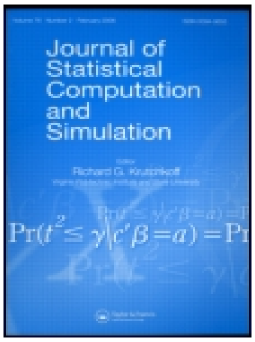


Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Comparisons of various types of normality tests

B. W. Yap & C. H. Sim

To cite this article: B. W. Yap & C. H. Sim (2011) Comparisons of various types of normality tests, Journal of Statistical Computation and Simulation, 81:12, 2141-2155, DOI: 10.1080/00949655.2010.520163

To link to this article: https://doi.org/10.1080/00949655.2010.520163

Published online: 18 May 2011.

Journal of Statistical Computation and Simulation
Vol. 81, No. 12, December 2011, 2141–2155

Comparisons of various types of normality tests

B.W. Yap[a]* and C.H. Sim[b]

[a] *Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia;* [b] *Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia*

(*Received 3 June 2010; final version received 29 August 2010*)

Normality tests can be classified into tests based on chi-squared, moments, empirical distribution, spacings, regression and correlation and other special tests. This paper studies and compares the power of eight selected normality tests: the Shapiro–Wilk test, the Kolmogorov–Smirnov test, the Lilliefors test, the Cramer–von Mises test, the Anderson–Darling test, the D'Agostino–Pearson test, the Jarque–Bera test and chi-squared test. Power comparisons of these eight tests were obtained via the Monte Carlo simulation of sample data generated from alternative distributions that follow symmetric short-tailed, symmetric long-tailed and asymmetric distributions. Our simulation results show that for symmetric short-tailed distributions, D'Agostino and Shapiro–Wilk tests have better power. For symmetric long-tailed distributions, the power of Jarque–Bera and D'Agostino tests is quite comparable with the Shapiro–Wilk test. As for asymmetric distributions, the Shapiro–Wilk test is the most powerful test followed by the Anderson–Darling test.

Keywords: normality tests; Monte Carlo simulation; skewness; kurtosis; generalized lambda distribution

1. Introduction

The importance of normal distribution is undeniable since it is an underlying assumption of many statistical procedures. It is also the most frequently used distribution in statistical theory and applications. Therefore, when carrying out statistical analysis using parametric methods, validating the assumption of normality is of fundamental concern for the analyst. An analyst often concludes that the distribution of the data 'is normal' or 'not normal' based on the graphical exploration (Q–Q plot, histogram or box plot) and formal test of normality. Even though graphical methods are useful in checking the normality of a sample data, they are unable to provide formal conclusive evidence that the normal assumption holds. Graphical method is subjective as what seems like a 'normal distribution' to one may not necessarily be so to others. In addition, vast experience and good statistical knowledge are required to interpret the graph properly. Therefore, in most cases, formal statistical tests are required to confirm the conclusion from graphical methods.

There are a significant number of tests of normality available in the literature. D'Agostino and Stephens [1] provided a detailed description of various normality tests. Some of these tests

*Corresponding author. Email: yapbeewah@yahoo.com

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2010.520163
http://www.tandfonline.com

are constructed to be applied under certain conditions or assumptions. Extensive studies on the Type I error rate and power of these normality tests have been discussed in [1–9]. Most of these comparisons were carried out using selected normality tests and selected small sample sizes. Some use tabulated critical values while others use simulated critical values. Consequently, there are still contradicting results as to which is the optimal or best test and these may mislead and often confuse practitioners as to which test should be used for a given sample size.

A search on normality tests available in statistical software packages such as SAS, SPSS, MINITAB, SPLUS, STATISTICA, STATGRAPHICS, STATA, IMSL library, MATLAB and R revealed that the commonly available normality tests in these software are: Pearson's chi-squared (CSQ) goodness-of-fit test, the Cramer–von Mises (CVM) test, the Kolmogorov–Smirnov test (referred to as KS henceforth), the Anderson–Darling (AD) test, the Shapiro–Wilk (SW) test, the Lilliefors (LL) test, the Shapiro–Francia test, the Ryan-Joiner test and the Jarque–Bera (JB) test.

Table 1 lists the normality test available for these statistical software packages. SAS provides the SW, KS, AD and CVM tests while MINITAB provides only the AD, Ryan–Joiner (similar to the SW test) and KS tests. We found that while the basic SPLUS package provides only the KS test and CSQ goodness-of-fit test, the ENVIRONMENTALSTATS for SPLUS (an add-on module) includes the SW and Shapiro–Francia tests. In SPSS, the significance of the SW statistic is calculated by linearly interpolating within the range of simulated critical values given in Shapiro and Wilk [10]. However, SAS, SPLUS, STATISTICA, STATA, MATLAB and R used the AS R94 algorithm for the SW test provided by Royston [11]. The LL test in SPSS and SPLUS used corrected critical values provided by Dallal and Wilkinson [12]. The individual and overall skewness–kurtosis test is provided only by STATA while STATGRAPHICS provides the standardized-skewness and standardized-kurtosis z-scores.

Assume that we have a random sample $X_1, X_2, \ldots, X_n$ of independently and identically distributed random variables from a continuous univariate distribution with an unknown probability density function (PDF) $f(x, \Theta)$, where $\Theta = (\theta_1, \theta_2, \ldots, \theta_p)'$ is a vector of real-valued parameters. Then the formal testing of whether the observed sample comes from a population with a normal distribution can be formulated as that of testing a composite hypothesis:

$$H_0 : f(x, \Theta) \in N(\mu, \sigma^2) \text{ against } H_a : f(x, \Theta) \notin N(\mu, \sigma^2).$$

A test is said to be powerful when it has a high probability of rejecting the null hypothesis of normality when the sample under study is taken from a non-normal distribution. In making comparison, all tests should have the same probability of rejecting the null hypothesis when the distribution is truly normal (i.e. they have to have the same Type I error which is α, the significance

Table 1. Normality tests available in statistical software packages.

	Test									
Software	SW	SF	KS	LL	CVM	AD	JB	CSQ	RJ	SKKU
SAS	✓		✓		✓	✓				
SPSS	✓			✓						
SPLUS	✓	✓	✓					✓		
STATISTICA	✓		✓	✓				✓		
STATA	✓	✓								✓
STATGRAPHICS	✓		✓	✓	✓	✓		✓		✓
MINITAB			✓			✓			✓	
MATLAB		✓	✓	✓	✓	✓	✓	✓		
R	✓	✓	✓	✓	✓	✓	✓	✓		
IMSL Library	✓		✓	✓				✓		

Notes: SW, Shapiro–Wilk test; SF, Shapiro–Francia test; KS, Kolomogorov–Smirnov test; LL, Lilliefors test; CVM, Cramer–Von Mises test; AD, Anderson–Darling test; JB, Jarque–Bera test; CSQ, chi-squared test; RJ, Ryan–Joiner test; SKKU, skewness–kurtosis test.

level). Using Monte Carlo simulation, 10,000 samples from a given non-normal distribution are generated and the power of the test is the proportion of samples which the test rejected the null hypothesis of normality. This simulation study focuses on the performance of eight selected normality tests: the SW test, the KS test, the LL test, the AD test, the JB test, the CVM test, the CSQ test and the D'Agostino–Pearson (DP) test. D'Agostino *et al.* [13] pointed out that DP [14] K_2 test which combined skewness ($\sqrt{b_1}$) and kurtosis (b_2) has good power properties over a broad range of non-normal distributions.

