

Geometric aspects of robust testing for normality and sphericity

Wolf-Dieter Richter^a, Luboš Střelec^b, Hamid Ahmadinezhad^c and Milan Stehlík^{de}

^aInstitute of Mathematics, University of Rostock, Rostock, Germany;

^bDepartment of Statistics and Operation Analysis, Mendel University in Brno, Brno, Czech Republic;

^cDepartment of Mathematical Sciences, Loughborough University, Loughborough, UK;

^dDepartment of Applied Statistics, Johannes-Kepler-University Linz, Austria;

^eInstitute of Statistics, University of Valparaíso, Valparaíso, Chile

ARTICLE HISTORY

Compiled October 30, 2016

Geometric aspects of robust testing for normality and sphericity

ABSTRACT

Stochastic Robustness of Control Systems under random excitation motivates challenging developments in geometric approach to robustness. The assumption of normality is rarely met when analyzing real data and thus the use of classic parametric methods with violated assumptions can result in the inaccurate computation of p -values, effect sizes, and confidence intervals. Therefore, quite naturally, research on robust testing for normality has become a new trend. Robust testing for normality can have counter-intuitive behavior, some of the problems have been introduced in [46]. Here we concentrate on explanation of small-sample effects of normality testing and its robust properties, and embedding these questions into the more general question of testing for sphericity. We give geometric explanations for the critical tests. It turns out that the tests are robust against changes of the density generating function within the class of all continuous spherical sample distributions.

KEYWORDS

Huberization, trimming, Lehman-Bickel functional, Monte Carlo simulations, power comparison, robust tests for normality, normality, sphericity

Classification codes: 62H15, 62H10

1. Introduction

Classic parametric statistical significance tests, such as analysis of variance and least squares regression, are widely used by researchers in many disciplines of chemistry, economics or social sciences. For classic parametric tests to produce accurate results, the assumptions underlying them (e.g. normality and homoscedasticity) must be satisfied. These assumptions are rarely met when analyzing real data. The use of classic parametric methods with violated assumptions can result in the inaccurate computation of p -values, effect sizes, and confidence intervals. This may lead to substantive errors in the interpretation of data.

Recently, several articles on robust testing for normality have been written (see for example [6], [9], [14], [15], [45] and [46]). However, deeper understanding of geometry of normality tests, normality and deviations from normality is still ongoing problem. To address these topics is the main objective of this paper. In particular we compare the recently introduced robust tests on the base of RT class, thus justifying RT class to be a good base for robust tests based on robustification of first four moments. In particular we show that selected RT class of tests accommodate the alternatives which are problematic for JB test: bimodal, Weibull and uniform alternatives – see [49]. It is worth a note that since JB test has no finite exponential moments, there were no asymptotical efficiency assessment to the best knowledge of the authors. From systematic study of RT class the superiority of Geary and Uthoff tests has been found for bimodal alternatives.

We emphasize that in this paper we use the assumption that possible contamination is only due to outliers. Thus we use three techniques for outliers filtering: trimming, huberization and functional approach introduced by [3]. For normality testing when the whole distribution may be contaminated see e.g. [1].

Testing for normality is a very important issue, for example in chemometrics. It is also notable that several books in the literature, e.g. in analytical chemistry, devote a section to normality testing; see for example [27]. However, robust testing for normality is still not completely communicated to the community of chemometricians. In the next section we illustrate necessity of robust testing for normality in chemometrics.

In section 3 the M-estimates are discussed. In section 4 the general RT class of tests is presented. We also list already known tests belonging to RT class, together with the most significant new tests. We also show on convenient examples that tests from RT class improve substantially the weaknesses of classical JB test. Sections 5, 6 and 7 contain theoretical explanation on geometrical behavior of tests and algebraic manipulations on main quadratic. In section 8 the comparative study of powers of robust tests is conducted. Discussion concludes the paper. Therein also practical guidelines for robust testing of normality are given. To maintain the continuity of explanation proofs and technicalities are put into Appendix.

2. Illustrative examples – necessity of robust testing for normality in chemometrics

In this section we illustrate the necessity of robust testing for normality in chemometrics. We present a few illustrative examples where testing for normality plays an important role.

2.1. Catalytic Isomerization: influence of truncation

In several chemical application truncation of normal distribution can have a severe impact on statistical decision. For illustration, let us consider kinetics of catalytic isomerization of n-Pentane (see [7]). Therein we have rates $r_1 = \frac{\gamma K_2(x_2-x_3|K)}{1+K_1x_1+K_2x_2+K_3x_3}$ and $r_2 = \frac{\gamma' K_2(x_2-x_3|K)}{(1+K_1x_1+K_2x_2+K_3x_3)^2}$, where K_i are adsorption equilibrium constants, x_i are partial pressures, K is equilibrium constant, γ is a constant dependent on catalyst and temperature T and $i = 1, 2, 3$ are indices for hydrogen, n -pentane and isopentane, respectively.

[7] introduced variables $y_1 = \left(\frac{x_2-x_3|K}{r_i}\right)$, $y_2 = (y_1)^{\frac{1}{2}}$ and tests for importance of higher factors, later is this discussed in [22]. In [44] we pointed out that we should be aware of the fact, that the values y_i , which are realizations of $y = \gamma K_2/(1 + \sum K_i x'_i)$, are not only positive but strictly over some positive constant (see their measured values in table III of [7] or realize their chemical meaning, i.e. k_i are equilibrium adsorption constants, x_i particular pressures of hydrogen, n -pentane, isopentane, respectively). Thus the proper test should be based on truncated normals. Therefore alongside the "outlier importance" discussed in [29], there is a further issue concerning the distributional deviations from F -distribution. For more discussion from point of view of chemometrics see [46].

2.2. Example: Control Charts for Chemical processes (positive kurtosis)

Here, we suppose some alternative distributions with zero skewness and positive excess (kurtosis). For the purpose of tests comparison we suppose the following alternatives commonly used in chemometrics and related fields:

- Laplace (t, s) distribution, where t and s are the location and scale parameters, respectively, defined as Laplace(0;1), with $SK = 0$ and $K_{exc} = 3$ where SK is skewness and K_{exc} is excess (kurtosis).
- t -Student(ν) distribution, where ν is the number of degrees of freedom, with $SK = 0$ for $\nu > 3$ and $K_{exc} = \frac{6}{\nu-4}$ for $\nu > 4$ where SK is skewness and K_{exc} is

kurtosis.

- Logistic (t, s) distribution, where t and s are the location and scale parameters, respectively, defined as $\text{Logistic}(0;1)$, with $SK = 0$ and $K_{exc} = 6/5$ where SK is skewness and K_{exc} is kurtosis.

Table 1 presents the power of selected classical and robust tests for normality against the mentioned alternatives (note that these results are based on results presented in [46]). Based on these results we can conclude that the best tests are SJ_{dir} , RJB and from RT class $TTRT2$ tests, which powerfull outperform the other tests. For example, power of the SJ_{dir} test against Laplace distribution with $SK = 0$ and $K_{exc} = 3$ for $n = 20$ is 0.391, RJB test has power 0.357 and finally the power of $TTRT2$ test is 0.345. On the other hand, commonly used the SW test has power only 0.260 and the classical JB test 0.307. If we suppose large sample size $n = 100$ the power of mentioned tests is following: $SJ_{dir} = 0.942$, $RJB = 0.889$, $TTRT2 = 0.862$, $SW = 0.796$ and finally $JB = 0.802$.

Similarly, if we suppose t -Student(ν) distribution $t(7)$ with $SK = 0$ and $K_{exc} = 2$ and $n = 100$, the results are following: SW test has power 0.363, $JB = 0.454$, $SJ_{dir} = 0.445$, $TTRT2 = 0.468$ and finally RJB test has power 0.477. Similar results are also achieved for t -Student(ν) distribution $t(5)$ and logistic distribution.

From the mentioned results we can see that power between the most powerful RJB , $SJ_{dir} = 0.445$ and $TTRT1$ tests vanished for distributions with excess kurtosis close to kurtosis of Gaussian normal distribution.

	$n = 20$				$n = 100$			
	Laplace	t_5	t_7	logistic	Laplace	t_5	t_7	logistic
<i>AD</i>	0.276	0.172	0.116	0.107	0.824	0.483	0.283	0.241
<i>DT</i>	0.290	0.223	0.160	0.142	0.725	0.587	0.394	0.332
<i>JB</i>	0.307	0.233	0.168	0.148	0.802	0.651	0.454	0.399
<i>JBU</i>	0.320	0.238	0.170	0.152	0.814	0.661	0.464	0.409
<i>LT</i>	0.232	0.139	0.093	0.083	0.699	0.329	0.173	0.154
<i>MC_{LR}</i>	0.058	0.051	0.052	0.049	0.130	0.058	0.054	0.050
<i>RJB</i>	0.357	0.241	0.169	0.153	0.889	0.678	0.477	0.424
<i>SJ_{dir}</i>	0.391	0.231	0.160	0.144	0.942	0.659	0.445	0.419
<i>SW</i>	0.260	0.186	0.133	0.116	0.796	0.569	0.363	0.301
<i>MMRT1</i>	0.268	0.208	0.148	0.130	0.790	0.631	0.436	0.375
<i>MMRT2</i>	0.263	0.204	0.142	0.126	0.787	0.626	0.429	0.368
<i>TTRT1</i>	0.099	0.068	0.064	0.063	0.341	0.306	0.156	0.107
<i>TTRT2</i>	0.345	0.235	0.165	0.151	0.862	0.667	0.468	0.413

Table 1. Power of the selected tests for normality against various heavy-tailed alternatives commonly used in chemometrics and related fields for $n = 20$ and $n = 100$

3. M-estimator of location

First of all, the M-estimates are a generalisation of the maximum likelihood (ML) estimates (see Maronna et al. 2006). Some popular robust M-estimates are Huber-estimate, Hampel estimate, Andrews' wave, Tukey's biweight. A lot of literature is available for robust M-estimates, for example, Huber (1981), Hampel et al. (1986), Rousseeuw and Leroy (1987), Staudte and Sheather (1990), Wilcox (1997) and

Maronna et al. (2006). We will use a function $\psi(x, C) = x$ for $|x| < C$ and $\psi(x, C) = C \text{sign}(x)$ for $|x| \geq C$.

