

Power Comparisons of Normality Tests Based on L-moments and Classical Tests

Ivana Malá*, Václav Sládek, Diana Bílková

Faculty of Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic

Received May 21, 2021; Revised November 11, 2021; Accepted November 21, 2021

Cite This Paper in the following Citation Styles

(a): [1] Ivana Malá, Václav Sládek, Diana Bílková, "Power Comparisons of Normality Tests Based on L-moments and Classical Tests," *Mathematics and Statistics*, Vol.9, No.6, pp. 994-1003, 2021. DOI: 10.13189/ms.2021.090615

(b): Ivana Malá, Václav Sládek, Diana Bílková, (2021). *Power Comparisons of Normality Tests Based on L-moments and Classical Tests. Mathematics and Statistics*, 9(6), 994-1003. DOI: 10.13189/ms.2021.090615

Copyright ©2021 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Normality tests are used in the statistical analysis to determine whether a normal distribution is acceptable as a model for the data analysed. A wide range of available tests employs different properties of normal distribution to compare empirical and theoretical distributions. In the present paper, we perform the Monte Carlo simulation to analyse test power. We compare commonly known and applied tests (standard and robust versions of the Jarque-Bera test, Lilliefors test, chi-square goodness-of-fit test, Shapiro-Francia test, Cramer-von Mises goodness-of-fit test, Shapiro-Wilk test, D'Agostino test, and Anderson-Darling test) to the test based on robust L-moments. In the text, in Jarque-Bera type test the moment characteristics of skewness and kurtosis are replaced with their robust versions - L-skewness and L-kurtosis. The distributions with heavy tails (lognormal, Weibull, loglogistic and Student) are used to draw random samples to show the performance of tests when applied on data with outliers. Small sample properties (from 10 observations) are analysed up to large samples of 200 observations. Our results concerning the properties of the classical tests are in line with the conclusion of other recent articles. We concentrate on properties of the test based on L-moments. This normality test is comparable to well-performing and reliable tests; however, it is outperformed by the most powerful Shapiro-Wilks and Shapiro-Francia tests. It works well for Student (symmetric) distribution, comparably with the most frequently used Jarque-Bera tests. As expected, the test is robust to the presence of outliers in comparison with sensitive tests based on product moments or correlations. The test turns out to be very universally reliable.

Keywords L-moments, Tests for Normality, Simulation, Power of Test, Robustness

1 Introduction

In parametric statistics, the assumption about the probability distribution of data is crucial. The optimal or at least acceptable choice of a probability distribution allows a wide range of powerful parametric procedures to be applied, inappropriate choice potentially resulting in misleading or completely erroneous and even dangerous outcomes. There are numerous statistical methods based on the normal distribution. The issue can be approached theoretically, using the theoretical basis of the data, or experimentally, exploring available data. In addition to exploratory graphs describing the empirical distribution, statistical tests have been designed to determine whether the random sample is drawn from the selected distribution (the normal one in this paper), or if this assumption is at least reasonable. These tests make it possible to assess the suitability of the selected distributions in the form of probabilistic models for the analysed data, not choosing the best of multiple options. Some of them are intended for the normal distribution only (based on its specific properties), others are more generally applicable goodness-of-fit tests.

If the goal is to select a proper test, we can employ a commonly used application-specific one along with practice-proven knowledge and software. It is necessary to consider individually the power of each test, depending on the real shape of the distribution or the random sample size. However, it is very difficult to assess the power of tests analytically, the recommendations thus being often supported by simulation studies. In the present paper, we use the Monte Carlo simulation to compare the power of the most commonly used goodness-of-fit test for a normal distribution. A normality test based on robust L-moments as a robust version of the Jarque-Bera test is compared to more frequently applied tests. To investigate the power of a test, in our simulation, we draw data from differ-

ent – symmetric short- and long-tailed as well as asymmetric – distributions and evaluate the relative frequency of rejection of the hypothesis of normality.

Many statistical tests have been developed to verify whether a sample is drawn from a normal distribution. In our simulation, we compare the performance of ten commonly used tests: in addition to the test relying on L-moments (part 3.1) and standard [JB] and robust versions [RJB] of the Jarque-Bera test, we use the Lilliefors test [LF], chi-square goodness-of-fit test, Shapiro-Francia test [SF], Cramer-von Mises goodness-of-fit test [CM], Shapiro-Wilk test [SW], D’Agostino test [DA] and Anderson-Darling [AD] test. We give original references to these tests in part 3.2 together with some remarks of properties (type one errors and power of tests).

The L-moment-based test, both Jarque-Bera tests, the Lilliefors, Anderson-Darling, D’Agostino, Shapiro-Francia and Shapiro-Wilk tests are intended only for testing normality, while the chi-square goodness-of-fit, Anderson-Darling and Cramer-von Mises tests are generally applicable as goodness-of-fit tests for testing a fully specified distribution (with given parameters) or a family of distributions. We compare the test based on L-moments (part 3.1), tests comparing sample skewness and kurtosis characteristics (Jarque-Bera, robust Jarque-Bera and D’Agostino), empirical and theoretical cumulative distribution functions (Lilliefors, Anderson-Darling) and Cramer based on the correlation coefficient (Shapiro-Francia and Shapiro Wilks).

