Journal of the Nigerian Statistical Association, Vol. **29**, 2017, 58– 70

A new affine invariant test for multivariate normality based on beta probability plots

M. S. Madukaife*
Department of Statistics, University of Nigeria, Nsukka, Nigeria

A new technique for assessing multivariate normality (MVN) is proposed in this work based on a beta transform of the multivariate normal data set. The statistic is the sum of interpoint squared distances between an ordered set of the transformed observations and the set of the beta population pth quantiles. We showed that the statistic is affine invariant. The critical values of the test were evaluated for different sample sizes and different random vector dimensions through extensive simulations. For some selected sample sizes and random vector dimensions, the empirical type-I-error rates and powers of the proposed test were compared with those of other already in use tests for MVN. The results showed that the test is a good and competitive tool for testing MVN.

Keywords: beta transform; Monte Carlo simulation; population pth quantile; empirical critical values; empirical power of a test; Q-Q plot.

1. Introduction

Let $\mathbf{x} = (x_1, x_2, ..., x_d)^T \in R^d$ be a d-component random vector whose sample realizations give rise to a multivariate data set. It is described completely by a probability law, which like in the univariate case, is either discrete or continuous. The basic central probability law (distribution) and building block in classical multivariate analysis is the multivariate normal distribution, Muirhead (2005, p. 1). This is because most often, most multivariate observations are at least approximately normally distributed. Also and more importantly, most techniques for multivariate analysis such as MANOVA, MANCOVA, multivariate regression analysis, canonical correlation, maximum likelihood discriminant analysis and maximum likelihood factor analysis depend on the distributional assumption of multivariate normality (MVN). It is therefore a matter of utmost importance to conduct a test of fit to a multivariate normal distribution on a multivariate data set prior to any meaningful statistical analysis especially when any of the techniques whose applicability depends on MVN is required to be employed.

Suppose the d-component random vector $\mathbf{x} \in R^d$ is defined by a distribution function $F(\mathbf{x})$. Let $F_0(\mathbf{x})$ be a distribution function of a multivariate normal population having mean vector μ and covariance matrix $\mathbf{\Sigma}$. Suppose a sample of n independent and identically distributed (iid) observation vectors $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ is available from an unknown continuous distribution function $F(\mathbf{x})$. Fan (1997) states the problem of assessing MVN of the random vector $\mathbf{x}$ on the basis of the iid observation vectors as that of testing the goodness-of-fit hypothesis:

$$H_0 : F(\mathbf{x}) = F_0(\mathbf{x}) \text{against} H_1 : F(\mathbf{x}) \neq F_0(\mathbf{x}). \quad (1)$$

Weiss (1958) proposed a test of fit for multivariate distributions based on stochastic convergence of an empirical function derived from the sample data to a weighted integral of the true multivariate probability density function. Since then, several formal techniques for assessing MVN have been proposed. These tests are based on diverse characterizations of the multivariate normal distribution such as measures of skewness and kurtosis, measures

*Corresponding author. Email: mbanefo.madukaife@unn.edu.ng

of entropy, empirical distribution function, empirical characteristic function and various transformation properties. Some of the techniques include Mardia (1970, 1974), Malkovich and Afifi (1973), Hawkins (1981), Royston (1983), Baringhaus and Henze (1988), Henze and Zirkler (1990), Cox and Wermuth (1994), Liang et al (2009), Cardoso de Oliveira and Ferreira (2010). Apart from these and many other formal techniques for assessing MVN, procedures that are based on graphical plots have also been proposed, for instance Healy (1968), Small (1978) and Scrucca (2000).

Suppose $\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}$ is a random sample of n observation vectors from $N_d(\mu, \mathbf{\Sigma})$. Let $\overline{\mathbf{x}}$ and $\mathbf{S}$ be estimators of μ and $\mathbf{\Sigma}$ respectively. It is known that the squared radii

$$y_j = (\mathbf{x_j} - \overline{\mathbf{x}})^T \mathbf{S^{-1}} (\mathbf{x_j} - \overline{\mathbf{x}}). \tag{2}$$

Healy (1968) obtained a graphical plot of the ordered squared radii, $y_{(j)}; j = 1, 2, ..., n$ versus the approximate expected order statistics from the chi-squared distribution with d degrees of freedom. As a means of assessing MVN of the data set, he stated that the MVN of the data set may be rejected if the plot fails to be approximately linear. He also suggested the use of square root or cube root normalizing transformations of $y_{(j)}$, that is $u_{(j)} = \sqrt{y_{(j)}}$ or $v_{(j)} = 3\sqrt{y_{(j)}}$. The transform $u_{(j)}$ or $v_{(j)}$ can then be plotted against the expected normal order statistics which is obtained as the inverse distribution function $F^{-1}(p_j)$ of the normal distribution with $p_j = n^{-1}(j - 0.5)$. Again, MVN of the data set is rejected if the graphical plot fails to be approximately linear. According to him, the transformed normal plot may be expected to work better than the chi-square plot as d increases.

Under the null hypothesis of multivariate normality, Gnanadesikan and Kettenring (1972) transformed y_j in (2) as

$$z_j = n(n-1)^{-2} y_j; j = 1, 2, ..., n. \tag{3}$$

Based on a theorem in Wilks (1962, p. 562), they stated that the $z_j s$ are independent observations from beta distribution of the first kind with parameters $a = d/2$ and $b = (n - d - 1)/2$, where n and d have their usual meanings. Using the z_j transform, Small (1978) proposed that the MVN of a data set may be rejected if the graphical plot of the order statistics $z_{(j)}$ against the approximate expected order statistics $c_j; j = 1, 2, ..., n$ from $B(a, b)$, with plotting points $P_j(z_{(j)}, c_j)$, is not approximately linear. The approximate expected order statistics z_j, based on Blom (1958), is obtained according to Small (1978) and Hanusz and Tarasinska (2012) by

$$F(c_j) = \int_0^{c_j} \frac{z^{a-1}(1-z)^{b-1}}{B(a,b)} dz = \frac{j - \alpha}{n - \alpha - \beta + 1}; \qquad j = 1, 2, ..., n \tag{4}$$

where $\alpha = (a-1)/2a$ and $\beta = (b-1)/2b$. This gives c_j in (4) as the inverse distribution function $F^{-1}(p_j)$ of the beta distribution with parameters a and b, where $p_j = [b(2aj - a + 1)]/(2abn + a + b)$.

