n=100$			
s	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$
1	1968	–	99	260	55	–	0	3	0	–	0	0
2	0	–	1535	1635	0	–	169	220	0	–	0	2
3	61	–	980	189	3	–	154	22	0	–	0	0
4	18	–	1561	5	0	–	242	0	0	–	0	0
5	0	89	216	406	0	1	18	95	0	0	0	2
6	143	436	127	865	15	19	27	447	0	0	1	65
7	6657	7791	44	438	8266	8311	15	161	7329	7329	1	19
8	752	1074	37	167	700	708	6	28	501	501	0	2
9	7	17	2	17	3	3	0	4	0	0	0	3
10	134	139	24	151	184	184	27	114	50	50	0	31
11	129	131	89	347	247	247	276	261	302	302	26	62
12	49	53	12	693	33	33	5	523	20	20	0	200
13	2	5	2458	380	0	0	3572	482	1	1	2055	317
14	51	51	264	1443	71	71	273	1945	82	82	195	1443
15	32	32	2552	3004	423	423	5216	5695	1715	1715	7722	7854

Table 10 Counts of $\{S(1)=s\}$, $\{S(5)=s\}$, $\{\mathbf{S1}=s\}$ and $\{\tilde{S}=s\}$ under the alternative 'uniform(0,1)', $d(n)=15$, based on 10000 samples in each case.

	$n=25$				$n=50$				$n=100$			
s	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$
1	9985	–	7476	9950	9970	–	3411	9992	9836	–	107	9999
2	0	–	266	43	0	–	34	8	0	–	0	1
3	0	–	789	5	0	–	906	0	0	–	184	0
4	0	–	1390	1	0	–	5565	0	0	–	9617	0
5	0	8301	36	0	0	8604	53	0	0	8319	71	0
6	3	864	16	0	2	675	21	0	1	484	18	0
7	2	403	9	0	2	262	1	0	0	93	1	0
8	0	152	11	0	1	41	9	0	0	12	2	0
9	1	31	3	1	0	10	0	0	0	3	0	0
10	2	83	1	0	1	83	0	0	0	123	0	0
11	3	103	0	0	20	261	0	0	135	831	0	0
12	3	57	0	0	4	57	0	0	19	127	0	0
13	1	5	3	0	0	7	0	0	1	8	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	0	0	0	0	0	0	0	0	0	0

Table 11 Counts of $\{S(1)=s\}$, $\{S(5)=s\}$, $\{\mathbf{S1}=s\}$ and $\{\tilde{S}=s\}$ under the alternative M(2), $d(n)=15$, based on 10000 samples in each case.

	$n=25$				$n=50$				$n=100$			
s	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$	$S(1)$	$S(5)$	**S1**	$\tilde{S}$
1	9946	–	6027	7734	9772	–	3682	6181	7529	–	957	2990
2	0	–	3291	1473	0	–	5656	2672	0	–	7429	4576
3	0	–	302	235	0	–	241	311	0	–	197	465
4	0	–	121	308	0	–	88	640	0	–	75	1786
5	0	6212	82	35	0	3772	43	48	0	591	21	63
6	0	327	15	28	0	78	4	23	0	3	0	18
7	29	1734	12	25	64	2209	28	12	153	1340	19	12
8	6	374	20	21	15	338	10	10	26	190	5	2
9	9	1133	52	53	132	3385	175	25	2148	7633	1184	0
10	7	129	5	5	13	154	10	10	58	167	44	6
11	1	41	0	4	2	44	6	5	12	47	9	3
12	1	13	0	11	1	6	0	14	1	5	1	11
13	1	30	53	14	1	7	44	16	6	20	56	24
14	0	5	19	17	0	5	13	9	0	4	3	18
15	0	2	1	37	0	2	0	24	0	0	0	26

Tables 6, 7 and 8 show that though the test based on $W_{S(1)}$ has nice asymptotic properties the power of $W_{S(1)}$ for small and moderate sample sizes ($n \leq 100$) is rather poor. This might again be explained by the fact that in a result of an estimation of a vector of means and a covariance matrix the first five components of the vector $T_{nk}(\hat{\beta})$ [see (2.14)] are very small. Therefore the selection rule $S(1)$ very rarely chooses values from the set $\{2,3,4,5\}$ [see Tables 9, 10 and 11] and for small sample sizes it has some difficulties in "jumping over this gap" to the dimension 6. It can be observed that the ability of $S(1)$ to choose values larger than 1 increases with an increase of the sample size [as it was proved by theoretical considerations] but for most of the considered alternatives the powers of $W_{S(1)}$ are not satisfactory even for sample size $n = 100$.