The choice of the most powerful test in various situations is of interest not only to statisticians, but also to users of statistical methods who only use the statistical tests in their research work. There is a huge spectrum of recommendations available in the statistical literature. The most common method how to approach the problem is a Monte Carlo simulation ([16], [26], [24], [4], [27], and [23]). The authors obtain results for a special choice of distributions and subsequently generalize the results to more general properties of distributions (symmetric, asymmetric, heavy-tailed distributions). Special attention should be paid to the sample size, as the qualities of tests depends strongly on it. Usually the idea behind a test causes the highest power for small or for large samples. Moreover, for small samples the asymptotic approximations are not applicable. According to Yap and Sin [26], descriptive and graphical information supplemented with formal normality tests can aid in making conclusion about the distribution of a variable. In their study, for symmetric short-tailed distributions, D’Agostino and Shapiro–Wilk tests showed the best results; the power of Jarque–Bera and D’Agostino tests being comparable to the Shapiro–Wilk for symmetric long-tailed distributions. For asymmetric distributions, the authors based on their research recommend also the Shapiro–Wilk test. Shapiro-Wilk test was also found well performing for small samples (10-150 observations) for various deviations from normality (Student,chi-squared or beta distribution). For symmetric distribution with high sample kurtosis (symmetric long-tailed), the researcher can use the JB, SW or AD test. Razali and Wah in [20] show that Shapiro-Wilk test is the most powerful normality test, followed by Anderson-Darling test, Lilliefors test and Kolmogorov-Smirnov test. However, the power

of all four tests is still low for small sample size.

In the paper, we investigate the above-mentioned normality test properties focusing on the L-moment-based test. An overview of L-moments and the underlying rationale are provided in section 2. In part 3, all tests and reasons for their construction are introduced. The results of the simulation are given in tables and figures in section 4. In the last part of the article, we make some suggestions emerging from the simulation. All computations are performed using the R [19] software.

2 Methods

Let X_1, X_2, \dots, X_n be a random sample of the size n from a continuous probability distribution of a random variable X with distribution function F and quantile function $Q(P) = F^{-1}(P)$. The hypothesis of normality ($X \sim N(\mu; \sigma^2)$) assumes that $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$, where Φ is the cumulative distribution function of the standard normal distribution. Suppose $X_{i:j}$ denotes the i -th order statistic in a sample of the size j . Let us denote by s^2 a sample variance (based on a random sample x_1, x_2, \dots, x_n with the sample size n)

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n [x_i - \bar{x}]^2 \tag{1}$$

and sample coefficients of skewness and kurtosis a_3 and a_4 , respectively, as

$$a_3 = \frac{\frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^3}{s^3}, a_4 = \frac{\frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^4}{s^4}, \tag{2}$$

where \bar{x} is the mean of the sample. The (theoretical) coefficient of skewness α_3 of all normal distributions is equal to 0, the coefficient of kurtosis α_4 equalling 3. The above properties are implemented in the paper, especially in the Jarque-Bera test.

2.1 L-moments

In [12], the formula for the evaluation of L -moments (for $k = 1, 2, \dots$) based on the quantile function $Q(\cdot)$ is given in the form

$$\lambda_k = \int_0^1 Q(u) P_{k-1}^*(u) du, \tag{3}$$

where P_r^* is the r -th shifted Legendre polynomial written as

$$P_r^*(u) = \sum_{l=0}^r p_{r,l}^* u^l, \tag{4}$$

polynomial coefficients $p_{r,l}^*$ being

$$p_{r,l}^* = (-1)^{r-l} \binom{r}{l} \binom{r+l}{l}, l \leq r. \tag{5}$$

If the expected value of the probability distribution is finite, then all L -moments are finite, [12]. It follows that for all distributions with a finite expected value, the second, third and fourth L -moments are also finite, allowing for application of

the L-skewness and L-kurtosis. Therefore, we can use L-moments and methods based on them even for the distributions without a finite variance or higher moments. In this paper, we employ the Student distribution with the degrees of freedom $\nu = 1 - 20$. The expected value is finite for $\nu > 1$, since for the Cauchy distribution for $\nu = 1$, the expected value is undefined. The L-moments of the Student distribution are finite for all degrees of freedom greater than 1, however, the variance is defined for $\nu > 2$, and skewness and kurtosis coefficients for $\nu > 3$ and $\nu > 4$, respectively.

For $k = 1, 2, 3, 4$ we obtain ([12], [15]) first four L-moments $\lambda_1, \lambda_2, \lambda_3,$ and λ_4 . As robust counterparts of the normalized product moment characteristics of variability, shape and concentration (variation coefficient, coefficients of skewness (variation coefficient, coefficients of skewness α_3 and kurtosis α_4) we define the following statistics based on L-moments – L-coefficient of variation τ , L-skewness τ_3 and L-kurtosis τ_4 given by the formula

$$\tau = \lambda_2/\lambda_1, \quad \tau_3 = \lambda_3/\lambda_2, \quad \tau_4 = \lambda_4/\lambda_2. \tag{6}$$

As previously described, these normalized values are correctly defined for all distributions with a finite expected value and according to [12] are bounded and we have

$$-1 \leq \tau_3 \leq 1, \quad 0.25(3\tau_3^2 - 1) \leq \tau_4 < 1. \tag{7}$$

The bounded area of all possible points (τ_3, τ_4) is shown in Figure 2.

The boundedness of L-moment ratios is an advantage. It is easier to interpret a statistic that is limited within the $[-1, 1]$ interval than the conventional skewness coefficient which may take arbitrarily large values. We obtain $\lambda_3 = \tau_3 = 0$ for all symmetric distributions with the finite expected value, this property corresponding to the standard coefficient of skewness for symmetric distributions. For all normal distributions, L-skewness equals 0 and L-kurtosis 0.1226; we can compare this property with the classical skewness and kurtosis coefficients of 0 and 3, respectively.

The sample counterpart (based on a random sample of the size n) to the L-moments are sample L-moments $l_k, k = 1, 2, \dots$ given in [12] as

$$l_k = \int_0^1 \hat{Q}(u) P_{k-1}^*(u) du, \tag{8}$$

where \hat{Q} is a sample quantile function. Then the sample L-skewness t_3 and sample L-kurtosis t_4 are defined as

$$t_3 = l_3/l_2, \quad t_4 = l_4/l_2. \tag{9}$$

Figure 2 allows to graphically represent the closeness of theoretical (in the case of constant normal values) and sample L-moments. The figure is the L-moment version of a skewness-kurtosis graph based on classical product moments and pairs (α_3, α_4) and (a_3, a_4) .

3 Tests of normality

Let us test the null hypothesis that the data are sampled from a normal distribution.

