

A Jarque-Bera type test for multivariate normality based on second-power skewness and kurtosis

Namhyun Kim^{1,a}

^aDepartment of Science, Hongik University, Korea

Abstract

Desgagné and de Micheaux (2018) proposed an alternative univariate normality test to the Jarque-Bera test. The proposed statistic is based on the sample second power skewness and kurtosis while the Jarque-Bera statistic uses sample Pearson's skewness and kurtosis that are the third and fourth standardized sample moments, respectively. In this paper, we generalize their statistic to a multivariate version based on orthogonalization or an empirical standardization of data. The proposed multivariate statistic follows chi-squared distribution approximately. A simulation study shows that the proposed statistic has good control of type I error even for a very small sample size when critical values from the approximate distribution are used. It has comparable power to the multivariate version of the Jarque-Bera test with exactly the same idea of the orthogonalization. It also shows much better power for some mixed normal alternatives.

Keywords: Goodness of fit test, Jarque-Bera test, second power kurtosis, second power skewness, multivariate normality, power comparison

1. Introduction

Assessing multivariate normality is an important issue in Statistics because the classical multivariate analysis usually requires the assumption of multinormality. Consequently, there are numerous test procedures devoted to this problem in the literature. Some references for a general review are Henze and Zirkler (1990), Henze (2002), Thode (2002), and Srivastava and Mudholkar (2003). Comparative power study is conducted in Horswell and Looney (1992), Romeu and Ozturk (1993), Mecklin and Mundfrom (2005), Farrell *et al.* (2007), and Hanusz *et al.* (2018).

In general, most multivariate tests are generalization of univariate normality tests. In univariate normality tests, the Shapiro and Wilk (1965) test and the Jarque and Bera (1980) test are widely used. The Shapiro-Wilk test is a regression and correlation test, and it is known as the most powerful test in the literature (D'Agostino and Stephens, 1986).

Its approximate tests are proposed by De Wet and Venter (1972), Shapiro and Francia (1972). The Shapiro-Wilk and its approximate tests have been generalized to multivariate cases by Malkovich and Afifi (1973), Royston (1983), Fattorini (1986), Srivastava and Hui (1987), Mudholkar *et al.* (1995), Kim and Bickel (2003), Kim (2004, 2005), and Villasenor-Alva and González-Estrada (2009). Zhou and Shao (2014) proposed a new test related to Fattorini (1986).

The Jarque-Bera test is based on sample Pearson's skewness and kurtosis, and it is popular especially in econometric fields. A related statistic is proposed by D'Agostino and Pearson (1973,

¹ Department of Science, Hongik University, 94 Wausan-Ro, Mapo-Gu, Seoul 04066, Korea.
E-mail: nhkim@hongik.ac.kr

1974). They used a transformation of D'Agostino (1970) for sample skewness and a transformation of Anscombe and Glynn (1983) for sample kurtosis. Jarque-Bera type tests are generalized to multivariate cases by Doornik and Hansen (2008), Kim (2015, 2016).

Mardia (1970, 1974)'s measure of multivariate skewness and kurtosis is one of the most popular tests for multivariate normality. Mardia and Foster (1983), Kim (2020) studied sum of transformed multivariate skewness and kurtosis. Henze and Zirkler (1990) test is often recommended as a formal test for multinormality because of its invariance and consistency. According to Mecklin and Mundfrom (2005), Farrell *et al.* (2007), it also has relatively good power over a wide range of alternatives.

Desgagné and de Micheaux (2018) introduced sample second power skewness and kurtosis. And they proposed a statistic on a combination of these measures as an alternative to the Jarque-Bera statistic. The Jarque-Bera test is based on sum of squares of the classical Pearson's skewness and kurtosis, which are the third and fourth standardized sample moments, respectively. In their terminology, the Jarque-Bera test uses the third power skewness and the fourth power kurtosis. In this paper, we generalize Desgagné and de Micheaux (2018)'s univariate statistic to a multivariate version using orthogonalization or an empirical standardization of data. A simulation is conducted to investigate the size of the test when the approximate distribution of the statistic is used. Also a power comparison is performed through a simulation.

2. Test statistics

Let X_1, \dots, X_n be a sample of independent and identically distributed (i.i.d.) random variables. The Jarque-Bera statistic is based on the classical Pearson's sample skewness $\sqrt{b_1}$ and kurtosis b_2 . The statistic is

$$JB = \frac{n}{6} \left(\sqrt{b_1} \right)^2 + \frac{n}{24} (b_2 - 3)^2. \quad (2.1)$$

Likewise, D'Agostino and Pearson (1973) proposed the statistic

$$JBT = \left(T_1 \left(\sqrt{b_1} \right) \right)^2 + \left(T_2(b_2) \right)^2, \quad (2.2)$$

to test univariate normality. T_1 is a transformation of the null distribution of the skewness $\sqrt{b_1}$ to normality using a Johnson's unbounded S_U curve proposed by D'Agostino (1970). T_2 is a normal approximation for the kurtosis b_2 by Anscombe and Glynn (1983).

