REGULAR ARTICLE

A test for normality and independence based on characteristic function

Wiktor Ejsmont[1] · **Bojana Milošević[2]** · **Marko Obradović[2]**

Abstract

In this article we prove a generalization of the Ejsmont characterization (Ejsmont in Stat Probab Lett 114:1–5, 2016) of the multivariate normal distribution. Based on it, we propose a new test for independence and normality. The test uses an integral of the squared modulus of the difference between the product of empirical characteristic functions and some constant. Special attention is given to the case of testing for univariate normality in which we derive the test statistic explicitly in terms of Bessel function and explore asymptotic properties. The simulation study also includes the cases of testing for bivariate and trivariate normality and independence, as well as multivariate normality. We show the quality performance of our test in comparison to some popular powerful competitors. The practical application of the proposed normality and independence test is discussed and illustrated using a real dataset.

✉ Wiktor Ejsmont
  wiktor.ejsmont@gmail.com

  Bojana Milošević
  bojana@matf.bg.ac.rs

  Marko Obradović
  marcone@matf.bg.ac.rs

[1] Department of Telecommunications and Teleinformatics, Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

[2] University of Belgrade, Faculty of Mathematics, Studentski trg 16, Belgrade, Serbia

1 Introduction

Some of classical and important problems in statistics are goodness-of-fit and independence tests. The traditional approach to testing for independence is based on Pearson's correlation coefficient, but its lack of robustness to outliers and departures from normality eventually led researchers to consider some alternative nonparametric procedures such as Savage, Spearman and van der Waerden, which in particular rely on linear rank statistics. The traditional approach for testing normality includes classical omnibus tests such as Kolmogorov–Smirnov and Anderson-Darling, and quantile-based tests such as Shapiro–Wilk and Shapiro–Francia.

In many circumstances, such as checking model assumptions, one needs to use both types of tests for the same sample. This is usually done separately and then one needs to be careful with multiple testing issues. In this paper we propose another way to test for the independence and normality simultaneously.

A theoretical framework to study this aspect was given by Ejsmont (2016). Ejsmont proved that the characterizations of a normal law are given by a certain invariance of the noncentral chi-square distribution. Namely in Ejsmont (2016, Corollary 3.3) it has been shown that if the random vectors $(X_1, \ldots, X_m, A)$ and $(Y_1, \ldots, Y_n, B)$ are independent and the distribution of $\sum_{i=1}^{m} X_i a_i + A + \sum_{j=1}^{n} Y_j b_j + B$ depends only on $\sum_{i=1}^{m} a_i^2 + \sum_{j=1}^{n} b_j^2$, then $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ are independent and have the same normal distribution. This result is obtained under the assumption that all moments exist. In the current paper we weaken this condition considerably and propose a new test of independence and normality.

The paper is organized as follows. In Sect. 2 we state and prove the main results of Ejsmont (2016) under the weakened assumption. Next, in Sect. 3, using this result we propose a new test for normality and explore its asymptotic properties. In Sect. 4 we present results of a wide empirical study in which we compare powers of our and competing tests. Finally, in Sect. 5 we provide the real-data example and present the list of potential application of the proposed tests. For all proofs, the readers are referred to the Appendix.

2 The theoretical base for the construction of a test

Notation. The scalar product of vectors $t, s \in \mathbb{R}^p$ is denoted by $\langle t, s \rangle$ and the Euclidean norm of t is $\|t\| = \sqrt{\langle t, t \rangle}$. Throughout this paper $X := (X_1, \ldots, X_m) \in \mathbb{R}^m$ and $Y := (Y_1, \ldots, Y_n) \in \mathbb{R}^n$ are random vectors, where m and n are positive integers. The characteristic functions of X and Y are denoted by $\varphi_X(\cdot) = E e^{i\langle \cdot, X \rangle}$ and $\varphi_Y(\cdot) = E e^{i\langle \cdot, Y \rangle}$, respectively. For complex-valued functions $f(\cdot)$, the complex conjugate of f is denoted by $\overline{f}$ and $|f|^2 = f\overline{f}$. In order to simplify the notation, we will denote $[n] = \{1, \ldots, n\}$. The concatenation of the vectors $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$ is denoted by $(a, b) \in \mathbb{R}^{m+n}$.

Our construction of a new test of normality is based on the following result. This is a generalization of the main result of Ejsmont (2016), under omitted moment assump-

tions (in Ejsmont (2016) we assume that random variables have all moments; the proof is also different).

Theorem 2.1 *Let* $(X_1, \ldots, X_m, A)$ *and* $(Y_1, \ldots, Y_n, B)$ *be independent random vectors, where* X_i *and* Y_j *are nondegenerate for* $i \in [m]$, $j \in [n]$, *and let statistic*

$$\langle a, X \rangle + \langle b, Y \rangle + A + B = \sum_{i=1}^{m} a_i X_i + \sum_{j=1}^{n} b_j Y_j + A + B,$$

have a distribution which depends only on $\|a\|^2 + \|b\|^2$, *where* $a \in \mathbb{R}^m$ *and* $b \in \mathbb{R}^n$. *Then random variables* $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ *are independent and have the same normal distribution with zero means.*

The construction of a new test is based directly on the Proposition below, that follows, in a sense, from Theorem 2.1; namely if $A = B = 0$, then Theorem 2.1 can be rewritten as follows.

Proposition 2.2 *Let* $(X_1, \ldots, X_m)$ *and* $(Y_1, \ldots, Y_n)$ *be independent random vectors, where* X_i *and* Y_j *are nondegenerate,* $E(X_i^2) = 1$, $E(Y_j^2) = 1$ *for* $i \in [m]$, $j \in [n]$. *Then the following statements are equivalent:*

(i) a statistic $\langle a, X \rangle + \langle b, Y \rangle$ *has a distribution which does not depend on*

$$(a_1, \ldots, a_m, b_1, \ldots, b_n),$$

whenever $\|a\|^2 + \|b\|^2 = 1$;

(ii) random variables $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ *are independent and have the same normal distribution* $N(0, 1)$.

3 The test statistic

In this section we propose a new class of test statistics for testing the null hypothesis that the sample comes from a multivariate normal distribution with independent components. In a univariate case it reduces to the normality null hypothesis.

Our methodology applied in this construction is based on distances between empirical and theoretical quantities. In the theory of hypothesis testing there are many types of distances that can be defined between statistical objects. One of the best known and mostly applied is the L^2 distance (used e.g. for the construction of Cramér (1928) and Anderson and Darling (1952) tests). If we want to test for multivariate normality, then we can also use the L^2 distance between empirical and theoretical characteristic function; see Baringhaus and Henze (1988) and Epps and Lawrence (1983). More recently, the characterization of a test for multivariate independence was given in Székely et al. (2007) and Székely and Rizzo (2009). Suppose that $X \in \mathbb{R}^m$, $Y \in \mathbb{R}^n$ are real-valued random vectors with characteristic functions φ_X and φ_Y, respectively. Then, for measuring independence, we can use the following

distance $\int_{\mathbb{R}^{m+n}} |\varphi_{X,Y}(t,s) - \varphi_X(t)\varphi_Y(s)|^2 w(t,s)dtds$, where $w(t,s)$ is an arbitrary positive weight function for which the integral above exists. We put forward a test that is also based on the distance between a function of the empirical characteristic function and some constant, and it was inspired by the articles (Baringhaus and Henze 1988; Székely et al. 2007; Székely and Rizzo 2009; Epps and Lawrence 1983). Our approach is based on the following reasoning.

The condition (i) from Proposition 2.2 simply means that we get the statement (ii) if the distribution of the statistic $\langle a, X \rangle + \langle b, Y \rangle$ is constant on the $(n+m)$-sphere with radius 1. This requirement can be rewritten using the characteristic function, namely, we get statement (ii) if and only if the function

$$E e^{i\langle a,X\rangle + i\langle b,Y\rangle} = \varphi_X(a)\varphi_Y(b)$$

is constant on the unit sphere $\|a\|^2 + \|b\|^2 = 1$, where $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$. We also know from the proof of Proposition 2.2 that this constant function must equal $e^{-\frac{1}{2}}$, namely

$$\varphi_X(a)\varphi_Y(b) - e^{-\frac{1}{2}} = 0$$

for all $\|a\|^2 + \|b\|^2 = 1$ or equivalently,

$$\int_{S_{n+m}} |\varphi_X(a)\varphi_Y(b) - e^{-\frac{1}{2}}|^2 dS_{n+m} = 0, \tag{1}$$

where $\int_{S_{n+m}} \cdot dS_{n+m}$ is the surface integral over $S_{n+m} = \{t \in \mathbb{R}^{n+m} \mid \|t\| = 1\}$. Finiteness of the integral above follows directly from $|\varphi_X(a)\varphi_Y(b)| \le 1$ and $e^{-\frac{1}{2}} < 1$, namely we see that

$$\int_{S_{n+m}} |\varphi_X(a)\varphi_Y(b) - e^{-\frac{1}{2}}|^2 dS_{n+m} \le (1 - e^{-\frac{1}{2}})^2 |S_{n+m}|.$$

