Power Comparison of Various Normality Tests

Fiaz Ahmad
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore, Pakistan
fiazcsaspu@gmail.com

Rehan Ahmad Khan Sherwani
College of Statistical and Actuarial Sciences
University of the Punjab, Lahore, Pakistan
stat4u@gmail.com

Abstract

Standard statistical procedures often require data to be normally distributed and the results of these methods will be inappropriate when the assumption of normality is not satisfied. Therefore, the postulation of normality is strictly required before proceeding statistical analysis. Although a number of criteria's have been available to assess the assumption of normality, but these tests do not have the same nature and power to diagnose the departures of a data set from normality, thus the choice of appropriate test always remains a key importance in the assessment of normality assumption. In the present study, power comparison of twelve standard normality tests was examined using simulated data generated from four distributions; Cauchy, Exponential, Weibull and Logistic under different sample sizes by using R codes. Results showed that under logistic distribution data, Geary test was observed most powerful test at the 5 % level of significance and Jarque Bera test at the 1 % level of significance. Under alternate Cauchy distribution, Shapiro Francia test performs well at the 5 % level of significance while at the 1 % level of significance, Shapiro Francia, Anderson Darling, Cramer von mises and Watson tests equally observed the power of a test. Shapiro Wilk test was highlighted as a more powerful test for data generated under Weibull distribution.

Keywords: Normality Tests, Power of Test.

1. Introduction

Many parametric methods (like correlation, regression, t – test, analysis of variance etc) require normality assumption. The assumption of normality is one of the most important assumptions of parametric procedures because of its extensive range of practical applications. In that respect is perhaps no distribution which is stated to be totally normal. The question is usually asked; however, whether or not the population from which dataset has been drawn can be adequately modeled with the normal distribution for the intended purpose.

Different procedures are available for testing the assumption of normality. The assumption of normality can be tested subjectively, like drawing box plot and Histogram etc. Appearing at a histogram and making the statement, "It looks normal to me," does not offer conclusive evidence that the normal assumption holds. A leptokurtic distribution looks symmetric and bell-shaped, but it is not normal in actuality. Many applications have underlying process distributions that are not normally distributed. The process may be better modeled with an Exponential, Weibull, Logistic, Cauchy and some other distribution. Power of test for normality is the ability of a test to detect the presence of

non-normality. Each of the tests for normality also varies in their ability to detect different types of departures from normality. Powerful test with large sample sizes will reject the normality assumption with only slight deviation from normality. This rejection may or may not answer the question of whether or not the normal approximation is adequate.

Researchers developed many tests for the comparison of normality assumption in different years, fathers' some tests were modified for attaining better performance. Firstly, Pearson (1900) developed chi-square test for detection of non-normality. Kolmogrov and Smirnov (1933) suggested formal test for normality. The chi square test, based on cumulative distribution function, and can be used for any univariant distribution. In certain situations, chi-square test not performs well, therefore the Kolmogorov Smirnov test, known as (KS test) used an alternative to Chi- square test. After two decades, Anderson and Darling proposed their test for normality. Kuiper (1960) brought out the test of normality. Afterwards, Shapiro and Wilk (1965) suggested test of normality. In (1968), Ajne, normality test was developed. After two years later, a modification of Kuiper and Ajne tests was proposed by Stephens (1970). D'Agostino (1972) introduced another test of normality. In the same year, modification of kolmogrov Smirnov was proposed by Stephens. Four years later, Vasicek's test of normality proposed by Vasicek (1976). Jarque and Bera designed test of normality in 1987.

Noughabi et al., (2014) introduced a goodness of fit test for Rayleigh distribution which based on Kullback Leibler discrimination methodology. They assess the accuracy of the proposed test using a simulation study; critical values and power for different tests was calculated. The simulation study indicates that the power of the tests affected by sample size and types of the alternate distribution. For a uniform alternative, the proposed test done well than the other tests. For all other substitutes, the Anderson Darling test had the greatest power. The proposed test obtained the maximum power at $m = 3$, $n = 10$ and $m = 4$, $n = 20$ With increasing n, the optimal choice of m increased. The use of the proposed test is revealed in a real example.

Ibrahim Al-Omari (2014) proposed three new entropy estimators which are the modification of the estimators propounded by Noughabi & Arghami (2011) and Ebrahimi, Pflughoeft & Soofi (1994). The new estimator of the continuous random variable is introduced by using ranked set sampling (RSS), double ranked set sampling (DRSS) and simple random sampling (SRS) techniques. The proposed estimators were compared with the estimator introduced by Vasicek's (1976). A numerical comparison was regarded in terms of biased values and root mean squared error (RMSE). The result reveals that the new estimators done well than Vasicek estimator and has a smaller root mean squared error. The results also reveal that the proposed estimator which based on the double ranked set sampling technique is more effective than ranked set sampling and simple random sampling.

