3-21-2014

A New method for Testing Normality based upon a Characterization of the Normal Distribution

Davayne A. Melbourne
davayne_melbourne@yahoo.com

DOI: 10.25148/etd.FI14040851

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

A NEW METHOD FOR TESTING NORMALITY BASED UPON A CHARACTERIZATION OF THE NORMAL DISTRIBUTION

A thesis submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Davayne Antoneil Melbourne

2014

To: Dean Kenneth G. Furton
College of Arts and Sciences

This thesis, written by Davayne Antoneil Melbourne, and entitled A New Method for Testing Normality Based upon a Characterization of the Normal Distribution, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this thesis and recommend that it be approved.

Florence George

Zhenmin Chen

Hassan Zahedi, Major Professor

Date of Defense: March 21, 2014

The thesis of Davayne Antoneil Melbourne is approved.

Dean Kenneth G. Furton
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2014

ACKNOWLEDGMENTS

I must thank the members of my committee for their continuous support, patience and insight without which I could not have completed my thesis. Dr. Florence George was always readily available to provide support and encouragement which reassured me of my ability to complete the research. Dr. Zhenmin Chen was a great guide as to the timely completion and providing resources for preparing my thesis in an excellent and approved manner. Finally, I would like to thank my major professor Dr. Hassan Zahedi, who was always confident in my abilities and open to discussion of different ideas and methods. He always ensured that I strived for excellence and provided firm but constructive criticism which was particularly helpful in both my thesis and coursework taken in the program.

The entire statistics curriculum has been stimulating and emphasized statistical theory while providing exposure to applications and use of technology to facilitate statistical analysis. It has equipped me with the tools to be competitive and competent in the working environment.

ABSTRACT OF THE THESIS

A NEW METHOD FOR TESTING NORMALITY BASED UPON A CHARACTERIZATION OF THE NORMAL DISTRIBUTION

by

Davayne Antoneil Melbourne

Florida International University, 2014

Miami, Florida

Professor Hassan Zahedi, Major Professor

The purposes of the thesis were to review some of the existing methods for testing normality and to investigate the use of generated data combined with observed to test for normality. The approach to testing for normality is in contrast to the existing methods which are derived from observed data only. The test of normality proposed follows a characterization theorem by Bernstein (1941) and uses a test statistic D*, which is the average of the Hoeffding's D-Statistic between linear combinations of the observed and generated data to test for normality.

Overall, the proposed method showed considerable potential and achieved adequate power for many of the alternative distributions investigated. The simulation results revealed that the power of the test was comparable to some of the most commonly used methods of testing for normality. The test is performed with the use of a computer-based statistical package and in general takes a longer time to run than some of the existing methods of testing for normality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The normal distribution is possibly the most important probability distribution function in the field of statistics. Much work has been done to investigate the behavior and properties of the normal distribution since many parametric statistical methods are formulated using the underlying assumption that data collected comes from a normal distribution. The need for determining normality has resulted in many tests being developed over the years to test whether a sample of observations can be modeled by the normal distribution. The assumption of normality is needed for many statistical tests which have implications or relevance to not just the field of statistics but across many disciplines, such as physical science, psychology, engineering, social sciences and many other subject areas.

Many of these tests detect deviations from normality when sample size is very large, and are usually formulated using a distance parameter which measures the deviation of the observed sample from normality. These methods all have some specific drawback ranging from performing poorly when sample size is small to having small power when the data follows a distribution that is similar in shape to the normal distribution. Because of the importance of the normal distribution, there is always a need to improve upon the currently available tests of normality or to develop new tests for testing normality.

The main purpose of the thesis is to first conduct a short review of existing methods for testing normality and then to investigate the use of simulated data combined with observed data to test for normality. The proposed procedure for testing normality will involve incorporating generated data and observed data to perform the test for normality. For conducting inference, the combination of observed and generated data is referred to as ‘enriched data’. The investigation is examines a specific characterization of the normal distribution. The characterization developed by Bernstein (1941) states, “If X_1 and X_2 are identically and independently distributed random variables and $U \equiv X_1 + X_2$ and $V \equiv X_1 - X_2$ are independent, then U, V, X_1 and X_2 are all normally distributed."

The method that is proposed is a new procedure for inference that is made possible because of the power of currently existing statistical packages which make it possible to produce large sets of random variates from a distribution. The proposed method is in contrast to the existing methods of inference which are based on observed data and expected data under the null hypothesis. After the test procedure is developed its power is compared to some of the most popular existing methods of testing for normality and observe whatever drawbacks may exist.

The organization of the thesis is as follows: A review of the literature on tests for normality and characterizations of the normal distribution are presented in Chapters 1 and 2. In Chapters 3 and 4, the test procedure and the main theorem are provided along with the simulation results and discussion.

II. LITERATURE REVIEW

Most parametric analyses assume that an observed data set can be modeled by a given distribution. Much effort has been placed over the years into developing methods for testing how well a set of data points can be modeled by a given distribution. These tests are known as goodness of fit tests, Conover (1999). Goodness-of-fit tests are used to assess whether data are consistent with a hypothesized null distribution. Typically measures of goodness of fit compute departures of the observed data from the expected values of the distribution under investigation.

Of all goodness-of-fit tests available, possibly the most commonly used and most important are those that test for normality. The amount of effort that has been devoted to testing for normality is warranted given the broad range of application of the normal distribution across disciplines. Some of the most commonly methods of testing for normality include Cramer-von Mises test (1929), Kolmogorov – Smirnoff (1939) translated to English by Massey (1951), Anderson Darling (1952), Shapiro and Wilk (1965), Lilliefors (1967) and Pearson, D'Agostino and Bowman (1973). Of all these the Shapiro-Wilk's is most commonly used and is generally regarded as the most robust method of testing normality.

1. Cramer-Von Mises test

The Cramer-Von Mises test is a goodness of fit test that was developed by Von Mises (1928) and Cramer (1931). The test statistic examines the distance between the empirical distribution and theoretical cumulative distribution function under the null hypothesis H_0. The statistic $\omega^2$is given by:

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \, dF^*(x) \qquad (1.1)$$

where, $F^*(x)$ is the theoretical distribution function and $F_n(x)$ is the empirical distribution function.

The test statistic can also be written as:

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n}[\frac{2i-1}{2n} - F(x_{(i)})]^2, \tag{1.2}$$

where, $x_{(i)}$ are ordered values of the sample.

A table of approximate critical values for the statistic under H_0 is given in Anderson and Darling (1952) and the bias and power of the test is also discussed by Thompson (1966). The test has been shown to be more powerful than Kolmogorov-Smirnov for certain types of hypothesized distributions. It is best suited for situations when it is expected that the alternative distribution deviates a little over the entire sample range rather than having large deviations over a small section of the sample. In the latter case the Kolmogorov – Smirnov is more suitable. Stephens (1974) provides a rather comprehensive comparison of various goodness-of-fit tests. In the Cramer-Von Mises test, when $F^*(x)$ is assumed to be normal then the test can be used as a test for normality.

2. Kolmogorov-Smirnov test

The Kolmogorov-Smirnoff is another general goodness-of-fit test which can be adopted to test for normality when the mean and variance are specified. The test statistic uses the largest vertical difference between the hypothesized and empirical distribution. The test statistic is defined as:

$$T = sup_x|F^*(x) - F_n(x)|, \tag{1.3}$$

where, $F^*(x)$ is the hypothesized distribution under H_0 and $F_n(x)$ is the empirical distribution.

When $F^*(x)$ is assumed to be a specified normal distribution the test can be used as a test for normality. In this case, if T exceeds the $1-\alpha$ significant point the null hypothesis of normality (μ, σ) is rejected at α level of significance. The distribution of T does not depend on the hypothesized distribution when the null distribution is a continuous distribution. Significance points are tabulated for different sample sizes and are given in Conover (1999) with additional details on the computation of the test statistic and some historical perspective on the method.

3. Anderson Darling

The Anderson-Darling test is a general goodness-of-fit test which tests whether the sample comes from a specified distribution. It tests the hypothesis that a sample has been drawn from a population with a specified continuous distribution function $F(x)$. Let $x_1, x_2 \ldots x_n$, be n sample observations under H_0, and let $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ be the n ordered sample observations. The test statistic is defined as:

$$W_n^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln u_j + \ln(1-u_{n-j+1})\right] \tag{1.4}$$

where, $u_i = F\left(x_{(i)}\right)$.

The null hypothesis is rejected for large values of the test statistic. The Anderson Darling's test is very efficient in detecting deviation of the true distribution from the hypothesized especially when it differs in the tails. Critical values for W_n^2 are not available for small sample sizes but asymptotic significant points are tabulated for large sample sizes in Anderson and

Darling (1952). When a significant number of ties exist in the sample, the test will frequently reject the null hypothesis, regardless of how well the data fit the distribution. The test can be adopted for testing for normality if $F(x)$ is assumed to be normal.

4. Shapiro-Wilk's Test

The Shapiro-Wilk's test of normality is built upon the test statistic W, derived from the sample itself and expected values of order statistics from a standard normal distribution. It tests the null hypothesis that the sample came from a normal distribution with unknown mean and variance. The test statistic is the square of the Pearson correlation coefficient computed between the order statistics $X_{(i)}$ and the scores a_i, which are the expected values of order statistics from a normal distribution. The W statistic is defined by:

$$W = \frac{1}{D}\left\lfloor \sum_{i=1}^{k} a_i\left(X_{(n-i+1)} - X_{(i)}\right) \right\rfloor^2 \tag{1.5}$$

where,

$$D = \sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{1.6}$$

and $X_{(i)}$ represent the i^{th} order statistic of the sample.

The values of W lie between 0 and 1 and small values of the statistic indicate departure from normality. Under H_0, W has a distribution that is independent of S^2 and $\bar{X}$ and is both scale and origin invariant. See Conover (1999) for further details and also a table of critical values of the test statistic.