Desgagné and de Micheaux (2018) defined the sample second power skewness B_2 and the sample second power kurtosis K_2 . While the Pearson's skewness $\sqrt{b_1}$ is the third standardized sample moment

$$\sqrt{b_1} = \frac{1}{n} \sum_{i=1}^n Z_i^3 = \frac{1}{n} \sum_{i=1}^n |Z_i|^3 \text{sign}(Z_i), \quad (2.3)$$

they considered the second power skewness B_2 ,

$$B_2 = \frac{1}{n} \sum_{i=1}^n |Z_i|^2 \text{sign}(Z_i), \quad (2.4)$$

where Z_i is the standardized observation, $Z_i = (X_i - \bar{X})/S$ with $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, $S^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

The Pearson's sample kurtosis b_2 is the fourth standardized sample moment,

$$b_2 = \frac{1}{n} \sum_{i=1}^n |Z_i|^4. \tag{2.5}$$

Since $n^{-1} \sum_{i=1}^n |Z_i|^2 = 1$, they considered the limit $n^{-1} \sum_{i=1}^n (|Z_i|^{2+\epsilon} - 1)/\epsilon$ as $\epsilon \rightarrow 0$. As a result, the second power kurtosis,

$$K_2 = \frac{1}{n} \sum_{i=1}^n Z_i^2 \log |Z_i|, \tag{2.6}$$

is obtained.

Desgagné and de Micheaux (2018) proposed a normality test statistic to test the null hypothesis $H_0 : X \sim N(\mu, \sigma)$ with μ and σ unspecified on a combination of their sample second power skewness and kurtosis defined in (2.4) and (2.6), while the Jarque-Bera statistic in (2.1) uses the third power skewness and the fourth power kurtosis according to their terminology. Their idea is to develop Rao's score test (Rao, 1948a) on the family of the asymmetric power distribution introduced by Komunjer (2007). They showed the resulting test statistic is a combination of their second power skewness and kurtosis in (2.4) and (2.6).

Desgagné and de Micheaux (2018) showed that the limit distribution of (B_2, K_2) follows,

$$\sqrt{n} \begin{pmatrix} -2B_2 \\ -2^{-1} \left[K_2 - \frac{(2-\log 2-\gamma)}{2} \right] \end{pmatrix} \xrightarrow{d} N_2 \left(\mathbf{0}, \begin{pmatrix} 4 \left(3 - \frac{8}{\pi} \right) & 0 \\ 0 & \frac{(3\pi^2-28)}{32} \end{pmatrix} \right),$$

using the central limit theorem with $\gamma = -\Gamma'(1)/\Gamma(1) = 0.577215665 \dots$, $\Gamma(x)$ is the gamma function. Hence the statistic

$$\frac{nB_2^2}{3 - \frac{8}{\pi}} + \frac{n \left(K_2 - \frac{(2-\log 2-\gamma)}{2} \right)^2}{\frac{(3\pi^2-28)}{8}}, \tag{2.7}$$

can be used as a test statistic for normality with the asymptotic chi-squared distribution with 2 degrees of freedom, χ_2^2 . Note that the above statistic has a very similar form with the Jarque-Bera statistic in (2.1). However, the asymptotic chi-squared distribution is not good enough for small sample sizes. Hence they proposed a modified version of the statistic in (2.7) so that the distribution can be approximated by a chi-squared even for small sample sizes. To do this, they defined the sample second power net kurtosis $K_2 - B_2^2$ and showed numerically the dependency between B_2 and $K_2 - B_2^2$ is negligible. Next, they found a transformation $Z_1(B_1)$ and $Z_2(K_2 - B_2^2)$ so that they follow approximately $N(0, 1)$.

Finally, their proposed statistic to test normality based on B_2 and K_2 in (2.4), (2.6) is

$$\begin{aligned} DX &= Z_1^2(B_2) + Z_2^2(K_2 - B_2^2), \\ &= \frac{nB_2^2}{\left(3 - \frac{8}{\pi} \right) \left(1 - \frac{1.9}{n} \right)} + \frac{n \left[\left(K_2 - B_2^2 \right)^{\frac{1}{3}} - \left(\frac{(2-\log 2-\gamma)}{2} \right)^{\frac{1}{3}} \left(1 - \frac{1.026}{n} \right) \right]^2}{72^{-1} \left(\frac{(2-\log 2-\gamma)}{2} \right)^{-\frac{4}{3}} (3\pi^2 - 28) \left(1 - \frac{2.25}{n^{0.8}} \right)}. \end{aligned} \tag{2.8}$$

The statistic DX in (2.8) follows approximately χ_2^2 , for all $n \geq 10$ with high numerical precision.

In this paper, we want to generalize the statistic $D\mathbf{X}$ in (2.8) to a multivariate version based on orthogonalization or an empirical standardization using the scaled residuals of the observations. Let $\mathbf{X}_1, \dots, \mathbf{X}_n$ be a p -dimensional i.i.d. random vectors, and let $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a p -variate multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We want to test the null hypothesis

$$H_0 : \mathbf{X}_1, \dots, \mathbf{X}_n \text{ is a sample from } N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ for some } \boldsymbol{\mu} \text{ and } \boldsymbol{\Sigma}.$$

Let

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j, \quad \mathbf{S} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})^T, \quad (2.9)$$

be a sample mean vector and a sample covariance matrix, respectively. T denotes a transpose.

If $\mathbf{X}_1, \dots, \mathbf{X}_n$ are a sample from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ with

$$\mathbf{Z}_i = \mathbf{S}^{*T} (\mathbf{X}_i - \bar{\mathbf{X}}), \quad i = 1, \dots, n, \quad (2.10)$$

follow $N_p(\mathbf{0}, \mathbf{I})$ asymptotically, with $\mathbf{0}$ the null vector of order p , \mathbf{I} the identity matrix of order $p \times p$. Here \mathbf{S}^* is a matrix satisfying $\mathbf{S}^{*T} \mathbf{S} \mathbf{S}^* = \mathbf{I}$. The \mathbf{Z}_i 's are called the scaled residuals, and each component of $\mathbf{Z}_i = (Z_{1i}, Z_{2i}, \dots, Z_{pi}), Z_{ki}, i = 1, \dots, n$ follows approximately independent univariate standard normal $N(0, 1)$ under the null hypothesis. We can construct a multivariate test statistic by adding up the univariate statistic for each coordinate of the scaled residuals in (2.10). The statistic is as follows,

$$DX_M = \sum_{k=1}^p DX(k), \quad (2.11)$$

where $DX(k)$ is the statistic in (2.8) for the coordinate $(Z_{k1}, Z_{k2}, \dots, Z_{kn}), k = 1, \dots, p$. The multivariate statistic in (2.11) follows χ_{2p}^2 approximately.