3.1 Test of normality based on L-moments

The test proposed in [11] is based on the transformations of L-skewness and L-kurtosis and a comparison with the values of normal distributions ($\tau_{3,4}$ in (13)). The transformed L-skewness

$$Z(\tau_3) = \frac{t_3}{\sqrt{\frac{0.1866}{n} + \frac{0.8000}{n^2}}} \tag{10}$$

has approximately the standard normal distribution which also applies to the transformed L-kurtosis

$$Z'(\tau_4) = \frac{t_4 - 0.1226}{\sqrt{\frac{0.0883}{n}}} \tag{11}$$

or more precisely

$$Z(\tau_4) = \frac{t_4 - 0.1226}{\sqrt{\frac{0.0883}{n} + \frac{0.6800}{n^2} + \frac{4.9000}{n^3}}}. \tag{12}$$

Moreover, the transformations are independent. It follows, that the test statistics

$$\tau_{3,4} = Z(\tau_3)^2 + Z(\tau_4)^2 \tag{13}$$

has an approximate chi-square distribution with two degrees of freedom.

In terms of the test power comparison, the authors in [11] show that the test relying on $\tau_{3,4}$ has the greatest power in distributions with high kurtosis in small samples and relatively high power in medium-sized or large samples. The test shows consistently high power, outperforming other normality tests for different distributions.

3.2 Tests of normality used in the simulation

The Jarque-Bera test for normality [14] is one of the most common goodness-of-fit tests utilized in economics not only in the distribution fitting but also in regression diagnostics and time series models. The test compares the sample coefficients of skewness and kurtosis with corresponding values for a normal distribution. This generally preferred test resembles an L-moment-based test, the latter using only robust L-moments instead of the classical product ones. The classical sample moments (sample variance, sample coefficients of skewness and kurtosis) are highly sensitive to outliers, the third and fourth powers of differences from the mean being used to evaluate them. The authors in [9] propose a modification of the Jarque-Bera test that utilizes a robust to outliers estimate of variance, using – instead of standard deviation s given as a root in (1) – the mean absolute deviation from the sample median \tilde{x}

$$\frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}|. \tag{14}$$

Simulation studies indicate that the robust Jarque-Berra test shows higher or similar power in detecting heavy-tailed alternatives compared to the classical JB test.

The Lilliefors test [18] is a variant of the Kolmogorov-Smirnov (KS) test for data normality testing without a fully

specified distribution. The limitation of the KS test is its high sensitivity to extreme values, the LF correction rendering the test less conservative. It is sometimes argued, however, that the test has a low power and therefore should not be seriously considered for testing normality [24].

The Shapiro-Wilk test [21] is based on the correlations of weighted order statistics from the sample and the corresponding normal scores. According to [24], some researchers recommend the SW test as the best choice for testing the normality of data. It is sensitive to non-normality even for samples smaller than 20 observations, asymmetric and long-tailed distributions in particular. Having compared the Shapiro-Wilk, Lilliefors and Anderson-Darling tests, the Monte Carlo simulation in [20] revealed that the SW test (followed by the AD) has the best power for a given level of significance.

The Shapiro-Francia test [22] is well-established and intuitively appealing for testing the departure from normality. The test statistic is the squared correlation between the ordered sample values and the (approximated) expected ordered quantiles from the standard normal distribution. The SF test statistic is one of the most powerful omnibus and purpose tests of non-normality available, its results being quick but not inferior.

The Cramer von Mises and Anderson-Darling [1] tests also compare the hypothetical and empirical distribution function with the test statistics, using the difference defined by the expected value of squared difference instead of the absolute value. The weighting function included in the test statistics allows to consider whether the emphasis is placed on the differences near the median or in the distribution tails. Thus, compared with the Cramer-von Mises distance, the Anderson-Darling distance puts more weight on observations in the tails of the distribution.

The asymptotic chi-square goodness-of-fit test is based on a comparison of hypothetical and empirical frequencies of the selected intervals. It is applicable to a fully specified and as well as estimated distribution, the situation affecting the degrees of freedom of the chi-squared distribution.

The D'Agostino test [3] is a test of normality applicable for samples of size fifty or larger which possesses the desirable omnibus property. Simulation results of the test power for possible alternatives if the sample size is 50 indicate that the test compares favourably with the Shapiro-Wilk test.

4 Results

We compare the power of tests based on a simulation study.

4.1 Simulation study

We use four continuous distributions with heavy tails in the simulation. The two-parametric lognormal distribution (LN) $LN(\mu; \sigma^2)$ is parameterized by the parameters μ ($E(\ln X)$) and σ^2 ($Var(\ln X)$), where $x > 0$, $\mu \in R$, and $\sigma^2 > 0$. The distribution is unimodal and positively skewed. The skewness of the distribution decreases with σ^2 (also see Figure 1 or Table 1). We selected $\mu = 0$ and $\sigma = 0.25, 0.3, 0.4, 0.5, 1$.

The Weibull (W) distribution with a positive shape parameter a and positive scale parameter δ is described by the probability density function

$$f_W(x) = \frac{a}{\delta} \left(\frac{x}{\delta}\right)^{a-1} \exp\left\{-\left(\frac{x}{\delta}\right)^a\right\}, x > 0. \quad (15)$$

The skewness increases with the shape parameter a , the kurtosis being also increasing. In the simulation, parameters $a = 0.25, 0.5, 1, 2$ and $\delta = 1$. The coefficient of skewness is lower than 3 for $a = 0.25, 0.5$ and 1 and higher for the choice $a = 2$ (Table 1).

The loglogistic (LL) distribution is defined by a positive scale parameter γ and a shape parameter δ

$$f_{LL}(x) = \frac{\gamma(x/\delta)^\gamma}{x(1+(x/\delta)^\gamma)^2}. \quad (16)$$

The parameters $\delta = 1$ and $\gamma = 1, 2, 5, 6$. No product moments are defined for $\gamma = 1$, the expected value is finite for $\gamma = 2$, variance for $\gamma = 3$, and third product moment for $\gamma = 3$. Figure 1 shows that the densities are decreasing (for $\gamma = 1$), unimodal and positively skewed for higher parameters.