5. **Lilliefor's test**

Lilliefor (1968) extended Kolmogorov's test for testing a composite hypothesis that the data come from a distribution with unknown shape or scale parameter. The test statistic is defined as

$$D = max_x |F^*(x) - S_n(x)| \tag{1.7}$$

where, $S_n(x)$ is the sample cumulative distribution function and $F^*(x)$ is the cumulative distribution function (CDF) of the null distribution.

Lilliefor's test is similar to the Kolmogorov-Smirnov test but the distribution of the test statistic under H_0 is different and hence critical values are different. The test can be used for testing normality when the distribution under H_0 is assumed to be normal with unknown mean and variance. See Conover (1999) for further reference and a table of critical values.

6. **D'Agostino's K-Squared tests**

D'Agostino's K-Squared tests use sample kurtosis and skewness to detect departures from normality. The drawback of these tests is that they have power only for the alternative hypothesis that the distribution is skewed and/or kurtic. The tests are derived from the third and fourth standardized moments given by:

$$\sqrt{\beta_1} = \frac{E\,(X-\mu)^3}{[E(X-\mu)^2]^{3/2}} = \frac{E(X-\mu)^3}{\sigma^3} \tag{1.8}$$

and,

$$\beta_2 = \frac{E\,(X-\mu)^4}{[E(X-\mu)^2]^2} = \frac{E(X-\mu)^4}{\sigma^4} \tag{1.9}$$

The sample estimates of $\sqrt{\beta_1}$ and β_2 are,

$$\sqrt{b_1} = \frac{m_3}{(m_2)^{\frac{3}{2}}} \quad \text{and} \quad b_2 = \frac{m_4}{(m_2)^2} \tag{1.10}$$

where, $$m_k = \frac{1}{n}\sum_i^n (X_i - \bar{X})^k \text{ respectively.} \tag{1.11}$$

I. Test of skewness $\sqrt{\beta_1}$:

Here the null is H_0: The data is normal $\sqrt{\beta_1} = 0$ versus

H_a: Non-normality as a result of skewness $\sqrt{\beta_1} \neq 0$

Under H_0, the test statistic $Z(\sqrt{b_1})$ is approximately normally distributed for $n > 8$ and is defined:

$$Z(\sqrt{b_1}) = \delta \ln\left({}^{Y}\!/_{\alpha} + \{({}^{Y}\!/_{\alpha})^2 + 1\}^{\frac{1}{2}}\right), \tag{1.12}$$

where, $$\alpha = \{2/(W^2 - 1)\}^{1/2}, \tag{1.13}$$

$$\delta = 1/\sqrt{\ln W}, \tag{1.14}$$

$$W^2 = -1 + \{2(\beta_2(\sqrt{b_1}) - 1)\}^{1/2}, \tag{1.15}$$

$$\beta_2(\sqrt{b_1}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \tag{1.16}$$

$$Y = \sqrt{b_1}\left\{\frac{(n+1)(n+3)}{6(n-2)}\right\}^{1/2} \tag{1.17}$$

II. Test of kurtosis $\boldsymbol{\beta_2}$:

Here the null is H_0: The data is normal $\beta_2 = 3$ versus

H_a: Non-normality as a result of kurtosis $\beta_2 \neq 3$

The test statistic $Z(b_2)$ is defined as:

$$Z(b_2) = \left(\left(1 - \frac{2}{9A}\right) - \left[\frac{1-2/A}{1+x\sqrt{2/(A-4)}}\right]^{\frac{1}{3}}\right)/\sqrt{2/9A} \tag{1.18}$$

where, $$A = 6 + \frac{8}{\sqrt{\beta_1(b_2)}}\left[\frac{2}{\sqrt{\beta_1(b_2)}} + \sqrt{(1 + \frac{4}{\beta_1(b_2)})}\right], \tag{1.19}$$

$$\sqrt{\beta_1(b_2)} = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}}, \tag{1.20}$$

and, $$x = (b_2 - E(b_2))/\sqrt{var(b_2)}. \tag{1.21}$$

It is also shown that under H_0 the test statistic $Z(b_2)$ is approximately normally distributed for n ≥ 20.

III. The omnibus test:

Pearson and D'agostino (1973) developed an omnibus statistic using $\sqrt{b_1}$ and b_2 which is able to detect deviations from normality as a result of kurtosis or skewness. They derived the test statistic

$$K^2 = Z^2(\sqrt{b_1}) + Z^2(b_2) \tag{1.22}$$

where, $Z(\sqrt{b_1})$ and $Z(b_2)$ are the normal approximations of skewness $\sqrt{\beta_1}$ and kurtosis β_2 defined in (1.8) and (1.9) above. Under H_0, the statistic K^2 has approximately a chi-square distribution with 2 degrees of freedom.

Characterizations of Normal distribution

A characterization of a distribution is a property that is unique to that distribution, and thus can be used to develop tests to determine whether a given sample is taken from that distribution. The earliest characterization of the normal distribution was by Gauss (1809). He showed that if the solution to the likelihood equation was the sample mean, for all possible samples and for all values of n, and the score function is continuous, then the underlying distribution must be normal.

Cramer's theorem (1936) states, if there exists a normally distributed random variable Z where $Z = X + Y$ (the sum of two independent random variables) then X and Y must be normally distributed as well. Further developments were made to the above characterization which eventually led to Bernstein's (1941) theorem which states, "Let X_1 and X_2 be independent random variables with finite variances. Then the sum $X_1 + X_2$ and the difference $X_1 - X_2$ are independent if and only if X_1 and X_2 are normally distributed."

Ahmad and Mugdadi (2003) developed a test for normality using this criterion. They test for normality by testing for independence of U = X_i - X_{i^*} and V = $X_i + X_{i^*}$, $i \neq i^*$. The kernel method of density estimation was used to estimate the joint density $h(u, v)$ of (U, V) and the marginal densities $U(h_1)$ and $V(h_2)$. The sample estimate of the distance parameter ∂ was used to test for normality where;

$$\partial = \iint_{-\infty}^{\infty} [h(u, v) - h_1(u)h_2(v)]^2 \; dv \, du \tag{2.1}$$

The parameter ∂ is a measure of departure from normality and is equal to zero if and only if the data follow a normal distribution. The derivation of kernel density estimates is very tedious and complicated and thus presents a drawback to this method.

Another well-known characterization is that the sample mean $\bar{X}$ and sample variance S^2 are independent if and only if the underlying population is normal. Similarly, $\bar{X}$ and $\frac{1}{n}\sum_i^n (X_i - \bar{X})^3$ are independent if and only if X is normal. Lin and Mudholkar (1980) proposed a test that examines the independence of $\bar{X}$ and S^2. They used a jackknife procedure (drawing subsets from the original n sample points) to estimate the correlation $\rho(\bar{X}; S^2)$, and used this for a test for normality against asymmetric alternatives. They also presented a test built on the independence of $\bar{X}$ and $\frac{1}{n}\sum_i^n(X_i - \bar{X})^3$ constructed using the same jackknife procedure. The authors named the tests the Z_2 test and Z_3 test respectively.

They obtained an expression for the test statistic Z_2 using Fisher's z-transform.

$$Z_2 = \frac{1}{2}\log\left(\frac{1+r_2}{1-r_2}\right) \tag{2.2}$$

where, r_2 is the sample correlation coefficient $r(\bar{X}; S^2)$. The statistic Z_2 is used to test for normality and can be used for both one sided and two sided tests.

For the Z_3 test they considered the mean $\bar{X}$ and the third central sample moment $\hat{\mu}_3 = \frac{1}{n}\sum_i^n(X_i - \bar{X})^3$. Using Fisher's z-transform they obtained a test statistic called Z_3 defined as:

$$Z_3 = \frac{1}{2}\log\left(\frac{1+r_3}{1-r_3}\right), \tag{2.3}$$

where, r_3 is the sample correlation coefficient between$(\bar{X}, \hat{\mu}_3)$. The statistic Z_3 is used to test for normality and can be used for both one sided and two sided tests.

Further characterizations of the normal distribution were made on the basis of order statistics, maximum-likelihood estimators and independence of variables. There are a plethora of theorems and characterizations of the normal and other distributions in the literature. Kagan, Linniks and Rao (1973) and Duerinckx, Ley and Swan (2012), covered a substantial review of many of the characterization results. Additionally books by Galambos and Kotz (1978) and Kakosyan, Keblanov and Melamed (1974) can be used for a detailed survey of many characterization results.

<u>III. BASIS FOR TEST</u>

The focus throughout the paper is to investigate the effect of testing for normality using observed data combined with additional simulated data (enriched data). The combination of generated and observed data will be used to develop a test statistic following the characterization of the normal distribution developed by Bernstein (1941). The proposed method is a new exploration into the use of simulated data that to the best of my knowledge has not been investigated before. If indeed the results of such explorations using the normal setting yield positive results, these tests could possibly be modified and applied to various other goodness-of-fit tests.

Main Theorem

Theorem 1: If X and Y are identically and independently distributed random variables and U ≡ X+Y and V ≡ X-Y are independent, then U, V, X and Y are all normally distributed.

The proof of Bernstein's theorem is rather complicated and is omitted here. For a detailed proof please see Bernstein (1941) or Wlodzimierz (2005). The theorem is used to develop our test of normality. In the most basic case we are interested in the following Goodness of Fit test.

Let $X_{1,}X_2 \ldots, X_n$ be the observed sample $\boldsymbol{X}$ from our population of interest. We wish to test:

H_0: **X** has a Normal distribution with specified mean μ_o and variance σ_o^2.

H_{a}: **X** does not have a Normal distribution.

Let $\mathbf{Y} = (Y_1, Y_2 \ldots, Y_n)$ be n randomly generated points from a Normal (μ_o, σ_o^2). Compute $\boldsymbol{U} \equiv \boldsymbol{X} + \boldsymbol{Y}$ and $\boldsymbol{V} \equiv \boldsymbol{X} - \boldsymbol{Y}$.

Since $\boldsymbol{Y}$ is normally distributed then by Bernstein's theorem, if $\boldsymbol{U}$ and $\boldsymbol{V}$ are independent then $\boldsymbol{X}$ must be normally distributed with mean μ_o and variance σ_o^2. Therefore the test of normality reduces to a test for independence between $\boldsymbol{U}$ and $\boldsymbol{V}$.