Malkovich and Afifi (1973) generalized the univariate skewness $\sqrt{b_1}$ and kurtosis b_2 in (2.3) and (2.5) to multivariate cases by Roy's union-intersection principle (Roy, 1953). That is based on the fact that $\mathbf{c}'\mathbf{X}$ follows a univariate normal for all $\mathbf{c}, \mathbf{c} \neq \mathbf{0}$, if \mathbf{X} follows a multivariate normal. Hence we need to investigate all the possible linear combinations that reduce to a normal under the null hypothesis, and find the direction of giving farthest away from a normal. However, a statistic by that principle is very hard to compute when the dimension of data is large. Hence we need some particular directions to check the normality. Srivastava and Hui (1987) suggested principal component approach for the direction. The statistic in (2.11) is closely related to their idea. Villasenor-Alva and González-Estrada (2009) generalized the Shapiro-Wilk's test to a multivariate version using the above idea. Kim (2015, 2016) also used the idea to generalize the univariate Jarque-Bera test and related statistics to multivariate cases.

For comparison, we refer to some multivariate normal test statistics. First of all, we consider the multivariate version of the Jarque-Bera type statistics JB_M, JBT_M proposed by Kim (2016). The statistics are a multivariate generalization of JB in (2.1) and JBT in (2.2) respectively, with exactly the same idea of the orthogonalization. They have the asymptotic distribution χ_{2p}^2 .

Mardia (1970, 1974) defined the multivariate measure of skewness $b_{1,p}$ and kurtosis $b_{2,p}$ as,

$$b_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \{(\mathbf{X}_i - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})\}^3, \quad (2.12)$$

$$b_{2,p} = \frac{1}{n} \sum_{j=1}^n \{(\mathbf{X}_j - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})\}^2, \quad (2.13)$$

where $\bar{\mathbf{X}}$ and \mathbf{S} are defined in (2.9). It is known (Mardia, 1970) that,

$$ms = \frac{nb_{1,p}}{6} \xrightarrow{d} \chi^2 \left(\frac{p(p+1)(p+2)}{6} \right), \quad (2.14)$$

$$mk = \frac{b_{2,p} - p(p+2)}{\sqrt{\frac{8p(p+2)}{n}}} \xrightarrow{d} N(0, 1). \quad (2.15)$$

Kim (2020) proposed omnibus statistics based on $b_{1,p}$, $b_{2,p}$ in (2.12), (2.13) for testing multivariate normality using some normalizing transformations. Among them, two statistics MN and NE are considered for comparison, where

$$MN = \frac{nb_{1,p}}{6} + \left(\frac{b_{2,p} - \frac{n-1}{n+1}p(p+2)}{\sqrt{\frac{(n-3)(n-p-1)(n-p+1)}{(n+1)^2(n+3)(n+5)}8p(p+2)}} \right)^2, \quad (2.16)$$

$$NE = N_s^2(b_{1,p}) + E_k^2(b_{2,p}). \quad (2.17)$$

The transformation is as follows,

$$N_s(b_{1,p}) = \frac{b_{1,p} - \frac{6f}{n}}{\frac{6}{n} \sqrt{2f}} \quad (2.18)$$

$$E_k(b_{2,p}) = \sqrt{\frac{n}{8p(p+2)}} \left[\frac{\exp(d(b_{2,p} - p(p+2))) - 1}{d} + \frac{2p(p+2)(1-2d)}{n} \right] \quad (2.19)$$

with $f = p(p+1)(p+2)/6$, $d = -(p+8)/(3p(p+2))$. The N_s in (2.18) is a normal approximation of a χ^2 variable for the skewness $b_{1,p}$, which is used in Mardia and Foster (1983). The E_k in (2.19) is a normalizing transformation for the kurtosis $b_{2,p}$ proposed by Enomoto *et al.* (2019). The statistic MN in (2.16) is almost the same as the statistic M_p by Doornik and Hansen (2008). The statistic MN has the asymptotic distribution χ_{f+1}^2 , and the statistic NE has χ_{2p}^2 . Kim (2020) recommended NE in (2.17) as an omnibus statistic because of its relatively good power in many alternatives.

As a formal test statistic for multivariate normality, the Henze and Zirkler (1990) statistic $T_{n,\beta}$ is frequently recommended. The statistic is as follows.

$$T_{n,\beta}(\mathbf{X}_1, \dots, \mathbf{X}_n) = n(4I(\mathbf{S} \text{ is singular}) + D_{n,\beta}I(\mathbf{S} \text{ is singular})), \quad (2.20)$$

where

$$D_{n,\beta} = \frac{1}{n^2} \sum_{i,j=1}^n \exp\left(-\frac{\beta^2}{2} \|\mathbf{Y}_i - \mathbf{Y}_j\|^2\right) - 2(1 + \beta^2)^{-\frac{n}{2}} \frac{1}{n} \sum_{j=1}^n \exp\left(-\frac{\beta^2}{2(1 + \beta^2)} \|\mathbf{Y}_j\|^2\right) + (1 + 2\beta^2)^{-\frac{n}{2}}$$