As it could be expected $W_{\mathbf{S1}}$ performs much better than $W_{S(1)}$. However its powers are still worse than powers of $W_{S(5)}$ [see alternative M(1), M(4) or M(6)]. Also the data driven smooth test for bivariate normality based on Hermite polynomials $H_{\tilde{S}}$ performs worse than $W_{S(5)}$. $H_{\tilde{S}}$ has especially great difficulties in detecting symmetric alternatives [M(1),'uniform(0,1)', TU(0.7) or LC(0.5,3)]. This is due to a fact that first four components of this statistic correspond to bivariate skewness measures. Thus to detect symmetric alternatives the selection rule $\tilde{S}$ would have to reach the dimension 5, what seems to be quite difficult under the considered sample sizes.

As it could be expected tests based on the statistics corresponding to the moments of some fixed order, as $b_{1,2}$, $b_{2,2}$ or K, are not able to detect certain alternatives to bivariate normality [see the alternatives M(1), 'uniform(0,1)', TU(0.7), LC(0.5,3) for $b_{1,2}$ and K and $M(2)$ and $\mathcal{X}$ for $b_{2,2}$]. Note that the test based on $b_{2,2}$ is equivalent to the test based on Shapiro-Wilk-Stephens statistic, recommended recently in a simulation study by Versluis (1996). Hence the above remark applies also to the second of these tests.

The results of simulations show that for the sample sizes $n \leq 100$ also tests based on W^* and M miss some symmetric alternatives [see again 'uniform(0,1)', TU(0.7) or LC(0.5,3)].

Tables 6–8 show that Bowman and Foster (1993) test based on ISE is the only test, except that based on $W_{S(5)}$, which has comparatively good power for all the considered alternatives. Let us mention that the results of simulations reported in Bowman and Foster (1993) show that "the integrated squared error statistic [ISE] has a very good power compared to sample entropy", also considered by Bowman and Foster (1993), to the radial distance test of Koziol (1982) and the omnibus test of Koziol (1983). From Tables 6–8 it is seen that the test based on ISE performs especially well for the alternatives obtained as the mixtures of bivariate normal distributions [see M(1) or M(4)] but in case of other alternatives its power is usually worse than power of $W_{S(5)}$.

Thus we conclude that the data driven smooth test for bivariate normality based on the statistic $W_{S(5)}$ can be recommended as an omnibus test for bivariate normality.[1]

6. Appendix

6.1. Independence of the matrix $\mathbf{R_k}(\beta)$ from β

Lemma 6.1 For each $\beta \in \Omega$ we have $R_k(\beta) = R_k(\beta_0)$, where $\beta_0 = (0,0,1,1,0)^T$.

Proof. The partial derivatives of $\log f(x,\beta)$ with respect to $\beta_1,\ldots,\beta_5$ are given by

$$\begin{pmatrix} \frac{\partial \log f(x,\beta)}{\partial \beta_1} \\ \frac{\partial \log f(x,\beta)}{\partial \beta_2} \end{pmatrix} = \Sigma^{-1}(x-\mu) \ , \quad x = (x_1,x_2)^T \ ,$$

$$\begin{aligned} \frac{\partial \log f(x,\beta)}{\partial \beta_3} &= \frac{1}{2|\Sigma|}\left(-\beta_4 + \beta_4(x-\mu)^T\Sigma^{-1}(x-\mu) - (x_2-\beta_2)^2\right) \ , \\ \frac{\partial \log f(x,\beta)}{\partial \beta_4} &= \frac{1}{2|\Sigma|}\left(-\beta_3 + \beta_3(x-\mu)^T\Sigma^{-1}(x-\mu) - (x_1-\beta_1)^2\right) \ , \\ \frac{\partial \log f(x,\beta)}{\partial \beta_5} &= \frac{1}{|\Sigma|}\left(\beta_5 - \beta_5(x-\mu)^T\Sigma^{-1}(x-\mu) + (x_1-\beta_1)(x_2-\beta_2)\right) \ , \end{aligned}$$

where μ and Σ are as in (2.3) and (2.4) , respectively.
Thus it is easy to check that

$$\begin{pmatrix} \frac{\partial \log f(X,\beta)}{\partial \beta_1} \\ \\ \vdots \\ \\ \frac{\partial \log f(X,\beta)}{\partial \beta_5} \end{pmatrix} = \tilde{L}(\beta) \begin{pmatrix} Y_1 \\ Y_2 \\ \frac{1}{2}(Y_1^2-1) \\ \frac{1}{2}(Y_2^2-1) \\ Y_1Y_2 \end{pmatrix} \ , \qquad (6.1)$$