In Section 2, we present the procedures for the eight normality tests considered in this study. The Monte Carlo simulation methodology is explained in Section 3. Results and comparisons of the power of the normality tests are discussed in Section 4. Finally a conclusion is given in Section 5.

2. Normality tests

Normality tests can be classified into tests based on regression and correlation (SW, Shapiro–Francia and Ryan–Joiner tests), CSQ test, empirical distribution test (such as KS, LL, AD and CVM), moment tests (skewness test, kurtosis test, D'Agostino test, JB test), spacings test (Rao's test, Greenwood test) and other special tests. In this section, we present the eight normality tests procedures investigated in this study.

2.1. *SW test*

The regression and correlation tests are based on the fact that a variable $Y \sim N(\mu, \sigma^2)$ can be expressed as $Y = \mu + \sigma X$, where $X \sim N(0, 1)$. The SW [10] test is the most well-known regression test and was originally restricted for sample size of $n \leq 50$. If $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denotes an ordered random sample of size n from a standard normal distribution ($\mu = 0, \sigma = 1$), let $\boldsymbol{m}' = (m_1, m_2, m_n)$ be the vector of expected values of the standard normal order statistics and let $\mathbf{V} = (v_{ij})$ be the $n \times n$ covariance matrix of these order statistics. Let $\boldsymbol{Y}' = (Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})$ denote a vector of ordered random observations from an arbitrary population. If $Y_{(i)}$'s are ordered observations from a *normal distribution* with unknown mean μ and unknown variance σ^2, then $Y_{(i)}$ may be expressed as $Y_{(i)} = \mu + \sigma X_{(i)}$ $(i = 1, 2, \ldots, n)$.

The SW test statistic for normality is defined as

$$\mathrm{SW} = \frac{\left[\sum_{i=1}^{n} a_i Y_{(i)}\right]^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}, \tag{1}$$

where

$$\boldsymbol{a}' = \boldsymbol{m}'\mathbf{V}^{-1}(\boldsymbol{m}'\mathbf{V}^{-1}\boldsymbol{m})^{-1/2}. \tag{2}$$

The a_i's are weights that can be obtained from Shapiro and Wilk [10] for sample size $n \leq 50$. The value of SW lies between zero and one. Small values of SW lead to the rejection of normality, whereas a value of one indicates normality of the data.

The SW test was then modified by Royston [15] to broaden the restriction of the sample size. He gave a normalizing transformation for SW as $Y = (1 - \mathrm{SW})^{\lambda}$ for some choices of λ. The parameter λ was estimated for 50 selected sample sizes and then smoothed with polynomials in $\log_e(n) - d$ where $d = 3$ for $7 \leq n \leq 20$ and $d = 5$ for $21 \leq n \leq 2000$. Royston [16,17] provided algorithm AS 181 in FORTRAN 66 for computing the SW test statistic and p-value for sample sizes 3–2000. Later, Royston [18] observed that SW's [10] approximation for the weights $\boldsymbol{a}$ used in the algorithms was inadequate for $n > 50$. He then gave an improved approximation to the weights and provided algorithm AS R94 [11] which can be used for any n in the range $3 \leq n \leq 5000$. This study used the algorithm provided by Royston [11].

A sample of variations of this test includes modifications suggested by Shapiro and Francia [19], Weisberg and Bingham [20] and Rahman and Govindarajulu [21].

2.2. *Empirical distribution function test*

The idea of the empirical distribution function (EDF) tests in testing normality of data is to compare the EDF which is estimated based on the data with the cumulative distribution function (CDF) of normal distribution to see if there is a good agreement between them. The most popular EDF tests are the ones developed by Kolmogorov–Smirnov [22], Cramer–von Mises [23] and Anderson–Darling [24].

2.2.1. *KS test*

Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be an ordered random sample and the distribution of X is $F(x)$. The EDF $F_n(x)$ is defined as the fraction of X_i's that are less than or equal to x for each x,

$$F_n(x) = \frac{\text{no. of observations} \leq x}{n} \quad -\infty < x < \infty. \tag{3}$$

The KS statistic belongs to the supremum class of EDF statistics and this class of statistics is based on the *largest* vertical difference between the hypothesized and empirical distribution. This test requires that the null distribution $F^*(x)$ be completely specified with known parameters. In KS test of normality, $F^*(x)$ is taken to be a normal distribution with known mean μ and standard deviation σ. The test statistics is defined differently for the following three different set of hypotheses.

For a right-tailed test $H_0 : F(x) = F^*(x)$ versus $H_a : F(x) > F^*(x)$, the test statistic $\text{KS}^+ = \sup_x[F^*(x) - F_n(x)]$ is the greatest vertical distance where the function $F^*(x)$ is above the function $F_n(x)$. Likewise, for the left-tailed test $H_0 : F(x) = F^*(x)$ versus $H_a : F(x) < F^*(x)$, the test statistic $\text{KS}^- = \sup_x[F_n(x) - F^*(x)]$ is the greatest vertical distance where the function $F_n(x)$ is above the function $F^*(x)$. The Kolmogorov–Statistic for a two-sided test, $H_0 : F(x) = F^*(x)$ versus $H_a : F(x) \neq F^*(x)$, is taken to be

$$\text{KS} = \max(\text{KS}^+, \text{KS}^-). \tag{4}$$

In this study, $F^*(x)$ is taken to be a normal distribution, and thus large values of KS indicate non-normality.

2.2.2. *LL test for normality*

The LL test is a modification of the KS test. This test was developed by Lilliefors [25] and is suitable when the unknown parameters of the null distribution must be estimated from the sample data. This test compares the empirical distribution of X with a normal distribution where its unknown μ and σ are estimated from the given sample data. The random sample of size $n, X_1, X_2, \ldots, X_n$ is assumed to be associated with a hypothesized distribution function $F(x)$ with unknown parameters.

The LL test statistic is again taken to be Equation (4), except that the values of μ and σ used are the sample mean and standard deviation.

The difference between the LL and KS test statistic is that the EDF $F_n(x)$ is obtained from the normalized sample (Z_i) while $F_n(x)$ in the KS test used the original X_i values. Then, Lilliefors [26] introduced the test for exponential distribution. This simulation study used the LILLF subroutine given in the FORTRAN IMSL libraries.