For the purposes of the paper the Hoeffding's D - statistic was used to test for independence. Before choosing the Hoeffding's test of independence as the method for our proposed test, simulations were conducted to compare the effectiveness of three common methods of testing for independence. The three methods compared were: Hoeffding's D-Statistic, Spearman's Rank Correlation Coefficient and Kendall's Tau. The results of simulation showed that the Hoeffding's test is able to capture more subtle departures from independence than the other methods. The greater power of the Hoeffding's test is a very important implication, which might be useful to other experimenters wishing to test independence when the sample size is not very large. If the aim of their experiment is to capture minute departures from independence of two samples then clearly Hoeffding's D test seems to be the logical choice.

Table 1 below provides a comparison of the power of the three methods for testing independence. The power is computed by calculating the percentage of times of 10,000 trials that independence between U = X + Y and V = X - Y is rejected.

Table 1: Comparison of power of 3 methods of testing independence at $\alpha = 0.05\%$ level of significance.

Distribution	Sample size	Spearman's Rank	Kendall's Tau	Hoeffding's-D
Normal (0,1)	10	0.070	0.063	0.091
	20	0.039	0.041	0.076
	50	0.052	0.052	0.059
	100	0.057	0.056	0.062
	200	0.064	0.059	0.058
	500	0.052	0.052	0.050
Chi-square 2	10	0.084	0.081	0.211
	20	0.091	0.087	0.408
	50	0.087	0.094	0.949
	100	0.088	0.095	1.000
	200	0.093	0.097	1.000
	500	0.099	0.112	1.000
Exponential	10	0.080	0.080	0.194
	20	0.101	0.101	0.414
	50	0.103	0.104	0.947

	100	0.103	0.103	1.000
	500	0.105	0.112	1.000
	1000	0.099	0.104	1.000
	2000	0.097	0.106	1.000
Uniform (0,1)	10	0.020	0.015	0.073
	20	0.006	0.015	0.042
	50	0.004	0.015	0.043
	100	0.005	0.015	0.097
	200	0.002	0.010	0.568
	500	0.001	0.007	1.000
T (7)	10	0.067	0.061	0.110
	20	0.064	0.062	0.072
	50	0.073	0.067	0.085
	100	0.074	0.062	0.074
	200	0.075	0.081	0.080
	1000	0.071	0.080	0.126
Gamma (4,5)	10	0.070	0.055	0.128
	20	0.047	0.047	0.126
	50	0.068	0.063	0.224
	100	0.056	0.056	0.522
	200	0.069	0.068	0.952
	500	0.047	0.051	1.000
	1000	0.054	0.056	1.000
Beta (2,2)	10	0.033	0.022	0.084
	20	0.036	0.027	0.066
	50	0.027	0.019	0.044
	100	0.026	0.017	0.038
	500	0.024	0.013	0.187
	1000	0.022	0.016	0.788
	2000	0.018	0.014	1.000
Cauchy (0,1)	10	0.121	0.128	0.171
	20	0.116	0.157	0.173
	50	0.111	0.171	0.263
	100	0.128	0.177	0.589
	500	0.182	0.222	1.000
	1000	0.098	0.156	1.000
	2000	0.114	0.162	1.000

Table 1

Description of Hoeffding's test:

Hoeffding's Dependence Coefficient D, is a nonparametric measure of association developed by Hoeffding (1948) that detects more general departures from independence. The statistic approximates a weighted sum over observations from a bivariate sample by placing ranks on observations. The D – statistic is defined by the SAS Institute (2010) as:

$$D = 30\frac{(n-2)(n-3)D_1+D_2-2(n-2)D_3}{n(n-1)(n-2)(n-3)(n-4)} \tag{3.1}$$

where,

$$D_1 = \sum(Q_i - 1)(Q_1 - 2), \tag{3.2}$$

$$D_2 = \sum(R_i - 1)(R_1 - 2)(S_i - 1)(S_i - 2), \tag{3.3}$$

and

$$D_3 = \sum(R_i - 2)(S_i - 2)(Q_i - 1), \tag{3.4}$$

where R_i is the rank of u_i, S_i is the rank of v_i and Q_i (also called the bivariate rank) is 1 plus the number of point with both $x\ and\ y$ values less than the ith point (u_i, v_i). That is for each sample point (u_i, v_i), Q_i is defined by:

$$Q_i = \sum_{j=1}^{n} I[(u_j, v_j) \leq (u_i, v_i)]$$

= 1 + number of sample points (u_j, v_j) which is less than $(u_i, v_i), i \neq j$

When no ties occur among data set, the D statistic values are generally between -0.5 and 1, with 1 indicating complete dependence. Generally for any D ≤ 0 it is safe not to reject H_0 and conclude there is independence. To test for independence, at $\alpha(0 < \alpha < 1)$ a given level of significance, let ρ_n, be the smallest number satisfying the inequality $P\{D > \rho_n\} < \alpha$ when $U\ and\ V$ are independent. Compute D from (3.1) and reject H_0 the hypothesis of independence if $D > \rho_n$. When n is large (n >10), the critical values of Hoeffding's D statistic are computed using the asymptotic distribution by Blum, Kiefer, and Rosenblatt (1961) for some selected values of significance. If the sample size is less than 10, refer to the tables in Hollander and Wolfe (1999) for the exact distribution of D.

In this chapter we covered and compared the power of some of the most common methods of testing for independence. On the basis of the simulation results the Hoeffding's test was selected as the preferred method of testing independence. In the next chapter we combine the concepts discussed in Chapters 1-3 to present the test for normality using enriched data. First we test when H_0 is a simple hypothesis, 'testing for a specified normal distribution' and then the procedure is extended to test for a composite hypothesis when the mean and variance are not specified. The proposed test of normality will be referred to as the 'Enriched Method'.

IV. TEST PROCEDURE

The first test that is investigated is for the simple hypothesis that:

H_o: *The data come from a specified normal distribution.*

H_a: *The data do not come from the specified normal distribution.*

The proposed test is similar to the Kolmogorov-Smirnov (KS) goodness-of-fit test which is widely used to test for a specified normal distribution. We will investigate how the simulated data and observed data can be combined to test for normality. The test procedure which uses linear combinations of the observed sample and the generated data is as follows.

Procedure for computing the test statistic D*:

3.1: Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a random sample from a population with mean μ_o and variance σ_o^2.

3.2: Randomly generate samples $\boldsymbol{Y_i} = (Y_{1i}, \dots, Y_{ni})$, $i = 1 \ to \ m$, of size n from a $N(\mu_o, \sigma_o^2)$.

3.3: Compute $\boldsymbol{U_i} = \boldsymbol{X} + \boldsymbol{Y_i}$ and $\boldsymbol{V_i} = \boldsymbol{X} - \boldsymbol{Y_i}$ for $i = 1 \ to \ m$. (m =100)

3.4: Compute D* = $\frac{1}{m}\sum_{i=1}^{m} D_i$

where, D_i is the Hoeffding's D-statistic computed between each $U_i = [(X_1 + Y_{i1}), (X_2 + Y_{i2}), \dots, (X_n + Y_{in})]$ and $V_i = [(X_1 - Y_{i1}), (X_2 - Y_{i2}), \dots, (X_n - Y_{in})]$, $i = 1 \ to \ m$.

3.5: Reject the null hypothesis of normality if D* > D_α , where D_α are critical points of the D* statistic developed under the null hypothesis of normality.

The table of significance points for D* statistic for α = 0.01, 0.05, 0.10, 0.50, 0.90 and 0.99 are given in table 2 below. These significance points were developed from 10,000 repetitions of the procedure given in 3.1 – 3.5 for generated samples from a $N(0,1)$ distribution and the various percentiles observed to develop the critical values for the various levels of significance.

Table 2: Significance points of D* statistic under normality when population mean and variance are known.

Sample size	p(0.01)	p(0.05)	p(0.10)	p(0.50)	p(0.90)	p(0.95)	p(0.99)
6	-0.13917	-0.09667	-0.07500	-0.00073	0.08000	0.13417	0.25083
10	-0.05020	-0.03841	-0.03200	-0.00053	0.03833	0.05782	0.11407
15	-0.02596	-0.02109	-0.01811	-0.00048	0.02292	0.03691	0.07898
20	-0.01733	-0.01428	-0.01242	-0.00040	0.01541	0.02504	0.04578
25	-0.01274	-0.01065	-0.00929	-0.00031	0.01239	0.02043	0.04040
30	-0.01005	-0.00852	-0.00755	-0.00026	0.01020	0.01637	0.02984
35	-0.00842	-0.00725	-0.00650	-0.00019	0.00867	0.01372	0.02886
40	-0.00701	-0.00602	-0.00543	-0.00010	0.00763	0.01239	0.02526
45	-0.00630	-0.00534	-0.00478	-0.00008	0.00654	0.01086	0.02199
50	-0.00547	-0.00474	-0.00423	-0.00007	0.00584	0.00975	0.01933
60	-0.00443	-0.00393	-0.00355	-0.00006	0.00499	0.00799	0.01529
70	-0.00386	-0.00331	-0.00297	-0.00005	0.00436	0.00694	0.01310
80	-0.00331	-0.00289	-0.00261	-0.00004	0.00355	0.00578	0.01049
90	-0.00296	-0.00257	-0.00232	-0.00002	0.00297	0.00486	0.00971
100	-0.00264	-0.00231	-0.00209	-0.00002	0.00282	0.00473	0.00902
200	-0.00125	-0.00108	-0.00098	-0.00005	0.00147	0.00237	0.00497
300	-0.00081	-0.00074	-0.00066	-0.00003	0.00098	0.00153	0.00289
400	-0.00061	-0.00055	-0.00050	-0.00002	0.00066	0.00100	0.00206
500	-0.00049	-0.00044	-0.00040	0.00000	0.00059	0.00099	0.00180
1000	-0.00024	-0.00022	-0.00020	0.00001	0.00026	0.00044	0.00079
2000	-0.00012	-0.00011	-0.00009	0.00001	0.00017	0.00029	0.00048

Table 2

It must be highlighted that the percentiles of D* do not depend on the mean or the variance of the specified normal distribution. That is they are invariant to the mean or the variance when the samples are from the same specified normal distribution. The independence of the percentiles is a direct result of Bernstein's theorem where independence holds as long as the mean and the variance of $\boldsymbol{X}$ and $\boldsymbol{Y_i}$ are the same. Additionally the choice of 100 samples for calculating D*, the average of the Hoeffding's D-statistics between $\boldsymbol{U_i}$ and $\boldsymbol{V_i}$, was done because such a size is large enough to ensure that the results are consistent. The simulations revealed that when the number of samples of $\boldsymbol{Y_i}$ is greater than 100, the distribution of D* becomes stable thus ensuring consistent results while avoiding unnecessary computation time for generating more samples.