Kohansal & Rezakhah (2014) introduced two new entropy estimators for the absolutely continuous random variable. Some properties of the new estimates are favored. Especially consistency of the first is proved. Moreover, two new tests of normality are

introduced which based on the new estimators. The power of these tests for normality was compared with the existing tests of normality. The results demonstrate that the new tests of normality were more reliable than the existing normality tests. Also the new estimator was compared with the existing estimator. By the simulation study, they confirmed that the proposed estimators and test statistics perform very well in estimating entropy and testing normality. It is also confirmed that the powers of the tests based on alternate distributions and sample size.

Chen (2014) suggested a test statistic that based on order statistics. The main principle of this document was to bring the goodness-of-fit test statistic for censored sample data framed by order statistic. Double, left and right censoring was used, which is the extended form of the work Chen & Ye (2009). Havva Alizadeh Noughabi & Noughabi (2013) proposed new estimator for continuous random variable attained by modify the estimator proposed by Ebrahimi, Pflughoeft, & Soofi (1994). The modified estimator was more viable then the parent test with respect to mean square error. From real data example, it was confirmed that the proposed estimator perform well. Ramos & Burgos (2013) compared the power of seven normality tests using alternate Ex-Gaussian distribution. Among these seven, the SW, CS and C test perform better taking Ex-Gaussian alternate distribution. They believed that Ex-Gaussian distribution is a suitable distribution to accommodate positively skewed distributions. A pre-assessment of the proposed tests by power across different types of distributions was performed via a meta-analysis. Zhao & Xu (2012) deduced some new goodness-of-fit tests based on the empirical distribution function (EDF) proposed by Rubin's. After introducing Rubin's EDF, it is substituting in Kolmogorov-Smirnov, Anderson-Darling and Cramer von Mises statistics in the place of classical EDF. The Rubin's EDF was based on randomized statistic, which drawn from the expectation and quantile as a test statistic. The proposed tests were persistent under straight forward hypothesis. Various comparisons of power was also performed to indicate that the modern tests are often additional authentic than the traditional ones.

Zamanzade & Arghami (2012) propounded new estimator which based on the absolutely continuous random variable. The performance of the estimator was confirmed by a simulation study taking various distributions as alternate, and found that the new estimator performed well in testing of normality and estimating entropy. They indicate that for various alternate distributions, various tests based on the different entropy estimator performed well. A study was conducted by Razali & Wah (2011), considering four formal normality tests. The power of all tests also assesses by comparing the critical value with each test statistic. Shapiro Wilk test was more reliable and gives better results than the other tests. However, for small sample sizes all tests give low power. Yap & Sim (2011) divided the tests of normality into different categories which are chi squared tests, empirical distribution function tests, moments tests, spacing's tests, regression and correlation tests and other special tests. They compare the power of eight different tests of normality via Monte Carlo simulation for various alternate distributions which are symmetric short-tailed, asymmetric and symmetric long-tailed distributions. They showed that Shapiro Wilk and D'Agostino perform well for alternate symmetric short-tailed distributions. D'Agostino and Jarque Bera test gives better results than Shapiro

Wilk test for symmetric long-tailed distributions and for asymmetric alternate distributions, Shapiro-Wilk test perform well followed by the Anderson-Darling test.

The research conducted by Hadi Alizadeh Noughabi & Arghami (2011) for comparison of power of seven normality tests. The Jarque Bera and Anderson Darling test, perform better if the presumed alternative distributions supported by (-∞, ∞) and are symmetric. The vasicek test of sample entropy estimator can be most powerful if the presumed alternatives distributions supported by (0, 1). Shapiro-Wilk and Vasicek test are most powerful against alternative with the support (0, ∞). If the assumed alternatives distributions support by (-∞, ∞) and are asymmetric than Shapiro-Wilk test is more powerful. Quessy & Mailhot (2011) explained the asymptotic power of seven normality tests. The procedures which consider in that study are based on either the empirical kurtosis or skewness. They compare the power of tests; Cramer-von Mises, Jarque Bera test, Anderson-Darling and Kolmogorov-Smirnov test. He compared the asymptotic relative effectiveness and local power curves of these tests in the light of a measure viewed by Berg & Quessy (2009). Four classes of local alternative distributions are considered in the present study that includes; Thadewald & Buning (2007) mixture of kurtosis and skewness alternatives, bimodal alternatives, heavy tailed alternatives and kurtosis alternatives.

2. Methodology

Twelve normality tests; Kolmogrove Smirnov, Lillifors, Kuiper, Cramer von mises, Anderson Darling, Shapiro Francia, Watson, Chi Square, Jarque Bera, Geary, D,Agostino D and Shapiro Wilk test are under investigation in this study. The description of each test is given in the next sections. In this study, compare the power of these normality tests by using different alternative distributions (Cauchy, Exponential, Logistic and Weibull) at different sample sizes (n=10, n=20, n=30, n=40, n=50, n=100, n=200 and 500).

2.1 Kolmogrov-Smirnov test

This is a well-known test of empirical distribution function (EDF). It is based on the distribution which is completely specified with known parameters. This test proposed by Kolmogrov and Smirnov (1933) and defined as

$$D^+ = \max\left\{\left(\frac{i}{n}\right) - zi\right\}, 1 \leq i \leq n;$$

$$D^- = \max\left\{zi - \frac{i-1}{n}\right\}, 1 \leq i \leq n;$$

$$D = \max(D^+, D^-)$$

where z_i is the cumulative probability of standard normal distribution and D_i is the difference between observed and expected values.

