ELSEVIER

Statistics & Probability Letters 35 (1997) 251–259

STATISTICS & PROBABILITY LETTERS

A multivariate Kolmogorov–Smirnov test of goodness of fit[1]

Ana Justel [a,*], Daniel Peña [b], Rubén Zamar [c]

[a] *Department of Mathematics, Universidad Autónoma de Madrid, Spain*
[b] *Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain*
[c] *Department of Statistics, University of British Columbia, BC, Canada*

Received 1 January 1996

Abstract

This paper presents a distribution-free multivariate Kolmogorov–Smirnov goodness-of-fit test. The test uses a statistic which is built using Rosenblatt's transformation and an algorithm is developed to compute it in the bivariate case. An approximate test, that can be easily computed in any dimension, is also presented. The power of these multivariate tests is studied in a simulation study. © 1997 Elsevier Science B.V.

Keywords: Bonferroni inequality; Empirical distribution function; Kolmogorov–Smirnov statistics; Rosenblatt's transformation

1. Introduction

Goodness-of-fit tests have been developed mostly for univariate distributions and, except for the case of multivariate normality, very few references can be found in the literature about multivariate goodness-of-fit tests (see, Krishnaiah, 1980; Kotz and Johnson, 1985 and D'Agostino and Stephens, 1986). A general approach based on empirical process theory can be found in Shorack and Wellner (1986) and Einmahl and Mason (1992).

In principle, the chi-square test can be applied for testing the goodness of fit of any multivariate distribution but it is unknown what is the best way to choose the cell limits and what is the best statistic to be used. Moore and Stubblebine (1981) suggested choosing as cell boundaries the concentric hyperellipses centered at the sample mean and with shape determined by the inverse of the covariance matrix. However, much work need to be done on the properties of this test.

The two most important classes of tests of goodness of fit based on the empirical distribution function of a random sample, the Kolmogorov–Smirnov statistic and the Cramer–von Mises group, have not been extended to the multivariate case. The problem is that the probability distribution of these multivariate

* Correspondence address: Departamento de Matematicas, Facultad de Ciencias, Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain. E-mail: ana.justel@uam.es.

[1] A. Justel and D. Peña were partially supported by PB93-0232 grant from DGICYT (Spain) and R. Zamar was partly supported by a grant from NSERC (Canada).

PII S0167-7152(97)00020-5

statistics is not distribution free as in the univariate case. Rosenblatt (1952) proposed a simple transformation of an absolutely continuous p-variate distribution into the uniform distribution on the p-dimensional hypercube and suggested using this transformation to build multivariate goodness-of-fit tests. Rincón-Gallardo, Quesenberry and O'Reilly (1979) consider this transformation for testing multinormality but a general multivariate Kolmogorov–Smirnov goodness-of-fit test, which is feasible to apply, has not yet been developed. The distribution function of the Cramer–von Mises statistic in the multivariate case has been studied by a number of authors (see Kotz and Johnson, 1985, pp. 35–39) but again, we do not have a general multivariate test of goodness of fit based on this statistic that can be readily applied.

Several procedures have been developed for testing multivariate normality. Among these procedures, the most often used are the tests based on multivariate measures of skewness and kurtosis (Mardia, 1970; Malkovich and Afifi, 1973; Small, 1980; Schwager and Margolin, 1982), the multivariate Shapiro–Wilks statistics (Royston, 1983), radii and angles test (Koziol, 1986) and tests based on the multivariate Box–Cox transformation (Velilla, 1995). Other procedures to check for multivariate normality have been proposed by Csörgõ (1986), Mudholkar et al. (1992) and Ghosh and Ruymgaart (1992).

In this paper we present two multivariate goodness-of-fit tests. In Section 2 we introduce a multivariate goodness-of-fit statistics which is distribution free and reduces to the Kolmogorov–Smirnov statistic in the univariate case. The computation of the proposed statistic is a problem in itself, and in this section we develop another statistic that can be easily computed for any dimension. Section 3 presents a procedure to compute the test statistics in the bivariate case. In Section 4 we explore the power of these two tests in testing multivariate normality and in assessing the fit of the Morgenstern distribution. The loss of power of the simplified statistic seems to be small, suggesting that it is a promising alternative for multivariate goodness-of-fit testing in any dimension. Finally, Section 5 includes some concluding remarks.