Using graphical plots to ascertain the MVN of a data set is highly subjective. As a result, a number of formal goodness-of-fit tests for MVN have been proposed from the graphical plots. One of such goodness-of-fit tests is by Hanusz and Tarasinska (2012). They obtained two geometric statistics as new measures for MVN test. The statistics were obtained as the sum of the areas formed between the plotting points and the zero intercept linear line $z = c$. In the first statistic, the plotting points are made up of the order statistics of the beta transform observations in (3) and their corresponding expected beta order statistics. In the second statistic, the plotting points are the ordered standardized principal component transformations which in each principal component are standard normally distributed and their corresponding expected standard normal order statistics. The statistic here is obtained as the sum of all the areas in all the d principal components. They concluded that large value of the statistics will lead to rejection of MVN of the data set under investigation and obtained empirical critical values under the null hypothesis of multivariate normality for the two statistics.

A very important drawback of these statistics is on their applicability. Some of the shapes whose areas are required in the statistic may be irregular and as such may not have easy-to-compute areas except with the use of special computer programs, as alluded to by the authors. Motivated by this drawback, Madukaife and Okafor (2017) converted the sum of areas of principal component observations of Hanusz and Tarasinska (2012) to sum of squared differences between jth observed order statistic in ith principal component $z_{i(j)}$ and jth expected standard normal order statistic c_j and summed them across the d principal components. Also, Madukaife and Okafor (2018) instead of converting the sum of areas obtained from beta transforms according to Hanusz and Tarasinska (2012), obtained the sum of squared differences between the chi squared order statistics according to the chi square transforms in (2) and their corresponding expected chi squared order statistics c_j. They stated that large values of the statistics lead to the rejection of the hypothesis of MVN of data sets and obtained the empirical critical values of the statistics at different sample sizes and different variable dimensions. They also showed that the statistics are at least as powerful as Hanusz and Tarasinska (2012) and recommended them in favour of the later due to their computational ease. Due to the exactness of the beta transform in (3), one would expect an adaptation of Madukaife and Okafor (2018) on beta transforms to give a better test. This is the main thrust of this work. The test is proposed in section 2 while the critical values of the proposed test are given in section 3. Section 4 gives the power performance of the test in comparison with powers of some other tests for MVN. Real life examples are given in section 5 while section 6 concludes the work.

2. The proposed test

Let $z_{(1)}, z_{(2)}, ..., z_{(n)}$ from an unknown distribution $F(z)$ arising from the z-transform of n independent multivariate data set in (3) be plotted against a set of n expected order statistics c_j from a known distribution $F_0(z)$ to form a Q-Q plot. Also, let the set of n expected order statistics c_j be plotted against itself, forming a Q-Q line $z = c$. Both plots are on the same axes. Then the size of the area of each enclosed shape is proportional to the distance of the plotted point away from the Q-Q line.

A measure of the distance between an observed jth point $P_j(z_{(j)}, c_j)$ and each of its orthogonal projections $P'_j(z_{(j)}, c_j)$ and $P_j"(z_{(j)}, c_j)$ on the straight line $z = c$ can be obtained as $|z_{(j)} - c_j|$. It is therefore easy to see that the area of the right-angled triangle formed by the observed point on the straight line $z = c$ is $1/2(z_{(j)} - c_j)^2$. This implies that $\triangle_j = (z_{(j)} - c_j)^2$ is an appropriate measure of the area of a square formed by the observed point. Hence, $D_n = \sum_{j=1}^{n}(z_{(j)} - c_j)^2$ which is the sum of the areas of all the squares formed by all the observed points is similar to the sum of the areas used by Hanusz and Tarasinska (2012) as a measure of a test of multinormality.

From the foregoing, it is proposed here to take the squared differences between the jth order statistic $z_{(j)}$ according to (3) and the jth expected beta order statistic c_j according to (4) for each $j(j = 1, 2, ..., n)$ and obtain the sum of the squared differences as a test statistic for assessing MVN of the data set.

$$D_n = \sum_{j=1}^{n}(z_{(j)} - c_j)^2 \qquad (5)$$

The statistic will reject MVN of the data set for large values of the statistic. This is because under the null hypothesis of multivariate normality of the data, the Q-Q plot for the beta distribution will be equivalent to that of multinormality (see Gnanadesikan and Kettenring

(1972); Small (1978)). Under this condition, observed values $z_{(j)}s$ will tend to the $c_j s$ such that $\triangle_j$ s will tend to zero and so D_n will tend to zero. The proposed statistic is less cumbersome and easier to apply than the Hanusz and Tarasinska (2012) statistic.

Our proposed statistic, D_n is also affine invariant because it is based on the Mahalanobis squared distance (transformation of the original $\mathbf{x}$ to y) and any test statistic which is based on the Mahalanobis squared distance would be found to be affine invariant, see Madukaife and Okafor (2018) and Henze (2002).