2.2 Lilliefors test

This test introduced by Lilliefors (1967) and it is the modification of the Kolmogrov-Smirnov test. Lilliefors test is suiTable when the distribution is not completely specified. When the original KS statistic is used in such situation, the probability of type 1 error tend to be smaller than the once given in the standard Table of KS (Lilliefors, 1967). Therefore in this situation, the Lilliefors test will be preferred over the KS test (Öztuna, Elhan et al. 2006).

2.3 Kuiper Test

Kolmogorov–Smirnov test and Kuiper test are closely related to each other. Kuiper test gives most positive D+ and negative D− differences between two cumulative distribution functions being compared.

This test was introduced by kuiper (1960), the test statistic is defined as under,

$$KUI = D^+ + D^-$$

$$\text{Where } D^+ = \sup_x |F_n(x) - F_0(x)|, \ D^- = \sup_x |F_0(x) - F_n(x)|$$

2.4 Cramer von Mises test

Conover (1999) said that this proposed by Cramer, Von Mises and Smirnov. This test is a goodness of fit test of a hypothesized distribution function compared with the empirical distribution function. The test statistic is defined as

$$CM = \frac{1}{12n} + \sum_{i=1}^{n} (Z_i - \frac{2i-1}{2n})^2$$

Where z_i is the cumulative distribution function of the specified distribution and $z_i = (X_{(i)} - \overline{X})/S$ and $\overline{X}$ and S is the sample mean and sample standard deviation.

2.5   Anderson-Darling test

Anderson Darling test gives more weight to the tails of the distribution than the Cramer von Mises test. Furthermore Cramer von Mises test is a distribution free test and Anderson Darling test makes use of the specific hypothesized distribution. It is proposed by Anderson and Darling (1954) and defined as follows

$$AD = -\frac{\left[\sum_{i=1}^{n} (2i-1)\left\{ \ln z_i + \ln(1 - z_{(n+1-i)}) \right\} \right]}{n} - n$$

2.6   Shapiro-Francia test

For large sample size, the computation of the inverse of the covariance matrix in Shapiro Wilk test is time consuming (Romao et al, 2010). Therefore a modification of the Shapiro-Wilk test proposed by the Shapiro and Francia (1972) and show that the power of this test is almost same as Shapiro and Wilk test but Shapiro Francia test is easier to calculate than the Shapiro Wilk test. The test statistic is defined as

$$W_{SF} = \frac{\left[\sum_{i=1}^{n} m_i x_i \right]^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} m_i^2}$$

The value of this test lies between 0 and 1. The values close to 1indicate the normality of the data and the value close to zero indicates the non-normality of the data.

2.7   Watson Test

This test introduced by Watson (1962), the test statistic is written as:

$$W = \frac{1}{12n} + \sum_{i=1}^{n}(Z_i - \frac{2i-1}{2n})^2 - n\left(\bar{z} - \frac{1}{2}\right)^2$$

Where $\bar{z} = \sum_{i=1}^{n} \frac{z_i}{n}$, $Z_i = \Phi_{?} ((X_{(i)} - \bar{X})/S)$, $\bar{X} = \sum_{i=1}^{n} X_i/n$, $S^2 = (\sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$

and in the null hypothesis $\Phi_{?}(x)$ represent the standardized hypothesized of normal distribution.

2.8 Chi-Square Test

For goodness of fit test it is the most popular and oldest goodness-of-fit test. This test proposed by Pearson (1900). It is generally used for discrete distribution only. The test statistic is described as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

The Chi-Square test statistic follows a Chi-Square distribution with degrees of freedom $n - k - 1$. Null hypothesis is rejected if the value of the test statistic is large.

2.9   Jarque-Bera test

It is the most popular test for the goodness of the fit test in the field of economic. It is mostly known from the proposal of Jarque and Bera (1987). The test statistic is denoted by JB and defined as:

$$JB = \frac{n}{6}\left(b_1 + \frac{(b_2 - 3)^2}{4}\right)$$

where b_1 and b_2 are the sample skewness and kurtosis.

2.10   Geary test

This test is the ratio of the mean deviation to the standard deviation. It is a lower tailed test is used to detect a symmetric long tailed alternative to the normality. This test proposed by Geary (1935) and defined as follows

$$w_n' = a(1) = \frac{\sum_{i=1}^{n}\left|x_i - \bar{x}\right|}{\left(n\sqrt{m_2}\right)}$$

It is the alternate of the kurtosis test. Geary shown that it is a better than Kurtosis.

2.11   D'Agostino D Test

It is the extension of the Shapiro-Wilk test in which we have no need to describe the vector of weights a . This test is proposed by D'Agostino (1971).

$$D = \frac{\sum_{i=1}^{n}\left(i-(n+1)/2\right).x_{(i)}}{n^2.\sqrt{m_2}}$$

Where n is the size of sample and m_2 is the sample variance of order statistics.