Table 1: Statistics and p -values for the Rao's bark deposit data

	(N, E, W, S)				(Y_1, Y_2, Y_3)			
	DX_M	ms	mk	HZ	DX_M	ms	mk	HZ
Statistics	17.7090	20.890	-0.398	0.9990	8.181	5.493	-0.696	0.679
p -value	0.0235	0.404	0.690	0.0111	0.225	0.856	0.486	0.276

with,

$$\|\mathbf{Y}_i - \mathbf{Y}_j\|^2 = (\mathbf{X}_i - \mathbf{X}_j)' \mathbf{S}^{-1} (\mathbf{X}_i - \mathbf{X}_j),$$

and,

$$\|\mathbf{Y}_j\|^2 = (\mathbf{X}_j - \bar{\mathbf{X}})' \mathbf{S}^{-1} (\mathbf{X}_j - \bar{\mathbf{X}}).$$

β is defined as,

$$\beta = \beta_p(n) = \frac{1}{\sqrt{2}} \left(\frac{2p+1}{4} \right)^{\frac{1}{p+4}} n^{\frac{1}{p+4}}.$$

In the next section, we compare the multivariate version of DX statistic, DX_M , and the referred ones through a simulation in terms of size of the tests and power.

3. Example and simulation study

3.1. Example

We consider the famous data set in Rao (1948b). It consists of the thickness of bark deposit on 28 cork trees measured by the weight of cork borings from the four directions, north (N), east (E), west (W), and south (S). The problem is to investigate if the thickness of bark deposit varies in the four directions. Hence, he selected three constraints to examine if the thickness of bark deposit varies in the four directions,

$$Y_1 = N - E - W + S, \quad Y_2 = S - W, \quad Y_3 = N - S,$$

by the reason he explained in the paper and tested $E(Y_i) = 0, i = 1, 2, 3$ for this purpose. The assumption of multivariate normality should be valid to test the problem by applying some techniques like Hotelling's T^2 test. Pearson (1956) examined the same dataset.

For the data set (N, E, W, S) and the contrasts (Y_1, Y_2, Y_3) , the statistics DX_M in (2.11), ms , mk in the left side of (2.14), (2.15) and the Henze and Zirkler (HZ) statistic in (2.20) are computed in Table 1. The corresponding p -values are also presented. We can use the R-package 'MVN' (see Korkmaz *et al.* (2014)) or 'mvnormalTest' for the statistic ms , mk , and HZ.

According to the result, the multivariate normality of the Rao's original data is rejected at the significance level 0.05 when we use the statistics DX_M or HZ. The Mardia's statistics ms , mk do not reject it. As for the Rao's data for the contrasts, the multivariate normality cannot be rejected. The data set is also analyzed in Mardia (1975), Srivastava and Hui (1987), and Kim (2015, 2016, 2020).

3.2. Comparison of test size

A simulation is performed to compare the proposed DX_M in (2.11) with the referred statistics. First of all, we check the test size of the statistics when we use the critical values of the corresponding limit distributions.

Table 2: Test size using the approximate distribution for $p = 2$

Statistic	α	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
DX_M	0.05	0.0546	0.0515	0.0538	0.0536	0.0520	0.0488
JB_M	0.05	0.0078	0.0284	0.0402	0.0426	0.0480	0.0544
JBT_M	0.05	0.0678	0.0656	0.0620	0.0600	0.0702	0.0648
MN	0.05	0.0317	0.0490	0.0507	0.0567	0.0549	0.0553
NE	0.05	0.0059	0.0325	0.0361	0.0398	0.0448	0.0507
DX_M	0.10	0.1097	0.0975	0.1027	0.0998	0.1009	0.0960
JB_M	0.10	0.0158	0.0436	0.0532	0.0608	0.0698	0.0800
JBT_M	0.10	0.1056	0.1036	0.1058	0.1004	0.1148	0.1116
MN	0.10	0.0506	0.0713	0.0765	0.0861	0.0845	0.0896
NE	0.10	0.0182	0.0509	0.0579	0.0621	0.0669	0.0777

Table 3: Test size using the approximate distribution for $p = 5$

Statistic	α	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 100$
DX_M	0.05	0.0580	0.0472	0.0534	0.0512	0.0540	0.0526
JB_M	0.05	0.0020	0.0288	0.0416	0.0384	0.0536	0.0658
JBT_M	0.05	0.0724	0.0746	0.0710	0.0708	0.0704	0.0584
MN	0.05	0.0020	0.0215	0.0364	0.0422	0.0447	0.0506
NE	0.05	0.0286	0.0423	0.0481	0.0530	0.0574	0.0603
DX_M	0.10	0.1155	0.0929	0.1008	0.1005	0.1045	0.1017
JB_M	0.10	0.0052	0.0398	0.0618	0.0566	0.0760	0.0938
JBT_M	0.10	0.1106	0.1198	0.1196	0.1196	0.1198	0.1062
MN	0.10	0.0042	0.0365	0.0579	0.0659	0.0715	0.0907
NE	0.10	0.0917	0.0894	0.0954	0.0984	0.1028	0.1050

As we mentioned, all the statistics DX_M , JB_M , JBT_M , and NE have the limit distribution χ^2_{2p} , and MN has χ_{f+1} with $f = p(p + 1)(p + 2)/6$. The limit distribution of the Henze and Zirkler (1990) statistic in (2.20) is an infinite sum of the weighted χ^2_1 variable, and it is omitted for comparison.