Simulation Results

Monte Carlo simulations were done to compare the power of the 'Enriched method' (EM) with that of the Shapiro -Wilk (SW), Anderson Darling (AD) and Kolmogorov-Smirnov (KS) tests. To obtain the simulated power against a particular distribution at $\alpha = 5\%$ for each sample size, a total of 10,000 samples were generated and then the different tests applied to each sample using the SAS statistical package. The simulated power was then computed by finding the percentage of 10,000 trials that normality was rejected using each method.

For the case of the simple hypothesis stated above, the power of the 'Enriched Method' which is being investigated was compared to the power of the KS test since the KS test is the most widely used when testing for specified normality. The power was estimated for various alternative distributions, some of which are presented in table 3 with corresponding graphs in figure 1.

The EM method has greater estimated power than the KS for some of the distributions investigated at $\alpha = 5\%$ level of significance. The EM method has uniformly greater power than the KS test against a T distribution with $n = 7$ degrees of freedom. At a sample size of 500, the EM test has adequate power of 98% while the KS test has power of 57%. Similar results are seen for a T distribution with $n = 3$ degrees of freedom, where the EM had a power of 90% for a sample size of 100 whereas the KS had power of 71% for the same sample size. Against a Beta (2, 2) distribution, the EM performed better for smaller sample sizes than the KS test. However as the sample size increased the KS test converged in power to the EM method eventually surpassing the power of the EM method at sample size of 500. For many other distributions the two tests performed relatively similar.

Both methods had similar power against an Exponential (λ = 1) distribution with both achieving adequate power for relatively small sample size. Against a Chi-square distribution with n = 4 degrees of freedom the KS test had better power at smaller sample sizes but as the sample sized increased the EM method converged in power to the KS test. The same observation was seen against a Gamma (4, 5) and Uniform (0, 1) distribution. See Figures 1(a) – (f) for a graphical representation of the results.

Both methods typically require large sample sizes in order to achieve adequate power against most symmetric distributions as can seen from the simulations made against the Beta (2, 2), T (7) and Uniform (0,1). In contrast the tests are able to detect non-normality for relatively small sample sizes when the distributions are highly skewed. The simulations failed to reveal which of the two methods had a clear cut advantage over the other as none of the methods consistently outperformed the other. It would appear however that the KS test has an advantage against distributions that are highly skewed.

Table 3: Comparison of Power of Enriched method and Kolmogorov-Smirnov test against some alternative distributions for known mean and variance at α = 5% level of significance.

Distribution	**Sample size**	**Enriched Method**	**Kolmogorov - Smirnov**
Uniform (0,1)	10	0.023	0.078
	20	0.040	0.150

	50	0.166	0.306
	100	0.366	0.630
	200	0.942	0.958
	500	1.000	1.000
	1000	1.000	1.000
	2000	1.000	1.000
Chisquare (4)	10	0.118	0.161
	20	0.267	0.337
	50	0.745	0.689
	100	0.981	0.987
	200	1.000	1.000
	500	1.000	1.000
	1000	1.000	1.000
	2000	1.000	1.000
T (7)	10	0.078	0.066
	20	0.084	0.090
	50	0.121	0.127
	100	0.204	0.165
	200	0.527	0.278
	500	0.980	0.568
	1000	1.000	0.883
	2000	1.000	1.000
Beta (2,2)	10	0.041	0.044
	20	0.064	0.063
	50	0.123	0.114
	100	0.214	0.198
	200	0.648	0.398
	500	0.843	0.882
	1000	0.986	0.998
	2000	1.000	1.000
Gamma (4,5)	10	0.079	0.119
	20	0.110	0.210
	50	0.267	0.416
	100	0.571	0.727
	200	0.961	0.923
	500	1.000	1.000
	1000	1.000	1.000
	2000	1.000	1.000
Exponential (λ =1)	10	0.268	0.300
	20	0.566	0.584
	50	0.994	0.964

	100	1.000	1.000
	500	1.000	1.000
	1000	1.000	1.000
	2000	1.000	1.000
Weibull (2, 2)	10	0.073	0.053
	20	0.087	0.098
	50	0.123	0.244
	100	0.263	0.406
	200	0.553	0.748
	500	0.986	0.989
	1000	1.000	1.000
	2000	1.000	1.000
T (3)	10	0.204	0.190
	20	0.333	0.220
	50	0.652	0.491
	100	0.907	0.712
	200	1.000	0.957
	500	1.000	1.000
	1000	1.000	1.000
	2000	1.000	1.000
Chi-square (20)	10	0.064	0.063
	20	0.084	0.101
	50	0.111	0.196
	100	0.165	0.363
	200	0.383	0.552
	500	0.882	0.930
	1000	1.000	1.000
	2000	1.000	1.000

Table 3

Figure 1(a). Comparison of power of EM & KS tests against T (7) (α =0.05)

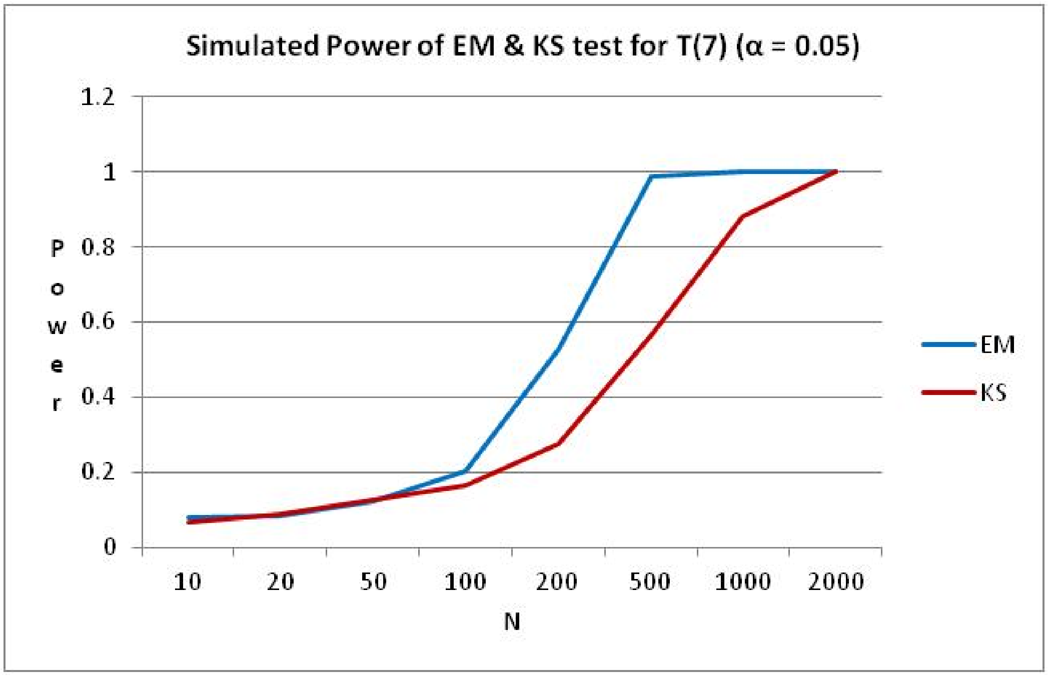


Figure 1(b). Comparison of power of EM & KS tests against T (3) (α=0.05)

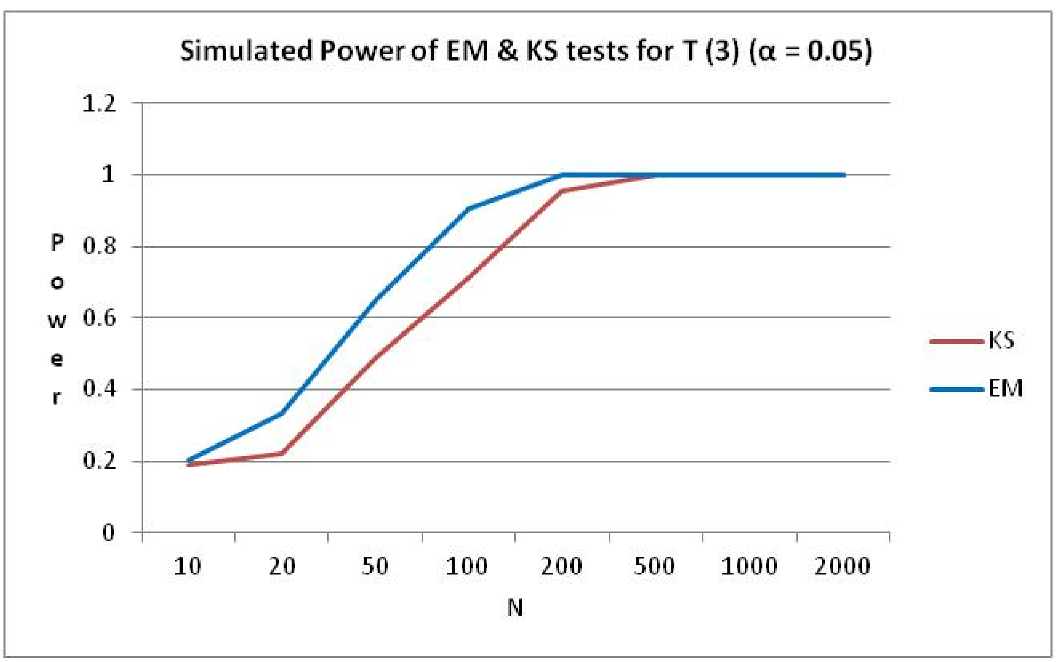


Figure 1(c). Comparison of power of EM & KS tests against Beta (2, 2) (α = 0.05)