We recall, that the most efficient location estimator, i.e. MLE, corresponds typically to the highest power at the alternative. Notice that MLE is a special form of M-estimate for a location parameter (since it is obtained for $\psi := \ln f$ and $\Psi = f'/f$, where $\ln f$ is log likelihood and the M-estimator can be expressed as $\arg \min \sum_{i=1}^n \psi(x_i - M)$, where M is the location parameter).

The robust inference emphasises the concepts of breakdown point and influence function of an estimator. The latter is then used to explore the efficiency and robustness properties of the estimator. Pure location problems are rare in practice, there is at least the need for simultaneous scale estimates to help in estimating the standard error of the location estimate. For our purpose (robustification of *JB* test) we robustify empirical variants of first four moments.

The reason so many classical procedures are nonrobust to outliers is that the parameters of the model are expressed in terms of moments, and their classical estimators are expressed in terms of sample moments, which are very sensitive to outliers. Another approach to robustness is to concentrate on the parameters of interest suggested by the problem under study. It may well turn out that these parameters can be expressed as functions of the underlying distribution independently of a particular parametric model; that is as descriptive measures. If these descriptive measures are judiciously chosen, their naturally induced estimators are robust to aberration in the data (see e.g. [3]).

4. General *RT* class

The general *RT* class is based on robustification of the classical Jarque-Bera test introduced by [4]. The general *RT* class test statistic is defined by [45] for purpose of robust testing for normality against Pareto tails and has the following general form

$$RT = \frac{k_1(n)}{C_1} \left(\frac{M_{j_1}^{\alpha_1}(r_1, T_{(i_1)}(s_1))}{M_{j_2}^{\alpha_2}(r_2, T_{(i_2)}(s_2))} - K_1 \right)^2 + \frac{k_2(n)}{C_2} \left(\frac{M_{j_3}^{\alpha_3}(r_3, T_{(i_3)}(s_3))}{M_{j_4}^{\alpha_4}(r_4, T_{(i_4)}(s_4))} - K_2 \right)^2, \quad (1)$$

where M_j are j th theoretical central moment estimators of the random variable defined as $M_j(r, T(F_n, s)) = \frac{1}{n-2r} \sum_{m=r+1}^{n-r} \varphi_j(X_{(m)} - T(F_n, s))$ for $j \in \{0, 1, 2, 3, 4\}$, where φ_j is a tractable and continuous function, where $\varphi_0(x) = \sqrt{\pi/2}|x|$ and $\varphi_j(x) = x^j$ for $j \in \{1, 2, 3, 4\}$, $X_{(m)}$ is the order statistic, $T(F_n, s)$ is a location functional applied to the sample X_1, X_2, \dots, X_n , r and s are the trimming constants for moments and location, respectively, K_1 and K_2 are small-sample variants of mean corrections, C_1 and C_2 are asymptotic constants, $\alpha_1, \alpha_2, \alpha_3$ and α_4 are exponents, and finally, $k_1(n)$ and $k_2(n)$ are functions of sample size n .

Note, that for the construction of the general *RT* class test statistic it was used location functional approach, which has been introduced by P. E. Bickel and E. L. Lehmann in a series of papers (e.g. [3]) and as was shown in [45] and [46] it looks to be playing a crucial role also by robust testing for normality. As it will be seen later, the power of *RT* class test mimics the effectiveness of location estimator in typical cases. Thus trade off between power and robustness is a typical issue here.

Definition 1. (Location Functional, see [3]) Let $T(F)$ be a function defined on the

set of distribution functions. We say that $T(F)$ is a location functional if the following conditions hold:

- (1) if G is stochastically larger than F then $T(G) \geq T(F)$,
- (2) $T(F_{aX+b}) = aT(F_X) + b$,
- (3) $T(F_{-X}) = -T(F_X)$.

Then we will call $\theta = T(F)$ a location parameter of F . Having a sample, we work with its empirical variant $\theta_n = T(F_n)$, F_n being an empirical *cdf* of a sample.

In the general *RT* class we used the following four different location estimators:

- mean: $T_{(0)} = \frac{1}{n} \sum_{i=1}^n X_i$,
- median: $T_{(1)} = F^{-1}(1/2)$,
- trimmed mean: $T_{(2)}(s) = \frac{1}{n-2s} \sum_{i=s+1}^{n-s} X_{(i)}$, where $X_{(i)}$ is the i -th order statistic of the sample and s is the trimming constant for location,
- pseudo-median: $T_{(3)} = \text{median}_{i \leq j} (X_i + X_j)/2$, i.e. the median of the set $\{(X_1 + X_1)/2, (X_1 + X_2)/2, (X_1 + X_3)/2, \dots, (X_1 + X_n)/2, (X_2 + X_2)/2, (X_2 + X_3)/2, \dots, (X_2 + X_n)/2, \dots, (X_{n-1} + X_n)/2, (X_n + X_n)/2\}$.

Note, that some theoretical results on consistency and asymptotic χ^2 -distribution of the general *RT* class test statistic can be found in [45] and [46].

In the *RT* class we have also a direct relation to M-estimation. From the classical M-estimation ([19], page 6) is proposed the odd function $\psi(x, c) = x$, for $|x| < c$ and $c \text{ signum}(x)$ otherwise. Notice, that a special case of *RT_{JB}* class fits to M-procedure based on

$$a_n \left[\sum_{i=1}^n \psi^3 \left(\frac{X_i - M_n}{\hat{\sigma}}, c \right) \right]^2 + b_n \left[\sum_{i=1}^n \psi^4 \left(\frac{X_i - M_n}{\hat{\sigma}}, c \right) - k \right]^2 \quad (2)$$

for a well choice of censoring, function ψ and $\hat{\sigma}$ being a robust estimator of variance.

The form of (2) has been already studied by [16]. Furthermore, we can use also variance functional construction given by Bickel-Lehman, similarly to the mean functional.

Remark 1. Constants C_1 and C_2 used in general *RT* class test statistic

As was noted in [46] choosing of appropriate constants C_1 and C_2 is the hardest aspect of the variants of *RT* class tests, because to obtain the constants C_1 and C_2 we need to find the expressions for $E(M_{n_1, n_2}^k)$ for a finite sample size. Such calculations are very tedious and therefore we obtained these constants from Monte Carlo simulations (see [45]). As was also mentioned in [46] the critical constant (for small and mid samples) under the trimming of moments ($r > 0$) are different from critical constants without trimming ($r = 0$), since only asymptotical distribution is normal (see [48]) in this case.

Remark 2. Trimming

Notice that two levels of trimming enter *RT* class test statistic: first trimming (with trimming constant s) enters trimming in the location estimator $T(F_n)$, the latter on trimming (with trimming constant r) enters $M_j(r, T(s))$. Amazing property of *RT* class and robust tests in general is, that power of *RT* class mimics the effectiveness of location estimator. Thus practitioner can tune how much of robustness is needed,

of course at price of the power. One should be really careful here: mechanical down-weighting of peculiar observations may divert attention from important clues to new discovery. Based on our simulation study we can suggest the suitable choosing of these constants: $s = r = 0.05n$.

Remark 3. Robustness of RT class

A statistical procedure is called robust, if its performance is insensitive to small deviations of the actual situation from the idealized theoretical model (see [20]). In particular, a robust procedure should be insensitive to the presence of a few "bad" observations (but at the same time the discordant minority of the observations might be prime source of information for improving the theoretical model). Here we assume that possible contamination is due to presence of outliers. Based on our research, we would like to point out the following robust aspects of RT class of tests:

- The classical JB test is not robust: it has a 0-breakdown point (see [6]).
- Note, that our tests are constructed on the robust Bickel-Lehman construction of location ([3]).
- The 2nd "version" of robust estimation is trimming of location with trimming constant s (so called "trimmed-mean tests") and moments with trimming constant r (so called "trim-trim tests"). Note that selected trim-trim tests have the same power level as Medcouple tests introduced by [6], so they are very robust.
- Normality testing procedure is typically a pre-test preceding some further testing or inference. Therefore one interesting issue is robustness with respect to other procedures, assuming that RT tests did not rejected normality. The influence of different shapes of distribution with the same first 4 moments on robustness has been discussed for sequential t -test by [30].

