

On the small sample properties of variants of Mardia's and Srivastava's kurtosis-based tests for multivariate normality

Zofia Hanusz, Joanna Tarasińska, Zbigniew Osypiuk

Department of Applied Mathematics and Computer Science, University of Life Sciences
in Lublin, Akademicka 13, 20-934 Lublin, Poland, e-mail: zofia.hanusz@up.lublin.pl;
joanna.tarasinska@up.lublin.pl, zbigniew.osypiuk@up.lublin.pl

SUMMARY

The kurtosis-based tests of Mardia and Srivastava for assessing multivariate normality (MVN) are considered. The asymptotic standard normal distribution of their test statistics, under normality, is often misused for too small samples. The purpose of this paper is to suggest mean-and-variance corrected versions of the Mardia and Srivastava test statistics. Simulation studies evaluating both the true sizes and the powers of original and corrected tests against selected alternatives are presented and compared to the size and the power of the Henze–Zirkler test. The proposed corrected statistics have empirical sizes closer to a nominal significance level than the original ones. It is also shown that the corrected versions of the tests can be more powerful than the original ones.

Key words: Henze–Zirkler test, true size studies, power studies, heavy-tailed distributions, light-tailed distributions

1. Introduction

Many methods of multivariate analysis, including MANOVA, discriminant analysis and multivariate regression, are based on the assumption of multivariate normality. Different types of statistical analysis assuming normality are sensitive to different types of departure from this assumption. It is known, for example (Mardia et al., 1979), that tests of means are more sensitive to skewness, while tests for equality of covariance matrices are affected by kurtosis (Layard, 1974). In addition, though tests of means such as t -tests and

ANOVA are considered to be robust, Tiku et al. (1986) noted that the power and the Type I error could be violated by skewness and kurtosis in the case of small samples.

In the literature can be found many tests for checking multivariate normality, in fact more than 50 (see e.g. Mecklin and Mundfrom, 2004). A first step for checking multivariate normality may be testing each variable for univariate normality, as univariate normality is necessary for multivariate normality. Of course it is not sufficient. Looney (1995) discussed an example (transformed Royston's hematology data) where univariate tests did not reject marginal normality but multivariate tests did. Additionally, the multiple testing problem appears in such a case.

Looney (1995) recommends using different tests for checking multivariate normality, as different tests are sensitive to different types of departures from normality. In general, practitioners prefer to use test statistics which have known asymptotic distributions, so there is no need either for using special tables or for determining critical values via simulation. Unfortunately, in practice these asymptotic distributions are frequently misused for too small samples.

In this paper we consider Mardia's and Srivastava's asymptotic kurtosis-based tests for multivariate normality, and suggest their mean-and-variance corrected versions. The small sample properties of these four test statistics are of special interest. The details of the tests are presented in Section 2. In Section 3 we evaluate the true sizes of the tests by a simulation study. The results of the simulation study on the power of the tests for selected alternatives are given in Section 4. In Section 5 some concluding remarks on comparison of the tests are presented. The consistent invariant Henze–Zirkler test (Henze and Zirkler, 1990), based on the distance between the empirical and theoretical characteristic functions, is also taken into account as a rival of four tests based on kurtosis. The choice of the Henze–Zirkler test is due to the fact that this is generally considered to be a very good test. Moreover, it is implemented in SAS Software.

All simulations and computations presented in the paper are carried out independently in the programs SAS/IML (SAS Institute Inc., 1989) and R (R Development Core Team, 2008).

2. Description of the tests

Let \mathbf{X} be a p -variate random variable with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We are interested in testing the multivariate normality of \mathbf{X} . In this section we describe the five tests which are investigated in the paper, namely two original tests proposed by Mardia and Srivastava, their two corrected versions, and the Henze–Zirkler test.