2. The multivariate Kolmogorov–Smirnov statistic

Given a sample $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ of i.i.d. random variables with distribution function F, consider the problem of testing $H_0: F = F_0$ versus $H_1: F \neq F_0$, where F_0 is some specified distribution function. In the univariate case, H_0 can be tested using the Kolmogorov–Smirnov statistic

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|,$$

where F_n is the empirical distribution function of the sample. It is also well known that this statistic is distribution free and it can be expressed as

$$D_n = \sup_{0\leqslant u\leqslant 1} |G_n(u) - u|, \tag{2.1}$$

where $G_n(u)$ is the empirical distribution function of the uniform 0–1 transformed sample $u_i = F_0(x_i)$, for $i = 1,\ldots,n$. The distribution-free property of the Kolmogorov–Smirnov statistic is derived from the result that any continuous random variable X with distribution function F can be transformed to a uniform random variable Y by the transformation $Y = F(X)$. A similar result holds for a continuous multivariate random variable $\boldsymbol{X}$, as it is shown in the following theorem, due to Rosenblatt (1952).

Theorem 1. *Let* $\boldsymbol{X} = (X_1,\ldots,X_p)$ *be a random vector with joint density*

$$f(x_1,\ldots,x_p) = f_1(x_1)f_2(x_2\,|\,x_1)\cdots f_p(x_p\,|\,x_1,\ldots,x_{p-1}),$$

and define the transformation $\boldsymbol{Y} = T(\boldsymbol{X})$ *by*

$$Y_1 = F(X_1),$$
$$Y_i = F(X_i|X_1,\ldots,X_{i-1}), \quad i=2,\ldots,p. \tag{2.2}$$

Then $Y_1,\ldots,Y_p$ *are i.i.d. uniform* 0–1.

The probability distribution function of the statistic

$$\sup_{\boldsymbol{x}\in\mathbb{R}^p} |F_n(\boldsymbol{x}) - F(x_1,\ldots,x_p)|,$$

where F_n is the empirical distribution function, is not distribution-free. However, we could use the transformation defined in Theorem 1 to test whether the values $(y_1,\ldots y_n)$ are a sample from a uniform distribution on the p-dimensional hypercube. The natural extension of the statistic (2.1) to the multivariate case is

$$d_n = \sup_{\boldsymbol{y}} |G_n(\boldsymbol{y}) - y_1 \cdots y_p|, \tag{2.3}$$

where G_n is the empirical distribution function of the transformed sample $\boldsymbol{y} = T(\boldsymbol{x})$. The statistic defined by (2.3) is not invariant because a relabelling of the components of $\boldsymbol{X}$ would give a different transformation (2.2) and, therefore, a different value of (2.3). We can take advantage of this lack of uniqueness to build a more powerful procedure as follows.

Let us define the sequence of transformations

$$y_1^j = F(z_1^j),$$
$$y_i^j = F(z_i^j\,|\,z_{i-1}^j,\ldots,z_1^j), \quad i=2,\ldots,p,$$

where $(z_1^j,\ldots,z_p^j)$, for $j=1,\ldots,p!$, is the jth permutation of the variables $(x_1,\ldots,x_p)$. Under the null hypothesis $d_n^j = \sup_{y^j} |G_n(\boldsymbol{y}^j) - y_1^j \cdots y_p^j|$, is distribution free in the class of continuous multivariate distributions. The $p!$ statistics d_n^j are not independent and if we use all of them to build the test we run into the standard problem of multiple testing. However, we can use the Bonferroni inequality to obtain an upper bound to the global significance level of a test in which the d_n^j statistics are used with a significant level of α. Therefore, we define the multivariate Kolmogorov–Smirnov statistic by

$$D_n = \max_{j=1,2,\ldots} d_n^j, \tag{2.4}$$

and if we wish to have a global significant level, α_g, the Bonferroni inequality leads to testing each d_n^j with a significant level $\alpha = \alpha_g/p!$.