[1] A ready-to-use program calculating empirical p-values of $W_{S(5)}$ for a given sample is available by *anonymous* ftp at banach.im.pwr.wroc.pl in the directory pub/contrib/mbogdan/binor .

where Y_1 and Y_2 are the components of the random vector $Y = (Y_1, Y_2)^T = L^T(X - \mu)$, $L = L(\beta)$ [cf. (2.8)], and

$$\tilde{L}(\beta) = \begin{pmatrix} l_1 & 0 & 0 & 0 & 0 \\ l_2 & l_3 & 0 & 0 & 0 \\ 0 & 0 & l_1^2 & 0 & 0 \\ 0 & 0 & l_2^2 & l_3^2 & l_2 l_3 \\ 0 & 0 & 2l_1 l_2 & 0 & l_1 l_3 \end{pmatrix} ,$$

with l_1, l_2, l_3 given by (2.8).

Observe that $\tilde{L}(\beta_0) = \mathbf{I}_{5\times5}$ and $L(\beta_0) = \mathbf{I}_{2\times2}$, where $\mathbf{I}_{k\times k}$ is the $k \times k$ identity matrix. Note also that $\forall\, j \in \mathbf{N}$

$$\gamma_j(X, \beta) = \gamma_j(Y, \beta_0) \ . \tag{6.2}$$

Moreover observe that if the random vector X is bivariate normal with the vector of parameters β then Y is also binormal with the vector of parameters β_0. Therefore by (6.1) we get

$$\begin{aligned} I_k(\beta) &= \mathbf{Cov}_{0,\beta} \left\{ \frac{\partial \log f(X,\beta)}{\partial \beta_t}, \gamma_j(X,\beta) \right\}_{t=1,\dots,5,\ j=1,\dots,k} \\ &= \tilde{L}(\beta)\mathbf{Cov}_{0,\beta_0} \left\{ \frac{\partial \log f(Y,\beta)}{\partial \beta_t} |_{\beta=\beta_0}\ , \gamma_j(Y,\beta_0) \right\}_{t=1,\dots,5,\ j=1,\dots,k} \\ &= \tilde{L}(\beta) I_k(\beta_0) \end{aligned} \tag{6.3}$$

and analogously

$$I_{\beta\beta} = \tilde{L}(\beta) I_{\beta_0\beta_0} (\tilde{L}(\beta))^T \ .$$

Thus

$$\begin{aligned} R_k(\beta) &= (I_k(\beta))^T \left(I_{\beta\beta} - I_k(\beta)(I_k(\beta))^T\right)^{-1} I_k(\beta) \\ &= (I_k(\beta_0))^T (\tilde{L}(\beta))^T [\tilde{L}(\beta)(I_{\beta_0\beta_0} - I_k(\beta_0)(I_k(\beta_0))^T)(\tilde{L}(\beta))^T]^{-1} \tilde{L}(\beta) I_k(\beta_0) \\ &= R_k(\beta_0) \ . \end{aligned}$$

□

6.2. Computational formula for the matrix $R_k(\beta)$

In this section we present the straightforward computational formula for the matrix $R_k(\beta)$.

Let

$$c_i = \int_{-\infty}^{\infty} b_i(\Phi(y)) y \phi(y)\, dy$$

and

$$e_i = \int_{-\infty}^{\infty} b_i(\Phi(y)) y^2 \phi(y)\, dy \ ,$$

where Φ and ϕ denote the d.f. and the density of the standard normal distribution, respectively, and b_i is the i-th normalized Legendre polynomial on $[0, 1]$.

Observe that c_i and e_i disappear for even and odd i, respectively. Below we give some values of c_i and e_i computed with *Mathematica*.

$$\begin{array}{ll} c_1 = 0.977205023801135 \ , & c_3 = 0.1830082402700861 \ , \\ c_5 = 0.0816989764273946 \ , & c_7 = 0.04772936798473241 \ , \\ e_2 = 1.232808888123174 \ , & e_4 = 0.5211245854593028 \ , \\ e_6 = 0.3045144697203598 \ , & e_8 = 0.2055889833015625 \ . \end{array} \tag{6.4}$$

For each $j \in \mathrm{N}$ let $i(j)$ and $k(j)$ be such that

$$\forall\, u = (u_1, u_2) \in [0,1]^2 \quad \tilde{\gamma}_j(u) = b_{i(j)}(u_1) b_{k(j)}(u_2) \ ,$$

where $\tilde{\gamma}$ is as in (2.6).