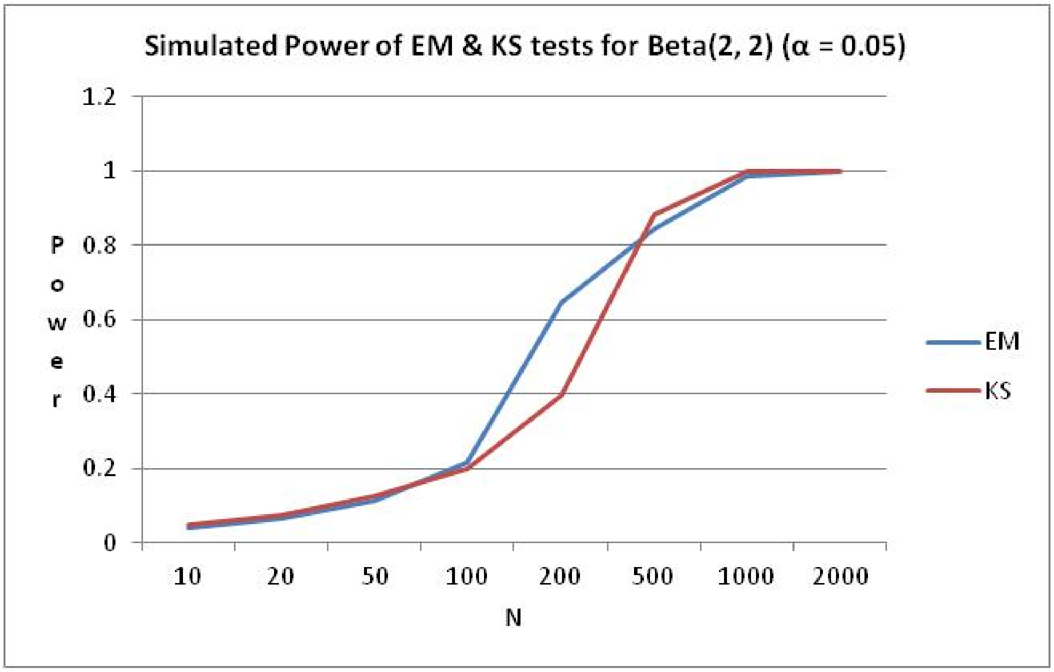


Figure 1(c). Comparison of power of EM & KS tests against Chi-square (4) (α = 0.05).

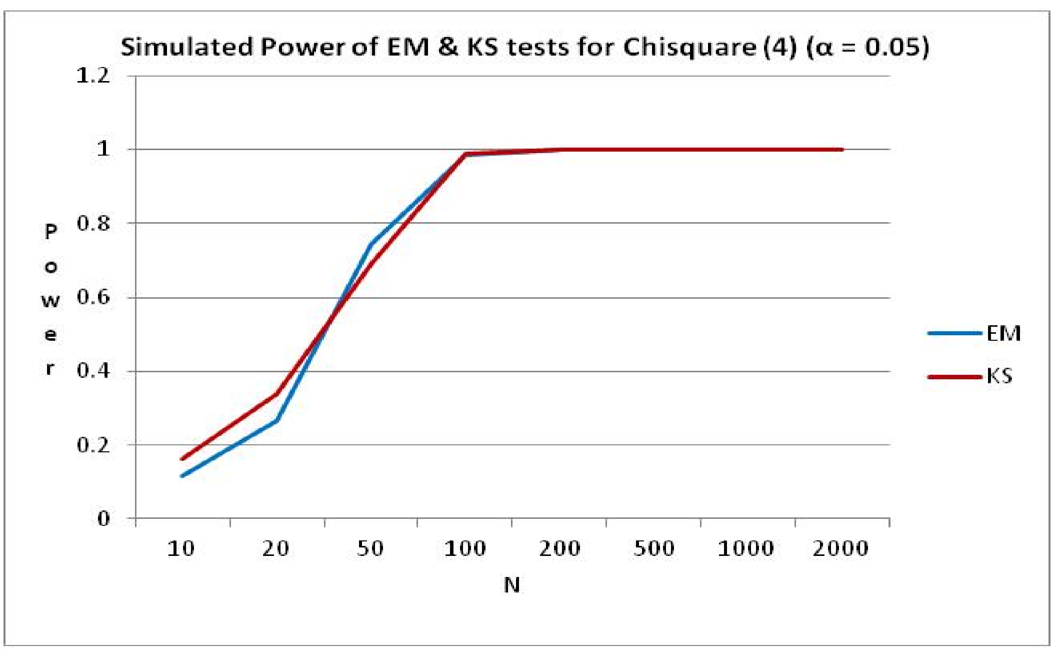


Figure 1(d). Comparison of power of EM & KS tests against Gamma (4, 5) (α = 5%)

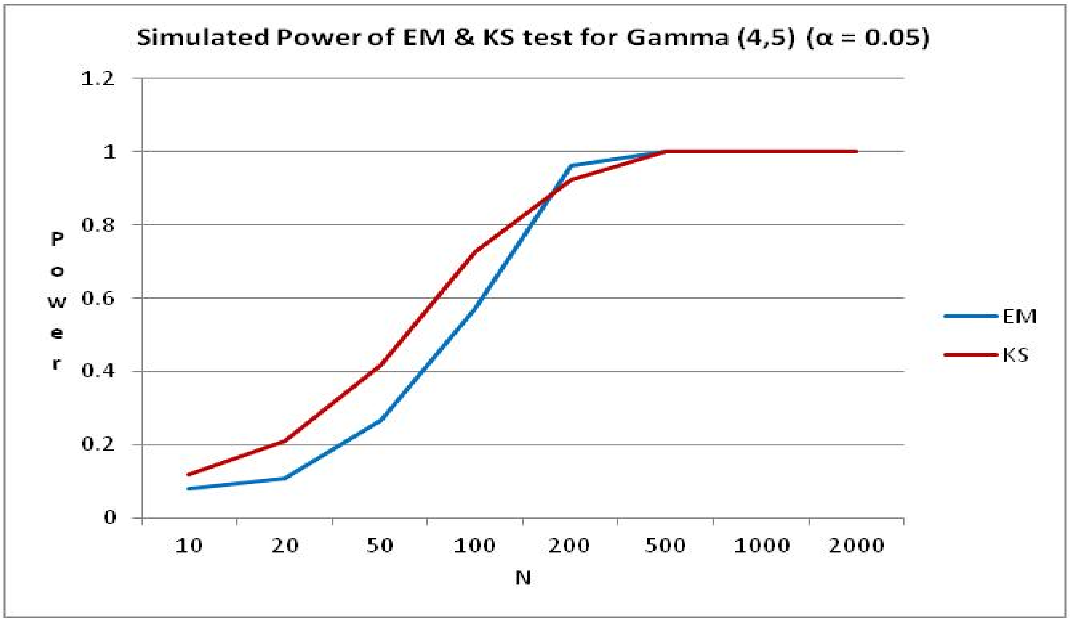


Figure 1(e). Comparison of power of EM & KS tests against a Uniform (0,1) (α = 5%)

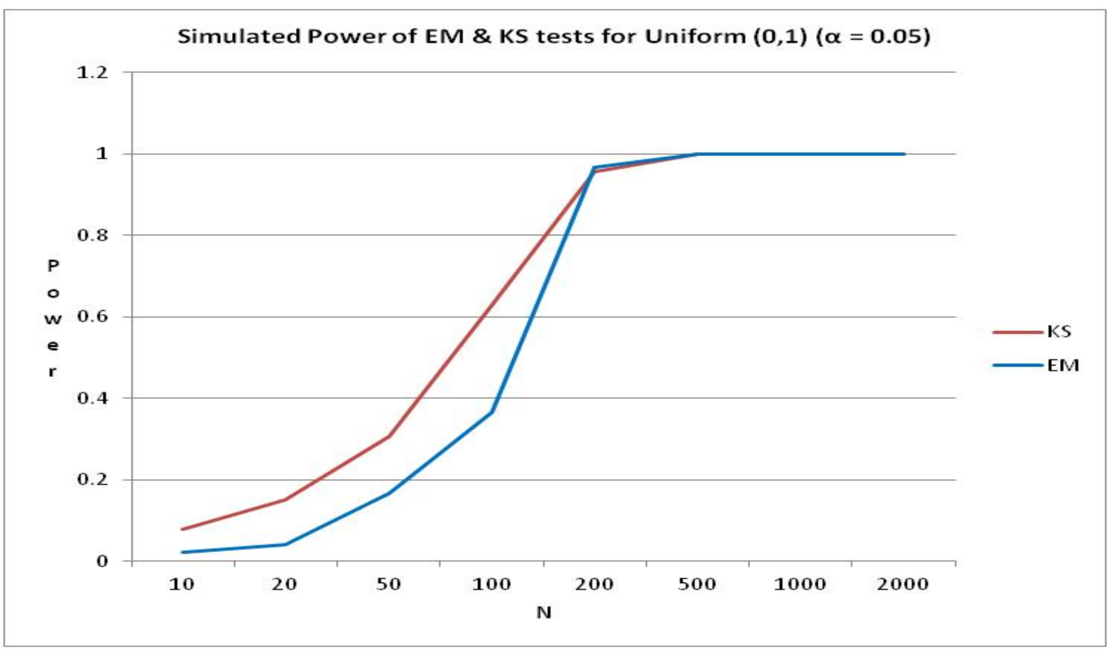
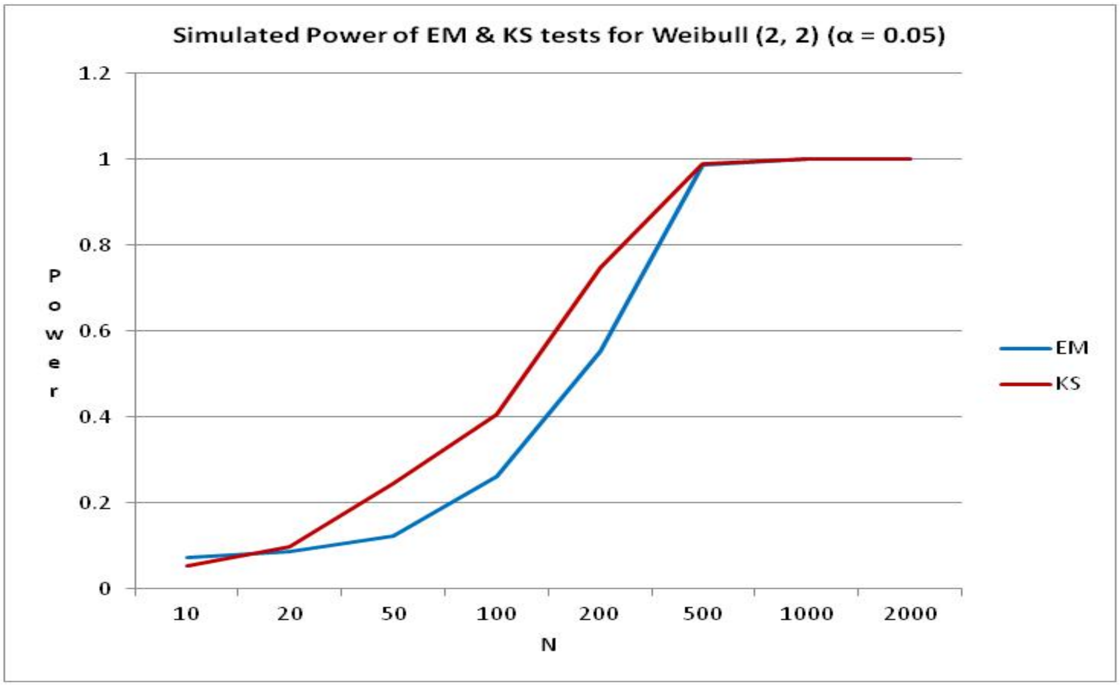