3. Critical values of the proposed test

The empirical critical values of the test for different combinations of the sample size n and the number of variables d are evaluated through extensive simulation studies. Precisely, the critical values at 0.5, 1, 2.5, 5 and 10 percent levels of significance are evaluated for $n = 10(2)20(5)50(10)100(50)300(100)500$ and $d = 2$, 3, 4 and 5. We generated $N = 100{,}000$ samples for $n < 100; N = 80,000$ samples for $100 \leq n < 300$ and $N = 50,000$ samples for $n \geq 300$ from a $N_d(\mathbf{0}, \mathbf{I})$; $d = 2, 3, 4, 5$. From each sample, we obtained the beta transform of the observation vectors as given in (3) and ordered the set of the transforms. We also obtained the corresponding expected beta order statistics as the pth quantiles of the beta distribution, with p given as the result in (4) for each j and calculated the sum of squared differences between each corresponding observed and expected order statistics. From the N sum of squared differences associated with each combination of n and d, we calculated the α-level critical value of the test for the n and d as the $100(1-\alpha)$ percentile of the N sum of squared differences. These percentile values are presented in Table 1. Based on this, we wish to reject multivariate normality of a data set with sample size n and dimension d if the calculated value of the proposed test statistic at level of significance α is greater than the corresponding empirical critical value.

4. Empirical power studies

In this section, we shall compare the power of the proposed test with the powers of some other time-honoured and highly regarded tests for MVN. The competing tests used here include the Mardia's skewness (MS) and kurtosis (MK) tests for MVN, Mardia (1970, 1974); Henze and Zirkler test for MVN, Henze and Zirkler (1990); Singh's classical test for MVN, Singh (1993); the combination test for MVN, Hwu et al (2002) and Madukaife and Okafor (T) and (G) tests for MVN, Madukaife and Okafor (2017, 2018). These tests are chosen from among all the numerous tests for multinormality because Mardia's tests (MS) and (MK) as well as the Henze and Zirkler test are among the most powerful tests for MVN in the literature and also among the most used (Mecklin and Mundfrom 2004). Again, Singh (1993) test is based on the same beta transforms of multivariate data and it has a good power. Also, the combination test of Hwu et al (2002) promised to have a highly competitive power when compared to other tests that preceded it. Also, Madukaife and Okafor (2017, 2018) tests have the same principle with the D_n test. The only difference between them is in the type of transformation. In this work, the competing tests shall be denoted by MS, MK, HZ, Scl, CT, T and G respectively while our proposed test shall be denoted by PT.

Since the null distribution of some of the test statistics for MVN is intractable, we used a Monte Carlo simulation study via empirical critical values throughout this study to ensure uniformity of power comparison of the tests. In the study, 10,000 data sets in each of the combinations of sample size $n = 10$, 20, 30, 50, 100 and dimension $d = 2$, 5 were generated from 11 different multivariate distributions, ranging from the standard multivariate normal to various departures from normality. For each of the combinations of n and d, we calculated the values of each of the eight statistics being compared in each of the 10,000 simulated samples and obtained the power of each test statistic as the percentage of the 10,000 samples

Table 1. Empirical critical values of the beta-probability plot sum of squared differences from the multivariate normal distribution

n			$d = 2$					$d = 3$		
	0.005	0.01	0.025	0.05	0.10	0.005	0.01	0.025	0.05	0.10
10	0.1952	0.1680	0.1289	0.0999	0.0729	0.1929	0.1617	0.1274	0.1023	0.0789
12	0.1750	0.1464	0.1086	0.0833	0.0594	0.1669	0.1432	0.1100	0.0874	0.0665
14	0.1511	0.1257	0.0926	0.0689	0.0487	0.1447	0.1231	0.0948	0.0749	0.0562
16	0.1331	0.1087	0.0781	0.0580	0.0408	0.1301	0.1099	0.0833	0.0645	0.0477
18	0.1152	0.0946	0.0681	0.0499	0.0345	0.1176	0.0978	0.0738	0.0565	0.0411
20	0.1031	0.0834	0.0595	0.0430	0.0295	0.1058	0.0869	0.0652	0.0495	0.0361
25	0.0764	0.0619	0.0433	0.0310	0.0215	0.0807	0.0658	0.0477	0.0360	0.0263
30	0.0588	0.0477	0.0329	0.0237	0.0164	0.0644	0.0518	0.0372	0.0279	0.0201
35	0.0464	0.0368	0.0255	0.0183	0.0128	0.0510	0.0415	0.0294	0.0220	0.0160
40	0.0382	0.0307	0.0212	0.0150	0.0104	0.0424	0.0341	0.0243	0.0180	0.0129
45	0.0318	0.0252	0.0174	0.0126	0.0087	0.0366	0.0291	0.0203	0.0150	0.0108
50	0.0273	0.0215	0.0147	0.0105	0.0074	0.0302	0.0243	0.0172	0.0128	0.0092
60	0.0199	0.0159	0.0109	0.0079	0.0055	0.0227	0.0182	0.0128	0.0096	0.0069
70	0.0157	0.0122	0.0085	0.0061	0.0043	0.0178	0.0139	0.0098	0.0073	0.0053
80	0.0122	0.0095	0.0066	0.0048	0.0034	0.0142	0.0114	0.0080	0.0059	0.0043
90	0.0101	0.0080	0.0056	0.0040	0.0028	0.0115	0.0092	0.0064	0.0048	0.0035
100	0.0085	0.0067	0.0046	0.0034	0.0024	0.0098	0.0077	0.0054	0.0040	0.0029
150	0.0043	0.0033	0.0022	0.0016	0.0012	0.0053	0.0042	0.0029	0.0022	0.0015
200	0.0024	0.0019	0.0013	0.0010	0.0007	0.0028	0.0022	0.0016	0.0012	0.0009
250	0.0016	0.0013	0.0009	0.0007	0.0005	0.0021	0.0017	0.0012	0.0009	0.0006
300	0.0012	0.0009	0.0006	0.0005	0.0004	0.0013	0.0011	0.0008	0.0006	0.0004
400	0.0007	0.0005	0.0004	0.0003	0.0002	0.0008	0.0006	0.0004	0.0003	0.0003
500	0.0005	0.0004	0.0003	0.0002	0.0001	0.0005	0.0004	0.0003	0.0002	0.0002

that is rejected by the statistic at 5 percent level of significance.