The values of parameters and classical and L-moment characteristics of the selected distributions are given in Tables 1 and 2. The densities of these distributions are shown in Figure 1, their position being plotted in the L-skewness L-kurtosis graph in Figure 2 (left). The classical skewness-kurtosis graph is shown in the right part of the figure. The allocation of pairs (L-skewness, L-kurtosis) in the bounded subset of the rectangle $[-1,1] \times [-0.25,1]$ shown in the left part of Figure 2 and it is easier to assess compliance or difference between probability distributions.

In the case of lognormal and Weibull distributions, all tests can be applied to the selected data. The choice of parameters for the loglogistic distribution were made to violate the assumptions of some test (see Table 2).

Using the Student distribution (t) with degrees of freedom ν for $\nu = 1, 2, \dots, 20$, we obtain $E(X) = 0, \nu > 1$. From the formula for the variance of the Student distribution $Var(X) = \nu/(\nu - 2), \nu > 2$, we obtain the values from 3 for $\nu = 3$ to 0.11 for $\nu = 20$. The coefficient of skewness is 0 for $\nu > 3$ as the distribution is symmetric. For $\nu > 4$, the coefficient of kurtosis is equal to $\alpha_4 = 6/(6 - \nu), \nu > 4$. For the values of α , we obtain those from 6 for $\nu = 5$ to 0.375 for $\nu = 20$. The Student distribution converges to normal with increasing degrees of freedom, and the approximation made for the values greater than 30 is assumed as acceptable.

We made 30,000 draws from the selected distributions and evaluated p -values for all tests applicable to particular data. The relative frequency of the rejection of the null hypothesis was recorded as a proxy to the power of the test, because the hypothesis was not valid.

All computations were made in the freeware R environment [19] providing a powerful tool for both statistical modelling and simulations. Packages *lmoments* [13], *lmom* [15] and *lmomco* [2] were used for the evaluation of L-moments, *moments* [17] for standard product moments and *actuart* [5] for

Table 1. Lognormal distribution (left) and Weibull distribution (right). CV denotes the coefficient of variation

value	$\mu = 0, \sigma$					$\delta = 1, a$			
	0.25	0.3	0.4	0.5	1	0.25	0.5	1	2
$E(X)$	1.03	1.05	1.08	1.13	1.65	24.00	2.00	1.00	0.89
$\sqrt{Var(X)}$	0.26	0.32	0.45	0.61	2.16	199.36	4.47	1.00	0.46
CV	0.25	0.31	0.42	0.53	1.31	8.31	2.24	1.00	0.52
α_3	0.78	0.95	1.32	1.75	6.18	0.000008	0.07	2.00	6.34
α_4	4.10	4.64	6.26	8.90	113.94	0.78	0.95	1.32	1.75
τ	0.14	0.17	0.22	0.28	0.52	0.94	0.75	0.50	0.29
τ_3	0.12	0.15	0.19	0.24	0.46	0.89	0.63	0.33	0.11
τ_4	0.13	0.14	0.15	0.19	0.29	0.78	0.40	0.17	0.11

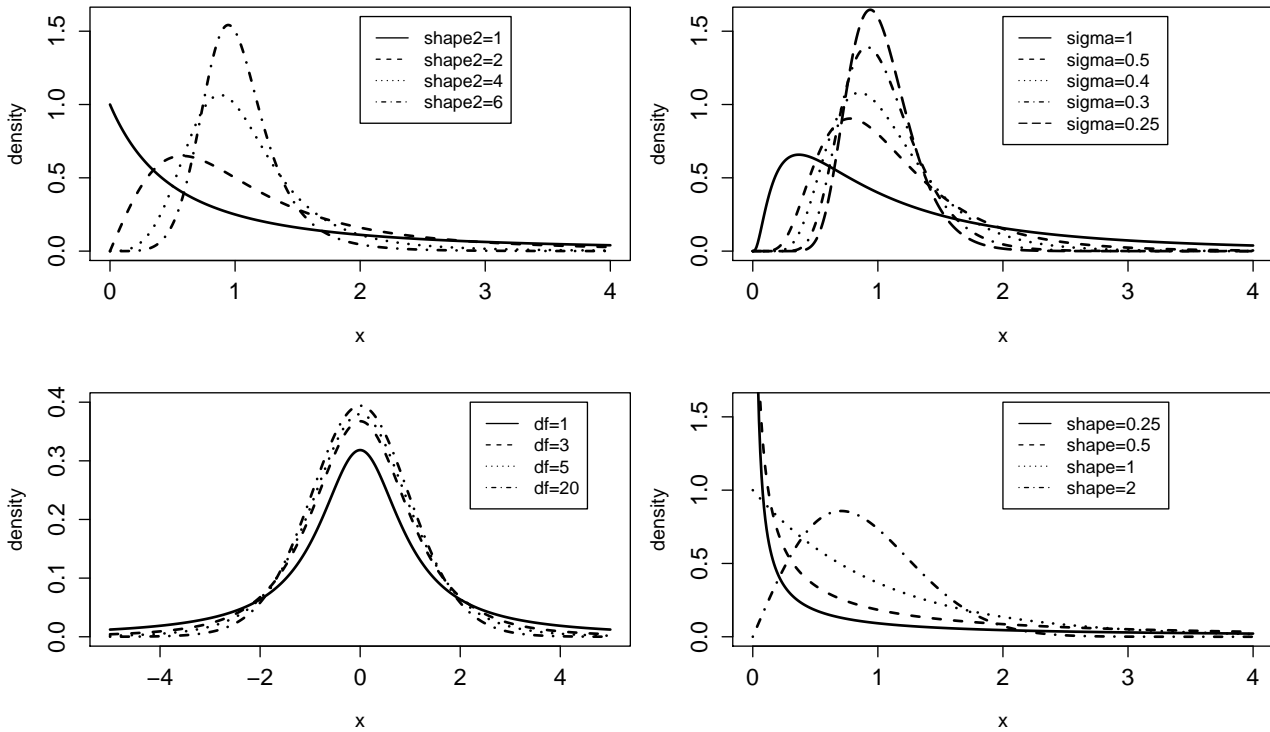


Figure 1. Densities of distributions in the simulation study. Loglogistic distribution (upper left), lognormal distribution (upper right), Student distribution (lower left), Weibull distribution (lower right).