2.1. Mardia type tests

Mardia (1970) introduced the measure of multivariate kurtosis as

$$\beta_{2,p} = E\left[(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \boldsymbol{\mu})\right]^2,$$

which is equal to $p(p+2)$ for a p -variate normal distribution. If $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ form a random sample from the same distribution as \mathbf{X} then we have the following estimate of $\beta_{2,p}$:

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

$$\text{where } \bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j \text{ and } \mathbf{S} = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'.$$

The statistic $b_{2,p}$ is invariant under linear transformations, thus there exists a null distribution that is independent of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Mardia (1970) and Mardia and Kanazawa (1983) gave formulae for moments of $b_{2,p}$ under multivariate normality, namely:

$$E(b_{2,p}) = \frac{p(p+2)(n-1)}{n+1},$$

$$\begin{aligned} \text{Var}(b_{2,p}) &= \frac{8p(p+2)(n-3)(n-p-1)(n-p+1)}{(n+1)^2(n+3)(n+5)}, \\ \mu_3(b_{2,p}) &= \frac{64p(p+2)(p+8)}{n^2} + O(n^{-3}). \end{aligned}$$

Thus, $b_{2,p}$ underestimates $\beta_{2,p}$ and its distribution is rather right-skewed, the skewness increasing with p . Mardia (1970) proved that, under multivariate normality, the statistic

$$M = \frac{b_{2,p} - \frac{p(p+2)(n-1)}{n+1}}{\sqrt{\frac{8p(p+2)}{n}}}$$

has the asymptotic standard normal distribution.

However, a frequently used test statistic for MVN based on Mardia's kurtosis is of the form:

$$M_1 = \frac{b_{2,p} - p(p+2)}{\sqrt{\frac{8p(p+2)}{n}}}.$$

Under the null hypothesis of a multivariate normal distribution, the asymptotic distribution of M_1 is standard normal. Mardia (1974) gave tables of critical values for $b_{2,p}$ and warned that even in the univariate case “ b_2 is very skewed for $n = 100$ and is hardly normal for $n = 1000$ ” (Mardia, 1980). In spite of this, many handbooks recommend using the asymptotic standard normal distribution of M_1 .

In the present work, instead of M_1 we propose to use the following mean- and-variance corrected version of Mardia's kurtosis test statistic:

$$M_2 = \frac{b_{2,p} - \frac{p(p+2)(n-1)}{n+1}}{\sqrt{\frac{8p(p+2)(n-3)(n-p-1)(n-p+1)}{(n+1)^2(n+3)(n+5)}}},$$

which takes into account the exact two first moments of $b_{2,p}$. Under the null hypothesis of a multivariate normal distribution, the asymptotic distribution of M_2 is standard normal.

To compare the original M_1 and corrected M_2 test statistics we generated 10 000 random samples from the multivariate normal distribution with null expectation vector and identity covariance matrix for different combinations of n and p . Example histograms of M_1 and M_2 for $n = 10$, $p = 2$ together with the standard normal density curve are given in Figure 1. It can be noticed that the histogram of M_1 is far from a normal curve. However, even for such small sample size the corrected version of the test statistic (i.e. M_2) improved the normality of the histogram, although it is still skewed.

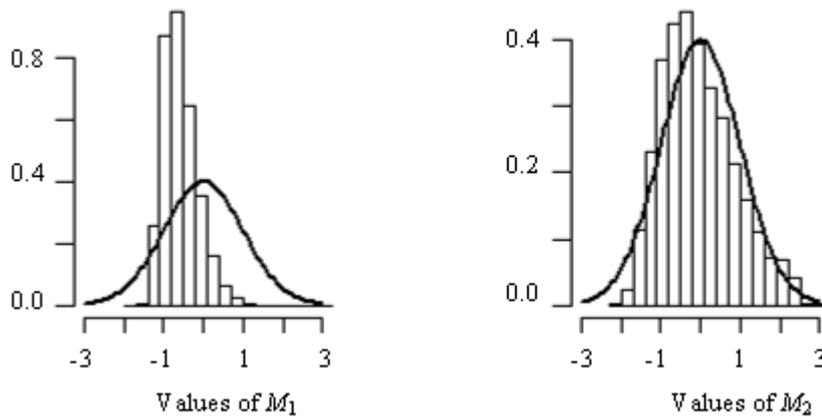


Figure 1. Histograms for statistics M_1 and M_2 obtained on the basis of 10 000 random samples generated from bivariate normal distribution, $n=10$

In power studies some authors (e.g. Henze, Zirkler, 1990) consider the asymptotic distribution of $b_{2,p}$ as normal with the expectation $p(p+2)(n-1)(n+1)^{-1}$ and the variance $8p(p+2)n^{-1}$, namely the statistic M . The histogram in Figure 2 shows that such correction is insufficient even for $n = 25$. The true size of the

test based on M , evaluated on the basis of 10 000 random samples, is 0.02 for $\alpha = 0.05$ and 0.03 for $\alpha = 0.1$, and is thus much lower than the postulated significance level α .