The first distribution we considered in this power comparison was the standard multivariate normal distribution. Since the null hypothesis is true in this case, the null hypothesis of multivariate normality should be rejected by each of the competing test statistics at about 5 percent level of significance. That is, each test should give a power of 5 percent. Mecklin and Mundfrom (2005) have stated that any of the statistics that gives power far above the 5 percent level would probably indicate a problem of high Type I error rate. The result of this study is presented in Table 2. From the result, it is seen that the MS test, the T and G tests and the PT test maintained approximately the same 5 percent error rate in all the considered combinations of sample size and dimension more than the rest of the statistics under consideration. Also, the variability in sample size n as well as in variable dimension d as measured by the standard deviation of the 10 observations (type 1 error rates) for each statistic shows that the proposed test is least affected, with respect to type 1 error rate, by changes in the sample size and variable dimension among all the 8 tests considered. This can be seen from its standard deviation $s_{PT} = 0.1350$ which is the least. The standard deviation of the observations for each of the other tests is obtained as $s_{MS} = 0.2914$; $s_{MK} = 0.3048$; $s_{HZ} = 1.3408$; $s_{CT} = 0.3917$; $s_{Scl} = 0.2700$; $s_T = 0.1958$ and $s_G = 0.2530$. Based on the standard deviations, the PT is least affected, followed by the T and G tests respectively while the HZ test is most affected with highest standard deviation value. Again, HZ test maintained type 1 error rates much lower than the nominal rate of 5 percent with a progressive increase towards 5 percent as the sample size n and variable dimension d increased.

There are several multivariate distributions from where multivariate data sets may be obtained other than the multinormal distribution. Some of these distributions are symmetric while some others are skewed. In this work, we categorized all the alternative multivariate

Table 1 continues

n			$d=4$					$d=5$		
	0.005	0.01	0.025	0.05	0.10	0.005	0.01	0.025	0.05	0.10
10	0.1925	0.1639	0.1293	0.1045	0.0816	0.1885	0.1614	0.1277	0.1040	0.0813
12	0.1664	0.1412	0.1111	0.0900	0.0703	0.1668	0.1435	0.1129	0.0917	0.0723
14	0.1443	0.1241	0.0979	0.0787	0.0606	0.1443	0.1244	0.0992	0.0809	0.0635
16	0.1304	0.1116	0.0870	0.0694	0.0533	0.1304	0.1121	0.0884	0.0719	0.0563
18	0.1174	0.0995	0.0764	0.0605	0.0461	0.1186	0.1024	0.0802	0.0643	0.0500
20	0.1056	0.0889	0.0678	0.0531	0.0404	0.1081	0.0913	0.0709	0.0571	0.0440
25	0.0836	0.0697	0.0523	0.0407	0.0301	0.0849	0.0723	0.0556	0.0439	0.0335
30	0.0666	0.0556	0.0408	0.0314	0.0234	0.0678	0.0572	0.0432	0.0339	0.0259
35	0.0553	0.0453	0.0330	0.0251	0.0187	0.0569	0.0473	0.0356	0.0279	0.0210
40	0.0451	0.0368	0.0271	0.0206	0.0151	0.0480	0.0391	0.0292	0.0227	0.0172
45	0.0379	0.0307	0.0224	0.0170	0.0127	0.0404	0.0329	0.0245	0.0191	0.0143
50	0.0321	0.0265	0.0191	0.0146	0.0108	0.0347	0.0285	0.0209	0.0163	0.0123
60	0.0241	0.0194	0.0143	0.0109	0.0080	0.0258	0.0213	0.0158	0.0122	0.0092
70	0.0190	0.0153	0.0111	0.0084	0.0063	0.0208	0.0170	0.0123	0.0095	0.0072
80	0.0155	0.0125	0.0089	0.0067	0.0050	0.0164	0.0134	0.0099	0.0076	0.0058
90	0.0126	0.0101	0.0073	0.0056	0.0041	0.0138	0.0111	0.0081	0.0063	0.0048
100	0.0106	0.0085	0.0061	0.0047	0.0035	0.0117	0.0094	0.0068	0.0053	0.0040
150	0.0061	0.0049	0.0036	0.0027	0.0019	0.0066	0.0054	0.0040	0.0030	0.0022
200	0.0046	0.0037	0.0028	0.0023	0.0018	0.0050	0.0042	0.0033	0.0027	0.0021
250	0.0024	0.0019	0.0014	0.0011	0.0008	0.0027	0.0023	0.0016	0.0012	0.0009
300	0.0015	0.0012	0.0009	0.0007	0.0005	0.0016	0.0013	0.0010	0.0008	0.0006
400	0.0009	0.0007	0.0005	0.0004	0.0003	0.0009	0.0008	0.0006	0.0004	0.0003
500	0.0006	0.0005	0.0003	0.0003	0.0002	0.0006	0.0005	0.0004	0.0003	0.0002

Table 2. Empirical type-I-error rates against multivariate normal distribution, nominal $\alpha = 5$ percent

Test	$n=10$		$n=20$		$n=30$		$n=50$		$n=100$	
	$d=2$	$d=5$	$d=2$	$d=5$	$d=2$	$d=5$	$d=2$	$d=5$	$d=2$	$d=5$
MS	4.8	4.9	4.6	4.9	5.0	4.8	5.2	5.3	5.4	4.5
MK	4.9	4.7	4.8	5.2	4.8	4.9	5.1	5.4	4.3	5.1
HZ	2.3	0.7	4.0	2.5	4.5	3.7	4.6	4.4	4.8	4.5
CT	9.7	9.9	9.2	9.5	9.0	9.7	8.7	9.3	9.7	9.0
Scl	4.6	4.9	4.8	5.2	5.0	4.8	4.3	5.1	5.1	5.0
T	4.9	5.2	5.2	5.4	5.2	4.8	4.9	4.9	5.1	4.9
G	4.8	5.4	5.3	4.7	5.1	4.7	5.1	5.1	4.8	4.8
PT	5.2	5.0	5.1	5.2	4.9	5.2	4.9	4.9	5.2	5.0

distributions considered into two, according to their symmetry. Hence, we have comparison of the test statistics for symmetric distributions as group 1 and comparison of the test statistics for skewed distributions as group 2. The symmetric distributions considered in group 1 include:

- The multivariate t distribution (MVt) with 2 degrees of freedom and identity matrices.
- Product of uniform distribution ($Unif(0,1)^d$) in the interval $(0,1)$, where $d = 2, 5$;
- Product of beta distribution ($Beta(2,2)^d$) with parameters $(2,2)$, where $d = 2, 5$ and
- Product of the arcsine distribution ($Arcsine^d$), where $d = 2, 5$.