Let us assume that we have a simple random sample $\mathbf{X} = (X_1, \ldots, X_N)$ from a multivariate distribution with m components, i.e. the data have the following structure:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,m} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,m} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ x_{N,1} & x_{N,2} & \ldots & x_{N,m} \end{bmatrix}. \tag{2}$$

We want to test the null hypothesis

$$\mathcal{H}_0^{(m)}: \qquad\qquad\qquad\qquad\qquad \mathcal{H}_1^{(m)}:$$

all columns of $\mathbf{X}$ are independent, and have   vs. $\mathcal{H}_0^{(m)}$ is not true.
normal $N(\mu_k, \sigma_k^2)$ distributions, $k = 1, \ldots, m$;

For this purpose we use Proposition 2.2 for $m = n$. Let $\widetilde{\mathbf{X}}$ denote the matrix obtained from $\mathbf{X}$ by column-wise standardization, i.e.

$$\widetilde{X}_{j,k} = \frac{x_{j,k} - \hat{\mu}_k}{\hat{\sigma}_k}, \quad k = 1, \ldots, m, \quad j = 1, \ldots, N, \tag{3}$$

where

$$\hat{\mu}_k = \frac{1}{N} \sum_{j=1}^{N} x_{j,k} \text{ and } \hat{\sigma}_k^2 = \frac{1}{N-1} \sum_{j=1}^{N} (x_{j,k} - \hat{\mu}_k)^2. \tag{4}$$

Let $\varphi_{\widetilde{\mathbf{X}}}(a)$ be the empirical characteristic function of $\widetilde{\mathbf{X}}$ defined by

$$\varphi_{\widetilde{\mathbf{X}}}(a) = \frac{1}{N} \sum_{k=1}^{N} e^{i \langle a, \widetilde{\mathbf{X}}_k \rangle},$$

where $\widetilde{\mathbf{X}}_k$ is the kth row of the matrix $\widetilde{\mathbf{X}}$. Similarly, the empirical counterpart of characteristic function of random variable $\langle a, X \rangle + \langle b, Y \rangle$ is

$$\frac{1}{N^2} \sum_{j,k} e^{i(\langle a, X_j \rangle + \langle b, X_k \rangle)},$$

where $a = (a_1, ..., a_m)^{\mathrm{T}}$ and $b = (b_1, ..., b_m)^{\mathrm{T}}$. Assuming that X and Y are both distributed as X_1, the natural test statistics based on (1) is

$$M_m = N \int \left| \frac{1}{N^2} \sum_{j,k} e^{i(\langle a, \widetilde{X}_j \rangle + \langle b, \widetilde{X}_k \rangle)} - e^{-\frac{1}{2}} \right|^2 dS_{2m}(a, b), \tag{5}$$

which can be further expressed as

$$M_m = N \int \left(\frac{1}{N^2} \sum_{j,k} (\cos (\langle a, \widetilde{X}_j \rangle + \langle b, \widetilde{X}_k \rangle) - e^{-\frac{1}{2}} \right)^2 \\ + \left(\frac{1}{N^2} \sum_{i,j} \sin (\langle a, \widetilde{X}_j \rangle + \langle b, \widetilde{X}_k \rangle) \right)^2 dS_{2m}(a, b), \tag{6}$$

where $a = (a_1, ..., a_m)^{\mathrm{T}}$ and $b = (b_1, ..., b_m)^{\mathrm{T}}$ such that $\langle a, a \rangle + \langle b, b \rangle = 1$.

We are interested in the one-sided test, i.e. the right-tailed test, because we see from the above construction that we reject the null hypothesis for large values of M_m. It is clear that the distribution of the test statistic, under the null hypothesis, does not depend on the null distribution location and scale parameters μ_k and σ_k, $k = 1, ..., m$, hence we may derive critical values using the Monte Carlo approach.

When the sample is univariate, the test statistic M_1 can be expressed in a closed form.

Proposition 3.1 *Let $\tilde{X} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ be the standardized sample. The statistic M_1 has the form*

$$M_1 = 2\pi N \left[\frac{1}{N^4} \sum_{n,j,k,l=1}^{N} J(d(\tilde{X}_n - \tilde{X}_k, \tilde{X}_j - \tilde{X}_l)) - e^{-\frac{1}{2}} \frac{2}{N^2} \sum_{n,j=1}^{N} J(d(\tilde{X}_n, \tilde{X}_j)) + e^{-1} \right].$$

(7)

where J is the Bessel function (of order zero) of the first kind, namely $J(z) = \sum_{k=0}^{\infty} (-1)^k \frac{(z^2/4)^k}{(k!)^2}$ and d is the distance from origin to point (x, y), i.e. $d(x, y) = \sqrt{x^2 + y^2}$.

3.1 Asymptotic properties

In this section we discuss some asymptotic properties of the proposed tests. First, we show their consistency against fixed alternatives.

Theorem 3.2 *Let $X_1, \ldots, X_N$ be an i.i.d. sample of m-variate random vectors with finite second moments. Denote the vectors of means and variances $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T$ and $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_m^2)^T$. Define $\varphi_Z(\cdot)$ as the characteristic function of the columnwise standardized random vector $Z = \mathrm{Diag}((\sigma_1^{-1}, \ldots, \sigma_m^{-1}))(X_1 - \boldsymbol{\mu})$. Then*

$$\frac{M_m}{N} \xrightarrow{a.s.} \Delta = \int |\varphi_Z(a)\varphi_Z(b) - e^{-\frac{1}{2}}|^2 dS_{2m}(a, b) \geq 0,$$

and $\Delta = 0$ if and only if the null hypothesis holds.

Next we examine the asymptotic distribution under the null hypothesis. The test statistic (6) is an integral of a sum of squares of two empirical processes. Denote them by $U_{1,N}^{(m)}$ and $U_{2,N}^{(m)}$, respectively. Then (6) can be expressed as

$$M_m = N \int \left| U_{1,N}^{(m)}(a, b) + i \cdot U_{2,N}^{(m)}(a, b) \right|^2 dS_{2m}(a, b),$$

(8)

which is convenient for exploring the asymptotic properties.

Here we focus on the case $m = 1$. The generalization to the multivariate case can be obtained analogously. Before formulating results about the limit null distribution we introduce the following notation.

Let $L^2[0, 2\pi]$ denote the Hilbert space of all complex-valued functions such that $\int_0^{2\pi} |g(t)|^2 dt < \infty$, with the inner product defined as

$$\langle g_1, g_2 \rangle = \int_0^{2\pi} g_1(t)\overline{g_2(t)} dt.$$

Let also $||\cdot||_{L^2}$ denote the norm in this space. Following the idea from Jammalamadaka et al. (2019), and moving to polar coordinates, our statistics can be expressed as

$$M_1 = ||Z_N||^2_{L^2},$$

where $Z_N(\alpha) = \sqrt{N}(U^{(1)}_{1,N}(\cos\alpha, \sin\alpha)) + i \cdot U^{(1)}_{2,N}(\cos\alpha, \sin\alpha)), \alpha \in [0, 2\pi]$.

Theorem 3.3 *Let $X_1, ..., X_N$ be an i.i.d. sample from normal $N(\mu, \sigma^2)$ distribution. Then $M_1 \xrightarrow{w} ||Z||^2_{L^2}$, where Z is a zero mean Gaussian random element from $L^2[0, 2\pi]$ with the covariance function*

$$K(\alpha_1, \alpha_2) = \lim_{N \to \infty} E Z_N(\alpha_1)\overline{Z_N(\alpha_2)} = E \Xi(X; \alpha_1)\overline{\Xi(X; \alpha_2)}, \tag{9}$$

where

$$\Xi(x; \alpha) = e^{-\frac{1}{2}\cos^2\alpha}\cos(x\sin\alpha) + e^{-\frac{1}{2}\sin^2\alpha}\cos(x\cos\alpha) - 2e^{-\frac{1}{2}} + \frac{1}{2}e^{-\frac{1}{2}}(x^2 - 1)$$
$$+ i \cdot \left(e^{-\frac{1}{2}\sin^2\alpha}\sin(x\cos\alpha) + e^{-\frac{1}{2}\cos^2\alpha}\sin(x\sin\alpha) - e^{-\frac{1}{2}}x\right). \tag{10}$$

From the Karhunen–Loève expansion of Gaussian process, the distribution of $||Z||^2_{L^2}$ can be further expressed as $\sum_{i=1}^{\infty} \lambda_i W_i^2$, where $\{\lambda_i\}$ is the sequence of positive eigenvalues of the integral operator with kernel (9) and $\{W_i\}$ is the i.i.d. sequence of standard normal random variables.

4 Simulation study

In this section we calculate empirical powers of the proposed tests and compare them with several competitors.