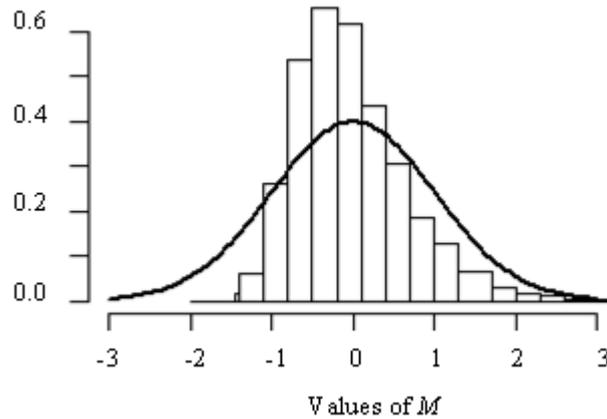


Figure 2. Histogram for statistic M obtained on the basis of 10 000 random samples generated from a bivariate normal distribution, $n=25$

2.2. Srivastava type tests

Srivastava (1984) proposed a test for assessing multivariate normality based on a coefficient of kurtosis that differs from Mardia's. Srivastava's sample kurtosis is the average of p univariate kurtosis coefficients for principal components. Let us recall his approach.

Let $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_p]$ be an orthogonal matrix such that $\mathbf{H}'\mathbf{S}\mathbf{H} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, and let $y_{ji} = \mathbf{h}_i' \mathbf{X}_j$ ($j = 1, \dots, n; i = 1, \dots, p$). Srivastava defines the sample kurtosis as follows:

$$b_{2,p}^S = \frac{1}{p} \sum_{i=1}^p \frac{1}{n \lambda_i^2} \sum_{j=1}^n (y_{ji} - \bar{y}_{\cdot i})^4 = \frac{1}{p} \sum_{i=1}^p b_{2(i)},$$

where $b_{2(i)}$ is the univariate kurtosis for the i th principal component. Under multivariate normality, the $b_{2(i)}$ ($i = 1, \dots, p$) are asymptotically independent and have the asymptotic normal distribution with mean 3 and variance $24n^{-1}$. Hence Srivastava proposes the following asymptotic test statistic:

$$S_1 = \left(b_{2,p}^S - 3 \right) \sqrt{\frac{np}{24}},$$

and states that under the null hypothesis of a multivariate normal distribution, the asymptotic distribution of S_1 is standard normal. Srivastava illustrates his test with an example with $n = 27$ and $p = 2$. It should be mentioned that Srivastava's test statistic is not invariant under linear transformations.

In this paper we propose the following mean-and-variance corrected version of S_1 :

$$S_2 = \frac{b_{2,p}^S - \frac{3(n-1)}{n+1}}{\sqrt{\frac{24n(n-3)(n-2)}{p(n+1)^2(n+3)(n+5)}}}.$$

Under the null hypothesis of a multivariate normal distribution, the asymptotic distribution of S_2 is standard normal. Figure 3 presents the histograms of S_1 and S_2 based on 10 000 random samples generated from a bivariate normal distribution for $n = 10$, together with the standard normal density curve.

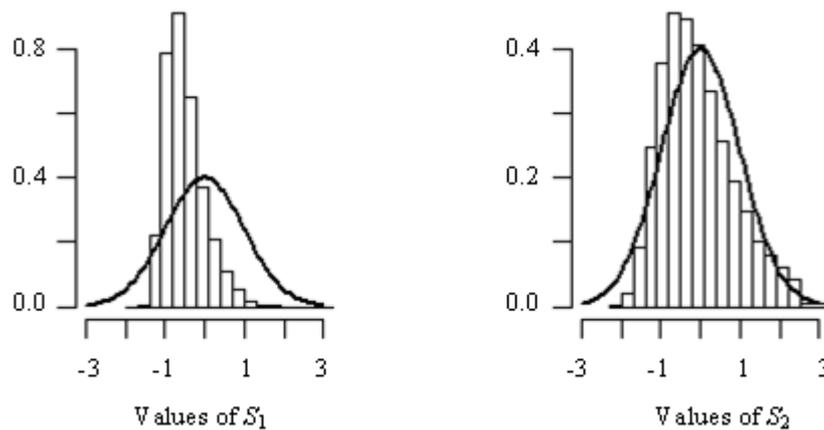


Figure 3. Histograms for statistics S_1 (left panel) and S_2 (right panel), obtained on the basis of 10 000 random samples generated from a bivariate normal distribution, $n=10$