From (6.1) by straightforward calculation we obtain that the elements of the matrix $I_k(\beta_0) = [a_{ij}]_{5\times k}$ are equal to

$$\begin{aligned} a_{1j} = &\ \mathbf{Cov}_{0,\beta_0}\left(\tfrac{\partial \log f(X,\beta)}{\partial \beta_1}, \gamma_j(X,\beta)\right) &= \begin{cases} c_{i(j)} \text{ when } k(j) = 0 \ , \\ 0 \text{ in other cases} \ , \end{cases} \\ a_{2j} = &\ \mathbf{Cov}_{0,\beta_0}\left(\tfrac{\partial \log f(X,\beta)}{\partial \beta_2}, \gamma_j(X,\beta)\right) &= \begin{cases} c_{k(j)} \text{ when } i(j) = 0 \ , \\ 0 \text{ in other cases} \ , \end{cases} \\ a_{3j} = &\ \mathbf{Cov}_{0,\beta_0}\left(\tfrac{\partial \log f(X,\beta)}{\partial \beta_3}, \gamma_j(X,\beta)\right) &= \begin{cases} \tfrac{1}{2} e_{i(j)} \text{ when } k(j) = 0 \ , \\ 0 \text{ in other cases} \ , \end{cases} \\ a_{4j} = &\ \mathbf{Cov}_{0,\beta_0}\left(\tfrac{\partial \log f(X,\beta)}{\partial \beta_4}, \gamma_j(X,\beta)\right) &= \begin{cases} \tfrac{1}{2} e_{k(j)} \text{ when } i(j) = 0 \ , \\ 0 \text{ in other cases} \ , \end{cases} \\ a_{5j} = &\ \mathbf{Cov}_{0,\beta_0}\left(\tfrac{\partial \log f(X,\beta)}{\partial \beta_5}, \gamma_j(X,\beta)\right) &= c_{i(j)} c_{k(j)} \ . \end{aligned} \tag{6.5}$$

Carrying out power simulations for the data driven smooth tests for bivariate normality we observed that in a case of sample sizes n not exceeding 200 increasing the maximal allowable number of components in the auxiliary vector $T_{nk}(\hat{\beta})$ [see (2.14)] above the level $d(n) = 20$ does not make any impact on the powers of these tests. Thus we conclude that in most of the practical situations it is enough to consider $d(n) \leq 20$. Therefore below we specify the elements of the matrix $I_{20}(\beta_0)$.

$$\begin{array}{lll} a_{1\,1} = a_{2\,2} = c_1 \ , & a_{1\,6} = a_{2\,7} = c_3 \ , & a_{1\,15} = a_{2\,16} = c_5 \ , \\ a_{3\,3} = a_{4\,4} = \dfrac{1}{2} e_2 \ , & a_{3\,10} = a_{4\,11} = \tfrac{1}{2} e_4 \ , & a_{5\,5} = c_1^2 \ , \\ & a_{5\,12} = a_{5\,13} = c_1 c_3 \ . & \end{array} \tag{6.6}$$

All other elements of $I_{20}(\beta_0)$ are equal to 0. Observe also that for each $k < 20$ the matrix $I_k(\beta_0)$ consists of k first columns of $I_{20}(\beta_0)$.

From (6.1) we get also that

$$I_{\beta_0\beta_0} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} . \tag{6.7}$$

From (6.5), (6.6) and (6.7) by straightforward calculation we obtain that the 5×5 matrix

$$D_k = \left(I_{\beta_0\beta_0} - I_k(\beta_0)(I_k(\beta_0))^T\right)^{-1}$$

is diagonal and for $k \le 20$ its elements are given by

$$\begin{aligned}
d_{11} &= \begin{cases} (1-c_1^2)^{-1} & \text{for} \quad 1 \le k \le 5 \\ (1-c_1^2-c_3^2)^{-1} & \text{for} \quad 5 < k \le 15 \\ (1-c_1^2-c_3^2-c_5^2)^{-1} & \text{for} \quad 15 < k \le 20 \ , \end{cases} \\
d_{22} &= \begin{cases} 1 & \text{for} \quad k = 1 \\ (1-c_1^2)^{-1} & \text{for} \quad 1 < k \le 6 \\ (1-c_1^2-c_3^2)^{-1} & \text{for} \quad 6 < k \le 16 \\ (1-c_1^2-c_3^2-c_5^2)^{-1} & \text{for} \quad 16 < k \le 20 \ , \end{cases} \\
d_{33} &= \begin{cases} 2 & \text{for} \quad 1 \le k \le 2 \\ (1/2-1/4e_2^2)^{-1} & \text{for} \quad 2 < k \le 9 \\ (1/2-1/4e_2^2-1/4e_4^2)^{-1} & \text{for} \quad 9 < k \le 20 \ , \end{cases} \\
d_{44} &= \begin{cases} 2 & \text{for} \quad 1 \le k \le 3 \\ (1/2-1/4e_2^2)^{-1} & \text{for} \quad 3 < k \le 10 \\ (1/2-1/4e_2^2-1/4e_4^2)^{-1} & \text{for} \quad 10 < k \le 20 \ , \end{cases} \\
d_{55} &= \begin{cases} 1 & \text{for} \quad 1 \le k \le 4 \\ (1-c_1^4)^{-1} & \text{for} \quad 4 < k \le 11 \\ (1-c_1^4-c_1^2c_3^2)^{-1} & \text{for} \quad k = 12 \\ (1-c_1^4-2c_1^2c_3^2)^{-1} & \text{for} \quad 13 < k \le 20 \ . \end{cases}
\end{aligned} \tag{6.8}$$