For each combination of $p = 2, 5$ and the sample size $n = 10, 20, 30, 40, 50, 100$, $N = 10,000$ random samples are generated, and the statistics are computed. The samples are from $N_p(\mathbf{0}, \mathbf{I})$ since the distributions of the statistics do not depend on μ and Σ . Table 2 and Table 3 show the test size of each statistic at the significance level $\alpha = 0.05, 0.10$. The size is computed by the proportion of the rejections based on the critical values from the limit distribution.

The simulation results indicate that the statistic DX_M has a good control of type I error, and satisfies the given significance level very closely for $n \geq 10$ when we use the critical values from the approximate distribution χ^2_{2p} . The statistics JB_M , MN, NE show downward size for small sample sizes, making the tests be conservative. JBT_M has a little bigger size. Jönsson (2011) considered the JB_M test statistic and suggested a modification to eliminate the size distortion. Apparently, DX_M follows the approximate distribution closely even for a small sample size. On the other hand, the other statistics do not have the right size especially for a small sample size.

3.3. Power comparison

Next, we compare power of the statistics through a simulation. Tables 4–7 represent the empirical power of DX_M and the referred statistics. The power of JB_M , JBT_M comes from Kim (2016), and the power of MN, NE and the Henze and Zirkler (HZ) statistic from Kim (2020). A simulation is conducted at the significance level $\alpha = 0.05$ for the combinations of dimensions $p = 2, 5$ and sample sizes $n = 20, 50$. As for DX_M , the critical values from the chi-squared distribution are used since Tables 2–3 indicate that the quantiles from χ^2_{2p} give the high accuracy. Except DX_M , we used the

Table 4: Power comparison of the statistics ($\alpha = 0.05, p = 2, n = 20$)

Alternative	DX_M	JB_M	JBT_M	MN	NE	HZ
$N(0, 1)^2$	0.05	0.04	0.04	0.05	0.06	0.05
Cauchy(0, 1) ²	0.97	0.94	0.96	0.97	0.94	0.97
Logistic(0, 1) ²	0.14	0.16	0.15	0.17	0.14	0.10
$(t_2)^2$	0.71	0.72	0.68	0.74	0.67	0.64
$(t_5)^2$	0.25	0.25	0.25	0.30	0.25	0.17
Beta(1, 1) ²	0.10	0.00	0.06	0.00	0.23	0.18
Beta(2, 2) ²	0.04	0.01	0.02	0.00	0.08	0.06
Beta(1, 2) ²	0.15	0.05	0.06	0.04	0.09	0.28
exp(1) ²	0.73	0.66	0.64	0.69	0.71	0.86
Lognormal(0, 0.5) ²	0.51	0.48	0.50	0.50	0.51	0.58
Gamma(0.5, 1) ²	0.92	0.87	0.85	0.90	0.93	0.99
Gamma(5, 1) ²	0.21	0.20	0.23	0.21	0.22	0.24
$(\chi_2^2)^2$	0.38	0.34	0.37	0.37	0.38	0.45
$(\chi_{15}^2)^2$	0.14	0.18	0.16	0.16	0.16	0.16
$N(0, 1) * t_5$	0.16	0.17	0.16	0.17	0.15	0.12
$N(0, 1) * \text{Beta}(1, 1)$	0.11	0.02	0.07	0.01	0.07	0.10
$N(0, 1) * \text{exp}(1)$	0.37	0.36	0.35	0.40	0.40	0.51
$N(0, 1) * \chi_5^2$	0.27	0.24	0.26	0.20	0.19	0.22
$NMIX_2(0.5, 4, 0, 0)$	0.89	0.04	0.49	0.01	0.15	0.52
$NMIX_2(0.5, 0, 0, 0.9)$	0.19	0.17	0.18	0.18	0.15	0.14
$NMIX_2(0.5, 4, 0, 0.9)$	0.82	0.22	0.50	0.23	0.27	0.82
$NMIX_2(0.9, 4, 0, 0)$	0.73	0.66	0.61	0.55	0.54	0.64
$NMIX_2(0.9, 0, 0, 0.9)$	0.06	0.07	0.07	0.06	0.06	0.05
$NMIX_2(0.9, 4, 0, 0.9)$	0.71	0.65	0.65	0.54	0.57	0.62

Table 5: Power comparison of the statistics ($\alpha = 0.05, p = 2, n = 50$)

Alternative	DX_M	JB_M	JBT_M	MN	NE	HZ
$N(0, 1)^2$	0.05	0.04	0.05	0.05	0.05	0.05
Cauchy(0, 1) ²	1.00	1.00	1.00	1.00	1.00	1.00
Logistic(0, 1) ²	0.25	0.28	0.27	0.32	0.26	0.15
$(t_2)^2$	0.97	0.94	0.94	0.97	0.94	0.95
$(t_5)^2$	0.50	0.51	0.49	0.56	0.48	0.32
Beta(1, 1) ²	0.32	0.00	0.37	0.00	0.81	0.68
Beta(2, 2) ²	0.10	0.00	0.10	0.00	0.31	0.17
Beta(1, 2) ²	0.45	0.10	0.20	0.12	0.29	0.80
exp(1) ²	0.98	0.94	0.96	1.00	1.00	1.00
Lognormal(0, 0.5) ²	0.90	0.85	0.84	0.95	0.95	0.94
Gamma(0.5, 1) ²	1.00	0.99	0.99	1.00	1.00	1.00
Gamma(5, 1) ²	0.53	0.46	0.48	0.56	0.61	0.53
$(\chi_2^2)^2$	0.80	0.74	0.75	0.86	0.88	0.87
$(\chi_{15}^2)^2$	0.39	0.33	0.36	0.41	0.45	0.36
$N(0, 1) * t_5$	0.32	0.34	0.30	0.32	0.28	0.17
$N(0, 1) * \text{Beta}(1, 1)$	0.43	0.03	0.54	0.01	0.17	0.32
$N(0, 1) * \text{exp}(1)$	0.80	0.73	0.73	0.88	0.90	0.92
$N(0, 1) * \chi_5^2$	0.72	0.59	0.59	0.54	0.56	0.52
$NMIX_2(0.5, 4, 0, 0)$	1.00	0.08	0.99	0.01	0.36	1.00
$NMIX_2(0.5, 0, 0, 0.9)$	0.48	0.41	0.33	0.28	0.21	0.32
$NMIX_2(0.5, 4, 0, 0.9)$	1.00	0.55	0.98	0.90	0.93	1.00
$NMIX_2(0.9, 4, 0, 0)$	0.99	0.99	0.98	0.97	0.98	0.95
$NMIX_2(0.9, 0, 0, 0.9)$	0.07	0.06	0.09	0.08	0.09	0.05
$NMIX_2(0.9, 4, 0, 0.9)$	0.99	0.98	0.98	0.98	0.98	0.94