2.3. Henze–Zirkler (H-Z) test

Henze and Zirkler (1990) proposed a class of invariant consistent tests for MVN based on the distance between the empirical and the normal distribution. The test statistic is of the form:

$$T_{n,\beta} = n(4I\{\mathbf{S} \text{ is singular}\} + D_{n,\beta}I\{\mathbf{S} \text{ is nonsingular}\}),$$

where \mathbf{S} is a sample covariance matrix,

$$D_{n,\beta} = \int_{R^p} \left| \Psi_n(\mathbf{t}) - \exp\left(-\frac{1}{2}\mathbf{t}'\mathbf{t}\right) \right|^2 \varphi_\beta(\mathbf{t}) d\mathbf{t},$$

$I\{\cdot\}$ is the indicator function,

$$\varphi_\beta(\mathbf{t}) = (2\pi\beta^2)^{-\frac{p}{2}} \exp\left(-\frac{\mathbf{t}'\mathbf{t}}{2\beta^2}\right), \text{ and } \Psi_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n \exp(it'\mathbf{Y}_j)$$

is the empirical characteristic function of scaled residuals $\mathbf{Y}_j = \mathbf{S}^{-1/2}(\mathbf{x}_j - \bar{\mathbf{X}})$, $j = 1, \dots, n$.

In this paper the parameter

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

is taken according to Henze and Zirkler's suggestion.

3. Simulation study of the true size of the tests

The aim of this section is to show that asymptotic normal distribution of the statistics M_1 and S_1 , described in Section 2, should not be recommended for testing multivariate normality. However, their corrected versions M_2 and S_2 work quite well.

For each combination of $n=10, 25, 50$ and $p=2, 3, 4$, ten thousand random samples from the multivariate normal distribution with null mean vector and

identity covariance matrix were generated, and the true sizes of tests based on M_1 , M_2 , S_1 , S_2 and the H-Z test were computed. The simulation results for significance levels $\alpha = 0.05$ and $\alpha = 0.1$ are given in Table 1. Tests based on M_1 , M_2 , S_1 , S_2 reject normality if the absolute value of the test statistics exceeds the $(1 - \alpha/2)$ th quantile of the standard normal distribution. For the Henze–Zirkler test normality is rejected if the test statistic exceeds a critical value given in Henze and Zirkler (1990, formula (3.4), p. 3605). It can be noticed that M_1 and S_1 give true sizes much below the nominal α even for $n = 50$, whereas the true sizes of M_2 and S_2 are close to the nominal value.

Table 1. True sizes (in percentages) of tests, obtained on the basis of 10 000 random samples generated from a p -variate normal distribution for significance level α

n	p	$\alpha = 0.05$					$\alpha = 0.1$				
		M_1	M_2	S_1	S_2	H-Z	M_1	M_2	S_1	S_2	H-Z
10	2	0	5	0	5	4	0	8	0	8	9
	3	0	5	0	5	3	0	9	0	9	8
	4	0	4	0	5	3	0	9	1	10	9
25	2	1	5	1	5	5	2	8	2	8	10
	3	1	5	1	5	5	3	9	3	8	10
	4	1	5	1	5	5	4	9	4	8	10
50	2	2	5	2	5	5	5	8	5	8	10
	3	2	5	2	5	5	5	8	5	8	10
	4	2	5	2	5	5	6	9	6	9	10

4. Simulation study on the power of the tests

To evaluate the power of the five tests being compared, 10 000 random samples from selected alternative distributions were generated for different combinations of n and p . Distributions with independent marginals, the mixtures of normal distributions and elliptically contoured distributions Pearson Type II (MPII) and Pearson Type VII (MPVII) were considered.

The Pearson Type II distribution with shape parameter $m > -1$ is a p -variate distribution with the following density function:

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{p}{2} + m + 1\right)}{\Gamma(m + 1)\pi^{\frac{p}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1 - (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]^m,$$

having support on the ellipsoid $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq 1$ (see Johnson, 1987). In the simulation study, MPPII with shape parameter $m = 0$ (i.e. the uniform distribution) and $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \mathbf{I}$ is considered; this is denoted by MPPII(0).