The statistic is computed sequentially and in general it will not be necessary to evaluate all the $p!$ transformations. The procedure is to check one by one the transformed samples, comparing the Kolmogorov–Smirnov statistics with the percentiles of the D_n distribution. The null hypothesis is rejected when one of the statistics is greater than the percentile. Otherwise, we continue with the next transformation. Unfortunately, and unlike in the univariate case, the computation of (2.3) is very involving, as it is shown in the next section, in which we present an algorithm for the case $p=2$. Although this algorithm could be extended to the $p>2$ case, the computation difficulties appear to be considerable. A much simpler approach is to build the test from the statistic $\tilde{d}_n$ which is defined as the supremum on the set of transformed sample points A,

$$\tilde{d}_n^j = \sup_{\boldsymbol{y}^j\in A} |G_n(\boldsymbol{y}^j) - y_1^j \cdots y_p^j|.$$

Table 1
Monte-Carlo approximation to the percentiles of the bivariate Kolmogorov–Smirnov statistic distribution D_n, with 10 000 replications

n	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
10	0.4024	0.4172	0.4368	0.4606	0.4964	0.5285	0.5717	0.5979	0.6167	0.6393
11	0.3862	0.4011	0.4191	0.4422	0.4766	0.5093	0.5473	0.5760	0.5944	0.6192
12	0.3714	0.3861	0.4035	0.4282	0.4630	0.4950	0.5297	0.5553	0.5793	0.6110
13	0.3564	0.3703	0.3874	0.4085	0.4419	0.4731	0.5055	0.5273	0.5455	0.5729
14	0.3464	0.3585	0.3745	0.3961	0.4284	0.4566	0.4901	0.5124	0.5341	0.5596
15	0.3339	0.3456	0.3611	0.3822	0.4141	0.4399	0.4705	0.4952	0.5116	0.5405
20	0.2922	0.3037	0.3170	0.3349	0.3618	0.3859	0.4154	0.4346	0.4505	0.4868
25	0.2643	0.2745	0.2862	0.3014	0.3254	0.3454	0.3770	0.4011	0.4150	0.4482
30	0.2420	0.2517	0.2621	0.2769	0.2989	0.3206	0.3427	0.3629	0.3785	0.4082
40	0.2103	0.2182	0.2284	0.2409	0.2596	0.2789	0.3009	0.3140	0.3334	0.3467
50	0.1892	0.1960	0.2048	0.2161	0.2350	0.2512	0.2679	0.2802	0.2909	0.3045
60	0.1744	0.1805	0.1879	0.1986	0.2148	0.2293	0.2473	0.2591	0.2740	0.2903
80	0.1506	0.1563	0.1637	0.1729	0.1869	0.1998	0.2139	0.2277	0.2392	0.2544
100	0.1364	0.1419	0.1477	0.1558	0.1675	0.1798	0.1926	0.2008	0.2090	0.2196
150	0.1110	0.1152	0.1203	0.1270	0.1376	0.1464	0.1572	0.1643	0.1710	0.1798
200	0.0965	0.1000	0.1044	0.1098	0.1187	0.1265	0.1378	0.1456	0.1516	0.1562
300	0.0792	0.0822	0.0858	0.0905	0.0977	0.1041	0.1121	0.1169	0.1229	0.1283

Table 2
Monte-Carlo approximation to the percentiles of the approximated bivariate Kolmogorov–Smirnov statistic distribution $\tilde{D}_n$, with 2000 replications

n	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
10	0.3083	0.3232	0.3422	0.3689	0.4095	0.4434	0.4870	0.5221	0.5551	0.5939
11	0.2943	0.3097	0.3290	0.3524	0.3921	0.4258	0.4624	0.4921	0.5080	0.5261
12	0.2863	0.3015	0.3184	0.3415	0.3786	0.4092	0.4536	0.4866	0.5046	0.5270
13	0.2744	0.2884	0.3062	0.3279	0.3599	0.3899	0.4268	0.4628	0.4864	0.5134
14	0.2656	0.2791	0.2963	0.3170	0.3519	0.3828	0.4228	0.4505	0.4790	0.5124
15	0.2575	0.2703	0.2863	0.3060	0.3365	0.3679	0.4080	0.4278	0.4435	0.4750
20	0.2274	0.2393	0.2530	0.2706	0.2974	0.3227	0.3545	0.3712	0.3946	0.4167
25	0.2071	0.2172	0.2308	0.2460	0.2698	0.2932	0.3230	0.3386	0.3564	0.3772
30	0.1911	0.2005	0.2120	0.2260	0.2488	0.2700	0.2966	0.3123	0.3343	0.3563
40	0.1676	0.1752	0.1849	0.1997	0.2190	0.2388	0.2602	0.2767	0.2934	0.3110
50	0.1531	0.1604	0.1693	0.1815	0.1986	0.2130	0.2334	0.2451	0.2584	0.2746
60	0.1418	0.1489	0.1572	0.1678	0.1829	0.1960	0.2155	0.2286	0.2416	0.2611
80	0.1248	0.1303	0.1377	0.1473	0.1608	0.1730	0.1884	0.1986	0.2091	0.2224
100	0.1132	0.1184	0.1248	0.1327	0.1455	0.1566	0.1710	0.1817	0.1873	0.1935
150	0.0935	0.0978	0.1029	0.1096	0.1201	0.1293	0.1393	0.1455	0.1511	0.1618
200	0.0826	0.0862	0.0903	0.0954	0.1042	0.1129	0.1237	0.1306	0.1382	0.1447
300	0.0687	0.0718	0.0752	0.0795	0.0869	0.0940	0.1015	0.1065	0.1133	0.1181