4.1 Testing for univariate normality

In Tables 1 and 2 we present power study results for the test based on M_1, with sample sizes $n = 20, 50$ and 100. The empirical sizes and powers are presented as percentages with '*' signifying 100%. The results are obtained using the Monte Carlo method with $N = 5000$ replications.

Among the plethora of normality tests, in order to evaluate the performance of our test versus the most popular normality tests, we selected the Shapiro–Wilk test (SW), see Shapiro and Wilk (1965), the Shapiro–Francia test (SF), see Shapiro and Francia (1972), and the Anderson-Darling test (AD), see Anderson and Darling (1952). Those tests are implemented in the R-package nortest by Gross and Ligges (2015). Additionally, we consider recent powerful tests based on empirical characteristic function (BHEP), see Henze and Wagner (1997), quantile correlation test based on the L_2-Wasserstein distance, see Del Barrio et al. (1999), the moment generating function

(HJG$_\beta$) proposed in Henze and Jiménez-Gamero (2019) and a test based on Stein's fixed point characterization proposed in Betsch and Ebner (2020).

The alternatives we consider are normal mixtures MixN(p, μ, σ^2) = $(1 - p)N(0, 1) + pN(\mu, \sigma^2)$, Student t_ν distribution, uniform $U(a, b)$ distribution, chi-squared χ^2_ν, beta B(a, b), gamma $\Gamma(a, b)$, Gumbel Gum(μ, σ) and lognormal LN(μ, σ), where all parameters are standard distribution parameters. This set of alternatives was also used in Betsch and Ebner (2020).

It can be seen from Tables 1 and 2 that the powers are reasonably high in comparison to other tests for all alternatives except for the uniform and normal mixtures. In the case of the Gumbel distribution our test outperforms the competitors, and for gamma and chi-squared it is one of the best.

4.2 Testing for normality and independence

Consider first a bivariate simple random sample $\mathbf{X} = (X_1, \ldots, X_N)$, where $X_j = (X_{j_1}, X_{j_2})$, $j = 1, \ldots, N$. Let $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_N)$ be its column-wise standardization (3). Here, the test statistic (5) becomes

$$
M_2 = \int \left(\frac{1}{N^2} \sum_{j,k} (\cos\left(a_1 \tilde{X}_{j_1} + a_2 \tilde{X}_{j_2} + b_1 \tilde{X}_{k_1} + b_2 \tilde{X}_{k_2}\right) - e^{-\frac{1}{2}} \right)^2
$$

$$
+ \left(\frac{1}{N^2} \sum_{i,j} \sin\left(a_1 \tilde{X}_{j_1} + a_2 \tilde{X}_{j_2} + b_1 \tilde{X}_{k_1} + b_2 \tilde{X}_{k_2}\right) \right)^2 dS_4(a, b),
$$

where $a = (a_1, a_2)^{\mathrm{T}}$ and $b = (b_1, b_2)^{\mathrm{T}}$ such that $\langle a, a \rangle + \langle b, b \rangle = 1$.

It is possible to further elaborate the expression above, similarly to the univariate case, however the resulting integrals do not have a convenient form, and from the computational point of view, the double sums are significantly faster then quadruple sums which would have been obtained otherwise. The integral over a sphere can be efficiently calculated using the SphericalCubature package from the R project (Nolan 2021).

In Table 3 and 4 we present powers of the new test. As competitors we chose the KS2 test, initially proposed in Koziol (1979) with data driven parameter selection introduced in Kallenberg et al. (1997), as well as the bivariate versions of BHEP and HJG$_\beta$ tests.

The reason behind choosing the KS2 test is that it is the only one in the literature known so far, to test for bivariate normality and independence. In Kallenberg et al. (1997) it is shown that it outperformsKolmogorov–Smirnov and Hoeffding's tests in most cases. The other tests are chosen since they are powerful for bivariate normality based on characteristic and moment generating functions. When they are applied to the column-wise standardized sample, they are suitable to test for normality and independence. The set of alternatives is taken from Kallenberg et al. (1997) for some choice of distribution parameters, and is given below. Unless stated otherwise, all distributions are defined for $x_i \in \mathbb{R}$, $i = 1, 2$. Distributions derived from the bivariate normal distribution inherit its parameter space $\mu_i \in \mathbb{R}$, $\sigma_i > 0$, $i = 1, 2$, $\rho \in [-1, 1]$.

Table 1 Power comparison for tests of univariate normality—Part I

Alt.	n	SW	BCMR	BHEP	AD	SF	$HJG_{2.5}$	HJG_5	$BE_1^{(1)}$	$BE_1^{(2)}$	M_1
N(1,4)	20	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5
MixN(0.3,1,0.25)	20	28	28	27	30	25	11	13	24	20	10
	50	60	60	62	68	57	16	26	56	48	20
	100	89	89	90	94	88	28	49	87	78	38
MixN(0.5,1,4)	20	40	43	42	46	48	34	33	36	33	19
	50	78	80	80	86	83	49	49	63	46	36
	100	97	98	98	99	98	69	68	91	66	57
t_3	20	34	37	34	33	40	38	37	34	34	36
	50	64	65	61	60	69	64	62	54	50	62
	100	88	89	86	85	91	86	84	76	67	83
t_5	20	19	20	18	17	22	22	22	19	19	22
	50	35	37	32	31	41	40	38	29	29	36
	100	56	58	50	48	63	59	55	41	37	56
t_{10}	20	10	11	9	9	12	12	12	10	10	12
	50	16	17	13	12	20	20	19	13	14	18
	100	22	24	16	15	28	28	26	16	15	24
$U(-\sqrt{3}, \sqrt{3})$	20	21	17	13	17	8	0	0	3	2	1
	50	75	70	55	58	47	0	0	7	2	0
	100	*	99	95	95	97	0	0	32	3	3
χ_5^2	20	43	44	42	38	42	33	36	44	44	42
	51	88	88	84	80	85	65	76	87	86	86
	100	*	*	99	99	*	91	98	99	99	*
χ_{15}^2	18	17	18	17	17	18	16	17	18	18	20
	50	43	42	39	34	40	31	37	45	46	47
	100	75	74	68	61	71	54	68	78	78	74
B(1, 4)	22	60	60	53	53	54	30	35	52	49	41
	50	98	98	94	95	97	57	76	94	92	90
	100	*	*	*	*	*	89	99	*	*	*
B(2, 5)	20	16	16	16	14	14	9	11	15	15	14
	50	50	47	45	39	40	16	25	44	42	38
	100	90	89	80	76	82	29	54	80	78	80

- a bivariate normal distribution $BivN(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ with density

$$g_1(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho)^2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)},$$

Table 2 Power comparison for tests of univariate normality—Part II

Alt.	n	SW	BCMR	BHEP	AD	SF	$HJG_{2.5}$	HJG_5	$BE_1^{(1)}$	$BE_1^{(2)}$	M_1
$\Gamma(1,5)$	20	83	83	77	77	80	57	63	78	76	72
	50	*	*	*	*	*	91	97	*	*	99
	100	*	*	*	*	*	*	*	*	*	*
$\Gamma(5,1)$	20	24	24	23	20	24	20	22	25	25	25
	50	60	59	53	49	58	42	50	63	62	62
	100	90	90	85	81	88	69	83	91	91	93
Gum(1,2)	20	31	32	31	28	32	28	30	33	33	34
	51	68	69	68	62	68	53	66	73	72	73
	100	95	95	93	88	95	84	90	94	96	97
LN(0,1)	20	93	93	91	90	91	78	83	91	90	87
	50	*	*	*	*	*	99	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*

- a mixture of bivariate normal distributions NMixA(ρ) with density

$$g_2(x_1, x_2, \rho) = \frac{1}{2}g_1(x_1, x_2; 0, 0, 1, 1, \rho) + \frac{1}{2}g_1(x_1, x_2; 1, 1, 1, 1, 0.9);$$

- a mixture of bivariate normal distributions NMixB(ρ) with density

$$g_3(x_1, x_2, \rho) = \frac{1}{2}g_1(x_1, x_2; 0, 0, 1, 1, \rho) + \frac{1}{2}g_1(x_1, x_2; 0, 0, 1, 1, -\rho);$$

- a bivariate lognormal distributions LogN(σ_1, σ_2, ρ) with density

$$g_4(x_1, x_2; \sigma_1, \sigma_2, \rho) = \frac{b_1 b_2}{(b_1 x_1 + a_1)(b_2 x_2 + a_2)}g_1(l_1, l_2; 0, 0, \sigma_1, \sigma_2, \rho), \quad x_i > -\frac{b_i}{a_i},$$

where $l_i = \log(b_i x_i + a_i)$, $a_i = e^{\sigma_i^2/2}$, $b_i = \sqrt{e^{2\sigma_i^2} - e^{\sigma_i^2}}$, $i = 1, 2$.
- a $\sinh^{-1}$-normal distribution $\sinh^{-1}$N($\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$) with density

$$g_5(x_1, x_2; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{b_1 b_2(w_1 + \sqrt{1 + w_1^2})(w_2 + \sqrt{1 + w_2^2})}{(1 + w_1^2 + w_1\sqrt{1 + w_1^2})(1 + w_2^2 + w_2\sqrt{1 + w_2^2})}$$
$$\times g_1(\sinh^{-1}(w_1), \sinh^{-1}(w_2); \mu_1, \mu_2, \sigma_1, \sigma_2, \rho),$$

where $w_i = b_i x_i + a_i$, $a_i = e^{\sigma_i^2/2}\sinh(\mu_i)$, $b_i = \sqrt{(e^{\sigma_i^2} - 1)(e^{\sigma_i^2}\cosh(2\mu_i) + 1)}$, $i = 1, 2$.