The Pearson Type VII distribution with shape parameter $\nu > 0$, called the general multivariate t distribution with ν degrees of freedom, is a p -variate distribution with the density function

$$f(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu + p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\pi\nu)^{\frac{p}{2}}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \left[1 + \nu^{-1} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]^{-\frac{\nu + p}{2}}.$$

The Pearson Type II and VII distributions were generated according to Johnson (1987) with the correction for MPVII type in the formula (6.9) from that book, namely $\mathbf{X} = (\sqrt{Y/\nu})^{-1} \mathbf{Z} + \boldsymbol{\mu}$, where \mathbf{Z} is $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ and independent of Y , which is χ_{ν}^2 . In the simulation study, the MPVII distribution with $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \mathbf{I}$ was taken; this is denoted by MPVII(ν).

The powers of four tests based on kurtosis were calculated as the percentage of samples for which the test statistics exceed the $\pm(1 - \alpha/2)$ th quantile of the standard normal distribution, where α is the significance level of the test. In the literature, some authors compare the power of tests by taking critical values which give the true size of the tests equal to nominal significance level. However, practitioners usually do not realize that tests do not maintain the nominal significance level, and take critical values ± 1.96 (for $\alpha = 0.05$) or ± 1.645 (for $\alpha = 0.1$).

The power of the Henze–Zirkler test was evaluated as the percentage of samples for which the test statistic exceeds the critical value given in Henze and Zirkler (1990, p. 3605).

As the corrected test statistics M_2 and S_2 usually give values shifted to the right and more scattered (compared to the original versions M_1 and S_1) it is obvious that the power of the mean-and-variance corrected tests improves for alternatives with kurtosis greater than normal (see Figure 4). For small samples the improvement can be huge, as Table 2 and Figure 4 show.

Table 2 gives the estimated powers obtained on the basis of 10 000 random samples generated from products of the t_1 , t_2 and t_7 distributions (denoted by t_v^p), and the multivariate t distributions with 2 and 7 degrees of freedom. These distributions have heavier tails than the normal distribution. For $v = 1$ and $v = 2$ degrees of freedom their tails are even so heavy that kurtosis does not exist. For $v > 4$ the kurtosis of the product of the t_v distributions is $\beta_{2,p} = p(p+2) + 6p(v-4)^{-1}$ and the kurtosis of MPVII(v) is $\beta_{2,p} = p(p+2) + 2p(p+2)(v-4)^{-1}$ (Henze, 1994).

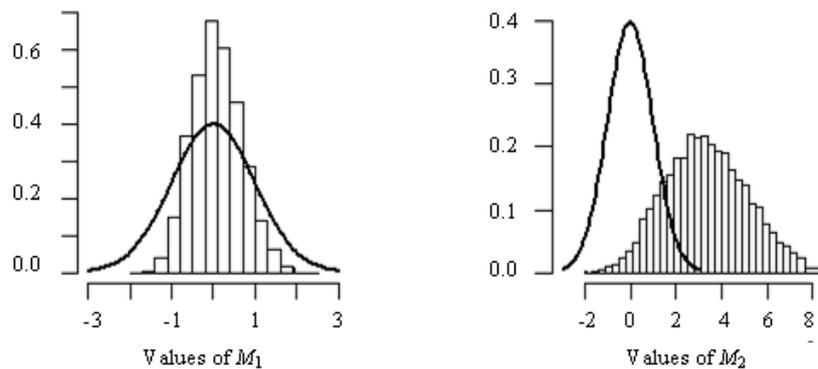


Figure 4. Histograms for statistics M_1 (left panel) and M_2 (right panel) obtained on the basis of 10 000 random samples generated from the product of t_1 distribution, $n=10$, $p=4$

Table 2. Power (in percentages) evaluated on the basis of 10 000 random samples generated from heavy-tailed distributions