When n is large,

$$\tilde{D}_n = \max_{j=1,2,\dots} \tilde{d}_n^j$$

will be close to D_n, as it is shown in the simulation results reported in Section 4.

The percentiles of the distribution of D_n and $\tilde{D}_n$ can be computed by Monte-Carlo simulation, sampling from independent uniforms 0–1. In this case, $y_i^j = F(x_i)$ for $i = 1,\dots,p$ and $j = 1,\dots,p!$. Therefore, D_n is equal to the Kolmogorov–Smirnov statistics for one transformation. Tables 1 and 2 present the percentiles of

Table 3
Monte-Carlo approximation to the bivariate Kolmogorov–Smirnov–Lilliefors statistic distribution D_n^*, with 2000 replications

n	0.25	0.2	0.15	0.1	0.05	0.025	0.01	0.005	0.0025	0.001
10	0.3060	0.3145	0.3249	0.3389	0.3610	0.3813	0.4010	0.4192	0.4363	0.4515
11	0.2941	0.3032	0.3138	0.3268	0.3474	0.3673	0.3887	0.4096	0.4204	0.4337
12	0.2823	0.2909	0.3012	0.3151	0.3367	0.3554	0.3791	0.3922	0.4055	0.4353
13	0.2732	0.2813	0.2906	0.3029	0.3210	0.3396	0.3616	0.3794	0.3904	0.4048
14	0.2638	0.2715	0.2803	0.2928	0.3122	0.3299	0.3508	0.3672	0.3798	0.4062
15	0.2557	0.2633	0.2720	0.2845	0.3023	0.3171	0.3352	0.3539	0.3663	0.3786
20	0.2244	0.2314	0.2395	0.2499	0.2659	0.2806	0.3012	0.3145	0.3295	0.3451
25	0.2033	0.2095	0.2167	0.2266	0.2417	0.2550	0.2706	0.2841	0.2947	0.3105
30	0.1862	0.1920	0.1987	0.2079	0.2218	0.2346	0.2474	0.2569	0.2674	0.2811
40	0.1636	0.1684	0.1746	0.1820	0.1927	0.2032	0.2159	0.2291	0.2364	0.2504
50	0.1471	0.1512	0.1563	0.1631	0.1735	0.1830	0.1940	0.2017	0.2140	0.2220
60	0.1347	0.1386	0.1433	0.1493	0.1599	0.1692	0.1799	0.1864	0.1915	0.2040
80	0.1174	0.1207	0.1252	0.1303	0.1382	0.1468	0.1561	0.1625	0.1669	0.1740
100	0.1055	0.1086	0.1126	0.1175	0.1250	0.1328	0.1402	0.1456	0.1511	0.1600
150	0.0869	0.0894	0.0926	0.0964	0.1027	0.1087	0.1155	0.1204	0.1245	0.1322
200	0.0818	0.0845	0.0877	0.0925	0.0994	0.1057	0.1138	0.1188	0.1241	0.1270
300	0.0598	0.0614	0.0635	0.0660	0.0699	0.0735	0.0774	0.0807	0.0834	0.0886

the statistics D_n and $\tilde{D}_n$ for the bivariate case, in the standard situation in which F_0 is completely specified by H_0.

In many cases, the parameters of the distribution F_0 are unknown and need to be estimated. Then the percentiles of Table 1 are not exact, and the distribution of the statistic when the parameter are estimated from the sample need to be computed again. Let D_n^* be this statistic. Note that, as in the univariate case, when the null hypothesis is rejected using Table 1, no further computations are necessary because the percentiles of D_n^* are always smaller than those of D_n. As an example, Table 3 presents the Monte-Carlo percentiles of the distribution of D_n^* in the particular case of testing normality and when the parameter are estimated by the sample mean and the sample covariance matrix. The statistic of this Table 3 can, therefore, be regarded as the multivariate generalization of the Kolmogorov–Smirnov–Lilliefors statistic.