Remark 4. Trade off between power and robustness

Two typical extremal behaviors occur in robust testing: the tests which are more robust have smaller power (since they are not affected by single outliers) and tests with higher power are typically less robust (because they are affected by single outliers). An example of the first extreme case are the Medcouple tests and selected RT class tests based on trimming and an example of the second extreme case are the RT tests based on mean-median combination. To be more precise, for instance the robust test of normality could be also obtained by removing the outliers from the data, using and outlier detection rule such as provided by the boxplot or a rule based on robust estimators of location and scale. When the majority of the data are instead normally distributed, this is a valuable alternative to the robust tests based on medcouple as both the boxplot and the most popular robust estimators of location and scale (such as M-estimators) are based on this normal assumptions and thus will indicate the correct set of outliers. However it becomes more complicated when even majority of the data points do not come from a normal distribution (see [6]). However, probably it will be in many practical situations more worth to conduct test with clear outliers adaptively, i.e. recognizing and deleting in the first step and then using the zero breakdown robust version of JB for normality testing. To illustrate this general framework, we use the right-location-standard normal distribution based on mixture of two normal distributions with various parameters. For this purpose we assume the right-contamination with distribution function: $F = (1 - p)N(\mu_1, \sigma_1^2) + pN(\mu_2, \sigma_2^2)$ for $p = 0.05$, $\mu_1 = 0$, $\mu_2 = 3$, $\sigma_1^2 = \sigma_2^2 = 1$, i.e. $F = 0.95N(0, 1) + 0.05N(3, 1)$, for large sample size $n = 100$. For mentioned contamination and sample size the power of the classical Jarque-Bera test is 0.744, power of SW test is 0.685 and $MMRT1$ has power 0.686. On the other

hand the $TTRT1$ test has power only 0.190, MC_{LR} test only 0.076, the Jarque-Bera and Shapiro-Wilk tests with deleting outliers based on boxplot in the first step have power 0.110 and 0.055, respectively. For complete comparison for right and central contamination see Table 1 in [46].

4.1. A special known tests of RT class

As can be seen from (1) there exist a vast amount of RT class tests, which we can obtain for a different settings of $r_i, T_{(i)}(s_i)$ etc. Next we refer to the RT class tests which has been studied already in the literature. These are i) classical JB test, ii) test of Urzua, iii) robust JB (RJB) test, iv) skewness test b_1 , v) the kurtosis test $\sqrt{b_2}$, vi) the Geary test a , vii) the Uthoff test U , and viii) the SJ test.

i) The classical Jarque-Bera test is a special case of RT test without trimming for $K_1 = 0, K_2 = 3, k_1(n) = n, k_2(n) = n, C_1 = 6, C_2 = 24, \alpha_1 = 1, \alpha_2 = 3/2, \alpha_3 = 1, \alpha_4 = 2, T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(0)}, j_1 = 3, j_2 = 2, j_3 = 4, j_4 = 2$. The Jarque-Bera test statistic JB is defined as

$$JB = \frac{n}{6} \left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)^2 + \frac{n}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2.$$

As pointed out by several authors (see for example [51]) the classical JB test behaves well in comparison with some other tests for normality if the alternatives belong to the Pearson family. However, the JB test behaves very badly for distributions with short tails and bimodal shape, sometimes it is even biased (see [49]).

ii) Test of Urzua (see [51]) is a special case of RT test without trimming for $K_1 = 0, K_2 = 3, k_1(n) = ((n+1)(n+3))/(n-2), k_2(n) = ((n+1)^2(n+3)(n+5))/(n(n-2)(n-3)), C_1 = 6, C_2 = 24, \alpha_1 = 1, \alpha_2 = 3/2, \alpha_3 = 1, \alpha_4 = 2, T_{(i_1)} = T_{(i_2)} = T_{(i_3)} = T_{(i_4)} = T_{(0)}, j_1 = 3, j_2 = 2, j_3 = 4, j_4 = 2$. The Urzua's Jarque-Bera test statistic JBU is defined as

$$JBU = \frac{(n+1)(n+3)}{6} \left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)^2 + \frac{(n+1)^2(n+3)(n+5)}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2.$$

iii) The robust Jarque-Bera test (see [14]) is a special case of RT test without trimming for $K_1 = 0, K_2 = 3, k_1(n) = n, k_2(n) = n, C_1 = 6, C_2 = 64, \alpha_1 = 1, \alpha_2 = 3, \alpha_3 = 1, \alpha_4 = 4, T_{(i_1)} = T_{(i_3)} = T_{(0)}, T_{(i_2)} = T_{(i_4)} = T_{(1)}, j_1 = 3, j_2 = 0, j_3 = 4, j_4 = 0$. The robust Jarque-Bera test statistic RJB is defined as

$$RJB = \frac{n}{C_1} \left(\frac{\hat{\mu}_3}{J_n^3} \right)^2 + \frac{n}{C_2} \left(\frac{\hat{\mu}_4}{J_n^4} - 3 \right)^2.$$

iv) The skewness test b_1 is a special case of RT test without trimming for $K_1 = 0,$

$K_2 = 0$, $k_1(n) = n$, $k_2(n) = 0$, $C_1 = 6$, $\alpha_1 = 1$, $\alpha_2 = 3/2$, $T_{(i_1)} = T_{(i_2)} = T_{(0)}$, $j_1 = 3$ and $j_2 = 2$. The skewness test statistic b_1 is defined as

$$b_1 = \frac{n}{6} \left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)^2.$$

v) The kurtosis test $\sqrt{b_2}$ is a special case of RT test without trimming for $K_1 = 3$, $K_2 = 0$, $k_1(n) = n$, $k_2(n) = 0$, $C_1 = 24$, $\alpha_1 = 1$, $\alpha_2 = 2$, $T_{(i_1)} = T_{(i_2)} = T_{(0)}$, $j_1 = 4$ and $j_2 = 2$. The kurtosis test statistic $\sqrt{b_2}$ is defined as

$$\sqrt{b_2} = \frac{n}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2.$$

vi) The Geary's test a (see [13]) is a special case of RT test without trimming for $K_1 = 0$, $K_2 = 0$, $k_1(n) = 1/n$, $k_2(n) = 0$, $C_1 = 1$, $\alpha_1 = 1/2$, $\alpha_2 = 1/4$, $T_{(i_1)} = T_{(i_2)} = T_{(0)}$, $\varphi_0(x) = |x|$, $j_1 = 0$, $j_2 = 2$ as an alternative to b_2 , because the small sample properties where more tractable for a . The Geary's test statistic a (originally denoted w'_n) is defined as

$$a = \frac{1}{n} \sum_{i=1}^n \frac{|X_i - \bar{X}|}{\sqrt{m_2}}.$$

vii) The Uthoff's test U (see [52]) is a special case of RT test without trimming for $K_1 = 0$, $K_2 = 0$, $k_1(n) = 1/n$, $k_2(n) = 0$, $C_1 = 1$, $\alpha_1 = 1/2$, $\alpha_2 = 1/4$, $T_{(i_1)} = T_{(i_2)} = T_{(1)}$, $\varphi_0(x) = |x|$, $j_1 = 0$ and $j_2 = 2$. The Uthoff's test statistic U is defined as

$$U = \frac{1}{n} \sum_{i=1}^n \frac{|X_i - M_n|}{\sqrt{m_2}}.$$

viii) SJ test (see [15]) is a special case of RT test without trimming for $K_1 = 0$, $K_2 = 0$, $k_1(n) = \sqrt{\pi/2}/n$, $k_2(n) = 0$, $C_1 = 1$, $\alpha_1 = 1/4$, $\alpha_2 = 1/2$, $T_{(i_1)} = T_{(0)}$, $T_{(i_2)} = T_{(1)}$, $\varphi_0(x) = |x|$, $j_1 = 2$ and $j_2 = 0$. The SJ test statistic SJ is defined as

$$SJ = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n \frac{\sqrt{m_2}}{|X_i - M_n|}.$$

4.2. Tractable approach to RT class

In [45] we introduced general RT class test statistic (see (1)) as well as RT_{JB} and RT_{RJB} subclasses. The RT_{JB} subclass test statistic is defined as follows

$$RT_{JB} = \frac{n}{C_1} \left(\frac{M_3(r_1, T_{(i_1)}(s_1))}{M_2^{3/2}(r_2, T_{(i_2)}(s_2))} - K_1 \right)^2 + \frac{n}{C_2} \left(\frac{M_4(r_3, T_{(i_3)}(s_3))}{M_2^2(r_4, T_{(i_4)}(s_4))} - K_2 \right)^2, \quad (3)$$

which is a special case of RT test statistics for $k_1(n) = n, k_2(n) = n, \alpha_1 = 1, \alpha_2 = 3/2, \alpha_3 = 1, \alpha_4 = 2, j_1 = 3, j_2 = 2, j_3 = 4, j_4 = 2$.

Similarly, RT_{RJB} subclass test statistic is based on robustification of the robust Jarque-Bera test introduced by [15] and is defined as follows

$$RT_{RJB} = \frac{n}{C_1} \left(\frac{M_3(r_1, T_{(i_1)}(s_1))}{M_0^3(r_2, T_{(i_2)}(s_2))} - K_1 \right)^2 + \frac{n}{C_2} \left(\frac{M_4(r_3, T_{(i_3)}(s_3))}{M_0^4(r_4, T_{(i_4)}(s_4))} - K_2 \right)^2, \quad (4)$$

which is a special case of RT test statistics for $k_1(n) = n, k_2(n) = n, \alpha_1 = 1, \alpha_2 = 3, \alpha_3 = 1, \alpha_4 = 4, j_1 = 3, j_2 = 0, j_3 = 4, j_4 = 0$.

By clustering based on power values from all analyzed tests of RT class the following representatives with good properties against most common alternatives has been obtained (for more details of these test statistics see section 3.2 in [46]):

- The mean-median $MMRT1$ test which is suitable for testing of normality against heavy- and light-tailed asymmetric alternatives.
- The mean-median $MMRT2$ test which is suitable for testing of normality against heavy- and light-tailed asymmetric alternatives as well as bimodal and short-tailed symmetric alternatives.
- The trim-trim $TTRT1$ test with trimming $s = r = 0.05n$ which is suitable for testing of normality against short-tailed symmetric alternatives and which is also more robust than most other tests of normality.
- The trim-trim $TTRT2$ test with trimming $s = r = 0.05n$ which is suitable for testing of normality against heavy-tailed symmetric alternatives and has good robust properties.