<i>n</i>	<i>p</i>	Alternative distribution	$\alpha = 0.05$					$\alpha = 0.1$				
			M_1	M_2	S_1	S_2	H-Z	M_1	M_2	S_1	S_2	H-Z
10	2	t_1^2	18	75	39	75	70	25	79	48	78	78
		t_2^2	3	41	10	43	33	5	47	15	48	44
		MPVII(2)	3	45	9	43	37	5	51	14	48	48
		t_7^2	0	11	0	12	7	0	15	1	16	14
		MPVII(7)	0	12	1	13	8	0	17	1	16	15
		t_1^3	4	78	42	82	70	9	83	51	85	79
	3	t_2^3	0	40	9	47	29	0	48	13	52	41
		MPVII(2)	0	46	8	48	36	1	54	12	53	47
		t_7^3	0	9	0	11	5	0	14	0	15	11
		MPVII(7)	0	11	0	15	7	0	17	1	19	14
		t_1^4	0	78	44	87	68	0	83	53	89	77
		t_2^4	0	37	8	49	24	0	45	12	55	38
25	4	MPVII(2)	0	46	8	53	32	0	55	12	58	45
		t_7^4	0	8	0	11	5	0	13	1	16	12
		MPVII(7)	0	12	0	16	7	0	18	1	21	14
		t_1^2	97	99	97	99	99	98	99	98	99	99
		t_2^2	69	83	70	81	73	74	86	75	84	79
		MPVII(2)	75	87	72	82	78	79	89	76	85	83
	3	t_7^2	10	23	11	21	12	14	28	14	26	20
		MPVII(7)	15	29	15	26	15	18	35	18	31	24
		t_1^3	99	100	99	100	100	99	100	99	100	100
		t_2^3	76	90	77	87	79	80	92	81	90	84
		MPVII(2)	83	94	78	88	86	87	96	83	92	90
		t_7^3	8	25	10	22	11	11	31	14	28	19
4	MPVII(7)	14	37	16	31	17	19	43	20	37	26	
	t_1^4	99	100	100	100	100	100	100	100	100	100	
	t_2^4	77	93	81	90	81	81	95	84	92	86	
	MPVII(2)	87	97	84	93	91	90	98	87	94	93	
	t_7^4	5	25	9	22	11	8	32	13	28	19	
	MPVII(7)	14	44	18	35	20	18	51	22	40	29	

For the alternative distributions with kurtosis less than normal, products of the uniform distributions with $\beta_{2,p} = p(p+2) - 1.2p$ and MPVII(*m*) with $\beta_{2,p} = p(p+2)(p+2m+2)(p+2m+4)^{-1}$ (Henze, 1994), the corrected tests can give less power. This is not surprising because for these distributions the test statistics M_1 and S_1 have negative values and the values of their corrected

versions M_2 and S_2 are greater. Additionally, it should be remembered that M_1 and S_1 give true test sizes far from the significance level.

Table 3 contains the estimated powers of the tests in the case of the product of the uniform distributions on $[0,1]$, denoted by $u[0,1]^p$, and the multivariate uniform distribution on the set $\{\mathbf{x} \in R^p : \mathbf{x}'\mathbf{x} \leq 1\}$, i.e. MPII(0). The kurtosis of these distributions does not differ very much from normal $p(p+2)$, the sample sizes $n = 25$ and 50 being considered. For $n = 10$ the power would be very small, not greater than α .

Table 3. Power (in percentages) evaluated on the basis of 10 000 random samples generated from light-tailed distributions

n	p	Alternative distribution	$\alpha = 0.05$					$\alpha = 0.1$				
			M_1	M_2	S_1	S_2	H-Z	M_1	M_2	S_1	S_2	H-Z
25	2	$u[0,1]^2$	2	10	1	5	24	23	33	12	18	42
		MPII(0)	4	18	1	6	27	37	49	17	25	43
	3	$u[0,1]^3$	10	17	3	5	21	41	40	15	16	37
		MPII(0)	26	38	5	10	23	67	66	28	29	39
50	4	$u[0,1]^4$	18	19	4	5	17	51	41	16	14	32
		MPII(0)	46	48	10	11	19	80	73	32	27	35
	2	$u[0,1]^2$	62	65	30	32	67	88	87	50	49	81
		MPII(0)	84	86	51	55	70	97	97	83	82	83
50	3	$u[0,1]^3$	78	75	26	24	64	94	91	49	44	78
		MPII(0)	98	97	59	56	71	100	100	85	81	83
	4	$u[0,1]^4$	84	77	24	20	57	96	91	46	38	73
		MPII(0)	100	99	58	52	69	100	100	81	74	81

Table 4 presents the simulated powers of the tests in the case of a mixture of normal distributions, i.e. the distribution of $\mathbf{Y} = K\mathbf{X}_1 + (1-K)\mathbf{X}_2$, where K , \mathbf{X}_1 and \mathbf{X}_2 are independent, $P(K = 1) = q$, $P(K = 0) = 1 - q$ and $\mathbf{X}_1 \sim N_p(\mathbf{0}, \mathbf{I})$, $\mathbf{X}_2 \sim N_p([4 \ 0]', \mathbf{I})$. Thus the distributions of \mathbf{X}_1 and \mathbf{X}_2 are at a Mahalanobis distance $\delta = 4$ from each other.