2.12   Shapiro-Wilk test

It is a well known and powerful regression test of normality. Proposed by Shapiro and Wilk (1965), and its gives better results for the size of the small sample. Accurracy is claimed for samples size from 3 to 5000. Sample size less than three will not produce a Shapiro-Wilk statistic. The test statistic is defined as

$$W = \frac{K\sigma^2}{(n-1)s^2} = \frac{a'x}{(n-1)s^2} = \frac{\left[\sum_{i=1}^{n}a_i x_{(i)}\right]^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

$$a' = \left(a_1,a_2,......,a_n\right) = m'V^{-1}\left[\left(m'V^{-1}\right)\left(V^{-1}m\right)\right]^{\frac{-1}{2}}$$

$$K = \frac{m'V^{-1}m}{m'V^{-1}V^{-1}m}$$

Where $x_{(i)}$ is the ith order statistic, $\bar{x}$ is the sample mean, $V = \left(v_{ij}\right)$ be the $n \times n$ covariance matrix of these ordered statistic and $m' = (m_1,m_2,........,m_n)$ be the vector of expected values of the standard normal order statistics. The value of this test lies between 0 and 1. The values of 1indicate the normality of the data and the value close to zero indicate the non-normality of the data.

3.   Results from Simulation Study

In this section of the simulation study, 10000 units are generated from normal distribution with different sample sizes n=10, n=20, n=30, n=40, n=50, n=100 n=200 and 500. 5% and 1% level of significance is considered. All tests are applied to the same samples with different sample sizes and the results obtained by simulation are summarized in Tables given below. The actual size of type 1 error rate and power of all the tests is estimated by the frequency of normally distributed populations.

Table 3.1 and 3.6 represents the actual size of the tests for 5% and 1% level of significance respectively. Table 3.2 -3.5 and 3.7-3.10 reveal the power of the normality tests which consider in the present study against selected alternate distributions for 5% and 1% level of significance respectively. Power of these normality tests also compare by graphically.

Type 1 error rate and power of 12 normality tests which included in the present study at 5% and 1% level of significance with different sample size and different alternate distributions are given as under.

Table 3.1:   Size of the test at alpha=0.05

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0498	0.0549	0.0515	0.0524	0.0519	0.0553	0.0515	0.0516
Lillifors	0.0500	0.0517	0.0500	0.0504	0.0516	0.0535	0.0504	0.0500
Kuiper	0.0472	0.0556	0.0501	0.0475	0.0473	0.0489	0.0551	0.0504
Cramer von mises	0.0510	0.0530	0.0484	0.0516	0.0499	0.0491	0.0509	0.0503
Anderson Darling	0.0516	0.0542	0.0503	0.0525	0.0497	0.0502	0.0505	0.0485
Shapiro Francia	0.0546	0.0573	0.0505	0.0534	0.0525	0.0540	0.0492	0.0508
Watson	0.0499	0.0527	0.0511	0.0498	0.0496	0.0490	0.0526	0.0507
Chi Square	0.0674	0.0496	0.0544	0.0575	0.0512	0.0533	0.0512	0.0516
Jarque Bera	0.009	0.0263	0.0313	0.0358	0.0349	0.0459	0.0440	0.0485
Geary	0.024	0.0355	0.0356	0.0408	0.0398	0.0437	0.0484	0.0486
D,Agostino D	0.0554	0.0556	0.0481	0.0502	0.0523	0.0518	0.0460	0.0489
Shapiro Wilk	0.0513	0.0539	0.0511	0.0523	0.0481	0.0496	0.0452	0.0492

Table 3.2:   Power results for different normality tests under alternate Cauchy distribution at (alpha=0.05)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.5763	0.8392	0.9450	0.9815	0.9933	1.0000	1.0000	1.0000
Lillifors	0.5777	0.8387	0.9443	0.9812	0.9933	1.0000	1.0000	1.0000
Kuiper	0.5875	0.8588	0.9562	0.9863	0.9961	1.0000	1.0000	1.0000
Cramer von mises	0.6152	0.8758	0.9622	0.9896	0.9972	1.0000	1.0000	1.0000
Anderson Darling	0.6100	0.8756	0.9649	0.9902	0.9975	1.0000	1.0000	1.0000
Shapiro Francia	0.6382	0.8918	0.9702	0.9916	0.9979	1.0000	1.0000	1.0000
Watson	0.6110	0.8747	0.9621	0.9892	0.9972	1.0000	1.0000	1.0000
Chi Square	0.5333	0.7729	0.9079	0.9625	0.9839	1.0000	1.0000	1.0000
Jarque Bera	0.4250	0.8147	0.9434	0.9814	0.9942	1.0000	1.0000	1.0000
Geary	0.4937	0.8756	0.9711	0.9936	0.9988	1.0000	1.0000	1.0000
D,Agostino D	0.5760	0.7701	0.8504	0.8849	0.9050	0.9516	0.9799	0.9894
Shapiro Wilk	0.5898	0.8629	0.9571	0.9876	0.9968	1.0000	1.0000	1.0000