- a generalized Burr-Pareto-logistic distribution GBPL(α, β) with standard normal marginals, with density

$$
\begin{aligned}
g_6(x_1, x_2; \alpha, \beta) = & \frac{(\alpha+1)\varphi(x_1)\varphi(x_2)}{\alpha \Phi(x_1)\Phi(x_2)} \left(\frac{1+\beta}{(\Phi(x_1))^{-\frac{1}{\alpha}} + (\Phi(x_2))^{-\frac{1}{\alpha}} - 1)^{\alpha+2}} \right. \\
& + \frac{4\beta}{2(\Phi(x_1))^{-\frac{1}{\alpha}} + 2(\Phi(x_2))^{-\frac{1}{\alpha}} - 3)^{\alpha+2}} \\
& - \frac{2\beta}{2(\Phi(x_1))^{-\frac{1}{\alpha}} + (\Phi(x_2))^{-\frac{1}{\alpha}} - 2)^{\alpha+2}} \\
& \left. - \frac{2\beta}{(\Phi(x_1))^{-\frac{1}{\alpha}} + 2(\Phi(x_2))^{-\frac{1}{\alpha}} - 2)^{\alpha+2}} \right), \quad \alpha > 0, \beta \in [-1, 1],
\end{aligned}
$$

where $\Phi(x)$ and $\varphi(x)$ are the standard normal distribution function and density;
- a Morgenstern distribution Morg(α), with standard normal marginals, with density

$$
g_7(x_1, x_2; \alpha) = \varphi(x_1)\varphi(x_2)\left(1 + \alpha\big(2\Phi(x_1) - 1\big)\big(2\Phi(x_1) - 1\big)\right), \quad \alpha \in [-1, 1];
$$

- a Pearson type VII distribution PearVII(α) with density

$$
g_8(x_1, x_2; \alpha) = \frac{\alpha}{2\pi}\left(1 + \frac{1}{2}(x_1^2 + x_2^2)\right)^{\alpha+1}, \quad \alpha > 0.
$$

Methods of generating random variates from these distributions are available in Johnson (1987) and Cook and Johnson (1986).

From Tables 3 and 4 we can see that our new test is the most powerful for the bivariate normal, normal mixtures (except for $\rho = \pm 0.5$ for the mixture B), GBPL and Morgenstern alternatives. Against the $\sinh^{-1}$ alternative it is the best for $\rho = \pm 0.5$, while for the lognormal alternative it is second best after the BHEP. It is interesting to note that in case of the normal mixture B our test performs better in comparison to others when ρ gets closer to zero, while in the case of the $\sinh^{-1}$ alternative it is the other way around, i.e. it performs better when ρ is far from zero.

Consider now testing for trivariate normality and independence.

In Table 5 we present the powers of our tests, as well as those of BHEP and HJG$_\beta$ tests. We performed testing against trivariate variants of some of the alternatives considered in the bivariate case. The labels are self-explanatory and the densities can be easily derived from the bivariate counterparts with possible exception of the trivariate Morgenstern distribution which can be found in Ota and Kimura (2021).

The conclusions are more or less similar to the bivariate case, for the alternatives considered. Worth mentioning is that in the case of the Morgenstern alternative our test significantly outperforms all competitors.

Table 3 Powers of tests for bivariate normality and independence—Part I

Alt.	ρ	n = 20					n = 50					n = 100				
		BHEP	HJG$_5$	HJG$_{2.5}$	KS2	M$_2$	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M$_2$	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M$_2$
BivN (0, 0, 1, 1, ρ)	0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	0.1	6	6	7	7	7	7	6	8	7	10	10	6	12	11	16
	0.3	13	14	24	27	24	28	19	46	41	53	52	28	77	74	82
	0.5	33	32	58	57	60	76	55	93	92	95	98	87	*	*	*
	− 0.1	5	6	7	7	7	7	6	9	7	10	8	7	12	10	15
	− 0.3	13	14	24	22	24	28	18	44	40	52	52	34	78	78	82
	− 0.5	34	32	59	58	61	75	53	92	91	95	97	92	*	*	*
NMixA(ρ)	0	63	43	68	71	76	97	69	95	96	98	*	96	*	*	*
	0.1	68	49	75	76	82	98	77	98	98	99	*	99	*	*	*
	0.3	79	63	88	88	92	*	91	*	*	*	*	*	*	*	*
	0.5	91	77	96	97	98	*	99	*	*	*	*	*	*	*	*
	− 0.1	59	38	61	64	69	95	62	91	93	96	*	92	*	*	*
	− 0.3	53	34	51	55	58	92	52	80	86	91	*	83	98	99	98
	− 0.5	51	31	42	48	49	91	46	69	79	83	*	74	93	98	98
NMixB(ρ)	0	9	6	10	10	13	17	6	17	17	27	29	8	37	39	50
	0.1	9	7	10	9	12	17	7	18	17	28	31	9	38	39	51
	0.3	13	9	13	13	16	25	11	22	20	32	47	17	44	42	59
	0.5	19	15	19	19	21	43	19	32	31	43	76	32	54	54	71
	− 0.1	9	7	11	10	13	17	6	17	17	27	31	10	38	39	51
	− 0.3	13	9	14	13	16	24	10	22	21	32	48	18	44	42	58
	− 0.5	19	13	19	19	21	43	19	31	31	42	76	32	55	52	72

Table 3 continued

Alt.	ρ	$n = 20$					$n = 50$					$n = 100$				
		BHEP	HJG$_5$	HJG$_{2.5}$	KS2	M_2	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M_2	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M_2
Sinh^{-1}N $(0, 0, 1, 1, \rho)$	0	51	50	40	31	36	87	84	76	69	73	99	98	96	91	95
	0.1	53	53	42	34	39	89	85	79	72	76	99	98	96	91	95
	0.3	59	58	53	42	51	93	89	89	80	89	*	*	99	97	99
	0.5	72	70	74	61	75	98	96	98	95	99	*	*	*	*	*
	− 0.1	54	51	41	34	38	88	86	79	72	76	99	98	96	91	96
	− 0.3	59	57	53	41	52	93	90	89	81	89	*	99	99	97	*
	− 0.5	72	71	74	62	75	98	96	98	96	99	*	*	*	*	*
Sinh^{-1}N $(0, 2, 1, 0.5, \rho)$	0	53	51	41	33	37	88	85	77	70	74	99	98	96	90	95
	0.1	53	52	42	33	39	89	86	79	71	76	99	98	96	92	96
	0.3	58	57	53	41	52	93	89	89	81	90	*	99	99	97	99
	0.5	71	70	74	62	75	98	95	98	95	99	*	*	*	*	*
	− 0.1	53	51	41	33	38	89	85	79	71	75	*	98	97	92	96
	− 0.3	59	58	53	42	51	93	90	89	80	89	*	99	99	97	99
	− 0.5	72	72	75	63	76	98	96	98	95	99	*	*	*	*	*

Table 4 Powers of tests for bivariate normality and independence—Part II

Alt.	ρ	$n=20$					$n=50$					$n=100$				
		BHEP	HJG_5	$HJG_{2.5}$	KS2	M_2	BHEP	$HJG_{2.5}$	HJG_5	KS2	M_2	BHEP	$HJG_{2.5}$	HJG_5	KS2	M_2
LogN $(1, 1, \rho)$	0	97	81	74	60	83	*	*	*	94	*	*	*	*	*	*
	0.1	97	81	75	60	84	*	*	*	94	*	*	*	*	*	*
	0.3	97	85	80	63	88	*	*	*	94	*	*	*	*	*	*
	0.5	98	89	87	71	93	*	*	*	97	*	*	*	*	*	*
	-0.1	96	80	75	60	82	*	99	*	96	*	*	*	*	*	*
	-0.3	97	81	79	61	86	*	*	*	96	*	*	*	*	*	*
	-0.5	98	85	88	68	93	*	*	*	98	*	*	*	*	*	*
LogN $(0.05, 0.5, \rho)$	0	97	81	74	60	82	*	*	*	94	*	*	*	*	*	*
	0.1	97	81	75	60	83	*	*	*	93	*	*	*	*	*	*
	0.3	97	84	79	63	88	*	*	*	94	*	*	*	*	*	*
	0.5	98	88	87	71	93	*	*	*	97	*	*	*	*	*	*
	-0.1	96	80	75	60	83	*	99	*	94	*	*	*	*	*	*
	-0.3	97	81	79	62	86	*	*	*	96	*	*	*	*	*	*
	-0.5	98	86	89	69	94	*	*	*	98	*	*	*	*	*	*