Figure 1(f). Comparison of power of EM & KS tests against Weibull (2, 2) (α=0.05)



Testing normality when population mean and variance are unknown

In many scenarios when testing normality, the population mean and variance are unknown to the experimenter and have to be estimated from the sample. In such a scenario, the previous test of normality cannot be used since it assumes that the population mean and variance are known. In this chapter we present a procedure for testing normality when the mean and variance are not specified.

That is given a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$ of size *n* we are testing:

H_o: $The\ data\ come\ from\ a\ normal\ distribution.$

H_a: $The\ data\ do\ not\ come\ from\ a\ normal\ distribution$

Indeed this test is more applicable to experimenters wishing to verify that data are normal regardless of mean or variance in order to perform many other parametric statistical tests. The test procedure follows directly from the case of known mean and variance. Significance points of the D* statistic are developed under the null hypothesis of normality similar to the case for known mean and variance except values were generated using the sample mean and sample variance of the observed samples. The table of critical values obtained is tabulated below for various sample sizes.

Table 4: Significance points of D* statistic under normality when population mean and variance are unknown.

Sample size	p(0.01)	p(0.05)	p(0.10)	p(0.50)	p(0.90)	p(0.95)	p(0.99)
6	-0.15333	-0.11333	-0.09167	-0.02169	0.043333	0.056667	0.081667
10	-0.05159	-0.04435	-0.03821	-0.01308	0.011190	0.018452	0.026607
15	-0.02677	-0.02337	-0.02087	-0.00868	0.004569	0.008104	0.015061
20	-0.01809	-0.01587	-0.01453	-0.00661	0.002184	0.005128	0.011119
25	-0.01328	-0.01172	-0.01069	-0.00507	0.001604	0.004065	0.009618
30	-0.01057	-0.00933	-0.00859	-0.00416	0.001236	0.003223	0.007971
35	-0.00875	-0.00783	-0.00724	-0.00364	0.000744	0.002564	0.007173
40	-0.00749	-0.00671	-0.00626	-0.00318	0.000620	0.002266	0.005799
45	-0.00651	-0.00588	-0.00542	-0.00286	0.000448	0.001969	0.004755
50	-0.00590	-0.00523	-0.00488	-0.00240	0.000705	0.001817	0.003992
60	-0.00478	-0.00430	-0.00400	-0.00214	0.000202	0.001236	0.003857
70	-0.00400	-0.00362	-0.00337	-0.00183	0.000172	0.001036	0.002997
80	-0.00346	-0.00312	-0.00292	-0.00157	0.000165	0.000922	0.002903
90	-0.00306	-0.00277	-0.00258	-0.00138	0.000158	0.000898	0.002496
100	-0.00279	-0.00248	-0.00234	-0.00126	0.000064	0.000577	0.002611
150	-0.00177	-0.00162	-0.00151	-0.00085	0.000048	0.000445	0.001443
200	-0.00134	-0.00119	-0.00113	-0.00061	0.000044	0.000406	0.001025
300	-0.00087	-0.00079	-0.00075	-0.00043	0.000031	0.000265	0.000643
400	-0.00066	-0.00062	-0.00057	-0.00032	0.000025	0.000151	0.000737
500	-0.00052	-0.00047	-0.00045	-0.000251	0.000016	0.000134	0.000474
1000	-0.00025	-0.00023	-0.00022	-0.000127	-0.000060	0.000062	0.000239
2000	-0.00013	-0.00012	-0.00011	-0.000064	-0.00009	0.000043	0.000108

Table 4

Procedure for computing D*:

4.1: Given a random sample $\boldsymbol{X} = (X_1, \ldots, X_n)$, compute the sample mean $\bar{x}$ and sample variance s^2 of $\boldsymbol{X}$.

4.2: Randomly generate samples $\boldsymbol{Y_i} = (Y_{1i}, \ldots, Y_{ni}),\ i = 1\ to\ m$ of size n from a normal distribution $N(\bar{x}\ , s^2)$.

4.3: Compute $\boldsymbol{U_i} = \boldsymbol{X} + \boldsymbol{Y_i}$ and $\boldsymbol{V_i} = \boldsymbol{X} - \boldsymbol{Y_i}$ for $i = 1\ to\ m$. (m =100)

4.4: Compute D* = $\frac{1}{m}\sum_{i=1}^{m} D_i$ where, D_i is the Hoeffding's D-statistic computed between each U_i and $V_i, i = 1\ to\ m.$

4.5: Reject the null hypothesis of normality if D* > $D_\propto$, where $D_\propto$ are critical points of the D* statistic developed under the null hypothesis of normality.

Simulation Results

The power of the Enriched method for testing normality when the population mean and variance was compared to the power of the Shapiro - Wilk's and the Anderson - Darling's test for normality. These are the two most popular and possibly most powerful tests for normality when the mean and variance of the population are unknown. The tests were compared for various alternative distributions by running 10,000 repetitions for each sample size and then the estimated power computed for α = 5%. Table 6 summarizes the simulated power for selected alternative distributions for $\alpha = 5\%$ level of significance with corresponding graphs in Figure 2.

From the table and graphs below it is clear that the test EM is on par with other traditional methods of testing for normality. The addition of the generated data to the observed data to develop the test statistic is justified from the results obtained. The EM method had better power than the AD test for some of the distributions investigated. Against a Beta (2, 2), the EM test had consistently better than the AD test and had better power for smaller sample sizes than the SW test. For a sample size of n =10, the EM method had a power of 12% while the AD and SW test had power of about 7%. The EM had power 80% for a sample size of 200 while the AD test was at 71% and the SW 92%. Figure 2(a) shows a graphical representation of the power of the three methods against the Beta (2, 2) distribution.

Against a Uniform (0, 1) distribution the EM method was uniformly more powerful than the AD test and was just as powerful as the Shapiro in rejecting normality. Figure 2(b) shows all three methods having adequate power for a sample size of 100 with the EM method having greater power for small sample sizes. Against the T distribution with n = 7 degrees of freedom, both the AD and SW tests have better power at smaller sample size than the EM test. A similar conclusion can be drawn from a Cauchy distribution where both the AD and SW tests have slightly better power than the EM.

Against asymmetric distributions, all of the methods in general require much smaller sample size to reject normality than against symmetric distributions. Against a Weibull (2, 2) distribution the EM test had uniformly better power than the AD test, but both methods still achieved less power than the SW test. Against a Chi-square (4) distribution both the AD and SW test have greater power than the Enriched method for all sample sizes. The same trend is seen against a Lognormal (0, 1) and exponential (λ = 1) distribution.

Table 5. Comparison of Power of the Enriched method, Shapiro-Wilk's and Anderson Darling's tests against some alternative symmetric distributions for (α = 5%) level of significance.