The mixture of two multinormal distributions in the form $\alpha[f_1(\mathbf{x})] + (1-\alpha)f_2(\mathbf{x}); \alpha = 0.4$ with arbitrary mean vectors $\mu_{\mathbf{1}} = (0,0)^T$, $\mu_{\mathbf{2}} = (-1,-1)^T$ for $d = 2$ and $\mu_{\mathbf{1}} = (0,0,0,0,0)^T$, $\mu_{\mathbf{2}} = (2,1,1,2,1)^T$ for $d = 5$ and arbitrarily chosen identity covariance matrices.

Also, the skewed distributions considered in group 2 include:

- Product of beta distribution ($Beta(1,5)^d$) with parameters $(1,5)$, where $d = 2, 5$;

- Product of the standard exponential distribution ($Exp(1)^d$), where $d = 2, 5$;
- Multivariate generalized extreme value distribution ($Mvgevd$) with parameter 0.72
- Product of beta with parameters $(1, 5)$ and the standard exponential distributions and
- Product of standard normal and standard exponential distributions.

The results are presented in Tables 3, 4, 5 and 6.

Table 3. Power comparison of tests for multivariate normality for various multivariate symmetric distributions at $\alpha = 5$ percent, $d = 2$

n	Distributions	MS	MK	HZ	CT	Scl	T	G	PT
10	MVNMIX	3.8	4.9	1.8	8.9	6.1	4.8	7.3	9.0
20	MVNMIX	4.5	5.0	3.6	8.6	5.2	4.5	4.7	10.6
30	MVNMIX	3.6	4.5	4.1	7.7	5.7	4.2	5.2	13.4
50	MVNMIX	3.8	5.3	4.9	6.9	6.1	4.3	5.1	17.4
100	MVNMIX	5.1	4.2	5.4	7.2	5.6	8.9	4.9	29.2
10	MVt	43.0	37.0	33.6	45.9	16.5	42.9	9.3	36.5
20	MVt	71.0	72.9	68.8	74.8	54.9	72.9	66.3	69.1
30	MVt	83.1	88.7	85.6	87.8	73.6	86.8	85.7	86.5
50	MVt	92.2	98.2	97.3	97.4	91.4	97.5	97.6	98.0
100	MVt	97.6	100.0	100.0	100.0	99.4	99.9	100.0	100.0
10	$Beta(2,2)^2$	1.8	6.9	1.3	8.9	6.4	3.2	10.2	10.3
20	$Beta(2,2)^2$	0.7	13.9	1.5	9.1	7.3	2.0	16.2	16.1
30	$Beta(2,2)^2$	0.4	24.0	7.3	9.8	9.2	1.8	20.4	25.5
50	$Beta(2,2)^2$	0.2	46.2	14.1	13.1	16.5	2.4	33.6	47.5
100	$Beta(2,2)^2$	0.1	83.5	37.5	25.8	38.8	9.8	69.7	87.2
10	$Arcsine^2$	1.7	20.6	4.4	13.1	11.9	7.2	25.1	18.6
20	$Arcsine^2$	0.2	66.1	42.0	26.5	22.0	13.3	61.6	59.1
30	$Arcsine^2$	0.0	91.5	82.0	41.2	40.6	21.4	84.9	88.3
50	$Arcsine^2$	0.0	99.8	99.5	68.4	76.7	34.8	99.0	99.7
100	$Arcsine^2$	0.0	100.0	100.0	96.6	99.7	72.7	100.0	100.0
10	$Unif(0,1)^2$	1.4	10.9	1.7	13.3	8.7	3.4	16.0	13.6
20	$Unif(0,1)^2$	0.2	35.2	9.2	15.9	13.5	3.7	34.6	34.1
30	$Unif(0,1)^2$	0.2	60.2	27.3	23.3	22.3	6.2	52.8	60.2
50	$Unif(0,1)^2$	0.0	91.6	61.7	40.2	45.8	17.2	82.0	90.1
100	$Unif(0,1)^2$	0.0	100.0	97.3	75.3	89.6	39.2	99.8	100.0

The power performance of a test for MVN which is the ability of the test to reject multivariate normality when the distribution is actually non multinormal, when compared with other alternative tests, is an accurate means of determining the relative goodness of the test. In this regard, from Tables 3 and 4, it is observed that the proposed test (PT) generally out-performed the Mardias skewness test in all the alternative distributions considered. Also, it out-performed the very highly regarded Henze-Zirkler test under sample size of 10 in almost all the distributions considered while it remained very competitive with the Henze-Zirkler test in the remaining sample sizes considered with outright superior performance in the multivariate t-distribution and the reverse in the arcsine distribution. When compared with combination test of Hwu et al (2002) and Singhs test, it showed a promising result with outright superior performance in almost all the symmetric distributions considered at large sample sizes of $n > 30$ except in the mixture of multivariate normal distributions (MVNMIX). The proposed test is however observed to be generally inferior to the Mardias kurtosis test for sample sizes between 10 and 50, except under the multivariate t-distribution. For the sample size of $n = 100$, the power of the proposed test was seen to be at par with the powers of any of the Mardia's kurtosis and the Henze-Zirkler tests in almost