By Lemma 6.1, for each $\beta \in \Omega$, we can compute the matrix $R_k(\beta)$ as follows

$$R_k(\beta) = R_k(\beta_0) = (I_k(\beta_0))^T D_k I_k(\beta_0) \ .$$

6.3. Bounds for the functions $\gamma_j(x, \beta)$ and their derivatives

To get some bounds for the functions $\gamma_j(x, \beta)$ and their derivatives we shall exploit some well known results on Legendre polynomials. To this end we introduce some notation. As before, for each natural j let $i(j)$, $k(j)$ and $m(j)$ be such that for every $u = (u_1, u_2)^T \in [0,1]^2$ it holds

$$\tilde{\gamma}_j(u) = b_{i(j)}(u_1) b_{k(j)}(u_2) \ ,$$

where $\tilde{\gamma}$ as in (2.6). Set $m(j) = i(j) + k(j)$,

$$\mathcal{A}_j = \sup_{u \in [0,1]^2} \max\{|b'_{i(j)}(u_1) b_{k(j)}(u_2)|, |b_{i(j)}(u_1) b'_{k(j)}(u_2)|\}$$

and

$$\mathcal{M}_j = \sup_{u \in [0,1]^2} \max\{|b''_{i(j)}(u_1) b_{k(j)}(u_2)|, |b'_{i(j)}(u_1) b'_{k(j)}(u_2)|, |b_{i(j)}(u_1) b''_{k(j)}(u_2)|\} \ .$$

The following lemma provides us with some bounds for $\gamma_j(x, \beta)$, $\mathcal{A}_j$ and $\mathcal{M}_j$.

Lemma 6.2 It holds

$$\sup_{x \in \mathbb{R}^2} |\gamma_j(x, \beta)| \le 2\sqrt{j} \ . \tag{6.9}$$

Moreover there exists a positive constant $\hat{c}_1$ such that

$$\mathcal{A}_j \leq \hat{c}_1 j^{3/2} \tag{6.10}$$

and

$$\mathcal{M}_j \leq \hat{c}_1 j^{5/2} \ . \tag{6.11}$$

Proof. Observe that for each natural l in the sequence of functions $\tilde{\gamma}$ there are $l+1$ consecutive functions $\tilde{\gamma}_j$ such that $m(j) = l$. Therefore we easily get that for each natural j

$$m(j)(m(j)+1)/2 \leq j \leq m(j)(m(j)+3)/2 \ . \tag{6.12}$$

For the orthonormal Legendre polynomials on [0,1] it holds [cf. Sansone (1959) pp. 181, 190 and 251]

$$\sup_{z\in[0,1]} |b_j(z)| = \sqrt{2j+1} \ ,$$

$$\sup_{z\in[0,1]} |b_j'(z)| \leq \bar{c}_1 j^{5/2}$$

and

$$\sup_{z\in[0,1]} |b_j''(z)| \leq \bar{c}_1 j^{9/2} \ ,$$

where $\bar{c}_1$ is a positive constant. Thus by simple calculation we get

$$\sup_{x\in\mathbf{R}^2} |\gamma_j(x,\beta)| \leq \sup_{u\in[0,1]^2} |b_{i(j)}(u_1) b_{k(j)}(u_2)| \leq m(j)+1 \ ,$$

$$\mathcal{A}_j < \sqrt{3}\bar{c}_1 m^3(j) \ ,$$

$$\mathcal{M}_j < \max\{\sqrt{3}\bar{c}_1 m^5(j), \bar{c}_1^2 m^5(j)\}$$

and by (6.12) the inequalities (6.9), (6.10) and (6.11) easily follow. □

Next lemma of this section shows that in some neighborhood of β_0 the first and the second order partial derivatives of $\gamma_j(x,\beta)$ with respect to $\beta_1,\ldots,\beta_5$ are bounded by the functions of x which are integrable with respect to $f(x,\beta)$. This fact was used in Section 3.1 to prove Lemma 3.3.