Table 3.3:   Power results for different normality tests under alternate Exponential distribution at (alpha=0.05)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.3163	0.5729	0.7775	0.9006	0.9636	1.0000	1.0000	1.0000
Lillifors	0.3049	0.5701	0.7778	0.8999	0.9641	1.0000	1.0000	1.0000
Kuiper	0.3676	0.6965	0.8814	0.9642	0.9914	1.0000	1.0000	1.0000
Cramer von mises	0.4056	0.7284	0.8956	0.9663	0.9922	1.0000	1.0000	1.0000
Anderson Darling	0.4287	0.7744	0.9322	0.9843	0.9966	1.0000	1.0000	1.0000
Shapiro Francia	0.4445	0.7956	0.9511	0.9903	0.9985	1.0000	1.0000	1.0000
Watson	0.3826	0.6939	0.8676	0.9470	0.9845	1.0000	1.0000	1.0000
Chi Square	0.3985	0.6553	0.8541	0.9542	0.9843	1.0000	1.0000	1.0000
Jarque Bera	0.1551	0.4781	0.7266	0.8738	0.9544	1.0000	1.0000	1.0000
Geary	0.0884	0.1985	0.2865	0.3567	0.4215	0.6382	0.8518	0.9899
D,Agostino D	0.3746	0.7001	0.8862	0.9595	0.9879	1.0000	1.0000	1.0000
Shapiro Wilk	0.4589	0.8331	0.9683	0.9961	0.9995	1.0000	1.0000	1.0000

Table 3.4: **Power results for different normality tests under alternate Weibull (Scale=2, Shape=3) distribution at (alpha=0.05)**

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0480	0.0496	0.0474	0.0488	0.0533	0.0688	0.0873	0.1709
Lillifors	0.0465	0.0472	0.0493	0.0493	0.0511	0.0694	0.0900	0.1717
Kuiper	0.0491	0.0503	0.0476	0.0497	0.0588	0.0677	0.0879	0.1605
Cramer von mises	0.0462	0.0492	0.0484	0.0473	0.0594	0.0706	0.1000	0.2085
Anderson Darling	0.0430	0.0475	0.0484	0.0470	0.0607	0.0740	0.1123	0.2640
Shapiro Francia	0.0447	0.0389	0.0404	0.0355	0.0413	0.0538	0.1011	0.3805
Watson	0.0466	0.0496	0.0500	0.0496	0.0576	0.0672	0.0964	0.1852
Chi Square	0.0631	0.0460	0.0537	0.0562	0.0583	0.0655	0.0711	0.1059
Jarque Bera	0.0064	0.0151	0.0182	0.017	0.0249	0.029	0.0563	0.3289
Geary	0.0179	0.0197	0.0198	0.0167	0.0157	0.0095	0.0049	0.0018
D,Agostino D	0.0381	0.0374	0.0387	0.0363	0.0443	0.0621	0.1080	0.2937
Shapiro Wilk	0.0447	0.0448	0.0504	0.0498	0.0566	0.0807	0.1589	0.5187

Table 3.5: **Power results for different normality tests under alternate Logistic (Location=0, Scale=1) distribution at (alpha=0.05)**

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0757	0.0772	0.0915	0.1014	0.1108	0.1544	0.2518	0.4976
Lillifors	0.0749	0.0767	0.0943	0.0997	0.1106	0.1594	0.2492	0.5064
Kuiper	0.0755	0.0823	0.1005	0.1138	0.1250	0.2093	0.3281	0.6422

Cramer von mises	0.0810	0.0911	0.1139	0.1288	0.1348	0.2141	0.3514	0.6870
Anderson Darling	0.0798	0.1011	0.1280	0.1450	0.1507	0.2499	0.4035	0.7642
Shapiro Francia	0.0968	0.1432	0.1820	0.2155	0.2448	0.3762	0.5681	0.8735
Watson	0.0807	0.0867	0.1114	0.1261	0.1270	0.2187	0.3656	0.7064
Chi Square	0.0827	0.0640	0.0688	0.0813	0.0777	0.0910	0.1200	0.2163
Jarque Bera	0.0260	0.0946	0.1435	0.1893	0.2238	0.3731	0.5787	0.8885
Geary	0.0484	0.1095	0.1589	0.1987	0.2384	0.3937	0.6127	0.9164
D,Agostino D	0.0991	0.1357	0.1580	0.1781	0.1841	0.2296	0.2584	0.2875
Shapiro Wilk	0.0816	0.1137	0.1475	0.1754	0.1939	0.3078	0.4944	0.8392

Table 3.6: Size of the test at alpha=0.01

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0079	0.0097	0.0111	0.0088	0.0094	0.0122	0.0105	0.0109
Lillifors	0.0106	0.0091	0.011	0.0081	0.0096	0.0117	0.0092	0.0096
Kuiper	0.0099	0.0105	0.0109	0.008	0.0105	0.012	0.0093	0.0116
Cramer von mises	0.0098	0.0091	0.0112	0.0081	0.0118	0.0112	0.0106	0.0121
Anderson Darling	0.0097	0.0099	0.0114	0.0082	0.0122	0.0109	0.0119	0.0110
Shapiro Francia	0.0118	0.0085	0.0111	0.0098	0.0121	0.0105	0.0108	0.0117
Watson	0.0097	0.0096	0.0109	0.0076	0.0114	0.0115	0.0112	0.0112
Chi Square	0.0114	0.0117	0.0094	0.0094	0.0110	0.0114	0.0108	0.0095
Jarque Bera	0.0024	0.0100	0.0149	0.0163	0.0184	0.0184	0.0163	0.0161
Geary	0.0056	0.0081	0.0082	0.0090	0.0112	0.0113	0.0091	0.0128
D,Agostino D	0.0133	0.0096	0.0102	0.0089	0.0114	0.0095	0.0095	0.0105
Shapiro Wilk	0.0120	0.0081	0.0099	0.0090	0.0106	0.0095	0.0101	0.0113