Distribution	**Sample size**	**SW**	**AD**	**EM**
Beta (2,2)	10	0.032	0.030	0.071
	20	0.072	0.068	0.128

	50	0.148	0.148	0.257
	100	0.448	0.334	0.558
	200	0.926	0.714	0.818
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
T (7)	10	0.087	0.092	0.047
	20	0.165	0.120	0.063
	50	0.243	0.184	0.130
	100	0.356	0.300	0.239
	200	0.577	0.421	0.425
	500	0.874	0.810	0.840
	1000	0.996	0.980	1.000
	2000	1.000	1.000	1.00
Uniform (0,1)	10	0.081	0.082	0.138
	20	0.190	0.154	0.367
	50	0.765	0.542	0.750
	100	0.996	0.960	0.988
	200	1.000	1.000	1.000
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Cauchy (0, 1)	10	0.602	0.628	0.401
	20	0.856	0.874	0.804
	50	0.998	0.998	0.996
	100	1.000	1.000	1.000
	200	1.000	1.000	1.000
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Chi-square (4)	10	0.242	0.238	0.113
	20	0.556	0.506	0.429
	50	0.935	0.928	0.886
	100	1.000	1.000	0.995
	200	1.000	1.000	1.000
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Lognormal (0, 1)	10	0.618	0.598	0.412
	20	0.938	0.918	0.890
	50	1.000	1.000	1.000

	100	1.000	1.000	1.000
	200	1.000	1.000	1.000
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Exponential (λ =1)	10	0.452	0.374	0.246
	20	0.842	0.790	0.753
	50	1.000	0.998	0.996
	100	1.000	1.000	1.000
	200	1.000	1.000	1.000
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Weibull (2, 2)	10	0.074	0.066	0.174
	20	0.166	0.140	0.270
	50	0.386	0.288	0.302
	100	0.804	0.616	0.633
	200	0.996	0.934	0.906
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000
Gamma (4, 5)	10	0.141	0.146	0.069
	20	0.292	0.252	0.207
	50	0.693	0.546	0.530
	100	0.952	0.904	0.888
	200	0.999	0.997	0.994
	500	1.000	1.000	1.000
	1000	1.000	1.000	1.000
	2000	1.000	1.000	1.000

Table 5

Figure 2(a). Comparison of power of various tests against Beta (2, 2) (α = 0.05)

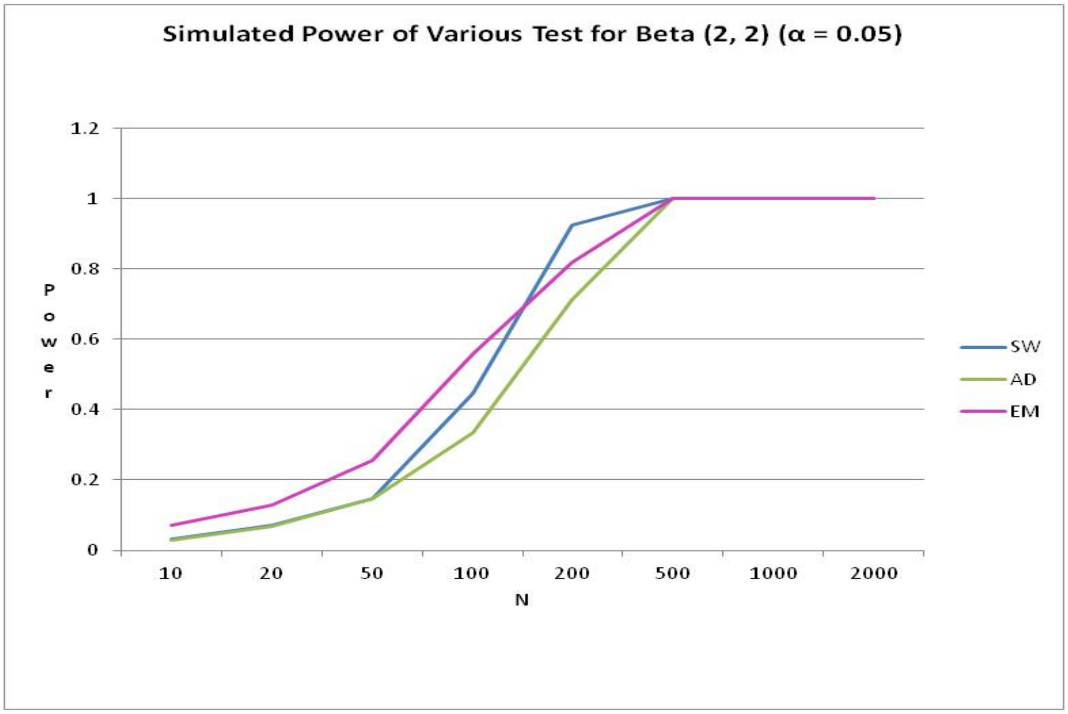


Figure 2(b). Comparison of power of various tests against T (7) (α = 0.05)

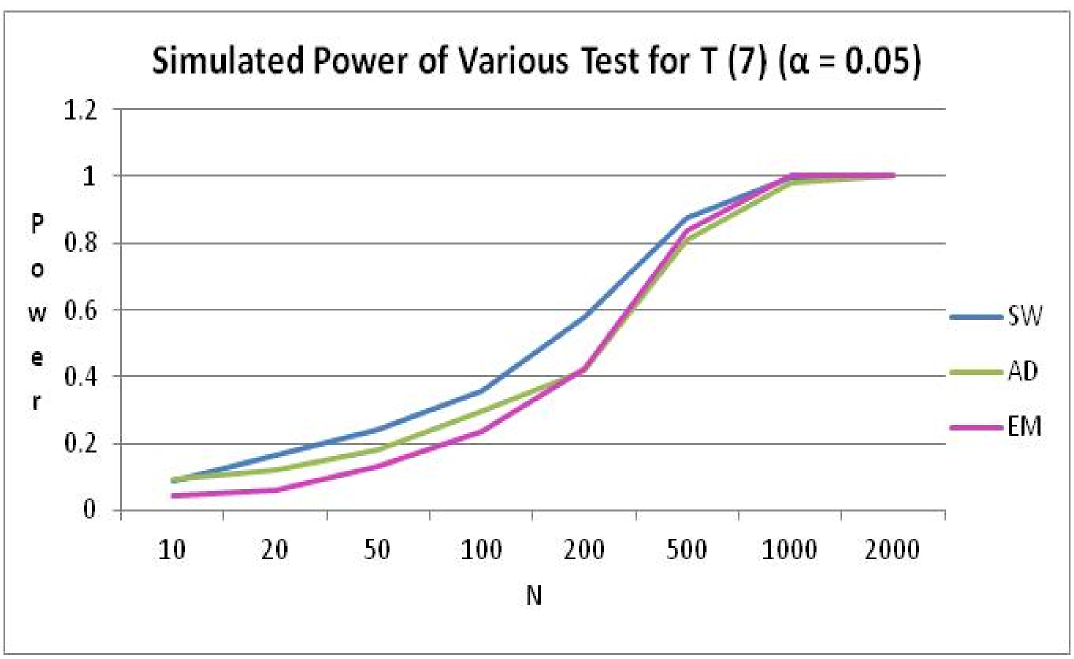


Figure 2(c). Comparison of power of various tests against Uniform (0, 1) (α = 0.05)

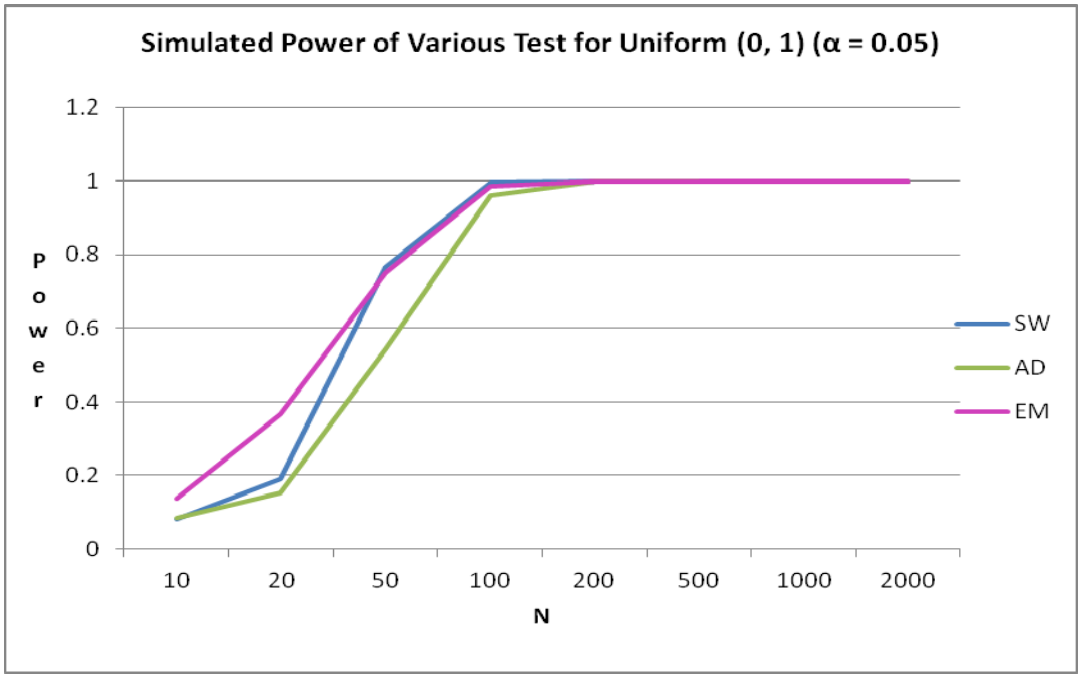


Figure 2(d). Comparison of power of various tests against Gamma (4, 5) (α = 0.05)