2.2.3. *CVM test*

Conover [27] stated that the CVM test was developed by Cramer [23], von Mises [28] and Smirnov [29]. The CVM test judges the goodness of fit of a hypothesized distribution $F^*(x)$ compared with the EDF $F_n(x)$ based on the statistic defined as

$$nw^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \, \mathrm{d}F(x), \tag{5}$$

which, like the KS statistic, is distribution-free, i.e. its distribution does not depend on the hypothesized distribution, $F^*(x)$. The CVM test is an alternative to the KS test. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the ordered observations of a sample of size n. The test statistic Equation (5) may be computed as [30,31]

$$\mathrm{CVM} = \frac{1}{12n} + \sum_{i=1}^{n} \left(Z_i - \frac{2i-1}{2n} \right)^2, \tag{6}$$

where $Z_i = \Phi((X_{(i)} - \bar{X})/S)$, $\bar{X} = \left(\sum_{i=1}^{n} X_i\right)/n$, $S^2 = \left(\sum_{i=1}^{n} (X_i - \bar{X})^2\right)/(n-1)$, $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ are the ordered observations and $\Phi(x)$ is the standardized hypothesized normal distribution of the null hypothesis.

2.2.4. *AD test*

The AD [24] test is actually a modification of the CVM test. It differs from the CVM test in such a way that it gives more weight to the tails of the distribution than does the CVM test. Unlike the CVM test which is distribution-free, the AD test makes use of the specific hypothesized distribution when calculating its critical values. Therefore, this test is more sensitive in comparison with the CVM test. A drawback of this test is that the critical values have to be calculated for each specified distribution.

The AD test statistic

$$\mathrm{AD} = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \Psi(F(x)) \, \mathrm{d}F(x), \tag{7}$$

is a weighted average of the squared discrepancy $[F_n(x) - F(x)]^2$, weighted by $\Psi(F(x)) = \{F(x)(1 - F(x))\}^{-1}$. Note that by taking $\Psi(F(x)) = 1$, the AD statistic reduces to the CVM statistic in Equation (5).

Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the ordered observations in a sample of size n. The AD statistic is computed as:

$$\mathrm{AD} = -\sum \left[\frac{(2i-1)\{\log P_i + \log(1 - P_{n+1-i})\}}{n} \right] - n, \tag{8}$$

where P_i is the CDF of the specified distribution and log is log base e. To test if the distribution is normal, $P_i = \Phi(Y_{(i)})$, where $Y_{(i)} = (X_{(i)} - \bar{X})/S$ and $\bar{X}$ and S is the sample mean and standard deviation, respectively. The AD statistic may also be computed from the modified statistic [1, p. 373]

$$\mathrm{AD}^* = \mathrm{AD}\left(1 + \frac{0.75}{n} + \frac{2.24}{n^2}\right). \tag{9}$$

2.3. *CSQ test*

The oldest and most well-known goodness-of-fit test is the CSQ test for goodness of fit, first presented by Pearson [32] ([27, p. 240]). However, the CSQ test is not highly recommended for continuous distribution since it uses only the counts of observations in each cell rather than the observations themselves in computing the test statistic. Let O_j denote the number of observations in cell j, for $j = 1, 2, \ldots, c$. Let p_j^* be the probability of a random observation being in cell j, under the assumption that the null hypothesis is true. Then, the expected number of observations in cell j is defined as $E_j = p_j^* n$ and n is the sample size. The CSQ test statistic CSQ is given by

$$\mathrm{CSQ} = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j}. \tag{10}$$

By considering equiprobable cells, $p_j = 1/c$, $j = 1, 2, \ldots, c$, and the CSQ test statistic is then reduced to

$$\mathrm{CSQ} = \frac{c}{n}\sum_{j=1}^{c}\left(n_j - \frac{n}{c}\right)^2, \tag{11}$$

where n_j is the number of observations that fall in the jth cell and c is the number of equiprobable cell.

Schorr [33] found that for large sample size, the optimum number of cells c should be smaller than $M = 4(2n^2/z_\alpha^2)^{1/5}$, where z_p is the $100(1-\alpha)$th percentile of the standard normal distribution. If k parameters of the distribution of X need to be estimated, then distribution of CSQ follows approximately a CSQ distribution with $c - k - 1$ degrees of freedom.

2.4. *Moment tests*

Normality tests based on moments include skewness ($\sqrt{b_1}$) test, kurtosis (b_2) test, the DP test and the JB test. The procedures for the skewness and kurtosis test can be found in D'Agostino and Stephens [1] and D'Agostino *et al.* [13]. These two tests are not included in this study as they are not available in major statistical software and not commonly used.

2.4.1. *JB test*

Jarque and Bera [34] used the Lagrange multiplier procedure on the Pearson family of distributions to obtain tests for normality of observations and regression residuals. They claimed that the test have optimum asymptotic power properties and good finite sample performance. The JB statistic is based on sample skewness and kurtosis and is given as

$$\mathrm{JB} = n\left(\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24}\right). \tag{12}$$

The JB statistic is actually the test statistic suggested by Bowman and Shenton [35]. The JB statistic follows approximately a CSQ distribution with two degrees of freedom. The JB statistic equals zero when the distribution has zero skewness and kurtosis is 3. The null hypothesis is that the skewness is zero and kurtosis is 3. Large values of skewness and kurtosis values far from 3 lead to the rejection of the null hypothesis of normality.

2.4.2. *The DP omnibus test*

The sample skewness and kurtosis $\sqrt{b_1}$ and b_2 are used separately in the skewness and kurtosis tests in testing the hypothesis if random samples are taken from a normal population. To overcome this drawback, D'Agostino and Pearson [14] proposed the following test statistic

$$\mathrm{DP} = Z^2(\sqrt{b_1}) + Z^2(b_2), \tag{13}$$

that takes into consideration both values of $\sqrt{b_1}$ and b_2, where $Z(\sqrt{b_1})$ and $Z(b_2)$ are the normal approximations to $\sqrt{b_1}$ and b_2, respectively. The DP statistic follows approximately a CSQ distribution with 2df when a population is normally distributed. It is often referred to as an omnibus test where 'omnibus' means it is able to detect deviations from normality due to either skewness or kurtosis. Bowman and Shenton [35] used the Johnson system, S_U and S_B as approximate normalizing distributions to set up contours in the $(\sqrt{b_1}, b_2)$ plane of the K^2 test statistic for sample sizes ranging from 20 to 1000. They noted that besides K^2 another composite test statistic is $(\sqrt{b_1})^2/\sigma_1^2 + (b_2 - 3)^2/\sigma_2^2$, where $\sigma_1^2 = 6/n$ and $\sigma_2^2 = 24/n$ are the asymptotic variances of $\sqrt{b_1}$ and b_2, respectively.