Lemma 6.3 There exists a neighborhood $\mathcal{V}(\beta_0)$ of $\beta_0 = (0,0,1,1,0)^T$ and a positive constant c_1 such that for each $x \in \mathbf{R}^2$ and $\beta \in \mathcal{V}(\beta_0)$ it holds

$$\max_{1\leq t\leq 5} \left| \frac{\partial \gamma_j(x,\beta)}{\partial \beta_t} \right| \leq c_1 j^{3/2}(1+ \| x \|) \tag{6.13}$$

and

$$\max_{1\leq t,u\leq 5} \left| \frac{\partial^2 \gamma_j(x,\beta)}{\partial \beta_t \partial \beta_u} \right| \leq c_1 j^{5/2}(1+ \| x \|^2) \ . \tag{6.14}$$

Proof. The first and second order derivatives of $\gamma_j(x,\beta)$ with respect to β are given by

$$\frac{\partial \gamma_j(x,\beta)}{\partial \beta_t} = b'_{i(j)}(\Phi(y_1))b_{k(i)}(\Phi(y_2))\frac{\partial \Phi(y_1)}{\partial \beta_t} + b_{i(j)}(\Phi(y_1))b'_{k(i)}(\Phi(y_2))\frac{\partial \Phi(y_2)}{\partial \beta_t} \tag{6.15}$$

and

$$\begin{aligned}\frac{\partial^2 \gamma_j(x,\beta)}{\partial \beta_t \partial \beta_u} = {} & b_{i(j)}(\Phi(y_1))\left(b''_{k(j)}(\Phi(y_2))\frac{\partial \Phi(y_2)}{\partial \beta_t}\frac{\partial \Phi(y_2)}{\partial \beta_u} + b'_{k(j)}(\Phi(y_2))\frac{\partial^2 \Phi(y_2)}{\partial \beta_t \partial \beta_u}\right) \\ & + b_{k(j)}(\Phi(y_2))\left(b''_{i(j)}(\Phi(y_1))\frac{\partial \Phi(y_1)}{\partial \beta_t}\frac{\partial \Phi(y_1)}{\partial \beta_u} + b'_{i(j)}(\Phi(y_1))\frac{\partial^2 \Phi(y_1)}{\partial \beta_t \partial \beta_u}\right) \\ & + b'_{i(j)}(\Phi(y_1))b'_{k(j)}(\Phi(y_2))\left(\frac{\partial \Phi(y_1)}{\partial \beta_t}\frac{\partial \Phi(y_2)}{\partial \beta_u} + \frac{\partial \Phi(y_1)}{\partial \beta_u}\frac{\partial \Phi(y_2)}{\partial \beta_t}\right)\end{aligned} \quad ,(6.16)$$

where y_1 and y_2 are the functions of x and β [cf. (2.9)].
Note that for $m \in \{1,2\}$ it holds

$$\frac{\partial \Phi(y_m)}{\partial \beta_t} = \phi(y_m)\frac{\partial y_m}{\partial \beta_t} \tag{6.17}$$

and

$$\frac{\partial^2 \Phi(y_m)}{\partial \beta_t \partial \beta_u} = \phi'(y_m)\frac{\partial y_m}{\partial \beta_t}\frac{\partial y_m}{\partial \beta_u} + \phi(y_m)\frac{\partial^2 y_m}{\partial \beta_t \partial \beta_u} \ , \tag{6.18}$$

where $\phi(\cdot)$ denotes the density of standard normal variable.

For any matrix $M(t_1,\ldots,t_s) = [m_{ij}(t_1,\ldots,t_s)]$ we denote by $\frac{\partial M}{\partial t_m}$ the matrix of the partial derivatives with respect to t_m of the elements of M, i.e. $\frac{\partial M}{\partial t_m} = [\frac{\partial m_{ij}}{\partial t_m}]$. Then the derivatives of $y = (y_1, y_2)^T$ with respect to β are given by

$$\frac{\partial y}{\partial \beta_t} = \frac{\partial L^T}{\partial \beta_t}(x-\mu) + L^T\frac{\partial (x-\mu)}{\partial \beta_t} \tag{6.19}$$

and

$$\frac{\partial^2 y}{\partial \beta_t \partial \beta_u} = \frac{\partial^2 L^T}{\partial \beta_t \partial \beta_u}(x-\mu) + \frac{\partial L^T}{\partial \beta_t}\frac{\partial (x-\mu)}{\partial \beta_u} + \frac{\partial L^T}{\partial \beta_u}\frac{\partial (x-\mu)}{\partial \beta_t} \ . \tag{6.20}$$

Since for each bounded neighborhood $\mathcal{V}(\beta_0)$ of β_0 such that its closure is a subset of Ω it holds that

$$\sup_{\beta \in \mathcal{V}(\beta_0)} \max\left\{|l_i|, |\frac{\partial l_i}{\partial \beta_t}|, |\frac{\partial^2 l_i}{\partial \beta_t \partial \beta_u}|,\ i \in \{1,2,3\},\ t,u \in \{3,4,5\}\right\} = C_1 < \infty \tag{6.21}$$

then (6.13) and (6.14) easily follow from Lemma 6.2 and the equations (6.15)–(6.20).