Our pilot simulation study showed that mean-median $MMRT1$ and $MMRT2$ tests have comparable power with the most powerful tests (SW and AD) against heavy- and light-tailed asymmetric alternatives like one side Cauchy, Weibull and exponential. Similarly, $TTRT2$ test has comparable power with the most powerful tests (SJ_{dir} and RJB) against heavy-tailed symmetric alternatives like Cauchy and t_5 . Consequently, $MMRT2$ and $TTRT1$ tests have comparable power with the most powerful tests (SW, AD and DT) against short-tailed symmetric alternatives like beta and uniform. Finally, $MMRT2$ test is also suitable for bimodal alternatives. Note that especially $TTRT1$ test has good robustness properties – this test is more robust than most other tests of normality.

Power of analyzed normality tests against symmetric and asymmetric heavy-, light- and short-tailed alternatives was presented in [46]. We found that SJ_{dir}, RJB and $TTRT2$ tests are the best tests against symmetric heavy-tailed alternatives like Cauchy, Laplace and t_3 . For moderately heavy-tailed symmetric alternatives like t_5, t_7 and logistic the RJB and $TTRT2$ test are most powerful. If we suppose the light-tailed asymmetric alternatives like lognormal, exponential, Burr and Weibull, the $SW, AD, MMRT1$ and $MMRT2$ tests perform well. For the short-tailed symmetric alternatives

like beta and uniform, the *SW*, *AD*, *MMRT2* and *TTRT1* tests have reasonably high power. On the other hand, the *JB*, *JBU*, *RJB* and *SJ_{dir}* tests are mostly biased for small sample size $n = 20$ (*SJ_{dir}* is biased even for large sample size $n = 100$).

5. Geometric interpretation of skewness and kurtosis based test statistics

This section is aimed to prepare the reader for the derivation of the exact cumulative distribution functions (cdfs) of skewness and kurtosis based statistics which will be given in the next section.

Let us recall that if x_1, \dots, x_n is a concrete sample of size n then the skewness test statistic based upon this sample allows the representation

$$b_1 = \frac{n}{6} \left(\frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \right)^2 = \frac{n}{6} \left[\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2}{\frac{1}{n} \left(\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right)^{3/2}} \right]^2. \quad (5)$$

Replacing the sample mean in nominator and denominator of this ratio with different location estimations of Lehman-Bickel type, we consider in the following statistic

$$\tilde{b}_1 = \frac{n}{6} \left[\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x}_z)^3}{\frac{1}{n} \left(\sum_{i=1}^n (x_i - \tilde{x}_z)^2 \right)^{3/2}} \right]^2, \quad \tilde{x}_z = \sum_{i=1}^n w_i x_i, \quad w_i \geq 0, \quad \sum_{i=1}^n w_i = 1, \quad \tilde{x}_n = \sum_{i=1}^n a_i x_i. \quad (6)$$

Let \tilde{B}_1 the population based version of the sample based quantity \tilde{b}_1 . Because we are interested in the probability $P(\tilde{B}_1 < t)$, we are interested in a geometric understanding of the subset $\{x \in \mathbb{R}^n : \tilde{b}_1 < t\}$, $t \in \mathbb{R}$ of the n -dimensional sample space \mathbb{R}^n . It turns out that the set $\{x \in \mathbb{R}^n : \tilde{b}_1 < t\}$, $t \in \mathbb{R}$ is a cone having its vertex in the origin. However, for making things as much visible as possible, our discussion of the exact distribution of \tilde{B}_1 is restricted here to the case

$$n = 2, w_1 = \Theta, w_2 = (1 - \Theta), 0 < \Theta < 1, a_1 = 1, a_2 = a > 0 \text{ and } x_1 = x, x_2 = y. \quad (7)$$

The cone $\{x \in \mathbb{R}^2 : \tilde{b}_1 < t\}$, $t \in \mathbb{R}$ can be closer described then in an easy way because the statistic \tilde{b}_1 can be reformulated then as follows

$$\begin{aligned} \tilde{B}_1 &= \frac{1}{3} \left[\frac{\frac{1}{2}((x - \Theta x - (1 - \Theta)y)^3 + (y - \Theta x - (1 - \Theta)y)^3)}{(\frac{1}{2}((x - x - ay)^2 + (y - x - ay)^2))^{3/2}} \right]^2 = \\ &= \frac{\frac{2}{3} [(1 - \Theta)^3(x - y)^3 + \Theta^3(x - y)^3]^2}{[a^2 y^2 + ((1 - a)y - x)^2]^3} \\ &= \frac{\frac{2}{3}(x - y)^6 [(1 - \Theta)^3 - \Theta^3]^2}{[a^2 y^2 + ((1 - a)y - x)^2]^3} = \frac{\frac{2}{3}(1 - z)^6 [(1 - \Theta)^3 - \Theta^3]^2}{[a^2 z^2 + ((1 - a)z - 1)^2]^3} \end{aligned}$$

where $z = \frac{y}{x}$. Analogously, the excess (or kurtosis) test statistic based upon the sample x_1, \dots, x_n allows the representations

$$\sqrt{b_2} = \frac{n}{24} \left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 \right)^2 = \frac{n}{24} \left[\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \right)^2} - 3 \right]^2. \quad (8)$$

Replacing the sample mean with \tilde{x}_z and \tilde{x}_n as before, we consider now the statistic

$$\sqrt{\tilde{b}_2} = \frac{n}{24} \left[\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x}_z)^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x}_n)^2 \right)^2} - 3 \right]^2. \quad (9)$$

Under the restrictions in (7), this statistic reads as

$$\begin{aligned} \sqrt{\tilde{B}_2} &= \frac{1}{12} \left[\frac{\frac{1}{2}((x - \Theta x - (1 - \Theta)y)^4 + (y - \Theta x - (1 - \Theta)y)^4)}{\left(\frac{1}{2}((x - x - ay)^2 + (y - x - ay)^2) \right)^2} - 3 \right]^2 = \\ &= \frac{1}{12} \left[\frac{2((1 - \Theta)^4(x - y)^4 + \Theta^4(y - x)^4)}{(a^2y^2 + ((1 - a)y - x)^2)^2} - 3 \right]^2 = \frac{1}{12} \left[\frac{\gamma(1 - z)^4}{(a^2z^2 + ((1 - a)z - 1)^2)^2} - 3 \right]^2 \end{aligned}$$

where

$$\gamma = 2 \left[(1 - \Theta)^4 + \Theta^4 \right], \quad z = \frac{y}{x}.$$

The sets

$$\tilde{C}_1(t) = \{(x, y) \in \mathbb{R}^2 : \tilde{b}_1 < t\} \quad \text{and} \quad \tilde{C}_2(t) = \{(x, y) \in \mathbb{R}^2 : \sqrt{\tilde{b}_2} < t\}, \quad t > 0 \quad (10)$$

can be proved to be cones in \mathbb{R}^2 having their vertex in the origin.

Let us remark the interesting fact that it is known that a value of Student's distribution can be interpreted as being the value of the standard Gaussian measure taken at a suitably defined cone having vertex in the origin, see in [34]. Moreover, such a value does not depend on the dgf of a spherical measure, and is therefore the same for all spherical measures. We will follow this idea and represent the exact distribution of \tilde{B}_1 in Section 7. For a p -generalization of this observation, see [35].

We determine now the boundary lines of the cone $\tilde{C}_1(t)$. The points of these lines satisfy the equation

$$(1 - z)^2 = \sqrt[3]{\frac{t}{\beta}} \left[(a^2z^2 + ((1 - a)z - 1)^2) \right] \quad (11)$$

where $\beta = \frac{2}{3}[(1 - \theta)^3 - \theta^3]^2$. The points which belong the the inner of this cone satisfy the corresponding inequality

$$(1 - z)^2 < \sqrt[3]{\frac{t}{\beta}} \left[(a^2z^2 + ((1 - a)z - 1)^2) \right]. \quad (12)$$

Solving equation (11) leads to

$$1 - 2z + z^2 = \sqrt[3]{\frac{t}{\beta}} [(a^2 z^2 + ((1-a)^2 z^2 - 2(1-a)z + 1)]$$

and

$$z^2 \left[1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}} \right] + z \left[-2 + 2(1-a) \sqrt[3]{\frac{t}{\beta}} \right] + 1 - \sqrt[3]{\frac{t}{\beta}} = 0.$$

Before division by $\left[1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}} \right]$ we need to take care that $a \neq \frac{\sqrt[3]{\frac{t}{\beta}} + \sqrt{-(\sqrt[3]{\frac{t}{\beta}})^2 + 2\sqrt[3]{\frac{t}{\beta}}}}{2\sqrt[3]{\frac{t}{\beta}}}$ and $a \neq -\frac{-\sqrt[3]{\frac{t}{\beta}} + \sqrt{-(\sqrt[3]{\frac{t}{\beta}})^2 + 2\sqrt[3]{\frac{t}{\beta}}}}{2\sqrt[3]{\frac{t}{\beta}}}$.