Alt.	α	$n=20$					$n=50$					$n=100$				
		BHEP	HJG_5	$HJG_{2.5}$	KS2	M_2	BHEP	$HJG_{2.5}$	HJG_5	KS2	M_2	BHEP	$HJG_{2.5}$	HJG_5	KS2	M_2
GBPL $(\alpha, -1)$	1	22	22	32	30	34	45	36	64	59	69	*	*	*	*	*
	2	8	10	11	9	9	10	12	15	12	13	*	95	*	*	*
	5	8	6	8	7	9	14	6	8	8	13	92	69	98	98	99
	10	12	7	12	11	15	22	6	19	19	33	84	48	94	95	98

Table 4 continued

Alt.	ρ	n = 20					n = 50					n = 100				
		BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M$_2$	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M$_2$	BHEP	HJG$_{2.5}$	HJG$_5$	KS2	M$_2$
GBPL (α, 1)	1	71	85	56	86	89	99	88	*	*	*	*	*	*	*	*
	2	44	63	35	63	69	87	61	95	96	98	*	95	*	*	*
	5	27	40	20	39	46	61	33	77	77	85	92	69	98	98	99
	10	24	32	16	31	39	50	21	63	63	76	44	10	42	46	61
Morg(α)	0.5	8	9	6	8	10	12	6	13	13	19	22	8	25	26	36
	0.75	13	14	7	13	18	23	8	25	24	38	46	13	53	55	68
	1	18	21	9	20	28	41	12	44	44	62	73	24	81	85	91
	− 0.5	8	8	6	8	10	12	6	12	12	19	21	7	23	24	35
	− 0.75	12	14	8	13	17	23	8	25	25	38	48	13	53	56	70
	− 1	17	21	9	20	28	40	12	44	45	63	74	24	82	85	92
PearVII(α)	1	72	69	74	65	53	97	96	94	93	92	*	*	*	99	*
	2	33	37	43	35	30	65	73	66	64	56	90	91	86	85	78
	5	10	16	19	15	11	16	29	24	24	16	24	44	35	36	22
	10	6	10	11	10	8	9	16	13	12	9	10	21	17	17	10

Table 5 Powers of tests for trivariate normality and independence

Alt.	ρ_{12}	ρ_{23}	ρ_{13}	$n=20$				$n=50$				$n=100$			
				BHEP	HJG$_5$	HJG$_{2.5}$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3
TrivN$(\rho_{12}, \rho_{13}, \rho_{23})$	0.1	0.1	0.1	7	6	7	8	9	6	11	14	12	7	19	27
	0.3	0.3	0.3	20	17	34	42	44	29	72	80	79	47	97	98
	0.5	0.5	0.5	55	47	79	86	94	79	99	*	*	99	*	*
	0.1	0.5	0.1	23	17	34	47	55	24	79	90	91	42	99	*
	0.1	0.5	0.5	47	28	62	79	92	51	98	*	*	86	*	*
NMixA$(\rho_{12}, \rho_{23}, \rho_{13})$	0.1	0.1	0.1	89	69	92	96	*	95	*	*	*	*	*	*
	0.3	0.3	0.3	95	81	98	99	*	99	*	*	*	*	*	*
	0.5	0.5	0.5	98	92	*	*	*	*	*	*	*	*	*	*
	− 0.1	− 0.1	− 0.1	84	57	82	88	99	88	99	*	*	*	*	*
	− 0.3	− 0.3	− 0.3	81	52	73	80	99	80	96	98	*	*	*	*
	− 0.5	− 0.5	− 0.5	88	50	67	74	*	77	93	97	*	*	*	*
	0.1	0.5	0.1	95	86	97	99	*	98	*	*	*	*	*	*
	0.1	0.5	0.5	98	85	99	*	*	*	*	*	*	*	*	*

Table 5 continued

Alt.	ρ_{12}	ρ_{23}	ρ_{13}	$n=20$				$n=50$				$n=100$			
				BHEP	HJG$_5$	HJG$_{2.5}$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3
NMixB ($\rho_{12}, \rho_{23}, \rho_{13}$)	0.1	0.1	0.1	13	6	13	21	29	7	29	49	58	11	62	82
	0.3	0.3	0.3	23	13	20	27	54	18	43	58	88	32	76	88
	0.5	0.5	0.5	53	24	32	38	96	38	62	75	*	63	90	97
	− 0.1	− 0.1	− 0.1	13	7	13	20	29	8	31	49	58	11	62	81
	− 0.3	− 0.3	− 0.3	23	13	19	26	53	18	43	57	88	31	75	89
	− 0.5	− 0.5	− 0.5	54	25	33	39	95	38	62	75	*	64	90	97
	0.1	0.5	0.1	22	12	19	25	52	17	42	58	87	28	74	89
	0.1	0.5	0.5	35	18	26	33	79	27	52	67	99	45	84	94

Alt.	α_1	α_2	α_3	α_4	BHEP	HJG$_5$	HJG$_{2.5}$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3
Morg ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$)	0.5	0.3	0	0.1	5	5	6	8	8	5	8	12	16	5	14	25
	0.8	0.1	0.1	0.1	9	6	8	12	16	6	15	28	34	6	33	57
	− 0.8	− 0.1	0.1	0.1	9	6	8	13	16	6	15	28	34	6	33	58
	0.9	0.1	0.1	0	11	6	10	16	20	6	18	35	43	6	43	70
	− 0.9	− 0.1	0.1	0	10	6	10	15	20	6	19	35	42	6	43	70
	0.5	0.5	0	0	6	5	6	8	9	5	8	12	15	6	14	25
	0.5	0.1	0.1	0	6	5	6	8	8	4	7	12	15	5	12	24
	0.5	0.3	0.1	0	7	5	6	8	9	6	8	12	15	6	14	24
	0.3	0.1	0.1	0	5	5	6	6	6	5	6	7	8	5	7	12

Alt.	α	BHEP	HJG$_5$	HJG$_{2.5}$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3
PearVII (α)	1	82	84	82	73	99	99	98	97	*	*	*	*
	2	42	54	50	37	78	84	82	66	97	97	96	89
	5	13	23	21	13	20	39	36	19	35	55	52	26
	10	8	13	12	8	9	19	17	8	11	26	23	10

4.3 Testing for multivariate normality

Statistical tests for multivariate normality are usually applied to the following standardization of the multivariate simple random sample $\mathbf{X}$. Let $\check{\mathbf{X}} = \widehat{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \mathbf{\Theta})$, where $\widehat{\Sigma}^{-\frac{1}{2}}$ is the unique square root of the inverse of the sample covariance matrix $\widehat{\Sigma}$.

Our test is no exception and if we apply it to $\check{\mathbf{X}}$ instead of $\widetilde{\mathbf{X}}$ we can use it to test multivariate normality.

It is easy to see that, like in the original case, the test statistic is affine invariant under the null hypothesis of multivariate normality, and hence the null distribution can be approximated using the Monte Carlo methods.

In Tables 6 and 7 we present the powers of our tests for testing bivariate and trivariate normality, respectively. We also present the powers of corresponding BHEP and HJG_β tests as competitors. The alternatives are chosen from the lists in Madukaife (2021) and Ebner and Henze (2020). The labels are self-explanatory and '$\times$' signifies that the components are independent.

From the tables we can see that no test is uniformly the best. Our test is never the least powerful, for most alternatives it is in the middle, and for couple of alternatives it is the most powerful. All the other tests have cases when they are the best as well as the cases when they are the worst. It is interesting to note that in the case of bivariate beta(0.5,0.5) alternative both our test and HJG tests have powers below the level of significance for $n = 50$. However, the power of our test exceeds this level for $n = 100$, while the HJG powers remain the same.

5 Applications

In this section we consider the application of our tests. We put emphasis on the normality and independence null hypothesis, as situations where we test for univariate or multivariate normality are quite common.

5.1 Model specification tests

In multivariate regression and time series analysis it is often the case that model errors and innovations are assumed to have a multivariate normal distribution with independent components. Such models are proposed and/or discussed in Breiman and Friedman (1997), Zivot and Wang (2006), and Francq and Zakoian (2019). Then our test for normality and independence can be used as a model specification test. The errors are unobservable and hence the testing must be applied to model residuals which are not an i.i.d. sample. To demonstrate the testing procedure would thus require some modification of the test in order to obtain p-values, which is beyond the scope of this paper but worth considering for future research.