3. Simulation methodology

Monte Carlo procedures were used to evaluate the power of SW, KS, LL, AD, CVM, DP, JB and CSQ test statistics in testing if a random sample of n independent observations come from a population with a normal $N(\mu, \sigma^2)$ distribution. The levels of significance, α considered were 5% and 10%. First, appropriate critical values were obtained for each test for 15 sample sizes n=10(5)100(100)500, 1000, 1500 and 2000. The critical values were obtained based on 50,000 simulated samples from a standard normal distribution. The 50,000 generated test statistics were then ordered to create an empirical distribution. As the SW, SF and RJ are left-tailed test, the critical values are the 100(α)th percentiles of the empirical distributions of these test statistics. The critical values for AD, KS, LL, CVM, DP and JB tests are the $100(1-\alpha)$th percentiles of the empirical distribution of the respective test statistics. Meanwhile, the SK and KU are two-tailed tests so the critical values are the $100(\alpha/2)$th and $100(1-\alpha/2)$th percentiles of the empirical distribution of the test statistics.

For the CSQ test, for a given sample size and the alternative distribution considered, the CSQ statistic was computed for various c (number of categories) that are less than $M = 4(2n^2/z_\alpha^2)^{1/5}$. The power of the CSQ test is then the highest power (proportion of rejected *samples*, i.e p-values less than α) among the c categories.

In our simulation study, 10,000 samples each of size n = 10(5)100(100)500, 1000, 1500 and 2000 are generated from each of the given alternative distributions. The alternative distributions are classified into symmetric short-tailed distributions, symmetric long-tailed distributions and asymmetric distributions. The six symmetric short-tailed distributions were (U(0,1), GLD(0,1, 0.25,0.25), GLD(0,1,0.5,0.5), GLD(0,1,0.75,0.75), GLD(0,1,1.25,1.25) and Trunc(−2,2). The eight symmetric long-tailed distribution include Laplace, logistic, GLD(0,1,−0.10,−0.10), GLD(0,1,−0.15,−0.15), t(10), t(15), ScConN(0.2,9) and ScConN(0.05,9). The 10 asymmetric distributions considered were gamma(4,5), Beta(2,1), Beta(3,2), CSQ(4), CSQ(10), CSQ(20), Weibull(3,1), Lognormal, LoConN(0.2,3) and LoConN(0.05,3). These distributions were selected

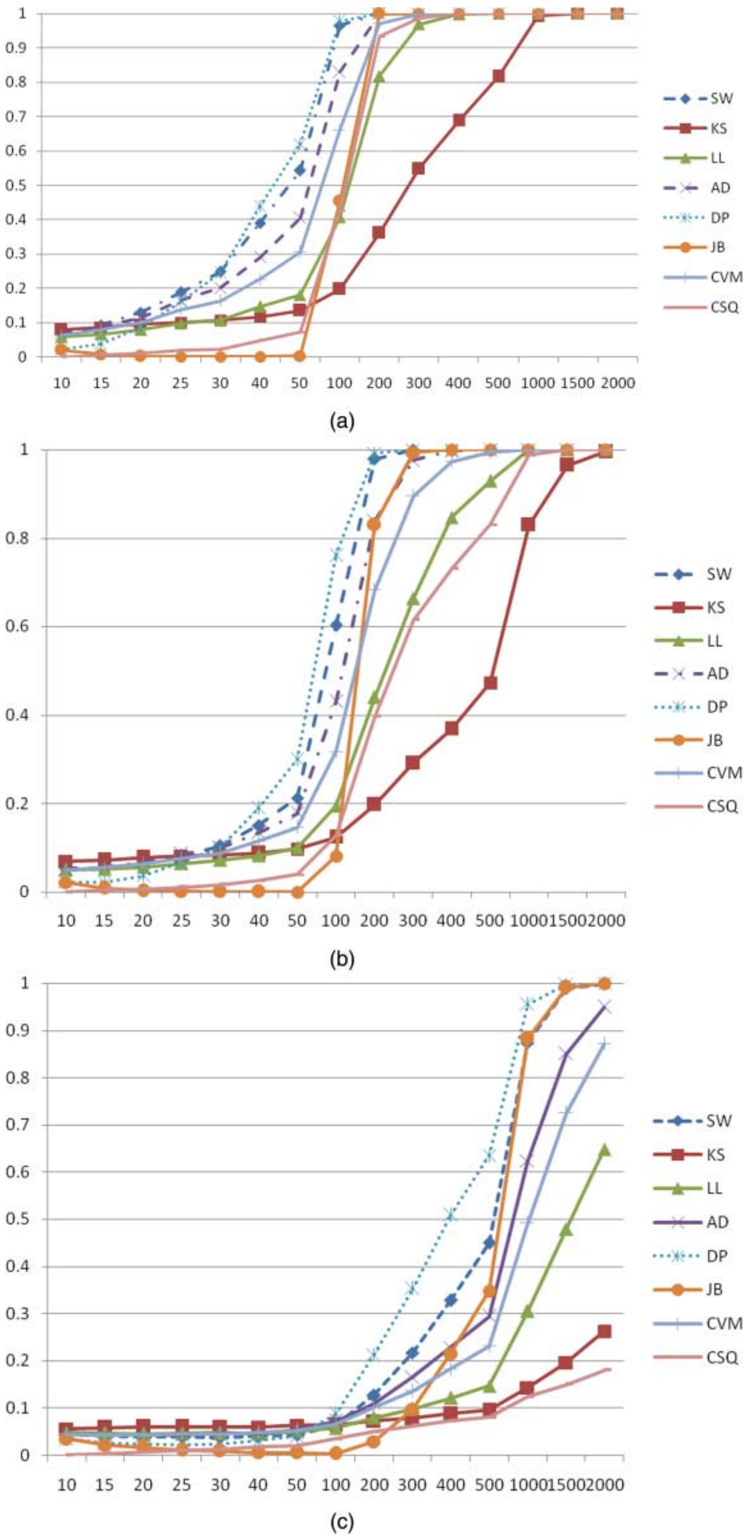


Figure 1. (a) Power comparisons for GLD(0.0,1.0,0.75,0.75) at 5% significance level (skewness = 0, kurtosis = 1.89). (b) Power comparisons for GLD(0.0,1.0,0.5,0.5) at 5% significance level (skewness = 0, kurtosis = 2.08). (c) Power comparisons for GLD(0.0,1.0,0.25,0.25) at 5% significance level (skewness = 0, kurtosis = 2.54).

to cover various standardized skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) values. The scale-contaminated normal distribution, denoted by ScConN(p, b) is a mixture of two normal distribution with probability p from a normal distribution $N(0, b^2)$ and probability $1 - p$ from $N(0, 1)$. The truncated normal distribution is denoted as TruncN(a, b). LoConN(p, a) denotes the distribution of a random variable that is sampled with probability p from a normal distribution with mean a and variance 1 and with probability $1 - p$ from a standard normal distribution.