The solutions of the equation $p_1(z) = 0$ where

$$p_1(z) = z^2 + z \frac{-2 + 2(1-a) \sqrt[3]{\frac{t}{\beta}}}{1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}}} + \frac{1 - \sqrt[3]{\frac{t}{\beta}}}{1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}}}$$

are

$$z_{1/2} = -p/2 \pm \sqrt{p^2/4 - q}$$

where

$$p = \frac{-2 + 2(1-a) \sqrt[3]{\frac{t}{\beta}}}{1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}}} \text{ and } q = \frac{1 - \sqrt[3]{\frac{t}{\beta}}}{1 - (a^2 + (1-a)^2) \sqrt[3]{\frac{t}{\beta}}}. \quad (13)$$

Thus the boundary lines of the cone $\tilde{C}_1(t)$ are

$$y_1 = (-p/2 + \sqrt{p^2/4 - q})x \text{ and } y_2 = (-p/2 - \sqrt{p^2/4 - q})x, \quad (14)$$

and the points $(x, y)^T$ from the inner part of this cone satisfy the inequality $p_1(y/x) < 0$. If we denote the angle between the boundary lines of the cone $\tilde{C}_1(t)$ by $\alpha_1(t)$ then

$$\alpha_1(t) = |\arctan(-p/2 + \sqrt{p^2/4 - q}) - \arctan(-p/2 - \sqrt{p^2/4 - q})|. \quad (15)$$

The following Figure 1 displays the boundary lines of cone $\tilde{C}_1(t)$ for $\theta = 1/3$, $a = 1$, $t = \frac{12250}{59049}$.

It can be shown in a similar manner that the boundary lines of the cone $\tilde{C}_2(t)$ are

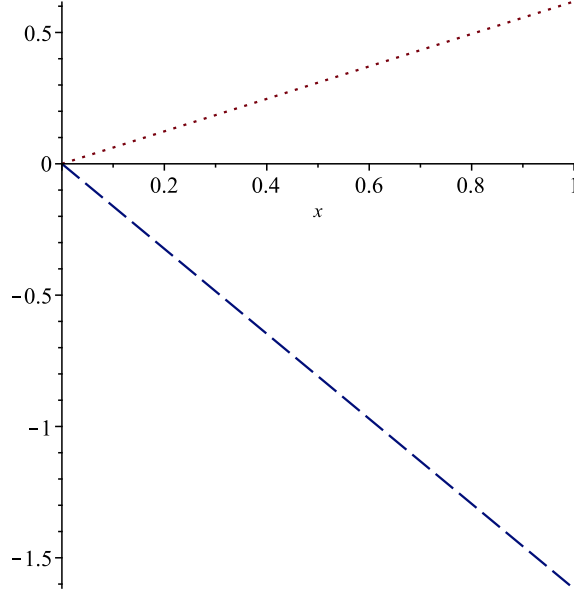


Figure 1. the boundary lines of cone $\tilde{C}_1(t)$ for $\theta = 1/3$, $a = 1$, $t = \frac{12250}{59049}$.

also given by (14) but where now

$$p = \frac{-2 + 2(1-a)\sqrt{(\sqrt{12}\cdot t + 3)/\gamma}}{1 - (a^2 + (1-a)^2)\sqrt{(\sqrt{12}\cdot t + 3)/\gamma}} \text{ and } q = \frac{1 - \sqrt{(\sqrt{12}\cdot t + 3)/\gamma}}{1 - (a^2 + (1-a)^2)\sqrt{(\sqrt{12}\cdot t + 3)/\gamma}}. \quad (16)$$

Thus, the angle between the boundary lines of $\tilde{C}_2(t)$ allows the representation (15), but with p and q from (16). Notice, that for same values for parameters θ, a and t there is no real cone $\tilde{C}_2(t)$.

6. Exact distributions of the skewness and kurtosis based test statistics

Let us assume throughout this section that the sample vector follows a multivariate Gaussian distribution. According to this assumption, the derivation of the exact distributions of the skewness and kurtosis based test statistics presented here is based upon an application of the following geometric measure formula in [33]

$$\Phi(A) = \frac{2^{1-n/2}}{\Gamma(n/2)} \int_0^\infty \mathfrak{F}(A, r) r e^{-r^2/2} dr \quad (17)$$

where the function

$$\mathfrak{F}(A, r) = \frac{AL([\frac{1}{r}A] \cap C)}{AL(C)}, r > 0$$

is called the Euclidean intersection proportion function of the set A . Here, C denotes the Euclidean unit circle and AL means Euclidean arc-length. Formula (17) has been proved for the special case of Gaussian distribution in [32] and was used later on in a series of subsequent papers.

If $n = 2$, $A = \tilde{C}_1(t)$ then

$$P(\tilde{B}_1 < t) = \int_0^\infty \mathfrak{F}(\tilde{C}_1(t), r) r e^{-r^2/2} dr. \quad (18)$$

We recall that, for arbitrary $r > 0$, $\tilde{C}_1(t)$ is a cone. Thus, the ipf of $\tilde{C}_1(t)$ does not depend on the radius r . In consequence,

$$P(\tilde{B}_1 < t) = \mathfrak{F}(\tilde{C}_1(t), 1) = \alpha_1(t)/(2\pi) \text{ and similarly, } P(\sqrt{\tilde{B}_2} < t) = \mathfrak{F}(\tilde{C}_2(t), 1) = \alpha_2(t)/(2\pi) \quad (19)$$

where α_1 and α_2 are chosen according to (15) with (13) and (16), respectively.

7. Representing the distribution of \tilde{B}_1 in terms of a Student distribution

As we have remarked in Section 5, the distribution of \tilde{B}_1 may be represented in terms of Student's distribution. For seeing that, we transform the representation of the cone into a representation which was used in [34]. Because the transformations $\mathfrak{T}_1, \mathfrak{T}_2$ where

$$\mathfrak{T}_1^{-1} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ and } \mathfrak{T}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

do not change the spherical measure of a set, it holds true that $\Phi(A) = \Phi(\mathfrak{T}_1 \mathfrak{T}_2 A)$. We note that

$$C_1^0(t) = \mathfrak{T}_1 \mathfrak{T}_2 \tilde{C}_1(t) = \{(\psi, \nu)^T \in \mathbb{R}^2 : \frac{|\psi|}{|\psi + \nu|} < \delta(t)\} \text{ where } \delta(t) = \sqrt{a(a-1)}(t/\beta)^{1/6}.$$

Thus

$$\Phi(C_1^0(t)) = 2\Phi(\{(\psi, \nu)^T \in \mathbb{R}^2 : \frac{\psi}{\nu} < \frac{\delta(t)}{1 - \delta(t)}\}).$$

Note that $\Phi(\{(\psi, \nu)^T \in \mathbb{R}^2 : \frac{\psi}{\nu} < \frac{\delta(t)}{1 - \delta(t)}\}) = F_1(\frac{\delta(t)}{1 - \delta(t)})$ where F_1 denotes the cdf of the Student cdf with one dgf.

8. Truncated data

In Section 6, we assumed that the sample vector X follows the multivariate standard Gaussian law. Here, we assume that the distribution law of X is a multivariate truncated Gaussian law having the pdf

$$f(x) = \frac{1}{\beta^*} \left(\prod_1^n \varphi(x_i) \right) I_B(x)$$

where B is a truncation set, $\beta^* = \Phi(B)$, and I_B denotes the indicator function of the set B .

In the special case that $B = [a, b]^{\times n}$ for some $a < b$ this means that X has i.i.d. marginal variables following the univariate truncated normal density $h(x) = \frac{\varphi(x)}{Z}$ where $Z = \Phi(b) - \Phi(a)$.

In the general truncation case it follows for the cdf of the test statistic $\tilde{B}_1(t)$ that

$$P(\tilde{B}_1 < t) = \frac{1}{\beta^*} \int_0^\infty \mathfrak{F}(\tilde{C}_1(t) \cap B, r) r e^{-r^2/2} dr, \quad (20)$$

Note that

$$\mathfrak{F}(\tilde{C}_1(t) \cap B, r) \begin{cases} = F(\tilde{C}_1(t), r) & \text{if } 0 < r \leq b \\ < F(\tilde{C}_1(t), r) & \text{if } b < r \leq \sqrt{b^2 + a^2}. \\ = 0 & \text{if } \sqrt{b^2 + a^2} < r \end{cases}$$

Thus the distribution of \tilde{B}_1 is as in Section 6.

In the special case that B is an Euclidean ball of radius ρ having its center at the origin it follows that $\mathfrak{F}(\tilde{C}_1(t) \cap B, r) = F(\tilde{C}_1(t), r)$, $r > 0$. Thus the distribution of \tilde{B}_1 differs from that in Section 6.

9. Exact distributions under non-standard model assumptions

A well known more general assumption w.r.t. the distribution the sample vector follows is that to assume it follows a spherical distribution. Recognize that a yet more general assumption is to let the sample vector follow an $l_{n,p}$ -symmetric distribution.

A general method of deriving exact distributions of statistics under non-standard model assumptions is outlined in [39]. Here, we follow the approach described there. Let the random vector $(X_1, X_2)^T$ follow the $l_{2,p}$ -symmetric density

$$\varphi_{g,p}(x) = C_{g,p} g(|x|_p^p), x \in \mathbb{R}^2, p > 0$$

where $|x|_p = (|x_1|^p + |x_2|^p)^{1/p}$ is a norm if $p \geq 1$ and an antinorm (see [28]) if $0 < p \leq 1$, $g : [0, \infty) \rightarrow [0, \infty)$ is a density generating function (dgf) which satisfies $0 < I_{g,p} < \infty$ for $I_{g,p} = \int_0^\infty r g(r^p) dr$, and $C_{g,p}$ is a suitable normalizing constant making $\varphi_{g,p}$ a density.

If $\Phi_{g,p}$ denotes the probability law corresponding to the density $\varphi_{g,p}$,

$$\Phi_{g,p}(A) = \int_A \varphi_{g,p}(x) dx, A \text{ is a measurable subset of } \mathbb{R}^2,$$

then formula(17) is a special case of the following geometric measure representation

$$\Phi_{g,p}(A) = \frac{1}{I_{g,p}} \int_0^\infty \mathfrak{F}_p(A, r) r g(r^p) dr \quad (21)$$

which was proved for the n -dimensional case in [38]. Note that

$$I_{g,p} = \frac{1}{2\pi(p)C_{g,p}}$$

where the constants $\pi(p)$ are generalizations of the famous circle number π which have been introduced and studied for convex and radially concave p -generalized circles (or $l_{2,p}$ -circles) in [36] and [37], respectively.