Other important applications, worth considering in the future, include novelty detection methods based on orthogonal expansion representation of random elements from a certain Hilbert space (see Rafajłowicz 2021); cluster analysis with finite Gaussian

Table 6 Power comparison of tests for bivariate normality

Alternative	n = 20				n = 50				n = 100			
	BHEP	$HJG_{2.5}$	HJG_5	M_3	BHEP	$HJG_{2.5}$	HJG_5	M_3	BHEP	$HJG_{2.5}$	HJG_5	M_3
$N_2(0, 5I)$	5	5	5	5	5	5	5	5	5	5	5	5
$t_2 \times t_2$	67	67	67	68	95	95	95	95	*	*	*	*
$B(0.5, 0.5) \times B(0.5, 0.5)$	48	1	1	1	99	0	0	1	*	0	0	11
$B(0.5, 0.5) \times t_2$	59	44	43	42	98	75	75	75	*	94	94	95
$B(3, 2) \times \chi^2_{10}$	17	11	12	15	45	23	29	43	83	39	60	83
$\Gamma(5, 1) \times W(2, 1)$	20	22	16	18	54	33	44	60	89	55	80	94
$\Gamma(5, 1) \times t_2$	51	51	51	52	87	82	84	87	88	55	80	94
$0.1N_2(0, I) + 0.9N_2(3, I)$	38	30	34	40	83	50	71	98	98	78	98	*
$0.5N_2(0, I) + 0.5N_2(0, 9I)$	46	39	40	41	87	56	57	61	99	76	78	87
PearVII(2)	33	41	41	39	65	70	71	68	90	90	90	88
PearVII(5)	11	16	16	15	18	28	28	24	27	42	41	33
PearVII(10)	6	10	10	9	8	15	15	12	10	21	21	15

Table 7 Power comparison of tests for trivariate normality

Alternative	$n=20$				$n=50$				$n=100$			
	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3	BHEP	HJG$_{2.5}$	HJG$_5$	M_3
$N_3(0,5I)$	5	5	5	5	5	5	5	5	5	5	5	5
$t_2 \times t_2 \times t_2$	72	77	78	79	*	*	*	*	*	*	*	*
$t_2 \times t_5 \times t_{10}$	38	48	47	47	75	81	81	78	96	96	96	95
$t_5 \times t_5 \times B(2,3)$	16	22	22	21	33	42	43	39	61	64	64	61
$0.1N_3(0,I)+0.9N_3(3,I)$	39	32	35	43	89	47	72	93	99	65	98	*
$0.5N_3(0,I)+0.5N_3(0,9I)$	63	53	54	59	98	72	77	83	*	89	93	98
PearVII(2)	39	51	51	51	78	82	84	82	97	97	97	97
PearVII(5)	12	20	20	19	20	36	37	33	38	54	54	46
PearVII(10)	7	11	11	11	9	17	17	14	12	26	26	19

mixtures (see Pan and Shen 2007; Yeung et al. 2001); and image processing (see Rafajłowicz and Rafajłowicz 2010; Rafajłowicz and Wietrzych 2010).

5.2 Random matrix

In probability theory and mathematical physics, a random matrix is a matrix-valued random variable, i.e., a matrix in which some or all elements are random variables. Many important properties of physical systems can be represented mathematically as matrix problems. For example, the thermal conductivity of a lattice can be computed from the dynamical matrix of the particle particle interactions within the lattice. One of the application is to approximate a covariance matrix, as long as the underlying distribution is normal. Below we give two different representations for using $\mathcal{H}_0^{(m)}$ hypothesis in the context of the correlation matrix $\mathbf{X}^T\mathbf{X}$.

Let Σ be a positive definite, symmetric $m \times m$ matrix, and let $\mathbf{X}^T$ be a random $m \times N$ matrix as in Eq. (2), whose columns are independent, identically distributed random vectors, each with the multivariate Gaussian distribution $N(0, \Sigma)$.

5.2.1 Wishart distribution

Assume that $N \geq m$. Then the random matrix $S = \mathbf{X}^T\mathbf{X}$ has the Wishart distribution $W(\Sigma, m, N)$, with density

$$\nu(dS) = \omega(m, N) \det(\Sigma^{-1}S)^{N/2} \exp\left(-\frac{1}{2}tr(\Sigma^{-1}S)\right),$$

for an appropriate normalizing constant $\omega(m, N)$. Under the null hypothesis $\mathcal{H}_0^{(m)}$ the matrix $\mathbf{X}^T\mathbf{X}$ has the Wishart distribution $W(I, m, N)$, where I is the $m \times m$ identity matrix.

5.2.2 Marchenko–Pastur distribution

The limit of the empirical spectral measure of Wishart matrices, with $\mathcal{H}_0^{(m)}$ assumption, was found in Marchenko and Pastur (1967) by Vladimir Marchenko and Leonid Pastur. Assume that $\Sigma \sim N(0, I)$. If $m, N \to \infty$ in such a way that $m/N \to \lambda \in (0, 1)$, then the empirical spectral distribution $\mathbf{X}^T\mathbf{X}$ converges weakly to the Marchenko–Pastur distribution with density (with respect to the Lebesgue measure)

$$\nu(dx) = \frac{1}{2\pi x} \sqrt{4\lambda - (x - (1 + \lambda))^2} \, dx,$$

supported on the interval $((1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2)$.

5.3 Real data example

As an illustrative example we consider the data from Stevens (2012) that appeared originally in Ambrose (1985). The data in Table 8 represent average ratings on three

Table 8 Performance aspects ratings

Group 1	Rhythm	3.2	3.7	4.7	4.1	2.4	2.7	3.4	4.1	4.2	1.9	2.9	3.5
	Intonation	4.2	3.9	5.0	3.5	3.4	3.1	4.0	4.2	3.4	3.5	3.2	3.1
	Tempo	2.8	3.1	2.9	2.8	2.8	2.7	2.7	3.7	4.2	3.7	3.3	3.6
Group 2	Rhythm	1.7	3.5	4.0	1.7	2.2	3.0	2.2	2.2	2.3	1.3	1.7	
	Intonation	1.7	1.8	1.8	1.6	3.1	3.3	1.8	3.4	4.3	2.5	2.8	
	Tempo	2.8	3.1	3.1	3.1	2.8	3.0	2.6	4.2	4.0	3.5	3.3	

performance aspects (rhythm, intonation and tempo) of two groups of elementary school children that received instruction in clarinet. The experimental group (1) had programmed instruction, while the control group (2) had traditional classroom instruction.

Suppose a researcher wants to examine the effects of instruction on these performance aspects. If these aspects are normally distributed and independent random variables, a separate univariate analysis of variance for each variable is recommended. However, if this is not the case, the researcher should either use some nonparametric procedure (in case of non-normality) or a multivariate analysis of variance (in presence of correlation). Hence, our test is handy to make a decision about the appropriate procedure.

Applying the test on data from Table 8 we get the p-values of 0.12 (group 1) and 0.19 (group 2) and therefore the recommendation for the researcher is to examine the effects using the univariate analysis of variance.

Appendix

Proof of Theorem 2.1 Our proof is based on the analysis of the characteristic function. We denote the characteristic function of W by $\varphi_W(\cdot)$. We write $(a, b) = r(\tilde{a}, \tilde{b})$ where $(\tilde{a}, \tilde{b})$ belongs to the unit sphere of $\mathbb{R}^{n+m}$, i.e. $r = \sqrt{\|a\|^2 + \|b\|^2}$.

Thus for $r > 0$ and $t \in \mathbb{R}$, we have

$$\varphi_{\frac{\langle a, \mathrm{X} \rangle + \langle b, \mathrm{Y} \rangle + A + B}{r}}(t) = \varphi_{\langle \tilde{a}, \mathrm{X} \rangle + \langle \tilde{b}, \mathrm{Y} \rangle + \frac{A+B}{r}}(t). \tag{11}$$

Under the hypothesis, the left hand side of (11) does not depend on $(\tilde{a}, \tilde{b})$ (equivalently, it depends on r), and thus the limit on the right hand side

$$\varphi_{\langle \tilde{a}, X \rangle + \langle \tilde{b}, Y \rangle + \frac{A+B}{r}}(t) \xrightarrow{r \to +\infty} \varphi_{\langle \tilde{a}, X \rangle + \langle \tilde{b}, Y \rangle}(t),$$

does not depend on $(\tilde{a}, \tilde{b})$. In particular, we have that the distribution of a statistic

$$\langle a, X \rangle + \langle b, Y \rangle = (\langle \tilde{a}, X \rangle + \langle \tilde{b}, Y \rangle) \sqrt{\|a\|^2 + \|b\|^2}$$

depends on $\|a\|^2 + \|b\|^2$ only. Let

$$h(\|a\|^2 + \|b\|^2) = E e^{i(\langle a, X \rangle + \langle b, Y \rangle)}.$$

Because of the independence of X and Y, we may write

$$h(\|a\|^2 + \|b\|^2) = E e^{i \langle a, X \rangle} E e^{i \langle b, Y \rangle} = \varphi_X(a) \varphi_Y(b). \tag{12}$$