3. An algorithm to compute the bivariate Kolmogorov-Smirnov statistic

In a one-dimensional sample, the empirical distribution changes only at the observed points, and the univariate Kolmogorov–Smirnov statistic is obtained by evaluating the distance between the empirical and theoretical distribution functions at these points. Nevertheless, when the dimension p is larger than one, the empirical distribution function jumps on an infinite number of points. For instance, suppose that $p=2$ and (x_1,y_1) is the observed point with the smallest first coordinate. Then the empirical distribution function changes at all the points (x,y_1) with $x \geqslant x_1$. Here we develop a procedure for calculating the Kolmogorov–Smirnov statistic (2.4) in the two-dimensional case by evaluating it on a finite set.

Since Theorem 1 holds, we may assume that $\boldsymbol{u}_1=(x_1,y_1),\dots,\boldsymbol{u}_n=(x_n,y_n)$ is a random sample from two independent uniform 0–1 distributions. In this context, the pair (x_j,y_i) is called *an intersection point* if $x_i<x_j$ and $y_i>y_j$. For $\boldsymbol{u}=(x,y)$ we define the *superior distance* $D_n^+(\boldsymbol{u})=(G_n(\boldsymbol{u})-G(\boldsymbol{u}))$ and the *inferior distance* $D_n^-(\boldsymbol{u})=(G(\boldsymbol{u})-G_n(\boldsymbol{u}))$, where G is the distribution function of two independent uniform random variables on $(0,1)$ and G_n is the empirical distribution function. Also, the left empirical distribution function in $\boldsymbol{u}$ is defined as $G_n(\boldsymbol{u}^-)=\lim_{\varepsilon\to 0} G_n(x-\varepsilon,y-\varepsilon)$. The proof is based on the behavior of the lateral Kolmogorov–Smirnov statistics $D_n^+=\sup_{\boldsymbol{u}} D_n^+(\boldsymbol{u})$ and $D_n^-=\sup_{\boldsymbol{u}} D_n^-(\boldsymbol{u})$.

Lemma 1. *If $x_0 = y_0 = 0$, then $D_n^+ = \max_{v \in I} D_n^+(v)$, where $I = \{(x_j, y_i) \mid x_i \leqslant x_j,\ y_i \geqslant y_j;\ i,j = 0,1,\dots,n\}$. Elements in the set I are: the pair $(0,0)$, the observed points and the intersection points.*

Proof. For each $u = (x, y)$ in the unit square, $x^u = x_{k_u} = \max_{i=0,1,\dots,n}\{x_i \mid x_i \leqslant x,\ \ y_i \leqslant y\}$ and $y^u = y_{p_u} = \max_{i=0,1,\dots,n}\{y_i \mid y_i \leqslant y,\ x_i \leqslant x\}$. The relationship between the coordinates is given by $x_{p_u} \leqslant \max\{x_i \leqslant x \mid y_i \leqslant y\} = x_{k_u}$ and $y_{k_u} \leqslant \max\{y_i \leqslant y \mid x_i \leqslant x\} = y_{p_u}$. Hence, $(x^u, y^u) \in I$. By the definition of (x^u, y^u), it is immediate that $G_n(x, y) = G_n(x^u, y) = G_n(x, y^u)$ and the set $\{x_i \mid x_i \in (x^u, x],\ \ y_i \in (y^u, y]\}$ is empty. Then $G_n(x^u, y^u) = G_n(x^u, y) + G_n(x, y^u) - G_n(x, y) = G_n(x, y)$ and, therefore, for each u

$$D_n^+(u) = G_n(x, y) - G(x, y) \leqslant G_n(x^u, y^u) - G(x^u, y^u) \leqslant \max_{v \in I} D_n^+(v).$$

Hence the lemma follows.

Lemma 2. *If $x_0 = 0$, $y_0 = 1$, $x_{n+1} = 1$ and $y_{n+1} = 0$, then $D_n^- = \max_{v \in P}(G(v) - G_n(v^-))$, where $P = \{(x_j, y_i) \mid x_j > x_i,\ \ y_j < y_i;\ i,j = 0,1,\dots,n+1\}$. Elements in set P are the pair $(1,1)$, the intersection points and the projections of the observed points on the right and on the top unit-square borders.*

Proof. For each $u = (x, y)$ in the unit square, $x^u = x_{k_u} = \min_{i=0,1,\dots,n+1}\{x_i \mid x_i > x\}$ and $y^u = y_{p_u} = \min_{i=0,1,\dots,n+1}\{y_i \mid y_i > y,\ x_i \leqslant x\}$. Obviously, $x_{k_u} > x \geqslant x_{p_u}$ and $y_{p_u} > y$. Hence, (x^u, y^u) is in the set $Q = \{(x_j, y_i) \mid x_j > x_i;\ i,j = 0,1,\dots,n+1\} = P \cup \{(x_j, y_i) \mid x_j > x_i,\ y_j > y_i;\ i,j = 1,\dots,n\}$.