Table 4. Power comparison of tests for multivariate normality for various multivariate symmetric distributions at $\alpha = 5$ percent, $d = 5$

n	Distributions	MS	MK	HZ	CT	Scl	T	G	PT
10	MVNMIX	4.7	4.7	0.9	11.1	4.7	3.4	6.0	3.2
20	MVNMIX	3.5	5.0	3.0	9.9	4.8	4.2	7.0	3.9
30	MVNMIX	3.0	6.4	6.6	7.9	4.3	6.3	6.9	3.8
50	MVNMIX	4.7	5.4	8.2	7.6	5.6	7.0	7.0	4.2
100	MVNMIX	3.3	9.2	22.9	8.0	5.8	15.7	6.1	5.2
10	MVt	46.5	34.0	17.7	48.5	3.9	56.4	0.3	40.8
20	MVt	93.3	93.6	84.2	94.0	43.9	88.0	53.3	94.6
30	MVt	98.7	99.3	98.0	98.7	83.5	97.1	95.2	99.4
50	MVt	100.0	100.0	100.0	99.9	98.3	99.9	100.0	100.0
100	MVt	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
10	$Beta(2,2)^5$	2.6	5.5	0.5	7.0	6.1	3.3	8.0	3.1
20	$Beta(2,2)^5$	0.6	12.4	2.1	8.5	7.0	2.7	17.6	3.7
30	$Beta(2,2)^5$	0.3	24.0	4.3	10.6	7.4	2.4	29.4	7.1
50	$Beta(2,2)^5$	0.1	48.9	10.8	15.1	10.6	1.6	50.4	18.6
100	$Beta(2,2)^5$	0.1	90.2	31.6	27.4	22.8	1.6	84.7	62.8
10	$Arcsine^5$	1.5	7.0	1.3	4.1	7.6	3.5	10.9	2.4
20	$Arcsine^5$	0.1	45.2	13.6	7.7	9.7	3.2	47.1	17.0
30	$Arcsine^5$	0.0	83.9	47.2	11.9	12.1	3.1	82.3	50.1
50	$Arcsine^5$	0.0	99.8	95.7	20.7	24.4	4.5	99.5	95.6
100	$Arcsine^5$	0.0	100.0	100.0	48.0	74.4	15.2	100.0	100.0
10	$Unif(0,1)^5$	1.6	6.1	0.8	5.1	6.7	3.0	9.2	2.7
20	$Unif(0,1)^5$	0.2	25.8	4.2	8.5	8.0	2.7	31.6	8.2
30	$Unif(0,1)^5$	0.0	53.6	12.3	12.8	10.1	2.1	56.8	20.8
50	$Unif(0,1)^5$	0.0	90.0	41.8	21.3	17.9	2.3	88.6	59.9
100	$Unif(0,1)^5$	0.0	100.0	93.2	43.3	50.7	5.1	100.0	99.2

all the symmetric alternative distributions considered. When compared with the two similar tests that are based on probability plots according to Madukaife and Okafor (2017, 2018), the proposed PT test performed reasonably well. Under the variable dimension $d = 2$, the proposed test remained superior, in terms of power performance, to both the T and G tests in all the symmetric distributions considered. Under $d = 5$ however, its power appeared to be inferior to that of the T test at large samples in most of the symmetric distributions considered except in MVt where the PT test performed better than the two similar tests, T and G. Based on these, the proposed test is therefore recommended as a good technique for carrying out test for MVN of data sets from symmetric alternative distributions.

From *Tables* 5 and 6, we observed the proposed test to be slightly inferior to the Mardias skewness (MS) test, Henze-Zirkler test at large samples, the combination test and the Madukaife and Okafor T test in most of the skewed alternative distributions considered. However, it is seen to be generally more powerful than the Mardias kurtosis (MK) test at all the sample sizes, the Henze-Zirkler test at the sample size of 10, the Singhs Scl test at all the sample sizes and Madukaife and Okafor G test at almost all the sample sizes. Also, there is no observed difference between the performance of the proposed test in $d = 2$ and $d = 5$. Even in those cases where the power of the proposed test is observed to be inferior, the proposed test still shows itself to be competitive especially at sample size of 100. It can therefore be recommended as a good technique for assessing MVN of data sets from skewed distributions.

Table 5. Power comparison of tests for multivariate normality for various multivariate skewed distributions at $\alpha = 5$ percent, $d = 2$

n	Distributions	MS	MK	HZ	CT	Scl	T	G	PT
10	$Beta(1,5)^2$	19.2	11.3	17.5	34.2	13.4	22.6	4.9	17.6
20	$Beta(1,5)^2$	49.0	19.7	58.9	53.8	26.1	51.7	19.5	29.1
30	$Beta(1,5)^2$	74.9	26.3	83.4	77.5	33.7	71.4	29.7	37.4
50	$Beta(1,5)^2$	96.6	35.5	98.4	96.8	45.6	87.2	43.5	49.5
100	$Beta(1,5)^2$	100.0	56.8	100.0	100.0	66.2	98.3	68.1	71.0
10	$Exp(1)^2$	36.7	22.5	34.1	40.6	22.1	41.9	7.5	32.0
20	$Exp(1)^2$	78.4	46.8	83.9	81.0	49.3	77.6	44.5	58.0
30	$Exp(1)^2$	95.0	63.2	97.0	95.8	66.6	90.8	66.7	72.6
50	$Exp(1)^2$	99.9	82.4	100.0	100.0	83.6	98.0	86.5	88.7
100	$Exp(1)^2$	100.0	98.0	100.0	100.0	97.8	100.0	98.7	98.9
10	$Mvgevd$	14.8	10.3	10.4	18.5	8.5	16.7	4.0	14.2
20	$Mvgevd$	34.0	22.0	30.4	37.5	22.1	33.8	20.4	29.7
30	$Mvgevd$	52.2	32.5	46.1	53.8	33.2	48.6	32.7	40.4
50	$Mvgevd$	76.2	47.3	68.8	78.4	48.4	69.8	51.1	56.3
100	$Mvgevd$	98.3	74.5	93.1	98.2	71.5	94.5	77.4	79.6
10	$Beta(1,5)Exp(1)$	29.3	16.5	24.7	33.1	17.2	42.8	5.2	24.9
20	$Beta(1,5)Exp(1)$	65.9	33.8	73.2	69.3	39.1	86.3	32.0	44.2
30	$Beta(1,5)Exp(1)$	89.1	45.8	92.6	90.1	51.8	98.2	48.8	58.1
50	$Beta(1,5)Exp(1)$	99.4	64.4	99.7	99.5	69.6	100.0	98.7	74.7
100	$Beta(1,5)Exp(1)$	100.0	87.5	100.0	100.0	90.8	100.0	100.0	93.9
10	$N(0,1)Exp(1)$	20.4	12.6	14.5	37.0	9.5	20.0	4.0	18.2
20	$N(0,1)Exp(1)$	46.1	24.2	46.1	51.1	25.7	43.7	22.7	32.9
30	$N(0,1)Exp(1)$	68.0	34.5	68.3	70.5	36.1	61.7	35.8	43.2
50	$N(0,1)Exp(1)$	92.2	49.3	92.4	92.9	49.6	81.9	54.2	59.0
100	$N(0,1)Exp(1)$	100.0	76.1	99.9	99.9	73.5	96.3	79.6	82.6