The function $x \rightarrow |x|_p$ is the Euclidean norm $|x|_2 = \sqrt{x_1^2 + x_2^2}$ if $p = 2$, the taxi cap norm $|x|_1 = |x_1| + |x_2|$ if $p = 1$, and, formally, the maximum norm $|x| = \max\{|x_1|, |x_2|\}$ if $p = \infty$.

The p -circles $C_p(r) = rC_p, C_p = \{x \in \mathbb{R}^2 : |x|_p = r\}$ are the level sets of the density of the random vector $(X_1, X_2)^T$. The function $r \rightarrow \mathfrak{F}_p(A, r)$ is a generalization of the ipf $r \rightarrow \mathfrak{F}_p(A, r)$ considered in Section 6. It is defined as

$$\mathfrak{F}_p(A, r) = \frac{\mathfrak{U}_p([\frac{1}{r}A] \cap C_p)}{2\pi(p)}$$

where the p -generalized arc-length measure \mathfrak{U}_p has been considered in [36, 37]. In the easiest non-Euclidean case, i.e. if $p = 1$, it holds

$$\mathfrak{F}_1(A, r) = \frac{AL([\frac{1}{r}A] \cap C_p)}{2\pi}.$$

We restrict our consideration therefore here as in [24] to the two cases $p = 2$ and $p = 1$.

Special density generating functions have been considered, e.g. in

[24]. If $g(r) = r^{M-1}e^{-\beta r^\gamma}$ with positive M, β, γ then $C_{g,1} = \frac{\gamma\beta^{(M+1)/\gamma}}{4\Gamma((M+1)/\gamma)}$, and if $g(r^2) = r^{2M-2}e^{-\beta r^{2\gamma}}$ with positive M, β, γ then $C_{g,2} = \frac{\gamma\beta^{M/\gamma}}{\pi\Gamma(M/\gamma)}$. Note, however, that in the given cases of test statistics \tilde{B}_1 and $\sqrt{\tilde{B}_2}$, the distributions do not depend of the dgf. This is due to the following much more general fact. If the ipf of a set A does not depend on the radius variable r , $\mathfrak{F}_p(A, r) = \mathfrak{F}_0$ say, then it follows from (21) that $\Phi_{g,p}(A) = \mathfrak{F}_0$.

It is shown in Section 5, that the ipfs of the sets $\tilde{C}_i(t), i = 1, 2$ generated by the statistics \tilde{B}_1 and $\sqrt{\tilde{B}_2}$, respectively, do not depend on the variable r . Thus the distributions of the statistics \tilde{B}_1 and $\sqrt{\tilde{B}_2}$ do not depend on the dgf, and are therefore the same as given in Section 6. Following a definition in [11], such statistics are called g -robust in [23], and in a more general setting in [39].

10. Exact joint two-dimensional distribution of the skewness and kurtosis based test statistics

Let us assume that the sample vector follows an n -dimensional spherical distribution $\Phi_{g,p}$. According to the previous results,

$$F(t_1, t_2) = P(\tilde{B}_1 < t_1, \sqrt{\tilde{B}_2} < t_2) = \Phi_{g,p}(\tilde{C}_1(t_1) \cap \tilde{C}_2(t_2))$$

Note that the intersection of two cones is a cone. The density of $\begin{pmatrix} \tilde{b}_1 \\ \sqrt{\tilde{b}_2} \end{pmatrix}$ is

$$f(t_1, t_2) = \frac{\partial^2}{\partial t_1 \partial t_2} F(t_1, t_2). \quad (22)$$

On using this density, we are going now to derive the density of $\tilde{B}_1 + \sqrt{\tilde{B}_2}$. To this end, let a map

$$\varphi = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix} : \mathbf{R}^- \times \mathbf{R}^+ \rightarrow \mathbf{R}^2$$

be defined by

$$\begin{pmatrix} \varphi_1(\tilde{B}_1, \sqrt{\tilde{B}_2}) \\ \varphi_2(\tilde{B}_1, \sqrt{\tilde{B}_2}) \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \tilde{B}_1 + \sqrt{\tilde{B}_2} \\ \tilde{B}_1 \end{pmatrix}.$$

Note that φ is 1-1 with inverse function $\psi = \varphi^{-1}$ given by

$$\begin{pmatrix} \tilde{B}_1 \\ \sqrt{\tilde{B}_2} \end{pmatrix} = \begin{pmatrix} Y_2 \\ Y_1 - Y_2 \end{pmatrix} = \begin{pmatrix} \psi_1(Y_1, Y_2) \\ \psi_2(Y_1, Y_2) \end{pmatrix}.$$

The absolute value of the Jacobian of this transformation is

$$I = \left| \det \begin{vmatrix} \frac{\partial \psi_1}{\partial y_1} & \frac{\partial \psi_1}{\partial y_2} \\ \frac{\partial \psi_2}{\partial y_1} & \frac{\partial \psi_2}{\partial y_2} \end{vmatrix} \right| = \left| \det \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} \right| = 1.$$

As a consequence, the density of $\begin{pmatrix} \tilde{B}_1 + \sqrt{\tilde{B}_2} \\ \tilde{B}_1 \end{pmatrix}$ is

$$f \begin{pmatrix} \tilde{B}_1 + \sqrt{\tilde{B}_2} \\ \tilde{B}_1 \end{pmatrix} (y_1, y_2) = f(\psi_1(y_1, y_2), \psi_2(y_1, y_2)) I = f(y_2, y_1 - y_2)$$

where f is chosen according to (22). Integrating this joint density w.r.t. the variable y_2 results in the desired density

$$f_{\tilde{B}_1 + \sqrt{\tilde{B}_2}}(y_1) = \int_{-\infty}^{\infty} f(y_2, y_1 - y_2) dy_2. \quad (23)$$

11. Tests based upon higher order moments

In this section we study testing based upon higher order moments. Let

$$\tilde{b}_3 = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \tilde{x}_z)^6}{\frac{1}{n} (\sum_{i=1}^n (x_i - \tilde{x}_N)^2)^3}$$

be the statistic considered in [26].

If $n = 2$, $\tilde{x}_z = \Theta x + (1 - \Theta)y$, $\tilde{x}_N = x + ay$ then

$$\tilde{B}_3 = 4 \frac{[(1 - \Theta)^6 - \Theta^6] (1 - z)^6}{[a^2 z^2 + ((1 - a)z - 1)^2]^3}, \quad z = \frac{y}{x}.$$

For every $t > 0$, the set

$$\tilde{C}_3(t) = \{(x, y) \in \mathbf{R}^2 : \tilde{b}_3 < t\}$$

is a cone with vertex in the origin and boundary

$$\partial \tilde{C}_3(t) = \{(x, y) \in \mathbf{R}^2 : [(1 - \Theta)^6 - \Theta^6]^{1/3} (1 - Z)^2 = a^2 z^2 + ((1 - a)z - 1)^2\}.$$

Because the $\Phi_{g,p}$ -value of such cones does not depend on the dgf g , the statistic \tilde{b}_3 is g -robust just like \tilde{b}_1 , \tilde{b}_2 and $\tilde{b}_1 + \sqrt{\tilde{b}_2}$ are g -robust. Note that the boundary lines of $\tilde{C}_3(t)$ satisfy the equations (14) but where now p and q are

$$p = \frac{-2C(\theta) + 2(1 - a)t}{C(\theta) - a^2 t - (1 - a)^2 t} \text{ and } q = \frac{C(\theta) - t}{C(\theta) - a^2 t - (1 - a)^2 t} \text{ with } C(\theta) = ((1 - \theta)^6 - \theta^6)^{1/3}. \quad (24)$$

We conclude this section with the following remark on a non- g -robust statistic. Let

$$\tilde{b}_k = \frac{\frac{1}{2} [(x - \frac{x+y}{2})^k + (y - \frac{x+y}{2})^k]}{[\frac{1}{2} \langle (x - \frac{x+y}{2})^2 + (y - \frac{x+y}{2})^2 \rangle]^{k/2}} = \begin{cases} \frac{|x-y|}{2^{k/2-1}}, & k \text{ is even} \\ 0, & k \text{ is odd} \end{cases}.$$

Then, for even k , the critical test region is

$$H_k(t) = \{(x, y) \in \mathbf{R}^2 : \left| \frac{x - y}{2} \right|^{k/2} > t\}$$

Denoting the acceptance region by $H_k^c(t)$, it turns out that $\Phi_{g,p}(H_k^c(t))$ depends on g and p , thus b_k is not g -robust.

12. Algebraic manipulations on main quadric

Throughout, we denote the set of real numbers by \mathbb{R} . As we can see from simulations, the power depends on choice of location functionals in both numerator and denominator of RT class tests. We conducted large simulation study for $n = 2$ and location functional in nominator given by weighted mean, i.e. $m(X_1, X_2) = \theta X_1 + (1 - \theta)X_2$, $\theta \in$

(0, 1). According to [31] weighted mean is an interesting class of location functionals. However, for $n = 2$ in order not to get constant skewness, we chose estimator $m * (X_1, X_2) = X_1 + aX_2$, $a \in (0, 1)$. Notice that $m*$ is biased for $\mu \neq 0$, $X_1, X_2 \sim N(\mu, \sigma^2)$. However it is an unbiased estimator of μ when $X_1, X_2 \sim N(0, \sigma^2)$. We can relate it here to concept of superefficiency (see [8], page 515 and references therein). This directly relates to Hodges' estimator (see e.g. [54]). Of course, measure of point $\mu = 0$ is zero, interestingly, the rejection cone is also degenerated (in real geometry). Thus, we have one degenerated and one non-degenerated cone, as it is proven in the next lines.