Table 2. Loglogistic distribution

value	$\delta = 1, \gamma$			
	1	2	4	6
$E(X)$	NA	1.57	1.11	1.05
$\sqrt{Var(X)}$	NA	NA	0.58	0.34
CV	NA	NA	0.52	0.32
α_3	NA	NA	4.28	1.82
α_4	NA	NA	NA	14.76
τ	NA	0.50	0.25	0.17
τ_3	NA	0.50	0.25	0.17
τ_4	NA	0.37	0.22	0.19

4.2 Results of the simulation study

The simulation results for lognormal, Weibull and loglogistic distributions are displayed in Tables 3, 4 and 5, respectively. We present the relative frequencies of the normality hypothesis rejection for all distributions, including the one that does not meet the requirements for the characteristics used in the test criterion calculation (for the Cauchy distribution, neither product moments nor L-moments exist, for all tests based on the skewness and kurtosis, the finite fourth product moment is assumed). The results for sample sizes 10, 20, 50, 100, and 200 are presented in the tables, including the chi-squared goodness-of-fit results, although it is an asymptotic test, the small sample test outcomes likely resulting from this asymptotics. In the last columns of all tables, the difference between the minimum and maximum hypothesis rejection relative frequencies is shown to demonstrate the difference between the best and the worst test performance for a particular distribution and sample size.

Burr distribution computations. For testing, we use *lmomco* for $\tau_{3,4}$ test, *KScorrect* [25] for the Lilliefors test, *normtest* [8] for the Jarque-Berra test, *lawstat* [7] for the robust Jarque-Berra test, *nortest* [10] for the chi-square, Shapiro-Francia and Anderson-Darling tests, and *moments* for the D’Agostino test.

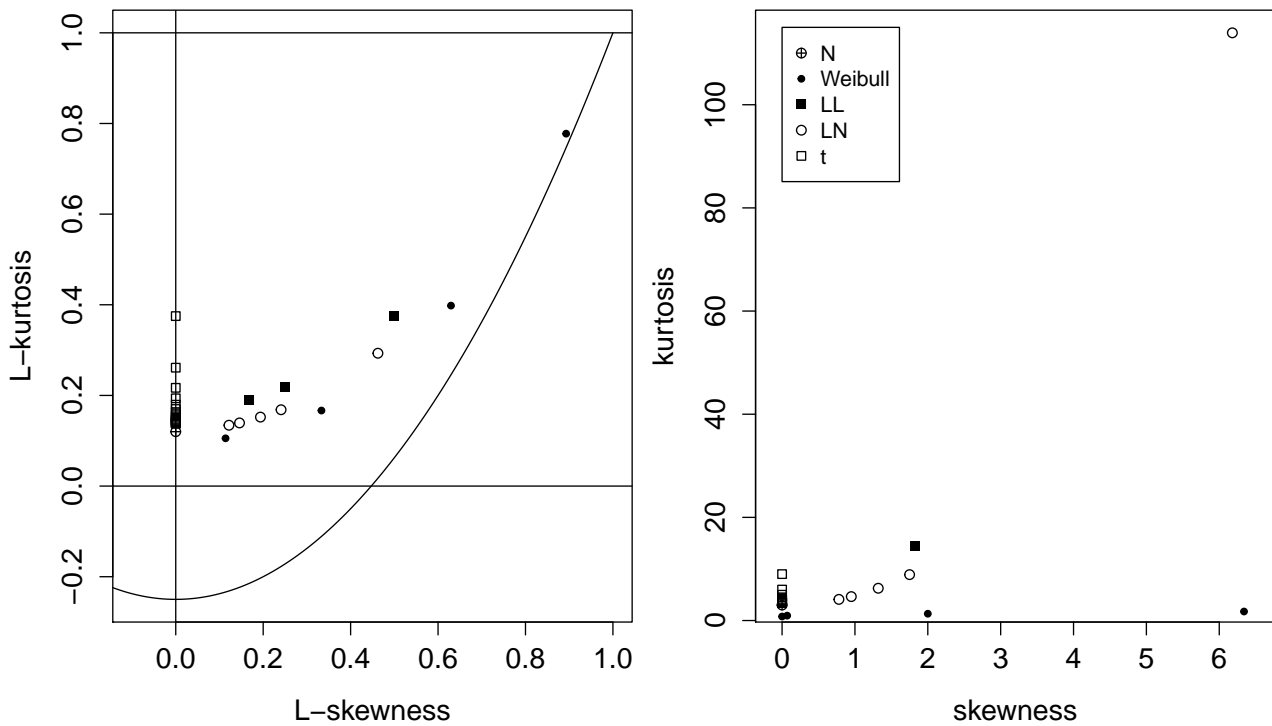


Figure 2. L-skewness and L-kurtosis graph for the distributions in the simulation study.

The lognormal distribution is unimodal and positively skewed with the coefficients of skewness and kurtosis increasing in the parameter σ (and independent of μ). Therefore, the power is lower for smaller parameters, especially for the test based on, inter alia, skewness. There are some tests with bad results (Lilliefors, chi-squared), a group of others that yield average outcomes (Cramer von Mises, robust Jarque-Bera, L-moments-based, Anderson-Darling), and the best performing ones (Shapiro-Wilks, Shapiro-Francia and D’Agostino, the latter exhibiting excellent properties). The Jarque-Bera test outperforms its robust version, the latter variant being of inferior quality to the robust test based on L-moments. For sample sizes of 100 and 200 observations, there are almost no type-two statistical errors and the frequencies of correct decisions are close to one.