Table 6: Power comparison of the statistics ($\alpha = 0.05, p = 5, n = 20$)

Alternative	DX_M	JB_M	JBT_M	MN	NE	HZ
$N(0, 1)^2$	0.05	0.05	0.04	0.05	0.05	0.04
Cauchy(0, 1) ²	1.00	0.99	0.99	1.00	0.99	0.99
Logistic(0, 1) ²	0.12	0.14	0.14	0.18	0.07	0.08
$(t_2)^2$	0.84	0.84	0.85	0.88	0.78	0.69
$(t_5)^2$	0.26	0.28	0.28	0.34	0.16	0.14
Beta(1, 1) ²	0.05	0.02	0.03	0.00	0.29	0.10
Beta(2, 2) ²	0.04	0.02	0.02	0.00	0.16	0.06
Beta(1, 2) ²	0.08	0.05	0.06	0.03	0.08	0.18
exp(1) ²	0.66	0.61	0.64	0.79	0.58	0.81
Lognormal(0, 0.5) ²	0.49	0.49	0.49	0.60	0.39	0.49
Gamma(0.5, 1) ²	0.91	0.86	0.88	0.97	0.91	0.99
Gamma(5, 1) ²	0.17	0.17	0.16	0.21	0.09	0.17
$(\chi^2_5)^2$	0.30	0.30	0.35	0.39	0.20	0.34
$(\chi^2_{15})^2$	0.12	0.14	0.14	0.15	0.06	0.12
$N(0, 1) * t_5$	0.10	0.12	0.12	0.10	0.06	0.06
$N(0, 1) * Beta(1, 1)$	0.05	0.03	0.04	0.03	0.07	0.06
$N(0, 1) * exp(1)$	0.14	0.17	0.15	0.18	0.08	0.13
$N(0, 1) * \chi^2_5$	0.19	0.18	0.17	0.11	0.05	0.08
$NMIX_5(0.5, 4, 0, 0)$	1.00	0.04	0.72	0.03	0.09	0.17
$NMIX_5(0.5, 0, 0, 0.9)$	0.52	0.43	0.44	0.65	0.41	0.54
$NMIX_5(0.5, 4, 0, 0.9)$	0.95	0.39	0.65	0.82	0.65	0.96
$NMIX_5(0.9, 4, 0, 0)$	0.86	0.70	0.66	0.40	0.24	0.32
$NMIX_5(0.9, 0, 0, 0.9)$	0.09	0.10	0.09	0.10	0.05	0.07
$NMIX_5(0.9, 4, 0, 0.9)$	0.84	0.75	0.70	0.43	0.22	0.40

Table 7: Power comparison of the statistics ($\alpha = 0.05, p = 5, n = 50$)

Alternative	DX_M	JB_M	JBT_M	MN	NE	HZ
$N(0, 1)^2$	0.05	0.05	0.06	0.04	0.05	0.04
Cauchy(0, 1) ²	1.00	1.00	1.00	1.00	1.00	1.00
Logistic(0, 1) ²	0.24	0.26	0.24	0.41	0.32	0.14
$(t_2)^2$	1.00	0.99	0.99	1.00	1.00	1.00
$(t_5)^2$	0.55	0.56	0.54	0.68	0.67	0.30
Beta(1, 1) ²	0.07	0.01	0.10	0.00	0.83	0.50
Beta(2, 2) ²	0.05	0.01	0.04	0.00	0.42	0.14
Beta(1, 2) ²	0.19	0.03	0.08	0.05	0.11	0.66
exp(1) ²	0.98	0.95	0.96	1.00	1.00	1.00
Lognormal(0, 0.5) ²	0.89	0.87	0.89	0.99	0.99	0.96
Gamma(0.5, 1) ²	1.00	0.99	1.00	1.00	1.00	1.00
Gamma(5, 1) ²	0.44	0.43	0.44	0.67	0.55	0.47
$(\chi^2_5)^2$	0.75	0.69	0.69	0.95	0.92	0.88
$(\chi^2_{15})^2$	0.30	0.33	0.32	0.49	0.36	0.28
$N(0, 1) * t_5$	0.22	0.25	0.21	0.20	0.15	0.09
$N(0, 1) * Beta(1, 1)$	0.15	0.06	0.24	0.02	0.07	0.10
$N(0, 1) * exp(1)$	0.35	0.35	0.37	0.55	0.44	0.39
$N(0, 1) * \chi^2_5$	0.54	0.42	0.46	0.25	0.20	0.16
$NMIX_5(0.5, 4, 0, 0)$	1.00	0.11	1.00	0.03	0.11	0.70
$NMIX_5(0.5, 0, 0, 0.9)$	0.96	0.84	0.80	0.90	0.86	0.98
$NMIX_5(0.5, 4, 0, 0.9)$	1.00	0.86	0.99	1.00	1.00	1.00
$NMIX_5(0.9, 4, 0, 0)$	1.00	0.98	0.97	0.88	0.81	0.92
$NMIX_5(0.9, 0, 0, 0.9)$	0.16	0.15	0.15	0.17	0.12	0.11
$NMIX_5(0.9, 4, 0, 0.9)$	0.99	1.00	1.00	0.99	0.96	0.91

critical values from the simulation, not from the asymptotic distributions to ensure the true level is $\alpha = 0.05$. Regarding the HZ statistic, we used the critical values given in Table 3.1 and 3.3 in Henze

and Zirkler (1990).