5. Real life examples

In this section, seven different cases were presented where the proposed test was applied to real-life data, generated from real-life experiments or investigations. The generated data sets are described as follows.

Nwagbata (2016) obtained two samples of 250 each for male and female babies delivered in a private hospital in Nigeria over a period of 5 years. Three basic measurements were taken from each of them at birth to represent their features at birth. The measurements, head circumference, body height and body weight, form two data sets from trivariate distributions which were hypothesized to be multinormal.

A combination of the two trivariate data sets above to have a 500 trivariate data points from a distribution which was hypothesized to be multinormal.

Fisher (1936)'s famous Iris-Setosa data; Iris-Versicolor data; and Iris-Virginica data, each data set with measurements, sepal length, sepal width, petal length, and petal width (all in cm) each of which was hypothesized to have come from a tetravariate normal population.

A combination of the three sets of tetravariate data from the three classes of Iris plant in Fisher (1936), which gives 150 tetravariate data points, hypothesized to have come from a tetravariate normal population.

The real-life data sets were tested for multivariate normality at 5 percent level of significance using our proposed statistic in comparison with the Mardias skewness and kurtosis tests, the Henze-Zirkler test, the Singh's classical test, the combination test, the T and G tests. The result is presented in *Table* 7. The result shows that the proposed test, the Mardia's tests, the Singh's test and the G test agreed in all the data sets considered. Also, the proposed test agreed with the T test and the combination test in all the data sets

Table 6. Power comparison of tests for multivariate normality for various multivariate skewed distributions at $\alpha = 5$ percent, $d = 5$

n	Distributions	MS	MK	HZ	CT	Scl	T	G	PT
10	$Beta(1,5)^5$	11.9	6.2	4.9	14.2	4.4	14.7	2.8	9.3
20	$Beta(1,5)^5$	41.9	18.6	43.6	42.9	13.0	32.9	1.5	27.5
30	$Beta(1,5)^5$	72.1	28.6	76.3	72.0	21.2	49.9	8.1	38.8
50	$Beta(1,5)^5$	98.2	44.5	98.1	98.0	36.4	73.6	29.6	54.7
100	$Beta(1,5)^5$	100.0	72.4	100.0	100.0	60.7	94.4	66.6	80.0
10	$Exp(1)^5$	25.4	13.4	12.1	26.5	4.3	34.5	1.0	18.9
20	$Exp(1)^5$	82.3	57.5	78.9	82.4	31.7	71.2	10.1	65.5
30	$Exp(1)^5$	98.0	78.7	97.4	98.2	59.5	87.8	52.8	85.0
50	$Exp(1)^5$	100.0	95.4	100.0	100.0	85.8	98.0	90.4	96.8
100	$Exp(1)^5$	100.0	100.0	100.0	100.0	99.1	100.0	99.9	99.9
10	$Mvgevd$	11.1	7.2	2.3	14.5	4.2	17 3	2.8	8.7
20	$Mvgevd$	37.1	23.6	19.7	38.8	11.2	39.7	2.0	30.0
30	$Mvgevd$	61.0	40.1	36.8	62.7	24.6	58.0	14.6	47.2
50	$Mvgevd$	89.3	64.6	66.4	90.3	44.9	82.7	48.8	69.5
100	$Mvgevd$	99.9	90.9	95.7	99.3	74.4	98.7	86.4	91.9
10	$Beta(1,5)^2Exp(1)^3$	18.6	9.3	7.9	21.3	4.3	35.4	2.1	13.9
20	$Beta(1,5)^2Exp(1)^3$	70.0	42.0	66.8	71.1	23.9	78.9	3.0	46.6
30	$Beta(1,5)^2Exp(1)^3$	93.7	63.5	93.0	93.7	46.6	94.8	25.6	67.3
50	$Beta(1,5)^2Exp(1)^3$	99.9	85.4	99.8	100.0	72.6	99.5	65.2	85.6
100	$Beta(1,5)^2Exp(1)^3$	100.0	99.2	100.0	100.0	95.5	100.0	95.0	98.4
10	$N(0,1)^3Exp(1)^2$	10.6	6.7	2.6	13.2	4.4	14.3	2.6	8.9
20	$N(0,1)^3Exp(1)^2$	36.4	19.2	24.3	38.7	14.0	30.3	1.9	28.5
30	$N(0,1)^3Exp(1)^2$	60.2	33.5	47.1	63.1	26.6	44.1	14.7	43.7
50	$N(0,1)^3Exp(1)^2$	91.8	54.9	82.2	91.0	46.0	65.5	43.1	63.2
100	$N(0,1)^3Exp(1)^2$	100.0	83.4	99.7	100.0	73.7	90.1	79.4	86.7

considered except in features of male babies at birth where the proposed test failed to reject MVN of the data which the T and CT tests both rejected. Again, the proposed test is in agreement with the HZ test in all the data sets considered except in the features of male babies at birth as well as in the Iris Setosa where the MVN was rejected by the HZ test against the decision of non rejection given by the proposed test. Consequent upon the outlined observations together with the computational ease of the proposed test, its use in real life can be highly recommended.