For the Weibull distribution within the selected parameterization, the skewness increases with the shape parameter a . The kurtosis is also increasing, but its coefficient is lower than 3 for all selected values (Table 1). Table 4 displays the results, the performance of all tests declining with increasing parameter. For high values of the parameter, larger samples are needed to distinguish between Weibull and normal distributions. In this case, the D’Agostino test is not comparable to the most powerful pair of the Shapiro-Wilks and Shapiro-Francia tests as is the case with the lognormal distribution. Both Jarque-Bera tests are in the group of poor-performing ones together with the chi-square (except for 10 observations and $a = 2$) and Lilliefors tests.

For the loglogistic distribution, no product moments are defined for $\sigma = 1$. Thus, we use the parameters implying only

finite expected value $\sigma = 2$, finite variance $\sigma = 3$, or a finite third product moment for $\sigma = 3$. Figure 1 shows that the densities are decreasing (for $\sigma = 1$), unimodal and positively skewed for higher parameters. Problems with the existence of product moments (Table 2) are to be considered carefully. Since we test normality and if the null hypothesis is valid, we nevertheless apply the test although this is problematic according to the nature of the given distribution. The Anderson-Darling performs well for small values of $\sigma = 2$. The three other tests – D’Agostino, Shapiro-Wilks and Shapiro-Francia – are also strong, the latter in particular maintaining a consistently high performance.

Figures 3, 4 and, 5 illustrates the test power dependence on the degrees of freedom of the Student distribution. Unlike the other distributions analysed, the Student distribution is symmetric, converging quickly to the standard normal distribution, although we still do not make a normal approximation of the highest value of freedom degrees $\nu = 20$. The graph lines of the same type referring to the test are shifted from the left (sample size 10) to the right (sample size 200). For small values of freedom degrees, it is easier to distinguish between the Student and normal distributions, the power of all tests being relatively high. With increasing degrees of freedom, the two distributions are increasingly similar, thus more difficult to distinguish. For small samples and heavy-tailed distributions, robust tests have the highest power, the one based on L-moments outperforming the robust Jarque-Bera test. The Shapiro-Francia and D’Agostino tests (the latter for higher freedom degree values) can handle heavy tail problems and the resulting data outliers.

Table 3. Power of test for lognormal distribution

n	shape	tau34	JB	RJB	SW	DA	LKS	AD	CR	PE	SF	length
10	0.25	0.989	0.917	0.923	0.994	0.933	0.968	0.993	0.990	0.991	0.992	0.077
10	0.5	0.853	0.711	0.693	0.896	0.750	0.756	0.875	0.855	0.876	0.877	0.203
10	1	0.379	0.338	0.305	0.446	0.369	0.301	0.412	0.384	0.395	0.435	0.145
10	2	0.077	0.079	0.073	0.084	0.086	0.072	0.081	0.076	0.090	0.087	0.019
20	0.25	1.000	0.999	0.998	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.002
20	0.5	0.997	0.961	0.945	0.999	0.978	0.982	0.997	0.995	0.995	0.998	0.054
20	1	0.772	0.631	0.568	0.836	0.706	0.576	0.775	0.727	0.655	0.798	0.268
20	2	0.120	0.120	0.104	0.153	0.146	0.098	0.129	0.117	0.078	0.142	0.075
50	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	1	0.998	0.976	0.932	1.000	0.988	0.960	0.997	0.990	0.983	0.999	0.067
50	2	0.328	0.264	0.215	0.410	0.365	0.205	0.306	0.261	0.137	0.353	0.273
100	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	2	0.682	0.563	0.445	0.790	0.697	0.398	0.610	0.513	0.262	0.719	0.529
200	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	2	0.961	0.947	0.894	0.993	0.959	0.709	0.932	0.853	0.558	0.984	0.435

Table 4. Power of test for Weibull distribution

n	shape	tau34	JB	RJB	SW	DA	LKS	AD	CR	PE	SF	length
10	0.25	0.989	0.917	0.923	0.994	0.933	0.968	0.993	0.990	0.991	0.992	0.077
10	0.5	0.853	0.711	0.693	0.896	0.750	0.756	0.875	0.855	0.876	0.877	0.203
10	1	0.379	0.338	0.305	0.446	0.369	0.301	0.412	0.384	0.395	0.435	0.145
10	2	0.077	0.079	0.073	0.084	0.086	0.072	0.081	0.076	0.090	0.087	0.019
20	0.25	1.000	0.999	0.998	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.002
20	0.5	0.997	0.961	0.945	0.999	0.978	0.982	0.997	0.995	0.995	0.998	0.054
20	1	0.772	0.631	0.568	0.836	0.706	0.576	0.775	0.727	0.655	0.798	0.268
20	2	0.120	0.120	0.104	0.153	0.146	0.098	0.129	0.117	0.078	0.142	0.075
50	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	1	0.998	0.976	0.932	1.000	0.988	0.960	0.997	0.990	0.983	0.999	0.067
50	2	0.328	0.264	0.215	0.410	0.365	0.205	0.306	0.261	0.137	0.353	0.273
100	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	2	0.682	0.563	0.445	0.790	0.697	0.398	0.610	0.513	0.262	0.719	0.529
200	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	2	0.961	0.947	0.894	0.993	0.959	0.709	0.932	0.853	0.558	0.984	0.435

5 Conclusion

Testing the normality of data is an essential part of the statistical analysis, the issue receiving considerable attention in the literature. In the present article, Monte Carlo simulations are employed to investigate the ability of commonly used normality tests to detect the abnormality of the data. We apply a L-moment-based test by Harri and Coble [11] and compare their properties to widely known and frequently used tests of normality.

Our results are in line with the recommendations of others, and we present qualities of the test based on L-moments. All tests in the study are well implemented in the program R.