The generalized lambda distribution GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, originally proposed by Ramberg and Schmeiser [36], is a four-parameter generalization of the two-parameter Tukey's Lambda family of distribution [37] and Karian and Dudewicz [38] has published tables that provide parameter $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ values for some given levels of skewness and kurtosis. The percentile function of GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is given as

$$Q(y) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}, \quad \text{where } 0 \leq y \leq 1, \tag{14}$$

where λ_1 is the location parameter, λ_2 is the scale parameter and λ_3 and λ_4 are the shape parameters that determine the skewness and kurtosis of the distribution. The PDF of GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is given as

$$f(x) = \frac{\lambda_2}{\lambda_3 y^{\lambda_3 - 1} - \lambda_4 (1-y)^{\lambda_4 - 1}} \text{ at } x = Q(y). \tag{15}$$

Karian and Dudewicz [38] provided the following results that give an explicit formulation of the first four centralized GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ moments. If X is GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ with $\lambda_3 > -1/4$

Table 2. Simulated power of normality tests for some symmetric short-tailed distributions ($\alpha = 0.05$).

	n	SW	KS	LL	AD	DP	JB	CVM	CSQ
$U(0.1)$:	10	0.0920	0.0858	0.0671	0.0847	0.0226	0.0199	0.0783	0.0001
$\sqrt{\beta_1} = 0, \beta_2 = 1.8$	20	0.2014	0.1074	0.1009	0.1708	0.1335	0.0025	0.1371	0.0160
	30	0.3858	0.1239	0.1445	0.3022	0.3636	0.0009	0.2330	0.0303
	50	0.7447	0.1618	0.2579	0.5817	0.7800	0.0127	0.4390	0.1164
	100	0.9970	0.2562	0.5797	0.9523	0.9965	0.7272	0.8411	0.7015
	300	1.0000	0.7045	0.9974	1.0000	1.0000	1.0000	1.0000	0.9999
	500	1.0000	0.9331	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	0.9996	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Tukey(0,1,1.25,1.25):	10	0.1057	0.0898	0.0733	0.0951	0.0241	0.02	0.0871	0.0001
$\sqrt{\beta_1} = 0, \beta_2 = 1.76$	15	0.1642	0.1004	0.0918	0.1397	0.0707	0.0077	0.121	0.0092
	20	0.2483	0.1137	0.1137	0.2054	0.1607	0.0025	0.1639	0.0185
	30	0.4585	0.1342	0.1687	0.3591	0.4238	0.001	0.2738	0.0364
	50	0.8241	0.1768	0.3015	0.6652	0.8414	0.0227	0.5162	0.1435
	100	0.9993	0.2914	0.6681	0.9790	0.9982	0.8254	0.9010	0.8136
	300	1.0000	0.7771	0.9993	1.0000	1.0000	1.0000	1.0000	1.0000
	500	1.0000	0.9624	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
TRUNC(−2,2):	10	0.0375	0.0485	0.041	0.0391	0.0249	0.0249	0.0407	0
$\sqrt{\beta_1} = 0, \beta_2 = 2.36$	20	0.0338	0.0492	0.0423	0.0392	0.0187	0.0082	0.0379	0.0065
	30	0.041	0.05	0.0449	0.0455	0.0246	0.0027	0.0442	0.0114
	50	0.0494	0.0474	0.0445	0.055	0.0608	0.0007	0.0526	0.0215
	100	0.1261	0.0484	0.062	0.0953	0.196	0.0033	0.0761	0.0346
	300	0.7884	0.0619	0.126	0.3297	0.819	0.4198	0.2137	0.1436
	500	0.9962	0.0814	0.2128	0.6284	0.9855	0.8961	0.3918	0.271
	1000	1	0.1325	0.5046	0.978	1	0.9999	0.7928	0.5226
	2000	1	0.3793	0.9081	1	1	1	0.9966	0.8356