Since F is continuous and increasing, the inferior distance is bounded by

$$\begin{aligned} D_n^-(u) &= G(u) - G_n(u) < G(x^u, y^u) - G_n(u) \\ &= G(x^u, y^u) - \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon) + \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon) - G_n(u) \\ &\leqslant \max_{v \in Q}(G(v) - G_n(v^-)) + \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon) - G_n(u). \end{aligned} \tag{3.1}$$

Because of the definition of (x^u, y^u), $G_n(u) = \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y) = \lim_{\varepsilon \to 0} G_n(x, y^u - \varepsilon)$ and the set $\{(x_i, y_i) \mid x_i \in (x, x^u),\ y_i \in (y, y^u)\}$ is empty. The left empirical distribution function verifies

$$\lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon) = \lim_{\varepsilon \to 0} G_n(x^u, y - \varepsilon) + \lim_{\varepsilon \to 0} G_n(x, y^u - \varepsilon) - G_n(u) = G_n(u). \tag{3.2}$$

In addition, if $(x^u, y^u) \in Q - P$, we define $x^w = \min\{x_i \mid x_i > x^u,\ \ y_i < y^u\}$. Then the pair (x^w, y^u) is in P, $G(x^u, y^u) < G(x^w, y^u)$ and the set $\{(x_i, y_i) \mid x_i \in [x^u, x^w),\ \ y_i \in (-\infty, y^u)\}$ is empty. Hence, $\lim_{\varepsilon \to 0} G_n(x^w - \varepsilon, y^u - \varepsilon) = \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon)$ and

$$G(x^u, y^u) - \lim_{\varepsilon \to 0} G_n(x^u - \varepsilon, y^u - \varepsilon) < \max_{v \in P}(G(v) - G_n(v^-)). \tag{3.3}$$

By (3.1)–(3.3), $\max_{v \in P}(G(v) - G_n(v^-))$ is a superior bound for $D_n^-(u)$.

Finally, let $u_0 = (x_0, y_0)$ be given by $u_0 = \arg\max_{v \in P}(G(v) - G_n(v^-))$, then

$$\begin{aligned} \max_{v \in P}(G(v) - G_n(v^-)) &= \lim_{\varepsilon \to 0}(G(x_0 - \varepsilon, y_0 - \varepsilon) - G_n(x_0 - \varepsilon, y_0 - \varepsilon)) \\ &= \lim_{\varepsilon \to 0} D_n^-(x_0 - \varepsilon, y_0 - \varepsilon). \end{aligned}$$

Hence, $D_n^- = \max_{v \in P}(G(v) - G_n(v^-)) = \sup_u D_n^-(u)$ and the lemma follows.

Theorem 2. *If $p = 2$, the Kolmogorov–Smirnov statistic* (2.4) *is*

$$D_n = \max_{u \in I,\, v \in P}\{G_n(u) - G(u), G(v) - G_n(v^-)\}.$$

The Kolmogorov–Smirnov statistics may be expressed as $D_n = \max\{D_n^+, D_n^-\}$ and the proof is straightforward by Lemmas 1 and 2. □

As a consequence of Theorem 2, D_n may be obtained by evaluating the distance in a finite amount of points which ranks from $3n$ to $3n + \binom{n}{2}$ depending on the sample configuration. The theorem leads to the following procedure to compute the Kolmogorov–Smirnov statistic (2.4): (1) compute the maximum distance in the observed points, $D_n^1 = \max_{i=1,\dots,n} D_n^+(\boldsymbol{u}_i)$; (2) compute the maximum and minimum distances in the intersection points, $D_n^2 = \max_{i,j=1,\dots,n}\{D_n^+(x_j, y_i) \mid x_j > x_i,\ y_j < y_i\}$ and $D_n^3 = 2/n - \min_{i,j=1,\dots,n}\{D_n^+(x_j, y_i) \mid x_j > x_i,\ y_j < y_i\}$; (3) compute the maximum distance among the projections of the observed points on the right unit-square border, $D_n^4 = 1/n - \min_{i=1,\dots,n} D_n^+(1, y_i)$; (4) compute the maximum distance among the projections of the observed points on the top unit-square border, $D_n^5 = 1/n - \min_{i=1,\dots,n} D_n^+(x_i, 1)$; and (5) compute the maximum $D_n = \max\{D_n^1, D_n^2, D_n^3, D_n^4, D_n^5\}$.