The simulation results show that for symmetric short-tailed distributions, the D’Agostino and Shapiro-Wilk tests are the most powerful. In terms of symmetric long-tailed distributions, the powers of the Jarque-Berra and D’Agostino tests are comparable with those of the Shapiro-Wilk test. As for asymmetric distributions, the Shapiro-Wilk test followed by the Anderson-

Darling test have the highest power. The Jarque-Berra and robust Jarque-Berra tests achieve similar performance for all simulated distributions except the Student one.

The test based on L-moments works similarly to average tests, being, however, outperformed by the Shapiro-Wilks and Shapiro-Francia tests. In the case of a symmetric distribution with long tails (Student distribution for small freedom degrees), the robust tests (L-moment-based and robust Jarque-Berra ones) exceed the performance of other tests. This property is expected, as the outliers occur in the samples, and some tests are sensitive to such values (especially test based on correlations or product moments).

The test relying on L-moments is suitable for all distributions analysed, and along with the robust Jarque-Berra it is the most powerful for the Student distribution, outperforming the latter test for asymmetric distributions and being comparable for long tails. Hence, we consider the L-moment-based test a generally acceptable high-performance option for long-tail data applicable to small samples as well.

Table 5. Power of test for loglogistic distribution

n	shape	tau34	JB	RJB	SW	DA	LKS	AD	CR	PE	SF	length
10	0.25	0.989	0.917	0.923	0.994	0.933	0.968	0.993	0.990	0.991	0.992	0.077
10	0.5	0.853	0.711	0.693	0.896	0.750	0.756	0.875	0.855	0.876	0.877	0.203
10	1	0.379	0.338	0.305	0.446	0.369	0.301	0.412	0.384	0.395	0.435	0.145
10	2	0.077	0.079	0.073	0.084	0.086	0.072	0.081	0.076	0.090	0.087	0.019
20	0.25	1.000	0.999	0.998	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.002
20	0.5	0.997	0.961	0.945	0.999	0.978	0.982	0.997	0.995	0.995	0.998	0.054
20	1	0.772	0.631	0.568	0.836	0.706	0.576	0.775	0.727	0.655	0.798	0.268
20	2	0.120	0.120	0.104	0.153	0.146	0.098	0.129	0.117	0.078	0.142	0.075
50	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
50	1	0.998	0.976	0.932	1.000	0.988	0.960	0.997	0.990	0.983	0.999	0.067
50	2	0.328	0.264	0.215	0.410	0.365	0.205	0.306	0.261	0.137	0.353	0.273
100	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
100	2	0.682	0.563	0.445	0.790	0.697	0.398	0.610	0.513	0.262	0.719	0.529
200	0.25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000
200	2	0.961	0.947	0.894	0.993	0.959	0.709	0.932	0.853	0.558	0.984	0.435

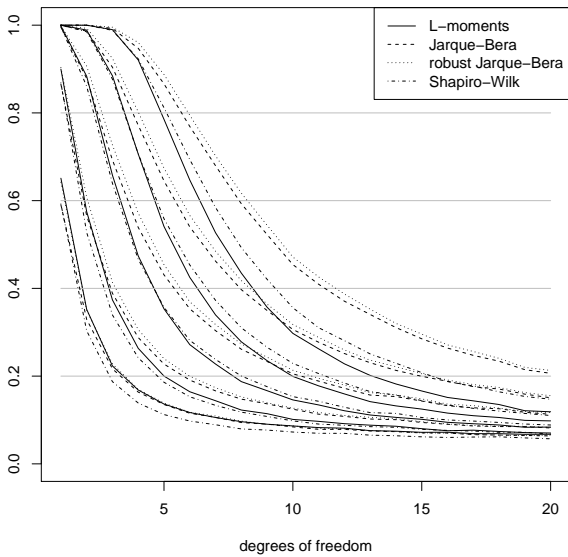


Figure 3. Student distribution, power of test a)

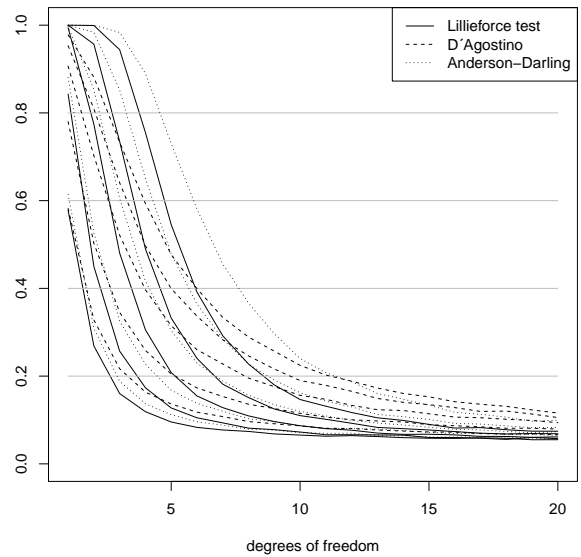


Figure 4. Student distribution, power of test b)

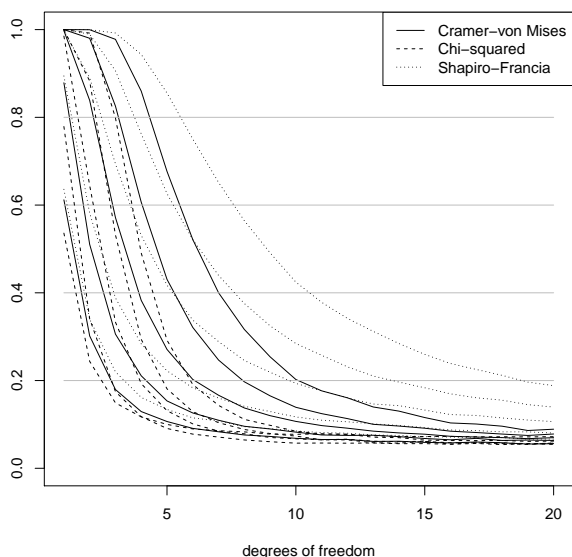


Figure 5. Student distribution, power of test c)

Acknowledgements

Institutional support from the funds for the long-term conceptual advancement of science and research, number IP 400 040, at the Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic is gratefully acknowledged.