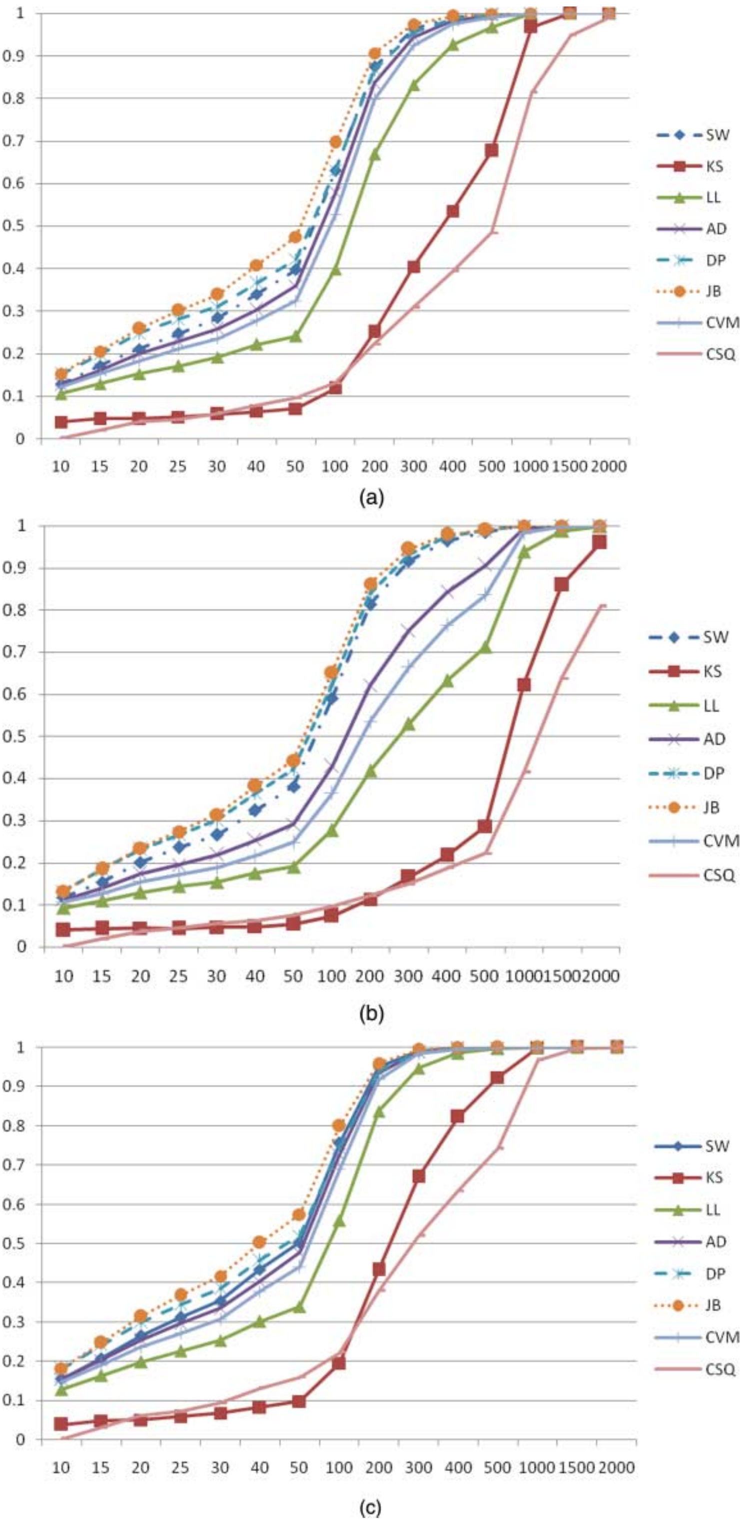


Figure 2. (a) Power comparisons for GLD(0.0,1.0,−0.10,−0.10) at 5% significance level (skewness = 0, kurtosis = 6.78). (b) Power comparisons for ScConN(0.05,3) at 5% significance level (skewness = 0, kurtosis = 7.65). (c) Power comparisons for GLD(0.0,1.0,−0.15,−0.15) at 5% significance level (skewness = 0, kurtosis = 10.36).

and $\lambda_4 > -1/4$ then its first four centralized GLD($\lambda_1, \lambda_2, \lambda_3, \lambda_4$) moments $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ (mean, variance, skewness, kurtosis) are given by

$$\alpha_1 = \mu = \lambda_1 + \frac{A}{\lambda_2}, \quad \alpha_2 = \sigma^2 = \lambda_1 + \frac{B - A^2}{\lambda_2^2}, \quad \alpha_3 = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3}$$

and

$$\alpha_4 = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4},$$

where

$$\begin{aligned}
A &= \frac{1}{1+\lambda_3} - \frac{1}{1+\lambda_4}\\
B &= \frac{1}{1+2\lambda_3} - \frac{1}{1+2\lambda_4} - 2\beta(1+\lambda_3, 1+\lambda_4)\\
C &= \frac{1}{1+3\lambda_3} - \frac{1}{1+3\lambda_4} - 3\beta(1+2\lambda_3, 1+\lambda_4) + 3\beta(1+\lambda_3, 1+2\lambda_4)\\
D &= \frac{1}{1+4\lambda_3} - \frac{1}{1+4\lambda_4} - 4\beta(1+3\lambda_3, 1+\lambda_4)\\
&\quad + 6\beta(1+2\lambda_3, 1+2\lambda_4) - 4\beta(1+\lambda_3, 1+3\lambda_4)\\
\beta(a,b) &= \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x.
\end{aligned}$$

Table 3. Simulated power of normality tests for symmetric long-tailed distribution ($\alpha = 0.05$).

	n	SW	KS	LL	AD	DP	JB	CVM	CSQ
$t(15)$:	10	0.0615	0.0469	0.0566	0.0607	0.0698	0.0699	0.0607	0.0001
$\sqrt{\beta_1} = 0, \beta_2 = 3.55$	20	0.0762	0.0453	0.0614	0.071	0.0906	0.0939	0.0679	0.0095
	30	0.0888	0.0496	0.0641	0.0748	0.101	0.1089	0.0692	0.0184
	50	0.1023	0.0459	0.0597	0.0787	0.1243	0.1421	0.0721	0.0215
	100	0.1412	0.046	0.0743	0.1044	0.1627	0.1942	0.0893	0.032
	300	0.2643	0.0574	0.1021	0.1644	0.2941	0.355	0.1368	0.0458
	500	0.3733	0.0703	0.1313	0.2244	0.4131	0.481	0.1832	0.0535
	1000	0.5977	0.0895	0.2227	0.4016	0.6521	0.7076	0.3331	0.0684
	2000	0.8677	0.1534	0.4178	0.6970	0.903	0.9240	0.6093	0.0993
Logistic:	10	0.0853	0.0428	0.0748	0.0822	0.1021	0.1006	0.0800	0.0002
$\sqrt{\beta_1} = 0, \beta_2 = 4.2$	20	0.1204	0.0458	0.0864	0.1054	0.1463	0.154	0.0978	0.0147
	30	0.1535	0.0493	0.1012	0.1295	0.1762	0.1923	0.1162	0.0252
	50	0.1930	0.0519	0.1083	0.1611	0.2224	0.2604	0.1403	0.0365
	100	0.2971	0.0589	0.1527	0.2367	0.3292	0.3866	0.2023	0.0460
	300	0.6355	0.1193	0.3384	0.5340	0.6479	0.7278	0.4726	0.0898
	500	0.8327	0.1861	0.5215	0.7567	0.8441	0.8870	0.6928	0.1305
	1000	0.9854	0.3745	0.8351	0.9652	0.9863	0.9920	0.9414	0.2359
	2000	1.0000	0.7388	0.9912	0.9998	1.0000	1.0000	0.9992	0.4571
Laplace:	10	0.1513	0.0392	0.1386	0.155	0.1754	0.1754	0.1533	0.0008
$\sqrt{\beta_1} = 0, \beta_2 = 6.0$	20	0.2675	0.0593	0.2286	0.2822	0.2851	0.3049	0.2735	0.0762
	30	0.3593	0.0831	0.298	0.3738	0.357	0.3955	0.3658	0.1054
	50	0.5193	0.1322	0.4243	0.5499	0.4862	0.553	0.5419	0.1862
	100	0.7966	0.2635	0.7027	0.8267	0.7127	0.7893	0.8224	0.3332
	300	0.9975	0.7934	0.9929	0.9991	0.9883	0.9946	0.9991	0.7428
	500	1.0000	0.9724	1.0000	1.0000	0.9996	0.9998	1.0000	0.9245
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

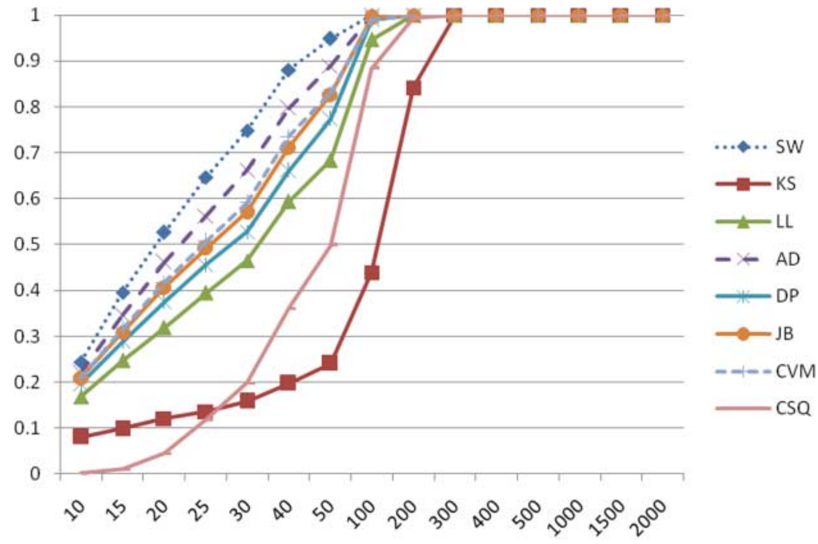
The procedure for generating a sample of size n taken from the GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ distribution is as follows:

Step 1: Set the $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ values.
Step 2: Generate $U_1, U_2, \ldots, U_n$ from the uniform distribution, $U(0, 1)$.
Step 3: Generate

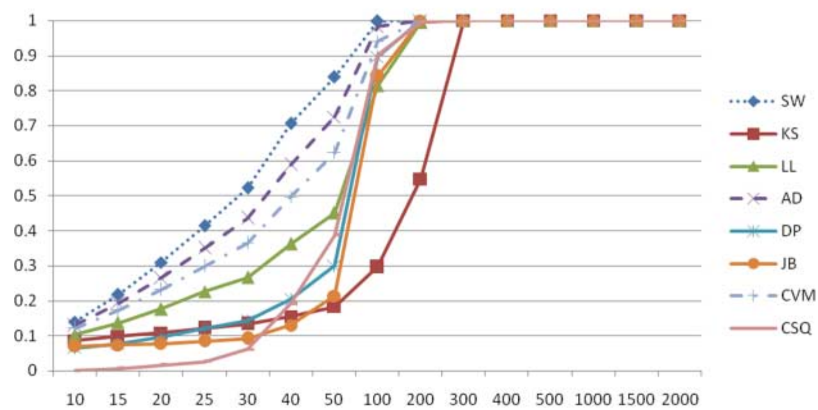
$$x_j = Q(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2}, \quad j = 1, 2, \ldots, n.$$

4. Discussion of results

This section discusses the results of the power of the normality tests for each of the three groups of distributions for $\alpha = 0.05$. The performance of the eight normality test statistics discussed in Section 2 in testing the goodness of fit of a normal distribution under selected alternative



(a)



(b)

Figure 3. (a) Power comparisons for CHI(4df) at 5% significance level (skewness = 1.41, kurtosis = 6.00). (b) Power comparisons for BETA(2,1) at 5% significance level. (skewness = −0.57, kurtosis = 2.40).

Table 4. Simulated power of normality tests for asymmetric distributions ($\alpha = 0.05$).

	n	SW	KS	LL	AD	DP	JB	CVM	CSQ
Weibull (3.1):	10	0.0493	0.0505	0.0493	0.0437	0.0434	0.0482	0.0442	0.0493
$\sqrt{\beta_1} = 0.17, \beta_2 = 2.73$	20	0.0475	0.0480	0.0465	0.0387	0.0354	0.0464	0.0400	0.0475
	30	0.0485	0.0535	0.0512	0.0369	0.0307	0.0511	0.0378	0.0485
	50	0.0546	0.0568	0.0582	0.0439	0.0321	0.0571	0.0412	0.0546
	100	0.0838	0.0653	0.077	0.0612	0.0357	0.0708	0.0549	0.0838
	300	0.2637	0.115	0.1595	0.1645	0.1263	0.1307	0.1664	0.2637
	500	0.513	0.1698	0.2781	0.2964	0.3343	0.2153	0.3609	0.513
	1000	0.9279	0.3298	0.5882	0.5908	0.8177	0.4407	0.867	0.9279
	2000	0.9999	0.6551	0.9348	0.912	0.9976	0.8121	0.9993	0.9999
Lognormal:	10	0.1461	0.0641	0.1092	0.1344	0.1361	0.1404	0.1279	0.0005
$\sqrt{\beta_1} = 1.07, \beta_2 = 5.10$	20	0.2944	0.0876	0.1886	0.2519	0.2578	0.2714	0.2276	0.0239
	30	0.4335	0.106	0.2646	0.3672	0.3561	0.379	0.3274	0.0689
	50	0.6577	0.1447	0.3862	0.5608	0.5464	0.5853	0.5025	0.1432
	100	0.9297	0.2372	0.6787	0.8586	0.8531	0.8798	0.8065	0.3148
	300	0.9999	0.622	0.9912	0.9998	0.9999	0.9999	0.9991	0.7858
	500	1.0000	0.8624	1.0000	1.0000	1.0000	1.0000	1.0000	0.9512
	1000	1.0000	0.9992	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
LoConN(0.2,3):	10	0.1381	0.0714	0.1188	0.1365	0.1111	0.1157	0.1329	0.0000
$\sqrt{\beta_1} = 0.68, \beta_2 = 3.09$	20	0.2558	0.0965	0.217	0.2634	0.1448	0.1615	0.2559	0.0279
	30	0.3874	0.1275	0.3183	0.4082	0.181	0.2093	0.3886	0.0832
	50	0.6149	0.189	0.5134	0.6502	0.2957	0.3651	0.6248	0.1966
	100	0.9285	0.3382	0.8463	0.9425	0.681	0.7476	0.9294	0.4663
	300	0.9999	0.8235	0.9995	1.0000	0.9994	0.9997	1.0000	0.9129
	500	1.0000	0.9789	1.0000	1.0000	1.0000	1.0000	1.0000	0.9885
	1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

symmetric short-tailed distributions is presented in Figure 1 and Table 2. Examining the results in Figure 1 and Table 2 reveal that DP and SW have better power compared with the other tests. The results also show that KS and CSQ tests perform poorly. Figure 2 and Table 3 reveal that for a symmetric long-tailed distribution, the power of JB and DP is quite comparable with the SW test. Simulated powers of the eight normality tests against selected asymmetric distributions are given in Figure 3 and Table 4. For asymmetric distributions, the SW test is the most powerful test followed by the AD test.

Similar patterns of results were obtained for $\alpha = 0.10$ and thus need not be repeated. Results of this simulation study support the findings of D'Agostino *et al*. [13] that KS and CSQ tests for normality have poor power properties. This study also support the findings by Oztuna *et al*. [6] that for a non-normal distribution, the SW test is the most powerful test. This study also shows that KS, modified KS (or Lillliefors test) and AD tests do not outperform the SW test.

5. Conclusion and recommendations

In conclusion, descriptive and graphical information supplemented with formal normality tests can aid in making the right conclusion about the distribution of a variable. Results of this simulation study indicated that the SW test has good power properties over a wide range of asymmetric distributions. If the researcher suspects that the distribution is asymmetric (i.e. skewed) then the SW test is the best test followed closely by the AD test. If the distribution is symmetric with low kurtosis values (i.e. symmetric short-tailed distribution), then the D'Agostino and SW tests have good power. For symmetric distribution with high sample kurtosis (symmetric long-tailed), the researcher can use the JB, SW or AD test. Work is in progress to compare the performance of goodness-of-fit test based on EDF and tests based on spacings such as Rao's spacing test [39,40].