Evaluating (12) first when $a = \mathbf{0} \in \mathbb{R}^m$ and then when $b = \mathbf{0} \in \mathbb{R}^n$, we get

$$h(\|b\|^2) = \varphi_Y(b) \text{ and } h(\|a\|^2) = \varphi_X(a),$$

respectively. Substituting this into (12), we obtain

$$h(\|a\|^2 + \|b\|^2) = h(\|a\|^2) h(\|b\|^2).$$

Note that $h(\cdot)$ is continuous, hence by multiplicative Cauchy functional equation we get

$$h\left(\|a\|^2 + \|b\|^2\right) = e^{c(\|a\|^2 + \|b\|^2)}.$$

Substituting $a = (a_1, 0, 0, \ldots, 0)$ and $b = \mathbf{0}$ in this equation, we see that it can be read as

$$E e^{i X_1 a_1} = e^{c a_1^2}, \text{ i.e. } X_1 \text{ has a normal distribution, with zero mean.}$$

Using this line of reasoning to other random variables, we see that X_i and Y_j have the same normal distribution, with zero mean. The independence of random variables $X_1, \ldots, X_m$ follows from the observation that for all $a = (a_1, \ldots, a_m) \in \mathbb{R}^m$

$$\varphi_X(a) = e^{c \sum_{j=1}^m a_j^2} = \varphi_{X_1}(a_1) \ldots \varphi_{X_m}(a_m).$$

$\square$

Proof of Proposition 2.2 $(i) \Rightarrow (ii)$. We see that the distribution of

$$\langle a, X \rangle + \langle b, Y \rangle = \sqrt{\|a\|^2 + \|b\|^2} \frac{\langle a, X \rangle + \langle b, Y \rangle}{\sqrt{\|a\|^2 + \|b\|^2}}$$

depends only on $\|a\|^2 + \|b\|^2$, which by Theorem 2.1 implies that X_i and Y_j are independent and have the same normal distribution $N(0, 1)$, because we assume that $E(X_i^2) = 1$, $E(Y_j^2) = 1$.

$(ii) \Rightarrow (i)$. We compute the characteristic function

$$E e^{i\langle a, X \rangle + i\langle b, Y \rangle} = e^{-(\|a\|^2 + \|b\|^2)/2},$$

from which we see that condition (i) is satisfied. $\qquad\square$

Proof of Proposition 3.1 Let us calculate the integral on the right hand side of (5).

$$M_1 = N \int \left| \frac{1}{N^2} \sum_{i,j} e^{i(a\tilde{X}_i + b\tilde{X}_j)} - e^{-\frac{1}{2}} \right|^2 dS_2(a, b)$$

$$= N \int V^2(a, b) dS_2(a, b).$$

Since

$$V^2(a, b) = \frac{1}{N^4} \sum_{i,j,k,l} \left(\left(\cos(a\tilde{X}_i + b\tilde{X}_j) - e^{-\frac{1}{2}} \right) \left(\cos(a\tilde{X}_k + b\tilde{X}_l) - e^{-\frac{1}{2}} \right) \right.$$

$$\left. + \sin(a\tilde{X}_i + b\tilde{X}_j) \sin(a\tilde{X}_k + b\tilde{X}_l) \right)$$

$$= \frac{1}{N^4} \sum_{i,j,k,l} (\cos(a(\tilde{X}_i - \tilde{X}_k) + b(\tilde{X}_j - \tilde{X}_l)) - \frac{2e^{-\frac{1}{2}}}{N^2} \sum_{i,j} \cos(a\tilde{X}_i + b\tilde{X}_j)$$

$$+ e^{-1},$$

switching to polar coordinates we obtain

$$V^2(a, b) = \frac{1}{N^4} \sum_{i,j,k,l} (\cos(\cos\alpha(\tilde{X}_i - \tilde{X}_k) + \sin\alpha(\tilde{X}_j - \tilde{X}_l))$$

$$- \frac{2e^{-\frac{1}{2}}}{N^2} \sum_{i,j} \cos(\cos\alpha\tilde{X}_i + b\sin\alpha\tilde{X}_j) + e^{-1}, \quad \alpha \in [0, 2\pi).$$

Since we need integration over S_2, we have to focus on computing the following integral

$$\int_0^{2\pi} \cos(x \cos \alpha + y \sin \alpha) d\alpha, \qquad \text{for } x, y \in \mathbb{R}.$$

By trigonometric identities the linear combination, or harmonic addition, of sine and cosine waves is equivalent to a single cosine wave with a phase shift and scaled amplitude, namely

$$x \cos \alpha + y \sin \alpha = \sqrt{x^2 + y^2} \cos(\alpha - \operatorname{atan2}(y, x)),$$

where $\operatorname{atan2}(y, x)$ is the generalization of $\arctan(y/x)$ that covers the entire circular range (we don't need a formal definition of atan2). Thus for $x, y \in \mathbb{R}$ and $xy \neq 0$ we get

$$
\begin{aligned}
\int_0^{2\pi} \cos(x \cos \alpha + y \sin \alpha) d\alpha &= \int_0^{2\pi} \cos\left(\sqrt{x^2 + y^2} \cos\left(\alpha - \operatorname{atan2}(y, x)\right)\right) d\alpha \\
&= \int_{-\operatorname{atan2}(y,x)}^{2\pi - \operatorname{atan2}(y,x)} \cos\left(\sqrt{x^2 + y^2} \cos t\right) dt \\
&= 2\pi J(\sqrt{x^2 + y^2}),
\end{aligned}
$$

where we used the following identity (see Abramowitz and Stegun 1972, p. 360):

$$2\pi J(z) = \int_0^{2\pi} e^{iz \cos \alpha} d\alpha = \int_0^{2\pi} \cos(z \cos \alpha) d\alpha = \int_0^{2\pi} \cos(z \sin \alpha) d\alpha. \quad (13)$$

If either $x = 0$ or $y = 0$, then the formula above is also true because we can use directly equation (13).

Therefore M_1 has the representation (7). □

Proof of Theorem 3.2 The proof follows from the strong consistency of the empirical characteristic function on bounded subsets of $\mathbb{R}^d$ (see Ushakov 2011, Chapter 3), as well as the strong consistency of the estimators (4) Then from (5) we get that

$$\frac{M_m}{N} \xrightarrow{a.s.} \Delta = \int |\varphi_Z(a)\varphi_Z(b) - e^{-\frac{1}{2}}|^2 dS_{2m}(a, b),$$

where Z is a random vector whose components have zero means and unit variances, and $\Delta = 0$ if and only if $|\varphi_Z(a)\varphi_Z(b) - e^{-\frac{1}{2}}| = 0$ for all a, b such that $||a||^2 + ||b||^2 = 1$, which occurs from Proposition 2.2 if and only if the components of Z are independent and normally distributed as $N(0, 1)$, i.e. when the null hypothesis is true. Hence, under any fixed alternative we have $\Delta > 0$ and the tests are consistent. □

Proof of Theorem 3.3 After switching to polar coordinates, as in the proof of Proposition 3.1, we get

$$M_1 = N \int_0^{2\pi} U_{1,N}^2(\alpha; \hat{\mu}, \hat{\sigma}) + U_{2,N}^2(\alpha; \hat{\mu}, \hat{\sigma}) d\alpha,$$

where

$$U_{1,N}(\alpha; \mu, \sigma) = U_{1,N}^{(1)}(\cos \alpha, \sin \alpha)$$
$$= \frac{1}{N^2} \sum_{i,j} \left(\cos \left(\cos \alpha \frac{X_i - \mu}{\sigma} + \sin \alpha \frac{X_j - \mu}{\sigma} \right) - e^{-\frac{1}{2}} \right)$$

$$U_{2,N}(\alpha; \mu, \sigma) = U_{2,N}^{(1)}(\cos \alpha, \sin \alpha) = \frac{1}{N^2} \sum_{i,j} \sin \left(\cos \alpha \frac{X_i - \mu}{\sigma} + \sin \alpha \frac{X_j - \mu}{\sigma} \right),$$

and $\hat{\mu}$ and $\hat{\sigma}$ are the sample mean and standard deviation, respectively.