6.4. Characterization of the class of alternatives to bivariate normality

Let $X_i = (X_{1i}, X_{2i})^T$ be the random vector distributed according to $\mathcal{P}$ on $\mathbf{R}^2$ and such that

$$(\mathbf{E}_{\mathcal{P}}X_{1i},\ \mathbf{E}_{\mathcal{P}}X_{2i},\ \mathbf{Var}_{\mathcal{P}}X_{1i},\ \mathbf{Var}_{\mathcal{P}}X_{2i},\ \mathbf{Cov}_{\mathcal{P}}(X_{1i},X_{2i}))^T = \beta_0 \ ,$$

where $\beta_0 = (0,0,1,1,0)^T$.

As before $\mathcal{N}(\beta_0)$ denotes bivariate normal distribution with the vector of parameters β_0.

Lemma 6.4 If

$$\mathcal{P} \neq \mathcal{N}(\beta_0) \ , \tag{6.22}$$

then there exists a natural j such that

$$\mathbf{E}_{\mathcal{P}}\gamma_j(X,\beta_0) \neq 0 \ . \tag{6.23}$$

Proof. Let Φ be the d.f. of standard normal distribution. Define a Borel measure ν on $[0,1]^2$ by

$$\nu(B) = \mathcal{P}(\{(x_1,x_2); (\Phi(x_1),\Phi(x_2)) \in B\}) \ . \tag{6.24}$$

Then (6.22) implies

$$\nu \neq \lambda \ , \tag{6.25}$$

where λ is the Lebesgue measure on $[0,1]^2$.

Let C_0 denote the space of continuous functions on $[0,1]^2$. By the fundamental Riesz-Markov-Kakutani Theorem, (6.25) yields that there exists a function $g \in C_0$ such that

$$\int_{[0,1]^2} g \, d(\nu - \lambda) \neq 0 \ . \tag{6.26}$$

From Stone-Weierstrass Theorem we have that the space of polynomials of the form $\sum_i \alpha_i u_1^j u_2^k$, where $u = (u_1,u_2) \in [0,1]^2$, $\alpha_i \in \mathbf{R}$ and j, $k \in \mathrm{N} \cup \{0\}$ is uniformly dense in C_0. Hence (6.26) implies that there exists a natural j such that

$$\int_{[0,1]^2} \tilde{\gamma}_j(u)\, d\nu(u) \neq 0 \ . \tag{6.27}$$

Since

$$\mathbf{E}_{\mathcal{P}}\gamma_j(X,\beta_0) = \int_{\mathbf{R}^2} \gamma_j(x,\beta_0)\, d\mathcal{P}(x) = \int_{[0,1]^2} \tilde{\gamma}_j(u)\, d\nu(u)$$

therefore (6.23) is an immediate consequence of (6.27). □

Acknowledgments. I would like to express my gratitude to Prof. T. Ledwina for introducing me into the problem, many helpful suggestions and a very careful and helpful reading of the manuscript. I am also indebted to S. Csörgő for useful discussion.

References

Bickel, P. J. and Ritov, Y. (1992). Testing for goodness of fit: A new approach. In *Nonparametric Statistics and Related Topics.* (A.K.Md.E. Saleh, ed.), North Holland, Amsterdam.

Bowman, A. W. and Foster, P. J. (1993). Adaptive smoothing and density-based tests of multivariate normality. *J. Amer. Statist. Assoc.* **88**, 529–537.

Cox, D. R. and Small, N. J. H. (1978). Testing multivariate normality. *Biometrika* **65**, 263–272.

Cox, D. R. and Wermuth, N. (1994). Tests of linearity, multivariate normality and the adequacy of linear scores. *Appl. Statist.* **43**, 347–355.

Csörgő, S. (1986). Testing for normality in arbitrary dimension. *Ann. Statist.* **14**, 708–723.

D'Agostino, R. B. (1986). Tests for the normal distribution. In *Goodness-of-Fit Techniques* (Ed. by R. B. D'Agostino and M. A. Stephens), *Statistics: Textbooks and Monographs,* Vol. 68, Marcel Dekker, New York.

Eubank, R. L., Hart, J. D. and LaRiccia, V. N. (1993). Testing goodness of fit via nonparametric function estimation techniques. *Comm. Statist. Theory Methods* **22**, 3327–3354.