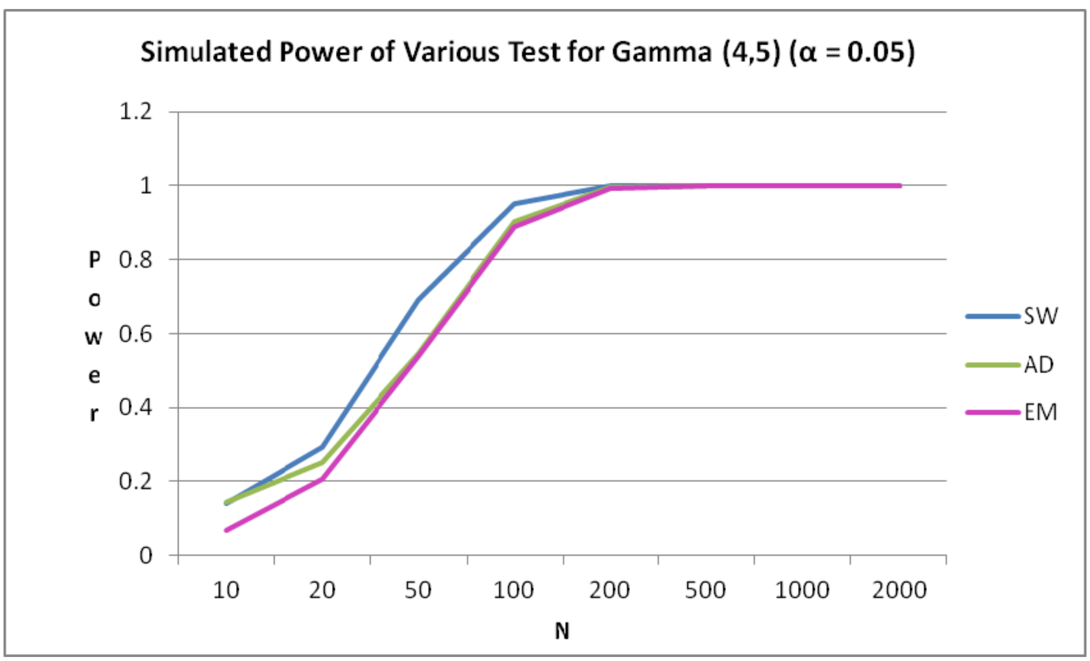
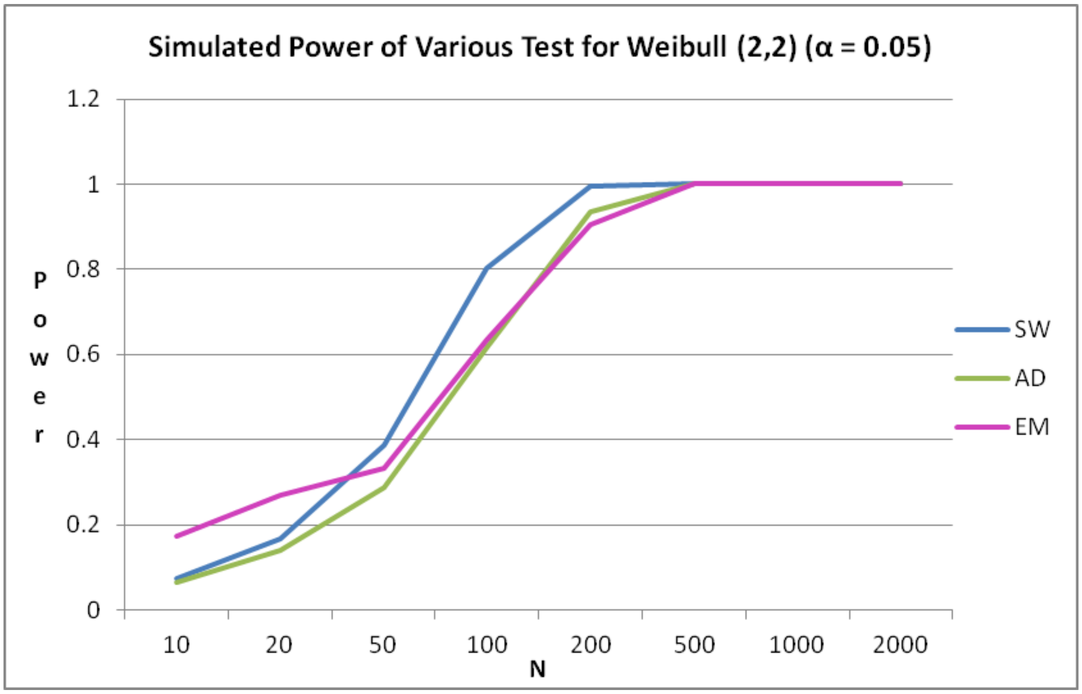


Figure 2(e). Comparison of power of various tests against Weibull (2, 2) (α = 0.05)



V. DISCUSSION

Overall, the simulation reveals the EM method having good potential when compared to the SW and AD tests. It also had comparative power to the KS test when testing for normality when population mean and variance are known. The method has better power than the AD test for many distributions that were investigated. However, it does not exclusively outperform the Anderson Darling's test and is less powerful than the Shapiro-Wilk's test for most of the distributions investigated. The test typically does not have great power for small sample sizes against symmetric distributions but achieves good power when sample size gets moderately large.

The test is not isolated from other tests of normality in its inability to achieve great power for small sample sizes as all of the methods of testing for normality typically require large sample size before they reach acceptable power against most symmetrical distributions. The inability to achieve good power for small sample size is a major limitation that poses concern to experimenters and is often the criticism of goodness-of-fit tests. In order to effectively test for normality these tests should be used in conjunction with graphical techniques to properly conclude normality or non-normality.

One observed limitation of incorporating simulated data to test for normality when the population mean and variance are known is an increase in the type 1 error. The result is due to the fact that generated data has to be incorporated with the observed data in order to perform the test. The impact is more profound at smaller sample sizes where there is a tendency to have slightly larger deviations in the sample mean and variance from the actual mean and variance of the population from which the sample was generated. Such deviations are in essence a slight breakdown of the theorem being used which assumes that the mean of the observed sample and that of the generated sample are the same.

When the deviation occurs, the test statistic tends to be larger since independence is marginally violated thus increasing the likelihood of rejecting the null hypothesis. Therefore even if the data is generated from a normal population with mean μ and variance σ^2 the test will sometimes reject normality if the generated data and the observed data have sample means and or sample variance which are far from the hypothesized population mean and variance. This error can be reduced by generating more samples and incorporating them with the observed sample.

Recommendations

The use of simulated data combined with observed data shows great potential. More work can be done to add to the investigations in this paper. In the case of testing normality when the mean and variance are unknown, instead of using the sample estimates of the mean and variance to generate the data one could estimate the parameters by using bootstrapping. This may lead to better generation results as these estimates may potentially be better estimates of the true population mean and variance than the sample mean and sample variance. Also, investigations could be done into the use of another measure to detect departure from independence other than the Hoeffding's -D that was used in this paper. If there is indeed a statistic that would detect more subtle deviations from independence than the Hoeffding's - D then this might serve to improve the power of testing when using the combination of simulated data and observed data.

Conclusion

The use of simulated data combined with observed data was investigated as a means of testing for normality. The investigation was done for two instances. The first instance when the population mean and variance are known and the other when they are unknown. The test was developed based upon the characterization theorem of the normal distribution by Bernstein (1941) and combines observed data and generated data to compute the test statistic D*. The statistic D* is based upon the Hoeffding's D-statistic which tests for independence of linear combinations of the observed and generated data.

Overall it was revealed that the use of Enriched data as a means of testing normality yielded some positive results. The test that was proposed achieved adequate power for many of the distributions investigated and was generally comparable to some of the existing methods of testing for normality. More work needs to be done to refine the method and possibly extend the use of simulated data to develop other goodness-of-fit tests. The test however cannot be carried out without the use of computer software and generally takes a longer time to run than existing methods.

VI. REFERENCES

1. Ahmad, I. and Mugdadi. A.R. (2003), "Testing Normality using Kernel Methods." *Journal of Non-Parametric Statistics*, Volume 15, Issue 3. pp 273-288.

2. Anderson T.W. and Darling D.A., (1954). "A Test of Goodness of Fit." *Journal of the American Statistical Association*, Vol. 49, No. 268, pp. 765-769.

3. Bernstein, S. (1941), "On a characteristic property of the normal law." *Leningrad Poly. Inst* (1941), pp. 21-22. *Collected Works 4* (1964), pp. 314-315.

4. Chen, L. and Shapiro, S. (1995). "An alternative test for normality based on normalized spacings." *J. Statist. Comput.* Simulation 53, 269-288

6 Conover, W.D. (1999), "Practical Non-Parametric Statistics." *Third edition, John Wiley and Sons*, Inc. New York, pp 428-433.

7 D'Agostino, R.B. and Pearson, E.S. (1973) "Test for Departure from Normality. Empirical results for the Distributions of b_2 and $\sqrt{b_1}$." *Biometrica*, Vol 60, No.3 pp. 613-622.

8 Duerinckx, M., Ley, C. and Swan, Y. (2012), "Maximum likelihood characterization of distributions," *Department of Mathematics,* University library of Bruxelles.

9 Galambos, J. and Kotz, S. (1978), "Characterizations of probability distributions," *Lecture Notes Math 675, Springer,* New York.

10 Hoeffding, W. (1948), "A nonparametric test of independence," *Ann Math. Stat.* Vol.19, pp. 546-557.

11 Hollander, M. and Wolfe, D.A. (1999), "Nonparametric Statistical Methods," *Wiley Series in Probability and Statistics*, Second Edition.

12 Kagan, A.M., Linnik, Y.V., Rao, C.R. (1973), "Characterization Problems in Mathematical Statistics," *Wiley*, ISBN 0-471-45421-4

13 Kakosyan, A.V., Klebanov, L.B and Melamed, J.A. (1984). "Characterization of distributions by the method of intensively monotone operators." *Lecture notes Math 1088, Springer-Verlag*, Berlin.

14 Lilliefor, H. W (1967), "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown." *Journal of the American Statistical Association*, Vol. 62, No. 318. pp. 399-402.

15 Lin, C.-C., Mudholkar, G.S. (1980), "A Simple Test for Normality against Asymmetric Alternatives" *Biometrika*, Vol. 67, pp. 455-461.

16 Massey, F. J. (1951), "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association*. Vol. 46, No. 253, pp. 68–78.

17 Quine, M.P. (1993), "On three Characterizations of the Normal distribution." *Probability and Mathematical Statistics*, Vol. 14, pp. 257-263.

18 SAS Institute Inc (2011), “Base SAS 9.3 Procedures Guide: Statistical Procedures.” *SAS Institute Inc*. *Cary, NC.* pp. 23.

19 Shapiro, S. S., and Wilk, M.B. (1965), “An Analysis of Variance Test for Normality (Complete Samples),” *Biometrica*, Vol. 52, pp. 591-611.

20 Stephens, M. A. (1974), “Edf statistics for goodness of fit and some comparisons,” *Journal of the American Statistical Association*, 69 (347):730–737.

21 Wlodzimierz, B. (2005), “Normal Distribution characterizations with applications,” *Lecture Notes in Statistics* 1995, Vol. 100. Revised June 7, 2005.