Table 7. Result of the tests for multivariate normality conducted on some real-life data sets

Data	n	PT	T	G	MS/MK	Scl	CT	HZ
Features of male babies at birth	250	Do not	Reject	Do not	Do not	Do not	Reject	Reject
Features of Female babies at birth	250	Reject	Reject	Reject	Reject	Reject	Reject	Reject
Features of babies at birth	500	Reject	Reject	Reject	Reject	Reject	Reject	Reject
Iris-Setosa	50	Do not	Do not	Do not	Do not	Do not	Do not	Reject
Iris-Versicolor	50	Do not	Do not	Do not	Do not	Do not	Do not	Do not
Iris-Virginica	50	Reject	Reject	Reject	Reject	Reject	Reject	Reject
Iris-Data	150	Reject	Reject	Reject	Reject	Reject	Reject	Reject

6. Conclusion

Mecklin and Mundfrom (2004) in a review of tests for MVN have stated that no test is universally the best. However, some tests have been observed to be generally weak in terms of power performance. From the power comparisons carried out in this work, the proposed test, no doubt, can be recommended for use as a good test for MVN. This is because it has a strong control over type-1error and it has a highly competitive power. However, what may appear to be a disadvantage of the proposed test is that its exact null distribution is unknown. Quite a number of other well recommended techniques also have unknown distributions and this makes them applicable only with the use of empirical critical values. This is the case with tests for MVN such as Hanusz and Tarasinska (2012), Hwu et al (2002), Ahn (1992), Liang et al (2009) and Singh (1993). Since the class of tests proposed here looks promising, further research should be done in the direction of the distribution of the test statistic, if it exists, in order to obtain a fully parametric test. Also, the computed value of the test statistic can be inflated with the presence of an outlier thereby leading to wrong decision of rejection even when it is not supposed to be so. It is therefore recommended that this proposed test be conducted along with a graphical probability plot to determine the presence of an outlier. Finally, since the test is based on Q-Q plot which is not peculiar to any distribution, it is recommended that the procedure be adapted for constructing goodness-of-fit tests of other multivariate and univariate distributions.

Acknowledgement

The author is thankful to the editorial team and the two anonymous referees for their valuable comments that greatly assisted in improving the presentation. Thanks are also due to Prof. F. C. Okafor for his very helpful comments in the original manuscript.

References

Ahn, S.K. (1992). F-probability plot and its application to multivariate normality. *Communications in Statistics-Theory and Methods*, 21, 997-1023.

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

Blom G. (1958). *Statistical estimates and transformed beta-variables.* Wiley. New York.

Cardoso de Oliveira, I.R.C. and Ferreira, D.F. (2010). Multivariate extension of chi-squared univariate normality test. *Journal of Statistical Computation and Simulation*, 80, 513-525.

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

Fan, Y. (1997). Goodness of fit tests for a multivariate distribution by the empirical characteristic function. *Journal of Multivariate Analysis*, 62, 36-63.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7, Part II, 179-188.

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

Hanusz, Z. and Tarasinska, J. (2012). New tests for multivariate normality based on Smalls and Srivastavas graphical methods. *Journal of Statistical Computation and Simulation*, 80, 513-526.

Hawkins, D. N. (1981). A new test for multivariate normality and homoscedasticity. *Technometrics*, 23, 105-110.

Healy M. J. R. (1968). Multivariate normal plotting. *Applied Statistics*, 17, 157-161.

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, 3595-3618.

Hwu, T., Han, C., and Rogers, K. J. (2002). The combination test for multivariate normality. *Journal of Statistical Computation and Simulation*, 72, 379-390.

Liang, J., Fang, M. L. and Chang, P. S. (2009). A generalized Shapiro-Wilk W statistic for testing high dimensional normality. *Computational Statistics and Data Analysis*, 53, 3883-3891.

Madukaife, M. S. and Okafor, F. C. (2017). A powerful affine invariant test for multivariate normality based on interpoint distances of principal components. *Communications in Statistics-Simulation and Computation*, DOI: 10.1080/03610918.2017.1309667.

Madukaife, M. S. and Okafor, F. C. (2018). A new large sample goodness of fit test for multivariate normality based on chi squared probability plots. *Communications in Statistics-Simulation and Computation*, DOI: 10.1080/03610-918.2017.1422749.

Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association*, 68, 176-179.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 573, 519-530.

Mecklin, C. J. and Mundfrom, D. J. (2004). An appraisal and bibliography of tests for multivariate normality. *International Statistical Review*, 72, 123-138.

Mecklin, C. J. and Mundfrom, D. J. (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.

Muirhead, R. J. (2005). *Aspects of multivariate statistical theory.* Wiley. New Jersey.

Nwagbata, A. C. (2016). A discriminant analysis of features of babies at birth. An unpublished project report, Department of Statistics, University of Nigeria, Nsukka.

Royston, J. P. (1983). Some techniques for assessing multivariate normality based on the Shapiro-Wilk W. *Applied Statistics*, 32, 121-133.

Scrucca, L. (2000). Assessing multivariate normality through interactive dynamic graphics. *Quaderni di statistica*, 2, 221-240.

Singh, A. (1993). Omnibus robust procedures for assessment of multivariate normality and detection of multivariate outliers, in *Multivariate Environmental Statistics*, eds. G.P. Patil and C.R. Rao. North Holland. Amsterdam.

Small, N. J. H. (1978). Plotting squared radii. *Biometrika*, 65, 657-658.

Weiss, I. (1958). A test of fit for multivariate distributions. *Annals of Mathematical Statistics*, 29, 595-599.

Wilks S. S. (1962). *Mathematical statistics*. Wiley. New York.