We generate $N = 5,000$ samples from each of the various alternative distributions. The alternatives in Kim (2020) are selected for comparison. They are the distributions with independent marginals and mixtures of normal distributions. $F_1 * F_2$ denotes the distribution with independent marginal distributions F_1 and F_2 . The product of p independent copies of F_1 is denoted F_1^p . $NMIX_p(\kappa, \mu, \rho_1, \rho_2)$ stands for the normal mixture

$$\kappa N_p(\mathbf{0}, \mathbf{R}_1) + (1 - \kappa) N_p(\mu \mathbf{1}, \mathbf{R}_2), \quad (3.1)$$

where \mathbf{R}_i is a matrix with diagonal elements equal to 1 and off-diagonal equal to ρ_i , $0 \leq \rho_i < 1$, $i = 1, 2$. According to Mecklin and Mundfrom (2005), the mixture of normal with $\kappa = 0.5$ is severely contaminated, symmetric and platykurtic (short-tailed), and the mixture with $\kappa = 0.9$ is mildly contaminated, skewed and leptokurtic (long-tailed).

The empirical power analysis shows the following. The remarkably best power is written in bold for distinction. First, all the statistics considered show similar behavior in power. They show better power against alternatives with symmetric thicker tailed marginals such as Cauchy and t -distributions than those with symmetric thinner tailed beta marginals. They show relatively good power against skewed marginal alternatives. The HZ statistic has the best power for alternatives with skewed marginals when $p = 2$, $n = 20$. However, this phenomena becomes weak when the sample size and the dimension of the data are bigger. Usually most test statistics for multivariate normality show extremely poor power against beta marginal alternatives. As Kim (2020) pointed out, the statistic NE shows relatively good power against symmetric light tailed marginal alternatives. The statistic HZ also has good power for these alternatives. The power of DX_M is lower than NE or HZ against the beta marginal alternatives. However, it is as good as JB_M or JBT_M .

Second, the proposed statistic DX_M shows outstanding power against some normal mixtures, where the contaminating distribution in (3.1) has a mean vector away from $\mathbf{0}$. Kim (2020) mentioned that the statistic MN has relatively good power against the normal mixture alternatives among the omnibus test statistics based on Mardia's skewness and kurtosis. The HZ statistic also shows good power for these alternatives. In the meantime, the DX_M statistic shows much better power than these statistics. Third, the statistic DX_M shows better or comparable power against almost all the alternatives considered, when we compare DX_M with JB_M , JBT_M . The two statistics JB_M , JBT_M are counterpart to DX_M in that they are based on the classical skewness and kurtosis, and they are generalized to a multivariate version by the exactly same way.

4. Conclusions

Desgagné and de Micheaux (2018) defined the sample second power skewness and kurtosis, and proposed a normality test statistic based on them. They suggested it as an alternative to the Jarque-Beta statistic, which is based on the classical Pearson's sample skewness and kurtosis. In this paper, we generalized the univariate normality test statistic proposed by Desgagné and de Micheaux (2018) to a multivariate version based on orthogonalization.

A simulation study reveals that the proposed statistic has the right test size even for a small sample size when we use the critical values from the approximate distribution, χ_{2p}^2 . It has very good control of type I error compared to the other referred statistics. So p -values of the test can be computed with χ_{2p}^2 distribution with high precision; we need no tables or simulated quantiles. In addition, the statistic has relatively good power against alternatives considered in the study, and shows the best power against some mixtures of multivariate normal distribution.

Acknowledgement

This work was supported by 2021 Hongik University Research Fund.