REFERENCES

- [1] Anderson, T.W., Darling, D.A., "A test of goodness of fit," *American Statistical Association Journal*, vol. 49, no. 268, pp. 765–769, 1954. DOI: 10.1080/01621459.1954.10501232
- [2] Asquith W.H., "lmomco: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions," R package, version 2.3.6., 2019. <https://cran.r-project.org/web/packages/lmomco/index.html>
- [3] D'Agostino, R.B., "An omnibus test of normality for moderate and large size samples," *Biometrika*, vol. 58, no. 234, pp. 341–348, 1971. DOI: 10.2307/2334522
- [4] D'Agostino, R.B., Belanger, A., D'Agostino, Jr. R.B., "A suggestion of using powerful and informative tests of normality," *The American Statistician*, vol. 44, no. 4, pp. 316–321 1993. DOI: 10.2307/2684359
- [5] Dutang, Ch., Goulet, V., Pigeon, M., "Actuar: An R Package for Actuarial Science," *Journal of Statistical Software*, vol. 25, no. 7, pp. 1–37, 2008. DOI: 10.18637/jss.v025.i07
- [6] Elamir, E.A.H., Seheult, A.H., "Trimmed L-Moments," *Computational Statistics & Data Analysis*, vol. 43, no. 3, pp. 299–314, 2013. DOI:10.1016/S0167-9473(02)00250-5
- [7] Gastwirth J.L., Gel Y.R., Wallace Hui W.L., "lawstat: An R Package for Law, Public Policy and Biostatistics," *Journal of Statistical Software*, vol. 62, no. 3, pp. 399–402, 2008. DOI: 10.18637/jss.v028.i03
- [8] Gavrilov, I., Pusev, R., "normtest: Tests for Normality", R package, version 1.1., 2004. <https://cran.r-project.org/web/packages/normtest/>
- [9] Gel, Y.R., Gastwirth, J.L., "A robust modification of the Jarque-Bera test of normality," *Economic letters*, vol. 99, no. 1, pp. 30–32, 2008. <https://ideas.repec.org/a/eee/econlet/v99y2008i1p30-32.html>
- [10] Gross, J., Ligges, U., "nortest: Tests for normality," R package, version 1.0-4, 2015. <https://cran.r-project.org/web/packages/nortest/index.html>
- [11] Harri, A., Coble, K.H., "Normality testing: two new tests using L-moments," *Journal of Applied Statistics*, vol. 38, no. 7, pp. 1369–1379, 2011. DOI: 10.1080/02664763.2010.498508
- [12] Hosking, J. R. M., "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 52, no. 1, pp. 105–124. 1990. <http://www.jstor.org/stable/2345653>
- [13] Hosking, J. R. M., "L-Moments," R package, version 2.8, 2019. <https://CRAN.R-project.org/package=lmom>
- [14] Jarque, C.M., Berra, A.K., "Efficient tests for normality, homoskedasticity and serial independence of regression residuals," *Economic letters*, vol. 6, no. 3, pp. 255–259, 1980. <https://EconPapers.repec.org/RePEc:eee:econlet/v:6:y:1980:i:3:p:255-259>
- [15] Karvanen, J., "Estimation of quantile mixtures via L-moments and trimmed L-moments," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 947–959. 2006. DOI:10.1016/j.csda.2005.09.014
- [16] Keskin, S., "Comparison of several univariate normality tests regarding Type I error rate and power of the test in simulation based on small samples," *Journal of Applied Sciences Research*, vol. 2, no. 3, pp. 296–300, 2006. <http://www.aensiweb.com/old/jasr/jasr/2006/296-300.pdf>
- [17] Komsta, L., Novomestky, F., "moments: Moments, cumulants, skewness, kurtosis and related tests," R package, version 0.14, 2019. <https://cran.r-project.org/web/packages/moments/index.html>
- [18] Lilliefors, H.W., "On the Kolmogorov-Smirnov test for normality with mean and variance unknown," *American Statistical Association Journal*, vol. 62, no. 318, pp. 399–402, 1967. DOI: 10.2307/2283970
- [19] R Core Team. "R: A language and environment for statistical computing. R Foundation for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2020. <https://www.R-project.org/>
- [20] Razali, N., Yap, W. B., "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modelling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011. <https://www.nrc.gov/docs/ML1714/ML17143A100.pdf>

- [21] Shapiro, S. S., Wilk, M. B., "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965. DOI: 10.2307/2333709
- [22] Shapiro, S. S., Francia, R. S., "An approximate analysis of variance test for normality," *Journal of the American Statistical Association*, vol. 67, no. 337, pp. 215—216, 1972. DOI: 10.1080/01621459.1972.10481232
- [23] Thadewald, T., Buning, H., "Jarque–Bera test and its competitors for testing normality – a power comparison," *Journal of Applied Statistics*, vol. 34, no. 1, pp. 87–105, 2007. DOI: 10.1080/02664760600994539
- [24] Thode, H.C., "Testing For Normality," CRC Press. 2002. ISBN 9780367447106
- [25] Yap, B. W., Sim, C. H., "KScorrect: Lilliefors-Corrected Kolmogorov-Smirnov Goodness-of-Fit Tests," R package, version 1.4.0, 2011. <https://cran.r-project.org/web/packages/kscorrect/index.html>
- [26] Yap, B. W., Sim, C. H., "Comparisons of various types of normality tests," *Journal of Statistical Computation and Simulation*, vol. 81, no. 12, pp. 2141–2155, 2011. DOI: 10.1080/00949655.2010.520163
- [27] Yazici, B., Yolacan, S., " A comparison of various tests of normality," *Journal of Statistical Computation and Simulation*, vol. 77, no. 2, pp. 175–183, 2007. DOI: 10.1080/10629360600678310