Eubank, R. L. and LaRiccia, V. N. (1992). Asymptotic comparison of Cramér von Mises and nonparametric function estimation techniques for testing goodness of fit. *Ann. Statist.* **20**, 2071–2086.

Fan, J. (1995). Tests of significance based on wavelet thresholding and Neyman's truncation. Manuscript.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations.* Wiley, New York.

Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika* **48**, 151–165.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1994). Power approximations to and power comparison of certain goodness-of-fit tests. *Scand. J. Statist.* **21**, 131–145.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1995). On selection rules, with an application to goodness-of-fit for composite hypotheses. Memorandum 1242, University of Twente, Faculty of Applied Mathematics.

Inglot, T., Kallenberg, W. C. M. and Ledwina, T. (1996). Data driven smooth tests for composite hypotheses. Submitted.

Inglot, T. and Ledwina, T. (1995). Asymptotic optimality of data driven Neyman's tests. *Ann. Statist.* to appear.

Javitz, H. S. (1975). Generalized smooth tests of goodness of fit, independence and equality of distributions. Unpublished thesis, Univ. of California, Berkeley.

Kallenberg, W. C. M. and Ledwina, T. (1995a). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Statist.* **23**, 1594–1608.

Kallenberg, W. C. M. and Ledwina, T. (1995b). On data driven Neyman's test. *Probab. Math. Statist.* **15**, 409–426.

Kallenberg, W. C. M. and Ledwina, T. (1995c). Data driven smooth test when the hypothesis is composite. Memorandum 1299, University of Twente, Faculty of Applied Mathematics.

Kallenberg, W. C. M. and Ledwina, T. (1995d). Data driven smooth tests for composite hypothesis: Comparison of powers. Submitted.

Kallenberg, W. C. M., Ledwina, T. and Rafajłowicz, E. (1996). Testing bivariate independence and normality. In preparation.

Koziol, J. A. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika* **69**, 423–427.

Koziol, J. A. (1983). On assessing multivariate normality. *J. R. Statist. Soc. B* **45**, 358–361.

Koziol, J. A. (1986). Assessing multivariate normality: a compendium. *Comm. Statist. Theory Methods* **15**, 2763–2783.

Koziol, J. A. (1987). An alternative formulation of Neyman's smooth goodness of fit tests under composite alternatives. *Metrika* **34**, 17–24.

Ledwina, T. (1994). Data driven version of Neyman's smooth test of fit. *J. Amer. Statist. Assoc.* **89**, 1000–1005.

Malkovich, J. F. and Afifi (1973). On tests for multivariate normality. *J. Amer. Statist. Assoc.* **68**, 176–179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Mardia, K. V. (1970). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhyā Ser. B* **36**, 115–128.

Mardia, K. V. (1980). Tests of univariate and multivariate normality. In *Handbook of Statistics 1. Analysis of Variance* (P. R. Krishnaiah, ed.), 279–320, North Holland, Amsterdam.

Marsaglia, G. and Zaman, A. (1991), A new class of random numbers generators. *Ann. Appl. Probab.* **1**, 462–480.

Neyman, J. (1937). 'Smooth test' for goodness of fit. *Skand. Aktuarietidskr.* **20**, 149–199.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, second edition,* John Wiley & Sons, Inc., New York.

Rayner, J. C. W. and Best, D. J. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press, New York.

Rubin, H. and Sethuraman, J. (1965). Probabilities of moderate deviations. *Sankhyā Ser. A* **27**, 325–346.

Sansone, G. (1959). *Orthogonal Functions.* Interscience, New York.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics.* Chapman and Hall, New York.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.

Shapiro, S. S. and Francia, R. S. (1972). An approximate analysis of variance test for normality. *J. Amer. Statist. Assoc.* **67**, 215–216.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.

Small, N. J. H. (1980). Marginal skewness and kurtosis in testing multivariate normality. *Appl. Statist.* **29**, 85–87.

Szkutnik, Z. (1987). On invariant tests for multidimensional normality. *Probab. Math. Statist.* **8**, 1–10.

Thomas, D. R. and Pierce, D. A. (1979). Neyman's smooth goodness-of-fit test when the hypothesis is composite. *J. Amer. Statist. Assoc.* **74**, 441–445.

Versluis, C. (1996). Comparison of tests for bivariate normality with unknown parameters by transformation to an univariate statistic. *Comm. Statist. Theory Methods* **25**, 647–665.

M.Bogdan
Institute of Mathematics
Technical University of Wrocław
50–370 Wrocław
Wybrzeże Wyspiańskiego 27
Poland