Large simulation study shows that most powerful test was given by $\theta_1 = 1/3$ and $a \in \{1/2, 1\}$. After some algebra for \tilde{B}_1 we receive the initial equation of the form

$$28(x^3 - 3x^2y + 3xy^2 - y^3) = 27a(y^2 - 2xy + 2x^2)\sqrt{2y^2 - 4xy + 4x^2} \quad (25)$$

We are interested in the set of real solutions to this equation. Below, we show that for a in the interior of the interval $I = \left[0, \frac{14\sqrt{2}}{27}\right]$ this solution set is two lines passing through the origin and it gives a degenerate answer, i.e. multiple line, on the boundary of I . Moreover, there are no real solutions for a taken outside I . For simplicity we put $f = y^2 - 2xy + 2x^2$. Taking the square of (25) we get

$$(x - y)^6 = 2 \left(\frac{27a}{28}\right)^2 f^3 \quad (26)$$

which factors through

$$((x - y)^2 - \alpha f)((x - y)^4 + \alpha f(x - y)^2 + \alpha^2 f^2) = 0 \quad (27)$$

where

$$\alpha = \sqrt[3]{2 \left(\frac{27a}{28}\right)^2}$$

Lemma 1. The equation $(x - y)^4 + \alpha f(x - y)^2 + \alpha^2 f^2 = 0$ has no nontrivial solution in \mathbb{R} for any nonzero value of α .

Proof. Note that if we substitute $z = (x - y)^2$ and $t = x^2$ in the equation we have

$$z^2 + \alpha(z + t)z + \alpha^2(z + t)^2 = 0$$

Obviously $x = y = 0$, and consequently $z = t = 0$, satisfies the equation. Assume, without loss of generality, that $z \neq 0$, and substitute $r = \frac{z+t}{z}$. The aim is to find real solutions to the quadratic equation

$$\alpha^2 r^2 + \alpha r + 1 = 0$$

The discriminant of this equation is $\Delta = -3\alpha^2$, which is never positive, and is zero if and only if $\alpha = 0$. Clearly if $\alpha = 0$, the solution set is the line $x = y$. \square

Lemma 2. The real solutions to the quadratic form $\{(x - y)^2 - \alpha(y^2 - 2xy + 2x^2) = 0\}$ for different values of α is

- (1) two lines $y = (1 + \sqrt{\frac{\alpha}{1-\alpha}})x$ and $y = (1 - \sqrt{\frac{\alpha}{1-\alpha}})x$, when $\alpha \in (0, 1)$,
- (2) the line $x = y$, when $\alpha = 0$,
- (3) the line $x = 0$, when $\alpha = 1$,
- (4) and it has no nontrivial real solution when $\alpha \notin [0, 1]$.

Proof. For simplicity, consider the substitution $z = x - y$. The equation now becomes

$$z^2 - \alpha(z^2 + x^2) = 0$$

which solves to

$$z = \pm \sqrt{\frac{\alpha}{1-\alpha}}x$$

The proof follows from this equality. □

Corollary 1. The real solutions to (25) depending on the value of the real parameter a are

- (1) two lines $y = (1 + \sqrt{\frac{\sqrt[3]{2(\frac{27a}{28})^2}}{1 - \sqrt[3]{2(\frac{27a}{28})^2}}})x$ and $y = (1 - \sqrt{\frac{\sqrt[3]{2(\frac{27a}{28})^2}}{1 - \sqrt[3]{2(\frac{27a}{28})^2}}})x$, when $a \in I$,
- (2) the line $x = y$, when $a = 0$,
- (3) the line $x = 0$, when $a = \frac{14\sqrt{2}}{27}$,
- (4) and it has no nontrivial real solution when $a \notin I$.

Note that when we solve a for α for substitution we do not consider the negative values of a . This is because we took the square of the equation (25) in the beginning.

13. Power comparisons – models for outliers

In this section we compare the power of the classical tests for normality and our omnibus test for normality. We have checked by conducting the thorough simulation study that all considered empirical tests hold the size for $\alpha = 0.05$ and for sample sizes $n = 20$ and $n = 100$. The alternatives we have considered are models for outliers. Here we consider two outlier models, as used in [2], i.e. location-outlier model and scale-outlier model.

- For the p -location-outlier model we consider X_1, \dots, X_{n-p} to be iid from $N(0, 1)$ and X_{n-p+1}, \dots, X_n to be iid from $N(\lambda, 1)$.
- For the p -scale-outlier model we consider X_1, \dots, X_{n-p} to be iid from $N(0, 1)$ and X_{n-p+1}, \dots, X_n to be iid from $N(0, \tau^2)$.

In this paper we consider only linear location estimators. In particular, median provides best protection against the presence of outlier in term of bias ([2]), but it comes at the cost of a higher MSE than some other estimators. The trimmed mean, linearly weighted mean and modified MLE turn out to be quite robust and efficient in general. In this paper we consider only median $T_{(1)}$, trimmed mean $T_{(2)}$ and pseudomedian $T_{(3)}$.

The results of our Monte Carlo simulations are summarized in Table 2. As it is shown in mentioned tables, the most commonly used tests of normality, such as the SW and

JB tests, are too strict in rejecting normality in the case of a small number of outliers, even when the sample size is large enough. To illustrate this general framework, notice that one of the most used SW test rejects the hypothesis of normality in 81% cases if we suppose the p -location outlier model for $p = 1$, $\lambda = 5$ and sample size $n = 100$. Similarly, the JB and RJB tests are too strict in rejecting normality. If we suppose the same p -location outlier model and sample size, the JB and RJB tests reject the hypothesis of normality in 88% and 86% cases, respectively. In contrast, the $TTRT1$ and MC_{LR} tests are more robust – $TTRT1$ test rejects normality in 29%, and MC_{LR} test even in 5%, which is consistent with chosen significance level.

	$n = 20$				$n = 100$			
	$p = 1$		$p = 5$		$p = 1$		$p = 5$	
	$\lambda = 3$	$\lambda = 5$	$\lambda = 3$	$\lambda = 5$	$\lambda = 3$	$\lambda = 5$	$\lambda = 3$	$\lambda = 5$
AD	0.173	0.630	0.273	0.922	0.086	0.374	0.546	1.000
DT	0.305	0.866	0.037	0.020	0.238	0.861	0.725	1.000
JB	0.314	0.872	0.056	0.115	0.269	0.884	0.754	1.000
JBU	0.319	0.878	0.033	0.016	0.273	0.886	0.751	1.000
LT	0.118	0.407	0.241	0.833	0.065	0.175	0.340	0.984
RJB	0.298	0.850	0.056	0.142	0.253	0.860	0.756	1.000
SJ_{dir}	0.234	0.728	0.051	0.190	0.160	0.630	0.550	0.998
SW	0.226	0.766	0.242	0.888	0.193	0.807	0.691	1.000
MC_{LR}	0.051	0.055	0.121	0.358	0.050	0.046	0.075	0.094
$MMRT1$	0.273	0.832	0.159	0.657	0.249	0.868	0.697	1.000
$MMRT2$	0.266	0.829	0.160	0.686	0.243	0.863	0.689	1.000
$TTRT1$	0.038	0.007	0.263	0.845	0.048	0.288	0.191	0.960
$TTRT2$	0.275	0.822	0.090	0.467	0.215	0.808	0.639	1.000

Table 2. Power of the selected tests for normality against outliers models for $n = 20$ and $n = 100$

Now, if we suppose a largest number of outliers in small sample size – e.g. the p -location outlier model for $p = 5$, $\lambda = 5$ and sample size $n = 20$ – we need the test with high power, because the hypothesis of normality is not sustainable. Based on our Monte Carlo simulations we can recommend the AD , SW and again $TTRT1$ tests – these test reject the hypothesis of normality in 92%, 89% and 85% cases, respectively. In contrast, the JB , JBU and RJB tests have small power against the mentioned p -location outlier model – the classical JB test reject the hypothesis of normality only in approximately 12% cases, the RJB test in 14% cases and finally the JBU reject the hypothesis of normality only in approximately 2% cases.

Based on these simulations we can conclude that especially the $TTRT1$ test is more robust against many different types of small numbers of outliers compared to sample size, while the SW , JB and RJB tests generally are non-robust. In other words, the $TTRT1$ test is more robust in case of small number of outliers than the classical normality tests such as SW and JB , and simultaneously have comparable power with the SW test in case of presence of large number of outliers compared to sample size.

14. Discussion and conclusions

This paper introduces the general RT class of robust tests for normality and discuss their properties, especially geometric ones. The further theoretical considerations of class RT will be of interest. In the simulation study we have focused on the power

study of selected tests from RT class. We have compared these tests with the selected tests for the normality on the large scale of alternatives. Some of the most important results (observations) are:

- The RT class includes a large number of tests with different properties, especially power and robustness.
- Some RT class tests (e.g. the classical JB test, RJB test, SJ test among others) have already been studied in the literature – these tests are special cases of RT class tests with/without trimming for the different settings of parameters of the general RT class test statistic.
- Based on our pilot simulation study we recommend four tests with good properties for the general use – $MMRT1$, $MMRT2$, $TTRT1$ and $TTRT2$ tests.
- $TTRT1$ test is more robust against many different types of small numbers of outliers.
- Many results of this paper are in the coherence or are extending the results of previous studies, e.g. [49].

Acknowledgements

Milan Stehlík acknowledges the support of Fondecyt Proyecto Regular No. 1151441. Luboš Střelec acknowledges the support of the Grant Agency of the Czech Republic (project No. 16-07089S).