Table 3.7: Power results for different normality tests under alternate Cauchy distribution at (alpha=0.01)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.4366	0.7519	0.8909	0.9576	0.9843	0.9999	1.0000	1.0000
Lillifors	0.4439	0.7483	0.8926	0.9563	0.9841	0.9999	1.0000	1.0000
Kuiper	0.4609	0.7842	0.9261	0.9727	0.9908	0.9999	1.0000	1.0000
Cramer von mises	0.4832	0.8046	0.9342	0.9768	0.9939	0.9999	1.0000	1.0000
Anderson Darling	0.4910	0.8092	0.9364	0.9789	0.9936	1.0000	1.0000	1.0000
Shapiro Francia	0.4912	0.8191	0.9378	0.9801	0.9937	1.0000	1.0000	1.0000
Watson	0.4805	0.8017	0.9357	0.9784	0.9942	0.9999	1.0000	1.0000
Chi Square	0.3633	0.6847	0.8405	0.9286	0.9674	0.9997	1.0000	1.0000
Jarque Bera	0.3382	0.7655	0.9184	0.9704	0.9904	1.0000	1.0000	1.0000

Geary	0.3593	0.7994	0.9449	0.9846	0.9969	1.0000	1.0000	1.0000
D,Agostino D	0.4391	0.6929	0.7854	0.8408	0.8758	0.9417	0.9698	0.9876
Shapiro Wilk	0.4753	0.7912	0.9244	0.9751	0.9925	1.0000	1.0000	1.0000

Table 3.8:   Power results for different normality tests under alternate Exponential distribution at (alpha=0.01)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.1340	0.3128	0.5247	0.7308	0.8520	0.9987	1.0000	1.0000
Lillifors	0.1324	0.3331	0.5406	0.7190	0.8518	0.9987	1.0000	1.0000
Kuiper	0.1863	0.4634	0.7368	0.9089	0.9621	1.0000	1.0000	1.0000
Cramer von mises	0.2008	0.4858	0.743	0.8948	0.9633	0.9998	1.0000	1.0000
Anderson Darling	0.2146	0.5465	0.8016	0.9405	0.9818	1.0000	1.0000	1.0000
Shapiro Francia	0.2107	0.5762	0.8289	0.9508	0.9877	1.0000	1.0000	1.0000
Watson	0.1907	0.4457	0.7039	0.8637	0.9417	0.9998	1.0000	1.0000
Chi Square	0.1265	0.4151	0.6517	0.7872	0.8691	0.9977	1.0000	1.0000
Jarque Bera	0.0841	0.3572	0.5819	0.7586	0.8747	0.9993	1.0000	1.0000
Geary	0.0358	0.1102	0.1764	0.2343	0.2829	0.4828	0.7419	0.9756
D,Agostino D	0.1783	0.4572	0.6857	0.8537	0.9357	0.9999	1.0000	1.0000
Shapiro Wilk	0.2346	0.6292	0.8733	0.9715	0.9942	1.0000	1.0000	1.0000

Table 3.9:   Power results for different normality tests under alternate Weibull (Scale=2, Shape=3) distribution at (alpha=0.01)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0087	0.0096	0.0086	0.0087	0.0131	0.0152	0.0194	0.0468
Lillifors	0.0093	0.0091	0.0077	0.0092	0.0105	0.0149	0.0212	0.0483
Kuiper	0.0101	0.0105	0.0092	0.011	0.0119	0.0135	0.0181	0.0462
Cramer von mises	0.0098	0.0084	0.0085	0.0102	0.0133	0.0159	0.0246	0.0672
Anderson Darling	0.0089	0.0089	0.0086	0.0102	0.0143	0.0161	0.0295	0.0996
Shapiro Francia	0.0085	0.0059	0.0061	0.0068	0.0072	0.0093	0.0190	0.1253
Watson	0.0095	0.0098	0.0080	0.0102	0.0132	0.0158	0.0236	0.0570
Chi Square	0.0134	0.0138	0.0110	0.0102	0.0125	0.0159	0.0167	0.0281
Jarque Bera	0.0018	0.0053	0.0078	0.0090	0.0106	0.0105	0.0128	0.0660
Geary	0.0029	0.0044	0.0039	0.0028	0.0031	0.0020	0.0011	-
D,Agostino D	0.0087	0.0050	0.0065	0.0068	0.0072	0.0107	0.0245	0.0974
Shapiro Wilk	0.0082	0.0072	0.0071	0.0092	0.0091	0.0157	0.0369	0.2139