Since the statistic is location-scale invariant, we may assume that the true parameter values are $\mu = 0$ and $\sigma = 1$. From the Taylor expansion of $U_{1,N}(\alpha; \hat{\mu}, \hat{\sigma})$ in the neighbourhood of $(\alpha; 0, 1)$ we get

$$\sqrt{N} U_{1,N}(\alpha; \hat{\mu}, \hat{\sigma})$$
$$= \sqrt{N} U_{1,N}(\alpha) + \sqrt{N} \hat{\mu} E(\sin(\cos \alpha X_1 + \sin \alpha X_2)(\cos \alpha + \sin \alpha))$$
$$+ \sqrt{N}(\hat{\sigma} - 1) E(\sin((\cos \alpha X_1 + \sin \alpha X_2)(\cos \alpha X_1 + \sin \alpha X_2)) + r_{1,N}(\alpha),$$
$$\tag{14}$$

where $U_{1,N}(\alpha) = U_{1,N}(\alpha; 0, 1)$, and $r_{1,N}(\alpha)$ is the Taylor's residual. When $N \to \infty$, $r_{1,N}(\alpha) = o_P(1)$ uniformly over α, which follows from the Bahadur representation of $(\hat{\mu}, \hat{\sigma})$, the finiteness of trigonometric functions and finite variance of normal random variables.

For a fixed α, $U_{1,N}(\alpha)$ is a V-statistic of order 2. The first projection of its kernel is

$$\psi_1(x; \alpha) = \frac{1}{2}(E(\cos(x \cos \alpha + X_2 \sin \alpha)) + \frac{1}{2} E(\cos(X_2 \cos \alpha + x \sin \alpha)) - e^{-\frac{1}{2}}$$
$$= \frac{1}{2}\left(e^{-\frac{1}{2}\cos^2 \alpha} \cos(x \sin \alpha) + e^{-\frac{1}{2}\sin^2 \alpha} \cos(x \cos \alpha)\right) - e^{-\frac{1}{2}}.$$

Under $\mathcal{H}_0^{(1)}$ we have $E \psi_1(X; \alpha) = 0$ and

$\text{Var}(\psi_1(X; \alpha)) < \infty$, therefore $U_{1,N}(\alpha)$ admits the Bahadur representation (see Korolyuk and Borovskich 2013)

$$\sqrt{N} U_{1,N}(\alpha) = \frac{2}{\sqrt{N}} \sum_{i=1}^{N} \psi_1(X_i; \alpha) + r_{2,N}(\alpha), \tag{15}$$

where $r_{2,N}(\alpha) = o_P(1)$ as $N \to \infty$. Since the parameter estimators can be expressed as

$$\sqrt{N}\hat{\mu} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i$$

$$\sqrt{N}(\hat{\sigma} - 1) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{1}{2}(X_i^2 - 1) + o_P(1),$$

combining (14) and (15), we get

$$\sqrt{N}U_{1,N}(\alpha; \hat{\mu}, \hat{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(2\psi_1(X_i; \alpha) + \frac{1}{2}(X_i^2 - 1)e^{-\frac{1}{2}}\right) + o_P(1).$$

Similarly,

$$\sqrt{N}U_{2,N}(\alpha; \hat{\mu}, \hat{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(2\psi_2(X_i; \alpha) - e^{-\frac{1}{2}}X_i\right) + o_P(1),$$

where

$$\psi_2(x; \alpha) = \frac{1}{2} \left(e^{-\frac{1}{2}\sin^2\alpha} \sin(x\cos\alpha) + e^{-\frac{1}{2}\cos^2\alpha} \sin(x\sin\alpha)\right).$$

Therefore

$$Z_N(\alpha) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Xi(X_i; \alpha) + R_N(\alpha),$$

where

$$\Xi(x; \alpha) = 2\psi_1(x; \alpha) + \frac{1}{2}e^{-\frac{1}{2}}(x^2 - 1) + i \cdot (2\psi_2(x; \alpha) - e^{-\frac{1}{2}}x),$$

which is equal to (10), and $||R_N||_{L^2}^2 = o_P(1)$, when $N \to \infty$.

It is easy to show that $E||\Xi(X; \cdot)||_{L^2}^2 < \infty$. Therefore from the central limit theorem in Hilbert spaces (van der Vaart and Wellner 1996, p. 50), we get that there exists a zero mean Gaussian random element Z from $L^2[0, 2\pi]$ with covariance function (9) such that $Z_N \xrightarrow{w} Z$.

Finally, from the continuous mapping theorem we get $M_1 \xrightarrow{w} ||Z||_{L^2}^2$, which ends the proof. $\qquad\square$

References

Abramowitz M, Stegun IA (eds) (1972) Bessel Functions J and Y, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover, New York, pp 358–364

Ambrose A (1985) The development and experimental application of programmed materials for teaching clarinet performance skills in college woodwind techniques courses. Unpublished doctoral dissertation, University of Cincinnati, OH

Anderson TW, Darling DA (1952) Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. Ann Math Stat 23:193–212

Baringhaus L, Henze N (1988) A consistent test for multivariate normality based on the empirical characteristic function. Metrika 35(6):339–348

Betsch S, Ebner B (2020) Testing normality via a distributional fixed point property in the Stein characterization. TEST 29:105–138

Breiman L, Friedman JH (1997) Predicting multivariate responses in multiple linear regression. J R Stat Soc B 59(1):3–54

Cook RD, Johnson ME (1986) Generalized Burr-Pareto-logistic distributions with applications to a uranium exploration data set. Technometrics 28(2):123–131

Cramér H (1928) On the composition of elementary errors: II. Statistical applications. Skandinavisk Aktuarietidskrift 11:141–180

Del Barrio E, Cuesta-Albertos JA, Matrán C, Rodríguez-Rodríguez JM (1999) Tests of goodness of fit based on the L_2-Wasserstein distance. Ann Stat 27(4):1230–1239

Ebner B, Henze N (2020) Tests for multivariate normality–a critical review with emphasis on weighted L2-statistics. TEST 29:845–892

Ejsmont W (2016) A characterization of the normal distribution by the independence of a pair of random vectors. Stat Probab Lett 114:1–5

Epps TW, Lawrence B (1983) A test for normality based on the empirical characteristic function. Biometrika 70(3):723–726

Francq C, Zakoian JM (2019) GARCH models: structure, statistical inference and financial applications. Wiley, New York

Gross J, Ligges U (2015) nortest: tests for normality. R package version 1.0-4

Henze N, Jiménez-Gamero MD (2019) A new class of tests for multinormality with i.i.d. and GARCH data based on the empirical moment generating function. TEST 28:499–521

Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. J Multivar Anal 62(1):1–23

Jammalamadaka SR, Jiménez-Gamero MD, Meintanis SG (2019) A class of goodness-of-fit tests for circular distributions based on trigonometric moments. SORT 43:317–336

Johnson ME (1987) Multivariate statistical distributions. Wiley, New York

Kallenberg WCM, Ledwina T, Rafajłowicz E (1997) Testing bivariate independence and normality. Sankhyā 59(1):42–59

Korolyuk VS, Borovskich YV (2013) Theory of U-statistics, vol 273. Springer, New York

Koziol JA (1979) A smooth test for bivariate independence. Sankhyā 41(3/4):260–269

Madukaife MS (2021) BHEP class of tests for multivariate normality: an empirical comparison. Asian J Probab Stat 15(4):185–195

Marchenko VA, Pastur LA (1967) Distribution of eigenvalues in certain sets of random matrices. Matematicheskii Sbornik. Novaya Seriya 72(144):507–536

Nolan JP (2021) SphericalCubature: Numerical integration over spheres and balls in n dimensions; multivariate polar coordinates. R package version 1:5

Ota S, Kimura M (2021) Effective estimation algorithm for parameters of multivariate Farlie-Gumbel-Morgenstern copula. Jpn J Stat Data Sci 4(2):1049–1078

Pan W, Shen X (2007) Penalized model-based clustering with application to variable selection. J Mach Learn Res 8(5):1145

Rafajłowicz W (2021) Learning novelty detection outside a class of random curves with application to COVID-19 growth. J Artif Intell Soft Comput Res 11(3):195–215

Rafajłowicz E, Rafajłowicz W (2010) Testing (non-) linearity of distributed-parameter systems from a video sequence. Asian J Control 12(2):146–158

Rafajłowicz E, Wietrzych J (2010) Recognition of finite structures with application to moving objects identification. International conference on artificial intelligence and soft computing. Springer, Berlin, pp 453–461

Shapiro SS, Francia RS (1972) An approximate analysis of variance test for normality. J Am Stat Assoc 67(337):215–216

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611

Stevens JP (2012) Applied multivariate statistics for the social sciences. Routledge, New York

Székely GJ, Rizzo ML (2009) Brownian distance covariance. Ann. Appl Stat 3(4):1236–1265

Székely GJ, Rizzo ML, Bakirov NK (2007) Measuring and testing dependence by correlation of distances. Ann Stat 35(2):2769–2794

Ushakov NG (2011) Selected topics in characteristic functions. De Gruyter, Berlin

van der Vaart AW, Wellner JA (1996) Weak convergence and empirical processes. Springer, New York

Yeung KY, Fraley C, Murua A, Raftery AE, Ruzzo WL (2001) Model-based clustering and data transformations for gene expression data. Bioinformatics 17(10):977–987

Zivot E, Wang J (2006) Vector autoregressive models for multivariate time series. Modeling financial time series with S-PLUS®, 385–429