4. Some simulation results

The power of the exact and the approximate multivariate Kolmogorov–Smirnov statistics when used as a normality test and as a general multivariate goodness-of-fit test have been investigated. In the first case, the null hypothesis is bivariate normal with mean $\boldsymbol{\mu} = \boldsymbol{0}$ and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

The alternative distribution is $(1-\varepsilon)\mathrm{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) + \varepsilon\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for several values of ε and $\boldsymbol{\mu}$. Table 4 shows the power of the normality test. As we may have expected, the power increases with n and is larger for the exact test than for the approximate one. However, for moderately large n ($n \geqslant 50$ say) the power of the approximate test is very close to that of the exact one. Table 4 shows that both tests are very powerful when n is large and $\varepsilon \geqslant 0.2$.

Table 5 shows the power of these statistics when the null distribution is Morgenstern (see Morgenstern, 1956). We have chosen this distribution because it is very flexible and may have fixed marginal distribution allowing different degrees of dependency. For instance, it can be used to generalize the Burr–Pareto-logistic class of distributions (see Johnson, 1987), and also to approximate the Plackett family of distributions. For some of its applications see Lai (1978). For the uniform marginal case, the joint density

Table 4
Empirical power of the Kolmogorov–Smirnov (D_n) and approximated Kolmogorov–Smirnov ($\tilde{D}_n$) tests with size $\alpha = 0.05$. The null hypothesis is a $\mathrm{N}(0, \Sigma)$ and the samples are generated from a normal mixture $(1-\varepsilon)\mathrm{N}(0,\Sigma) + \varepsilon\mathrm{N}(\mu, \Sigma)$

		$n=15$		$n=25$		$n=50$		$n=100$	
		D_n	$\tilde{D}_n$	D_n	$\tilde{D}_n$	D_n	$\tilde{D}_n$	D_n	$\tilde{D}_n$
$\mu = (3,3)'$	$\varepsilon = 0.1$	0.13	0.12	0.16	0.14	0.23	0.21	0.41	0.37
	$\varepsilon = 0.2$	0.27	0.26	0.40	0.38	0.67	0.66	0.94	0.94
	$\varepsilon = 0.4$	0.73	0.73	0.92	0.92	1.00	1.00	1.00	1.00
$\mu = (3,-1)'$	$\varepsilon = 0.1$	0.14	0.12	0.18	0.13	0.27	0.19	0.44	0.33
	$\varepsilon = 0.2$	0.41	0.21	0.45	0.32	0.73	0.58	0.96	0.90
	$\varepsilon = 0.4$	0.76	0.58	0.93	0.84	1.00	0.99	1.00	1.00

Table 5
Empirical power of the Kolmogorov–Smirnov (D_n) and the approximated Kolmogorov–Smirnov ($\tilde{D}_n$) tests of size α. The null hypothesis is a Morgenstern with parameter 0.5 and uniform marginals. The samples are generated from two independent Beta(a,b)

		$\alpha=0.05$		$\alpha=0.025$		$\alpha=0.005$	
	n	D_n	$\tilde{D}_n$	D_n	$\tilde{D}_n$	D_n	$\tilde{D}_n$
Beta(10, 10)	10	0.966	0.540	0.874	0.420	0.435	0.235
	20	1.000	0.881	1.000	0.802	0.998	0.645
	50	1.000	1.000	1.000	1.000	1.000	0.991
	100	1.000	1.000	1.000	1.000	1.000	1.000
Beta(3, 3)	10	0.294	0.269	0.159	0.156	0.028	0.045
	20	0.709	0.515	0.535	0.360	0.220	0.149
	50	0.997	0.957	0.983	0.883	0.818	0.638
	100	1.000	1.000	1.000	0.999	1.000	0.992
Beta(3, 2)	10	0.214	0.207	0.100	0.091	0.012	0.013
	20	0.601	0.522	0.411	0.326	0.121	0.082
	50	0.993	0.985	0.971	0.932	0.683	0.565
	100	1.000	1.000	1.000	1.000	0.999	0.996
Beta(0.5, 1)	10	0.624	0.666	0.493	0.526	0.246	0.284
	20	0.873	0.880	0.792	0.782	0.568	0.547
	50	0.997	0.998	0.991	0.990	0.928	0.929
	100	1.000	1.000	1.000	1.000	1.000	1.000
Beta(0.5, 0.5)	10	0.362	0.325	0.249	0.208	0.095	0.074
	20	0.577	0.497	0.437	0.351	0.213	0.147
	50	0.891	0.862	0.806	0.728	0.504	0.427
	100	0.997	0.994	0.983	0.975	0.914	0.881