Table 3.10:   Power results for different normality tests under alternate Logistic (Location=0, Scale=1) distribution at (alpha=0.01)

Normality Tests	n=10	n=20	n=30	n=40	n=50	n=100	n=200	n=500
Kolmogrove Smirnov	0.0158	0.0219	0.0243	0.0281	0.0301	0.0560	0.0988	0.2602
Lillifors	0.0197	0.0213	0.0260	0.0276	0.0294	0.0511	0.0886	0.2517
Kuiper	0.0165	0.0283	0.0311	0.0403	0.0403	0.0845	0.1544	0.4245
Cramer von mises	0.0201	0.0251	0.0335	0.0422	0.0504	0.0874	0.1765	0.4721
Anderson Darling	0.0218	0.0307	0.0420	0.0505	0.0606	0.1076	0.2064	0.5455
Shapiro Francia	0.0269	0.0498	0.0738	0.102	0.1118	0.2116	0.3735	0.7488
Watson	0.0193	0.0254	0.0323	0.0407	0.0505	0.0956	0.1874	0.5011
Chi Square	0.0146	0.0165	0.0166	0.0197	0.0204	0.0229	0.035	0.0757
Jarque Bera	0.0094	0.0526	0.095	0.1364	0.1498	0.2801	0.4596	0.8135
Geary	0.0114	0.0412	0.0658	0.0926	0.111	0.2211	0.4127	0.7967
D,Agostino D	0.0309	0.0473	0.068	0.0798	0.0812	0.1133	0.1400	0.1641
Shapiro Wilk	0.0253	0.0391	0.0623	0.0801	0.0901	0.1709	0.3160	0.7054

Under alternate Cauchy distribution almost all tests perform well at different sample sizes. By increasing sample size, power of all tests also increased. However Shapiro Francia gave maximum power followed by Cramer von mises test, Watson test and Anderson Darling test. For small sample size, Jarque Bera and Geary test and for large sample size D' Agostino D test had low power respectively (Table 3.2). Results of the power of the normality tests under alternate exponential distribution is given in Table 3.3. Shapiro Wilk test performed well under this alternate distribution for small and moderate sample sizes. Shapiro Francia, Anderson Darling and Cramer von mises test also perform well but it had slightly low power than Shapiro Wilk test. Greay test did not perform well among its competitors and generate low power for small and moderate sample sizes while for large sample sizes it provides better results. All the normality tests did not perform well at different sample sizes under Weibull distribution with scale parameter 2 and shape parameter 3 at 5% level of significance. For sample size 500, Shapiro Wilk, Shapiro Francia and Jarque Bera had better results than the other tests. Under this alternate distribution, Geary test also did not perform well and had low power for all sample sizes.

At five percent level of significance, when the alternate distribution is logistic, all tests had low power and not perform well for small and moderate sample sizes. For large sample size Geary test performed well and had maximum power than its competitors while the Chi Square test also had low power of the test. Jarque Bera test had provided better results followed by Shapiro Francia, Shapiro Wilk, Anderson Darling and Watson test. Table 3.6 showed the actual size of the type 1 error rate at one percent level of significance. It is clear from the results that less powerful test surrounded by those normality tests is the Jarque Bera test. It performed well only for the sample size 20. For the sample size less than 20 it had lowest significance level than the actual and for the sample size greater than 20 it had larger significance level than the actual. All other test almost had actual size of the test whatever the sample size.

For one percent level of significance under alternate Cauchy distribution, Shapiro Francia, Anderson darling, Cramer von mises and Shapiro Wilk test have better results. For moderate and large sample sizes almost all tests perform well and gives maximum power. For the small sample size 10, Jarque Bera and Geary test had low power than the other test but for the sample size larger than 10 it has better results than the D' Agostino D test and Chi Square test (Table 3.7). Under alternate Exponential distribution at 1% level of significance, Shapiro Wilk test perform well and gives better results and than the other test and Geary test is the less powerful test. Geary test also gives better results for the large sample size but its power remains low among all other test whatever the sample size (Table 3.8).

Table 3.9 showed the power of the tests under alternate Weibull distribution for different sample sizes at 1% level of significance. It is clear that all tests had low power for all sample sizes considered in the study. By increasing sample sizes, power of all tests also increases except the Geary test. For large sample size, Shapiro Wilk test gives better results followed by Shapiro Francia test. In Table 3.10, power of 12 different normality test under alternate Logistic distribution at 1% level of significance with location parameter 0 and scale parameter 1 was provided. The power of all test remain very low for small and moderate sample sizes and for large sample size few tests had better results. Under this alternate distribution Jarque Bera test is the most powerful test for moderate and large sample sizes followed by Geary test, Shapiro Francia test and Shapiro Wilk test.