References

[1] R.B. D'Agostino and M.A. Stephens, *Goodness-of-fit Techniques*, Marcel Dekker, NewYork, 1986.

[2] S.S. Shapiro, M.B. Wilk, and H.J. Chen, *A comparative study of various test for normality*, J. Amer. Statist. Assoc. 63(324) (1968), pp. 1343–1372.

[3] E.S. Pearson, R.B. D'Agostino, and K.O. Bowman, *Test for departure from normality: Comparison of powers*, Biometrika 64(2) (1977), pp. 231–246.

[4] W.K. Wong and C.H. Sim, *Goodness-of-fit based on empirical characteristic function*, J. Statist. Comput. Simul. 65 (2000), pp. 243–269.

[5] S. Keskin, *Comparison of several univariate normality tests regarding Type I error rate and power of the test in simulation based on small samples*, J. Appl. Sci. Res. 2(5) (2006), pp. 296–300.

[6] D. Oztuna, A.H. Elhan, and E. Tuccar, *Investigation of four different normality tests in terms of Type I error rate and power under different distributions*, Turk. J. Med. Sci. 36(3) (2006), pp. 171–176.

[7] P.J. Farrell and K. Rogers-Stewart, *Comprehensive study of tests for normality and symmetry: Extending the Spiegelhalter test*, J. Statist. Comput. Simul. 76(9) (2006), pp. 803–816.

[8] T. Thadewald and H. Buning, *Jarque–Bera test and its competitors for testing normality – a power comparison*, J. Appl. Statist. 34(1) (2007), pp. 87–105.

[9] B. Yazici and S. Yolacan, *A comparison of various tests of normality*, J. Statist. Comput. Simul. 77(2) (2007), pp. 175–183.

[10] S.S. Shapiro and M.B. Wilk, *An analysis of variance test for normality*(*complete samples*), Biometrika 52 (1965), pp. 591–611.

[11] J.P. Royston, *Remark AS R94: A remark on Algorithm AS 181: The W-test for normality*, Appl. Statist. 44 (1995), pp. 547–551.

[12] G.E. Dallal and L. Wilkinson, *An analytic approximation to the distribution of the Lilliefor's test statistic for normality*, Amer. Statist. 40 (1986), pp. 294–296.

[13] R.B. D'Agostino, A. Belanger, and R.B. D'Agostino Jr, *A suggestion for using powerful and informative tests of normality*, Amer. Statist. 44(4) (1990), pp. 316–321.

[14] R.B. D'Agostino and E.S. Pearson, *Testing for departures from normality. I. Fuller empirical results for the distribution of* b_2 and $\sqrt{b_1}$, Biometrika 60 (1973), pp. 613–622.

[15] J.P. Royston, *An extension of Shapiro and Wilk's W test for normality to large samples*, Appl. Statist. 31 (1982), pp. 115–124.

[16] J.P. Royston, *The W test for normality*, Appl. Stat. 31 (1982), pp. 176–180.

[17] J.P. Royston, *Expected normal order statistics* (*exact and approximate*), Appl. Statist. 31 (1982), pp. 161–165.

[18] J.P. Royston, *Approximating the Shapiro–Wilk W test for non-normality*, Stat. Comput. 2 (1992), pp. 117–119.

[19] S.S. Shapiro and R.S. Francia, *An approximate analysis of variance test for normality*, J. Amer. Statist. Assoc. 67(337) (1972), pp. 215–216.

[20] S. Weisberg and C. Bingham, *An approximate analysis of variance test for non-normality suitable for machine computation*, Technometrics 17 (1975), pp. 133–134.

[21] M.M. Rahman and Z. Govindarajulu, *A modification of the test of Shapiro and Wilk for normality*, J. Appl. Statist. 14(2) (1997), pp. 219–235.

[22] A.N. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione*, Giornale dell' Instituto Italiano degli Attuari 4 (1933), pp. 83–91.

[23] H. Cramer, *On the composition of elementary errors*, Skandinavisk Aktuarietidskrift 11 (1928), pp. 13–74, 141–180.

[24] T.W. Anderson and D.A. Darling, *A test of goodness of fit*, J. Amer. Statist. Assoc. 49(268) (1954), pp. 765–769.

[25] H.W. Lilliefors, *On the Kolmogorov–Smirnov test for normality with mean and variance unknown*, J. Amer. Statist. Assoc. 62 (1967), pp. 534–544.

[26] H.W. Lilliefors, *On the Kolmogorov–Smirnov test for the exponential distribution with mean unknown*, J. Amer. Statist. Assoc. 64 (1969), pp. 387–389.

[27] W.J. Conover, *Practical Nonparametric Statistics*, 3rd ed., John Wiley and Sons, New York, 1999.

[28] R. von Mises, *Wahrscheinlichkeitsrechnung und Ihre Anwendung in der Statistik und Theoretischen Physik*, F. Deuticke, Leipzig, Vol. 6.1, 1931.

[29] N.V. Smirnov, *Sui la distribution de* w^2 (*Criterium de M.R.v. Mises*), C.R. (Paris), 202 (1936), pp. 449–452 (6.1).

[30] A.W. Marshall, *The small sample distribution of* nw_n^2, Ann. Math. Statist. 29 (1958), pp. 307–309.

[31] M.A. Stephens and U.R. Maag, *Further percentage points for* W_N^2, Biometrika 55(2) (1968), pp. 428–430.

[32] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling*, Philos. Mag. 50(5) (1900), pp. 157–175.

[33] B. Schorr, *On the choice of the class intervals in the application of chi-square test*, Oper. Forsch. U. Stat. 5 (1974), pp. 357–377.

[34] C.M. Jarque and A.K. Bera, *A test for normality of observations and regression residuals*, Int. Stat. Rev. 55(2) (1987), pp. 163–172.

[35] K.O. Bowman and L.R. Shenton, *Omnibus test contours for departures from normality based on* $\sqrt{b_1}$ *and* b_2, Biometrika 62(2) (1975), pp. 243–250.
[36] J.S. Ramberg and B.W. Schmeiser, *An approximate method for generating asymmetric random variables*, Commun. ACM 17 (1974), pp. 78–82.
[37] C. Hastings, F. Mosteller, J.W. Tukey, and C.P. Winsor, *Low moments for small samples: A comparative study of statistics*, Ann. Math. Statist. 18 (1947), pp. 113–136.
[38] Z.A. Karian and E.J. Dudewicz,*Fitting Statistical Distributions: The Generalized Lambda Distribution and the Generalized Bootstrap Methods*, CRC Press, New York, 2000.
[39] J.S. Rao, *Some tests based on arc-lengths for the circle*, Sankhya B(4), 38,(1976), pp. 329–338.
[40] J.S. Rao and M. Kuo, *Asymptotic results on the Greenwood statistic and some of its generalizations*, J. R. Statist. Soc. Ser. B 46 (1984), pp. 228–237.