Table 4. Power (in percentages) evaluated on the basis of 10 000 random samples generated from mixtures of normal distributions

<i>n</i>	<i>q</i>	<i>p</i>	$\alpha = 0.05$				$\alpha = 0.1$					
			<i>M</i> ₁	<i>M</i> ₂	<i>S</i> ₁	<i>S</i> ₂	H-Z	<i>M</i> ₁	<i>M</i> ₂	<i>S</i> ₁	<i>S</i> ₂	H-Z
50	0.95	2	35	49	41	52	38	42	56	48	59	48
		3	19	36	28	41	25	26	43	35	48	35
		4	11	27	19	32	17	16	35	26	40	27
	0.90	2	23	38	30	43	66	31	46	38	52	75
		3	12	26	20	33	47	17	33	27	41	60
		4	6	19	14	26	31	10	26	19	34	44
	0.80	2	4	9	5	9	86	7	13	9	14	92
		3	3	7	4	8	64	6	11	7	12	76
		4	2	6	3	8	43	6	11	7	12	57
	0.50	2	8	9	16	18	72	26	25	38	37	85
		3	6	6	14	13	41	19	15	30	27	57
		4	6	4	13	11	25	17	12	27	22	39
100	0.95	2	68	76	75	81	60	75	82	80	85	70
		3	45	58	59	68	44	55	67	67	75	55
		4	30	45	47	57	32	38	54	56	66	43
	0.90	2	45	57	56	66	92	55	65	66	74	96
		3	26	39	42	52	81	34	48	52	62	88
		4	16	28	32	44	65	23	37	42	53	75
	0.80	2	5	8	6	9	100	9	12	11	14	100
		3	4	7	5	8	96	8	12	9	13	98
		4	3	6	4	7	84	8	10	8	11	91
	0.50	2	37	34	53	51	99	58	54	72	69	100
		3	20	15	36	32	87	36	29	53	48	93
		4	15	10	29	24	63	28	20	45	38	76

It is surprising that in the case of Mardia's and Srivastava's tests, greater disturbances are sometimes detected with less power than smaller disturbances. This is explained by the following formula for coefficient of kurtosis for a mixture (Henze, 1994):

$$\beta_{2,p} = p(p+2) + \lambda\delta^4 \frac{1-6\lambda}{(1+\lambda\delta^2)^2}, \quad \lambda = q(1-q).$$

The plot for $f(q) = \lambda \delta^4 (1 - 6\lambda)(1 + \lambda \delta^2)^{-2}$ with $\delta = 4$ is given in Figure 5. Thus the mixture with $q \cong 0.8$ has almost the same kurtosis as the normal distribution.

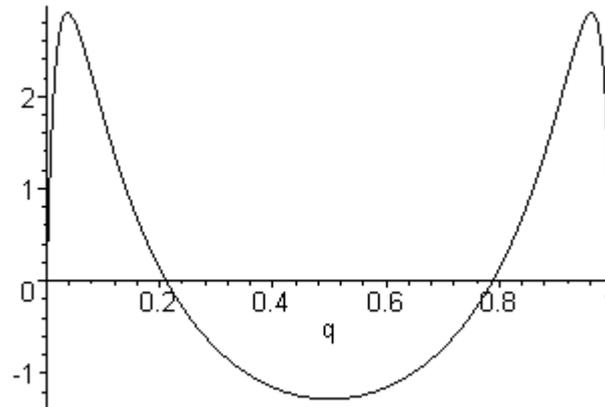


Figure 5. Plot of the function $f(q)$, $\delta = 4$

5. Conclusions

On the basis of the simulation studies we can state that none of the tests is uniformly the best. Nevertheless, some specific remarks can be formulated.

(1) The results in Table 1 show that the true sizes of tests based on M_1 and S_1 are much lower than the nominal significance level α . However, the true sizes of test procedures based on the corrected statistics M_2 and S_2 are very close to the significance level, especially for $\alpha = 0.05$. It should be mentioned that the true size of the Henze–Zirkler test for $n = 10$ is a little smaller than α .