References

- Anscombe FJ and Glynn WJ (1983). Distribution of the kurtosis statistic b_2 for normal statistics, *Biometrika*, **70**, 227–234.
- D’Agostino RB (1970). Transformation to normality of the null distribution of g_1 , *Biometrika*, **57**, 679–681.
- D’Agostino RB and Stephens MA (1986). *Goodness-of-Fit Techniques*, Marcel Dekker, New York.
- D’Agostino RB and Pearson ES (1973). Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, **60**, 613–622.
- D’Agostino RB and Pearson ES (1974). Correction and amendment: Tests for departure from normality: empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, **61**, 647.
- Desgagné A and de Micheaux PL (2018). A powerful and interpretable alternative to the Jarque-Bera test of normality based on 2nd-power skewness and kurtosis, using the Rao’s score test on the APD family, *Journal of Applied Statistics*, **45**, 2307–2327.
- De Wet T and Venter JH (1972). Asymptotic distributions of certain test criteria of normality, *South African Statistical Journal*, **6**, 135–149.
- Doornik JA and Hansen H (2008). An omnibus test for univariate and multivariate normality, *Oxford Bulletin of Economics and Statistics*, **70**, 927–939.
- Enomoto R, Hanusz Z, Hara A, and Seo T (2019). Multivariate normality test using normalizing transformation for Mardia’s multivariate kurtosis, *Communications in Statistics-Simulation and Computation*, **49**, 684–698.
- Farrell PJ, Salibian-Barrera M, and Naczk K (2007). On tests for multivariate normality and associated simulation studies, *Journal of Statistical Computation and Simulation*, **77**, 1065–1080.
- Fattorini L (1986). Remarks on the use of the Shapiro-Wilk statistic for testing multivariate normality, *Statistica*, **46**, 209–217.
- Hanusz Z, Enomoto R, Seo T, and Koizumi K (2018). A Monte Carlo comparison of Jarque-Bera type tests and Henze-Zirkler test of multivariate normality, *Communications in Statistics - Simulation and Computation*, **47**, 1439–1452.
- Henze N (2002) Invariant tests for multivariate normality: A critical review, *Statistical Papers*, **43**, 467–506.
- Henze N, and Zirkler B (1990). A class of invariant consistent tests for multivariate normality, *Communications in Statistics-Theory and Methods*, **19**, 3539–3617.
- Horswell RL and Looney SW (1992). A comparison of tests for multivariate normality that are based on measures of multivariate skewness and kurtosis, *Journal of Statistical Computation and Simulation*, **42**, 21–38.
- Jarque C and Bera A (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, **6**, 255–259.
- Jönsson K (2011). A robust test for multivariate normality, *Economics Letters*, **113**, 199–201.
- Kim N (2004). An approximate Shapiro-Wilk statistic for testing multivariate normality, *The Korean Journal of Applied Statistics*, **17**, 35–47.
- Kim N (2005). The limit distribution of an invariant test statistic for multivariate normality, *The Korean Communications in Statistics*, **12**, 71–86.
- Kim N (2015). Tests based on skewness and kurtosis for multivariate normality, *Communications for*

- Statistical Applications and Methods*, **22**, 361–375.
- Kim N (2016). A robustified Jarque-Bera test for multivariate normality, *Economics Letters*, **140**, 48–52.
- Kim N (2020). Omnibus tests for multivariate normality based on Mardia's skewness and kurtosis using normalizing transformation, *Communications for Statistical Applications and Methods*, **27**, 501–510.
- Kim N and Bickel PJ (2003). The limit distribution of a test statistic for bivariate normality, *Statistica Sinica*, **13**, 327–349.
- Komunjer I (2007). Asymmetric power distribution: theory and applications to risk measurement, *Journal of Applied Econometrics*, **22**, 891–921.
- Korkmaz S, Goksuluk D, and Zararsiz G (2014). MVN: An R package for assessing multivariate normality, *The R Journal*, **6**, 151–162.
- Malkovich JF and Afifi AA (1973). On tests for multivariate normality, *Journal of the American Statistical Association*, **68**, 176–179.
- Mardia KV (1970). Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**, 519–530.
- Mardia KV (1974). Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies, *Sankhya A*, **36**, 115–128.
- Mardia KV (1975). Assessment of multinormality and the robustness of Hotelling's T^2 test, *Applied Statistics*, **24**, 163–171.
- Mardia KV and Foster K (1983). Omnibus tests of multinormality based on skewness and kurtosis, *Communications of Statistics-Theory and Methods*, **12**, 207–221.
- Mecklin CJ and Mundfrom DJ (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality, *Journal of Statistical Computation and Simulation*, **75**, 93–107.
- Mudholkar GS, Srivastava DK, and Lin CT (1995). Some p -variate adaptations of the Shapiro-Wilk test of normality, *Communications of Statistics-Theory and Methods*, **24**, 953–985.
- Pearson ES (1956). Some aspects of the geometry of statistics, *Journal of the Royal Statistical Society, Series A*, **119**, 125–146.
- Rao CR (1948a). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Mathematical Proceedings of the Cambridge Philosophical Society*, **44**, 50–57.
- Rao CR (1948b). Test of significance in multivariate analysis, *Biometrika*, **35**, 58–79.
- Romeu JL and Ozturk A (1993). A comparative study of goodness-of-fit tests for multivariate normality, *Journal of Multivariate Analysis*, **46**, 309–334.
- Roy SN (1953). On a heuristic method of test construction and its use in multivariate analysis, *Annals of Mathematical Statistics*, **24**, 220–238.
- Royston JP (1983). Some techniques for accessing multivariate normality based on the Shapiro-Wilk W , *Applied Statistics*, **32**, 121–133.
- Shapiro SS, and Francia RS (1972). An approximate analysis of variance test for normality, *Journal of the American Statistical Association*, **67**, 215–216.
- Shapiro SS and Wilk MB (1965). An analysis of variance test for normality (complete samples), *Biometrika*, **52**, 591–611.
- Srivastava MS (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality, *Statistics & Probability Letters*, **2**, 263–267.
- Srivastava MS and Hui TK (1987). On assessing multivariate normality based on Shapiro-Wilk W

- statistic, *Statistics & Probability Letters*, **5**, 15–18.
- Srivastava DK and Mudholkar GS (2003). Goodness of fit tests for univariate and multivariate normal models (In Khattree R and Rao CR eds.), *Handbook of Statistics 22: Statistics in Industry*, Elsevier, North Holland, 869–906.
- Thode Jr. HC (2002). *Testing for Normality*, Marcel Dekker, New York.
- Villasenor-Alva, JA and González-Estrada E (2009). A generalization of Shapiro-Wilk's test for multivariate normality, *Communications in Statistics-Theory and Methods*, **38**, 1870–1883.
- Zhou M and Shao Y (2014). A power test for multivariate normality, *Journal of Applied Statistics*, **41**, 351–363.

Received February 12, 2021; Revised April 07, 2021; Accepted May 17, 2021