References

- [1] Alvarez-Esteban P.C., del Barrio E., Cuesta-Albertos J.A., and Matrán C. (2010). Assessing when a sample is mostly normal. *Computational Statistics & Data Analysis*, Vol. 54, No. 12, pp. 2914–2925.
- [2] Balakrishnan N. (2007) Permanents, Order Statistics, Outliers, and Robustness. *Revista Matemática Complutense*, Vol. 20, No. 1, pp. 7–107.
- [3] Bickel P.J., and Lehmann E.L. (1975). Descriptive Statistics for Nonparametric Models II. Location. *Annals of Statistics*, Vol. 3, No. 5, pp. 1045–1069.
- [4] Bera A., and Jarque C. (1981). Efficient tests for normality, heteroskedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters* Vol. 7, No. 4, pp. 313–318
- [5] Bowman K.O., and Shenton L.R. (1975). Omnibus contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika*, Vol. 62, No. 2, pp. 243–250.
- [6] Brys G., Hubert M., and Struyf, A. (2008). Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, Vol. 23, No. 3, pp. 429–442.
- [7] Carr N.L. (1960). Kinetics of Catalytic Isomerization of n-Pentane. *Industrial & Engineering Chemistry*, Vol. 52, No. 5, pp. 391–396.
- [8] Cassela G., and Berger R.L. (2002). *Statistical Inference*, 2nd Edition, Duxbury Advanced Series, Thomson Learning.
- [9] Coin D. (2008). Testing normality in the presence of outliers. *Statistical Methods and Applications*, Vol. 17, No. 1, pp. 3–12.
- [10] D’Agostino R., and Pearson E. (1973). Tests for departures from normality. Empirical results for the distribution of $\sqrt{b_1}$ and b_2 . *Biometrika*, Vol. 60, No. 3, pp. 613–622.
- [11] Fang K.T., Kotz S., and Ng K.W. (1990). *Symmetric multivariate and related distributions*. Chapman and Hall, London.

- [12] Gastwirth J.L. (1982). Statistical properties of a measure of tax assessment uniformity. *Journal of Statistical Planning and Inference*, Vol. 6, No. 1, pp. 1–12.
- [13] Geary R.C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, Vol. 27, No. 3–4, pp. 310–332.
- [14] Gel Y.R., and Gastwirth, J.L. (2008). A robust modification of the Jarque-Bera test of normality. *Economics Letters*, Vol. 99, No. 1, pp. 30–32.
- [15] Gel Y.R., Miao W., and Gastwirth J.L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Computational Statistics & Data Analysis*, Vol. 51, No. 5, pp. 2734–2746.
- [16] Hamzah N.A., and Sabri S.R.A. (2007). Robust Jacque-Bera Test of Normality. *Proceedings of The 9th Islamic Countries Conference on Statistical Sciences 2007*, ICCS-IX 12-14 Dec 2007.
- [17] Hettmansperger T.P., and McKean J.W. (1998). Robust Nonparametric Statistical Methods *Kendalls Library of Statistics*, No. 5, 37 p.
- [18] Hewitt Ch.C., and Lefkowitz B. (1979). Methods for fitting distributions to insurance loss data. *Proceedings of the Casualty Actuarial Society*, Vol. 66, pp. 139–160.
- [19] Huber P.J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, Vol. 35, No. 1, pp. 73–101.
- [20] Huber P.J. (1965). A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, Vol. 36, No. 6, pp. 1753–1758.
- [21] Erceg-Hurn D.M., and Mirosevich V.M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, Vol. 63, No. 7, pp. 591–601.
- [22] Gilmour S.G., and Trinca L.A. (2012). Optimum design of experiments for statistical inference. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 61, No. 3, pp. 345–401.
- [23] Ittrich, C., Krause, D. and W.-D.Richter (2000). Probabilities and large quantiles of noncentral generalized chi-square distributions. *Statistics* 34, 53-101.
- [24] Kalke, S., Richter, W.-D. and F. Thauer (2013). Linear combinations, products and ratios of simplicial or spherical variates. *Communications in Statistics: Theory and Methods*, 42:3, 505-527.
- [25] Keilson J., and Steutel F.W. (1974). Mixtures of Distributions, Moment Inequalities and Measures of Exponentiality and Normality. *The Annals of Probability*, Vol. 2, No. 1., pp. 112–130.
- [26] Lin, C.-C., Mudholkar, G.S.(1980). A simple test for normality against asymmetric alternatives. *Biometrika* 67, 455-461.
- [27] Miller J., and Miller J. (2005). *Statistics and Chemometrics for Analytical Chemistry*, 5th Edition, Prentice and Hall.
- [28] Moszyńska, M, and Richter, W.-D.(2012). Reverse triangle inequality. Antinorms and semi-antinorms. *Studia Scientiarum Mathematicarum Hungarica* 49(1), 120-138.
- [29] Müller W.G., and Stehlík M. (2012). Discussion of the paper by Steven G. Gilmour and Luzia A. Trinca. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 61, No. 3, pp. 345–401.
- [30] Nürnberg G., and Rasch D. (1983). The influence of different shapes of distributions with the same first four moments on robustness. *Robustness of statistical methods and nonparametric statistics*, Ed. by Dieter Rasch and Moti Lal Tiku, pp. 83–84
- [31] Obenchein R.L. (1971). Multivariate Procedures Invariant under Linear transformations, *Ann.Math. Stats.* 42 (5), 1569-1578.
- [32] Richter, W.-D. (1985). Laplace-Gauß integrals, Gaussian measure asymptotic behaviour and probabilities of moderate deviations. *Z. Anal. Anw.* Vol. 4, No. 3, pp. 257-267.
- [33] Richter, W.-D. (1991). Eine geometrische Methode in der Stochastik. *Rostock. Math. Kolloqu.*, Vol. 44, pp. 63-72.
- [34] Richter, W.-D. (1995). A geometric approach to the Gaussian law. In: *Symposia Gaussiana*, Conf. B, Eds: Mammitzsch/Schneeweiß. Walter de GruyterCo, Bln., pp. 25-45.

- [35] Richter, W.-D. (2007). Generalized spherical and simplicial coordinates. *J. Math. Anal. Appl.*, Vol. 336, pp. 1187-1202.
- [36] Richter, W.-D. (2008a). On $l_{2,p}$ -circle numbers. *Lithuanian Math. J.*, Vol. 48, No. 2, pp. 228-234.
- [37] Richter, W.-D. (2008b). On the π -function for nonconvex $l_{2,p}$ -circles. *Lithuanian Math. J.*, Vol 48, No. 3, pp. 332-338.
- [38] Richter, W.-D. (2009). Continuous $l_{n,p}$ -symmetric distributions. *Lithuanian Math. J.*, Vol. 49, No. 1, pp. 93-108.
- [39] Richter, W.-D. (2012). Exact distributions under non-standard model assumptions. *AIP Conf. Proc.*, Vol. 1479, pp. 442.
- [40] Richter, W.-D. (2014). Geometric disintegration and star-shaped distributions. *Journal Statist. Distrib. Appl.*, Vol. 20, No. 1, 24 pages.
- [41] Rodríguez-Díaz J.M., et al. (2012). Filling and D-optimal designs for the correlated generalized exponential models. *Chemometrics and Intelligent Laboratory Systems*, Vol. 114, No. 5, pp. 10–18.
- [42] Shapiro S.S., and Wilk M.B. (1965). An analysis of variance test for normality. *Biometrika*, Vol. 52, No. 3–4, pp. 591–611.
- [43] Stehlík M., and Střelec L. (2009). On normality assumptions for claims in insurance. *Acta Universitatis Agriculturae et Silviculturae Mendeleianae Brunensis*, Vol. 58, No. 3, pp. 141–146.
- [44] Stehlík M., and Střelec L. (2012). Discussion of the paper by Steven G. Gilmour and Luzia A. Trinca. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol. 61, No. 3, pp. 345–401.
- [45] Stehlík M., Fabián Z., and Střelec L. (2012). Small sample robust testing for Normality against Pareto tails. *Communications in Statistics – Simulation and Computation*, Vol. 41, No. 7, pp. 1167–1194.
- [46] Stehlík M., Střelec L., and Thulin M. (2014). On robust testing for normality in chemometrics. *Chemometrics and Intelligent Laboratory Systems*, Vol. 130, No. 1, pp. 98–108.
- [47] Stehlik, M., Economou, P., Kiseljak, J., and W.-D. Richter (2014). Kullback-Leibler life time testing. *Applied Mathematics and Computation*, Vol. 240, pp. 122–139.
- [48] Stigler S.M. (1973). The Asymptotic Distribution of the Trimmed Mean. *The Annals of Statistics*, Vol. 1, No. 3, pp. 472–477.
- [49] Thadewald T., and Bunning, H. (2007). Jarque-Bera test and its competitors for testing normality – a power comparison. *Journal of Applied Statistics*, Vol. 34, No. 1, pp. 87–105.
- [50] Thode H.C. (2002). *Testing for normality*. New York: Marcel Dekker, 2002. 368 p.
- [51] Urzua C.M. (1996). On the correct use of omnibus tests for normality. *Economics Letters*, Vol. 53, No. 3, pp. 247–251.
- [52] Uthoff V.A. (1973). The most powerful scale and location invariant test of the normal versus the double exponential. *The Annals of Statistics*, Vol. 1, No. 1, pp. 170–174.
- [53] New measurement of the rate coefficient for the reaction of OH with methane. Vaghjiani Ghanshyam.L., and Ravishankara, A.R. (1991). *Nature*, Vol. 350, No. 6317, pp. 406-409.
- [54] Vaart, A. W. van der (1998). *Asymptotic statistics*. Cambridge University Press.