(2) From the results in Table 2 we can conclude that the corrected M_2 and S_2 tests for heavy-tailed distributions, such as the product of the t distributions and the multivariate t distributions, are more powerful than the original M_1 and S_1 tests respectively. The power of M_2 and S_2 is greater than the power of the Henze–Zirkler test. For $n = 25$ M_2 is the best test, while for $n = 10$ S_2 is mostly the best. Generally, we can see that for all tests the disturbance from normality is more easily detected in the case of multivariate t than in the case of

a product of t distributions with the same degrees of freedom. This is not surprising if we consider the tests based on kurtosis, because MPVII distribution differs in kurtosis from the multivariate normal distribution more than the product of t distributions with the same degrees of freedom. However, we can notice that the same rule also applies to the Henze–Zirkler test.

(3) The results in Table 3 show that for light-tailed distributions, such as the product of the uniform distributions and the multivariate uniform distribution, the corrected tests M_2 and S_2 can be less powerful than the original M_1 and S_1 respectively. Mardia-type tests are much better than Srivastava-type ones, and mostly better than the Henze–Zirkler test. The power of Mardia-type tests increases with p , while the power of the Henze–Zirkler test decreases with p . All tests detect non-normality more easily for MP1I(0) than for the product of the uniform distributions.

(4) Regarding the results in Table 4 we can notice that for mixtures of normal distributions the corrected tests M_2 and S_2 are more powerful than M_1 and S_1 respectively, except for the case of $q = 0.5$. In this case the kurtosis for the mixture is lower than normal $p(p + 2)$ (see Figure 5). The Henze–Zirkler test is the best at detecting non-normality except for the case of $q = 0.95$, i.e. for a small disturbance of normality. In this case S_2 is the best. The tests based on kurtosis are completely useless when $q = 0.80$, as the kurtosis of such a mixture does not differ from the normal $p(p + 2)$. The power of all tests decreases with p .

(5) On the basis of the results presented, one more remark can be made. Horswell and Looney (1992) claimed that affine-invariant procedures are better than coordinant-dependent ones for correlated variables, whereas coordinant-dependent procedures are better when the variables are uncorrelated. This seems not to be valid when we observe the results in Tables 2–4. In all cases the variables are uncorrelated, and yet affine-invariant Mardia tests can be better than coordinant-dependent Srivastava tests. However, it should be remembered that Horswell and Looney (1992) did not consider Srivastava’s test in their paper.

REFERENCE

- Henze N., Zirkler B. (1990): A class of invariant consistent tests for multivariate normality. *Communication in Statistics – Theory and Methods* 19: 3595-3617.
- Henze N. (1994): On Mardia's kurtosis test for multivariate normality. *Communication in Statistics – Theory and Methods* 23: 1031-1945.
- Horswell R.L., Looney S.W. (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.
- Johnson M.E. (1987): *Multivariate Statistical Simulation*. New York: John Wiley & Sons.
- Layard M.W.J. (1974). A Monte Carlo comparison of tests for equality of covariance matrices. *Biometrika* 16: 461-465.
- Looney S.W. (1995): How to use tests for univariate normality to assess multivariate normality. *The American Statistician* 29: 64-70.
- Mardia K.V. (1970): Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57: 519-530.
- Mardia K.V. (1974): Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. *Sankhya B* 36:115-128.
- Mardia K.V. (1980): Tests of univariate and multivariate normality. In: *Handbook of Statistics 1*, ed. P.R. Krishnaiah, Amsterdam: North-Holland Publishing Company: 279-320.
- Mardia K.V., Kanazawa M. (1983): The null distribution of multivariate kurtosis. *Communication in Statistics – Simulation and Computation* 12: 569-576.
- Mardia K.V., Kent J.T., Bibby J.M. (1979): *Multivariate Analysis*. New York: Academic Press.
- Mecklin C.J., Mundfrom D.J. (2004): An appraisal and bibliography of tests for multivariate normality. *International Statistical Review* 72: 123-138.
- R Development Core Team (2008): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- SAS Institute Inc. (1989): *SAS/IML Software: Usage and Reference, Version 6 (First Edition)*, SAS Institute, Cary, NC.
- Srivastava M.S. (1984). A measure of skewness and kurtosis and a graphical method for assessing multivariate normality. *Statistics & Probability Letters* 2: 263-267.
- Tiku M.L., Tan W.Y., Balakrishnan N. (1986): *Robust Inference*. New York: Marcel Dekker.