function is

$$f(x_1,x_2)=1+\alpha(2x_1-1)(2x_2-1) \quad 0\leqslant x_1,x_2\leqslant 1 \quad -1\leqslant\alpha\leqslant 1,$$

and it is straightforward to show that for this distribution

$$F(x_1)=x_1 \quad 0\leqslant x_1\leqslant 1,$$

$$F(x_2\,|\,x_1)=(1-\alpha(2x_1-1))x_2+\alpha(2x_1-1)x_2^2 \quad 0\leqslant x_2\leqslant 1.$$

The alternative distributions are independent Beta distributions with several combination of shape parameters, to allow for different degrees of asymmetry. In this simulation study, we have chosen $\alpha=0.5$. Similar results were also found for other values of α. The results in Table 5 show that again, as one could expect, for small n ($n=10$) the power is very low unless the degree of kurtosis or asymmetry is high. The difference between the power of the exact and approximate test is negligible for large n ($n\leqslant 50$).

5. Concluding remarks

As in the univariate case, the multivariate Kolmogorov–Smirnov test presented in this paper may provide a general and flexible goodness-of-fit test, specially for situations when specific tests are yet to be developed. The main problem in the application of the test is the computation of the statistic in the case $p>2$. An extension of the computing algorithm developed in this paper may be possible, but still the numerical complications seem considerable. However, our simulation results show that the approximate Kolmogorov–Smirnov test statistics introduced in this paper, that is trivial to compute, seems to be a promising alternative when n is moderately large.

Acknowledgements

We are very grateful to Victor Yohai for useful discussion and comments.

References

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

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-fit Techniques. Marcel Dekker, New York.

Einmahl, J.H.J., Mason, D.M., 1992. Generalized quantile processes. Ann. Statist. 20, 1062–1078.

Ghosh, S., Ruymgaart, F.H., 1992. Applications of empirical characteristic functions in some multivariate problems. Can. J. Statist. 20, 429–440.

Johnson, M.E., 1987. Multivariate Statistics Simulation. Wiley, New York.

Kotz, S., Johnson, N.L., 1983. Encyclopedia of Statistical Sciences. Wiley, New York.

Kotz, S., Johnson, N.L., 1985. Encyclopedia of Statistical Sciences. Wiley, New York.

Koziol, J.A., 1986. Assessing multivariate normality: a compendium. Commun. Statist. Theoret. Meth. 15, 2763–2783.

Krishnaiah, P.R., 1980. Handbook of Statistics. North-Holland, Amsterdam.

Lai, C.D., 1978. Morgenstern's bivariate distribution and its application to point processes. J. Math. Anal. Appl. 65, 247–256.

Malkovich, J.F., Afifi, A.A., 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.

Moore, D.S., Stubblebine, J.B., 1981. Chi-square tests for multivariate normality, with application to common stock prices. Commun. Statist. Theoret. Meth. A 10, 713–733.

Morgenstern, D., 1956. Einfache beispiele zweidimensionaler verteilungen. Mitt. Math. Statist. 8, 234–235.

Mudholkar, G.S., McDermott, M., Srivastava, D.K., 1992. A test of p-variate normality. Biometrika 79 (4) 850–854.

Rincón-Gallardo, S., Quesenberry, C.P., O'Reilly, F.J., 1979. Conditional probability integral transformations and goodness-of-fit tests for multivariate normal distributions. Ann. Statist. 7, 1052–1057.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Ann. Math. Statist. 23, 470–472.

Royston, J.P., 1983. Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. Appl. Statist. 32, 121–133.

Schwager, S.J., Margolin, B.H., 1982. Detection of multivariate normal outliers. Ann. Statist. 10, 943–954.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

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

Velilla, S., 1995. Diagnostics and robust estimation in multivariate data transformation. J. Amer. Statist. Assoc. 90, 945–951.