Conclusion

A comprehensive study of the power of accessible normality tests has been performed and observed that the power of these tests has been affected by changing the sample size, level of significance and alternate distribution. By increasing the sample size and level of significance, power of these tests also increased. In general, Shapiro Francia test leads under alternate Cauchy distribution at 5% and 1% level of significance followed by Watson test and Anderson Darling test. Under alternate Exponential distribution Shapiro Wilk test had better results followed by Shapiro Francia, Anderson Darling and Cramer von miss a test at both levels of significance. When the alternate distribution is Weibull, performance of almost all tests was low, whatever the sample size. However, Shapiro Wilk test slighltly had better results than the other studied tests. Under alternate Logistic distribution, all tests had low power for small sample sizes but for moderate and large sample sizes Geary test perform well followed by Jarque Bera test, Shapiro Francia test and Shapiro Wilk test at the 5% level of significance. At the 1% level of significance under Logistic alternate distribution, Jarque Bera test performed well and had a maximum power among its competitors. Secondly, Geary test, Shapiro Francia test and Shapiro Wilk test also give better results. Finally, it is concluded that Shapiro Wilk and Shapiro Francia test performed well and had better results in almost all cases except the alternate Logistic distribution.

References

1.   Ajne, B. (1968). A simple test for uniformity of a circular distribution. *Biometrika, 55*(2), 343-354.

2.   Alizadeh Noughabi, R., Alizadeh Noughabi, H., & Ebrahimi Moghaddam Behabadi, A. (2014). An entropy test for the Rayleigh distribution and power comparison. *Journal of Statistical Computation and Simulation, 84*(1), 151-158.

3.   Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association, 49*(268), 765-769.

4.   Berg, D., & Quessy, J. F. (2009). Local Power Analyses of Goodness- of- fit Tests for Copulas. *Scandinavian Journal of Statistics, 36*(3), 389-412.

5.   Chen, Z. (2014). Goodness-of-Fit Test Based on Arbitrarily Selected Order Statistics.

6.   Chen, Z., & Ye, C. (2009). An alternative test for uniformity. *International Journal of Reliability, Quality and Safety Engineering, 16*(04), 343-356.

7.   Conover, W. (1999). Statistics of the Kolmogorov-Smirnov type. *Practical nonparametric statistics*, 428-473.

8.   D'Agostino, R. B. (1972). Small sample probability points for the D test of normality. *Biometrika, 59*(1), 219-221.

9.   D'Agostino, R. B. (1971). An omnibus test of normality for moderate and large size samples. *Biometrika, 58*(2), 341-348.

10.   Ebrahimi, N., Pflughoeft, K., & Soofi, E. S. (1994). Two measures of sample entropy. *Statistics & Probability Letters, 20*(3), 225-234.

11.   Geary, R. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika*, 310-332.

12.   Ibrahim Al-Omari, A. (2014). Estimation of entropy using random sampling. *Journal of Computational and Applied Mathematics, 261*, 95-102.

13.   Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.

14.   Kohansal, A., & Rezakhah, S. (2014). Two New Estimators of Entropy for Testing Normality. *arXiv preprint arXiv:1405.0321*.

15.   Kolmogorov, A. N. (1933). *Sulla determinazione empirica di una legge di distribuzione*: na.

16.   Kuiper, N. H. (1960). Tests concerning random points on a circle. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen-Series A, 63*, 38-47.

17.   Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62*(318), 399-402.

18.   Noughabi, H. A., & Arghami, N. R. (2011). Monte Carlo comparison of seven normality tests. *Journal of Statistical Computation and Simulation, 81*(8), 965-972.

19.   Noughabi, H. A., & Noughabi, R. A. (2013). On the entropy estimators. *Journal of Statistical Computation and Simulation, 83*(4), 784-792.

20. Öztuna, D., Elhan, A. H., & Tüccar, E. (2006). Investigation of Four Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. *Turkish Journal of Medical Sciences, 36*(3).

21. Pearson, K. (1900). X. 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 be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50*(302), 157-175.

22. Quessy, J.-F., & Mailhot, M. (2011). Asymptotic power of tests of normality under local alternatives. *Journal of Statistical Planning and Inference, 141*(8), 2787-2802.

23. Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

24. Romao, X., Delgado, R., & Costa, A. (2010). An empirical power comparison of univariate goodness-of-fit tests for normality. *Journal of Statistical Computation and Simulation, 80*(5), 545-591.

25. Thadewald, T., & Büning, H. (2007). Jarque–Bera test and its competitors for testing normality–a power comparison. *Journal of Applied Statistics, 34*(1), 87-105.

26. Shapiro, S. S., & Francia, R. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association, 67*(337), 215-216.

27. Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 591-611.

28. Stephens, M. A. (1970). Use of the Kolmogorov-Smirnov, Cramér-Von Mises and related statistics without extensive Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 115-122.

29. Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society. Series B (Methodological)*, 54-59.

30. Watson, G. S. (1962). Goodness-of-fit tests on a circle. II. *Biometrika*, 57-63.

31. Yap, B., & Sim, C. (2011). Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation, 81*(12), 2141-2155.

32. Zamanzade, E., & Arghami, N. R. (2012). Testing normality based on new entropy estimators. *Journal of Statistical Computation and Simulation, 82*(11), 1701-1713.

33. Zhao, J., & Xu, X. (2012). New tests based on Rubin's empirical distribution function. *Journal of Statistical Planning and Inference, 142*(6), 1356-1366.