11-14-2014

An Alternative Goodness-of-fit Test for Normality with Unknown Parameters

Weiling Shi
amandashi3315@gmail.com

DOI: 10.25148/etd.FI14110735

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

AN ALTERNATIVE GOODNESS-OF-FIT TEST FOR NORMALITY WITH

UNKNOWN PARAMETERS

A thesis submitted in partial fulfillment of

the requirements for the degree of

MASTER OF SCIENCE

in

STATISTICS

by

Weiling Shi

2014

To: Interim Dean Michael R. Heithaus
College of Arts and Sciences

This thesis, written by Weiling Shi, and entitled an Alternative Goodness-of-Fit Test for Normality with Unknown Parameters, 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.

Gauri Ghai

Florence George

Zhenmin Chen, Major Professor

Date of Defense: November 14, 2014

The thesis of Weiling Shi is approved.

Interim Dean Michael R.Heithaus
College of Arts and Sciences

Dean Lakshmi N. Reddi
University Graduate School

Florida International University, 2014

DEDICATION

I dedicate this thesis to my parents. The completion of this thesis would not be possible without their love, support and encouragement.

ACKNOWLEDGMENTS

I am deeply grateful to my major professor and mentor Dr. Zhenmin Chen. With Dr. Chen's consistent support, guidance and help I finished my master's study at Florida International University. Dr. Chen's profession, responsibility and caring exert an important influence in my life. I benefited from Dr. Chen's wisdom and approach to research. All his virtues will be a valuable treasure of my whole life in the future.

I would also like to thank the members of my thesis committee, Dr. Gauri Ghai and Dr. Florence George, for their time, advice and reviewing my thesis.

In the end, I would like to thank my parents for their understanding, encouragement and love throughout the years. They gave me the warmest family one has ever wished for.

ABSTRACT OF THE THESIS

AN ALTERNATIVE GOODNESS-OF-FIT TEST FOR NORMALITY WITH UNKNOWN PARAMETERS

by

Weiling Shi

Florida International University, 2014

Miami, Florida

Zhenmin Chen, Major Professor

Goodness-of-fit tests have been studied by many researchers. Among them, an alternative statistical test for uniformity was proposed by Chen and Ye (2009). The test was used by Xiong (2010) to test normality for the case that both location parameter and scale parameter of the normal distribution are known. The purpose of the present thesis is to extend the result to the case that the parameters are unknown. A table for the critical values of the test statistic is obtained using Monte Carlo simulation. The performance of the proposed test is compared with the Shapiro-Wilk test and the Kolmogorov-Smirnov test. Monte-Carlo simulation results show that proposed test performs better than the Kolmogorov-Smirnov test in many cases. The Shapiro Wilk test is still the most powerful test although in some cases the test proposed in the present research performs better.

TABLE OF CONTENTS

LISTS OF TABLES

TABLE PAGE

LISTS OF FIGURES

FIGURE PAGE

CHAPTER I INTRODUCTION

1.1 Introduction

The goodness-of-fit test is a particular useful statistical model for testing whether observed data are representative of a particular distribution. A goodness-of-fit test can summarize the discrepancy between observed values and the values expected under any given model. Numerous research papers have been published by scientists concerning these tests. There are many existing test statistics including some commonly used goodness-of-fit tests such as the Chi-squared test (Pearson, 1900), the Kolmogorov-Smirnov test (Kolmogorov, 1933 and Smirnov,1939), the Cramer-Von Mises test (Cramer,1928 and von Mises), and the Anderson-Darling test (Anderson and Darling, 1952). All these commonly used statistical tests can be used to test normality.

The Chi-squared test is the most important member of the nonparametric family of statistical tests because it has some attractive features including the fact that it can be applied to any univariate distribution and calculated much easier than other test statistics. It is used for quantitative and binned data. For non-binned data, a histogram or frequency table should be constructed to put the data into the categories before the Chi-squared test is used. However, the values of the Chi-squared test are affected by skewness and kurtosis. Plus, it is sensitive to the sample size. The Chi-squared test has reduced power especially for the small sample size under 50.

The Kolmogorov-Smirnov test (K-S test) is also a nonparametric test for the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution, or to compare two samples. The K-S test relies on the fact that the value of the sample cumulative density function is asymptotically normally distributed. The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. However, the K-S test tends to be more sensitive near the center of the distribution than it is at the tails of the distribution. Additionally, the most serious limitation is that the distribution must be fully specified. If location, scale, and shape parameters are estimated from the data, the critical region of the K-S test is no longer valid. The K-S test statistic typically must be determined by simulation. Various studies have found that, even in this corrected form, the test is less powerful for testing normality than the Shapiro-Wilk test or the Anderson–Darling test.

The Anderson-Darling test is a modification of the K-S test which gives more weight to the tails of the distribution than the K-S test. The K-S test is “distribution free” in the sense that the critical values do not depend on the specific distribution being tested, while the Anderson-Darling test makes use of the specific distribution in calculating critical values. The Anderson-Darling test has the advantage of allowing a more sensitive test than the K-S test and the disadvantage that critical values must be calculated for each distribution.

The Shapiro-Wilk test, proposed by Samuel Sanford Shapiro and Martin Wilk in 1965, is used for testing normality and lognormal distributions. It compares the observed cumulative frequency distribution curve with the expected cumulative frequency curve. The Shapiro-Wilk test is based on the ratio of the best estimator of the variance to the usual corrected sum of squares estimator of the variance. The Shapiro-Wilk test is not as affected by ties as the Anderson-Darling test, but is still biased by sample size.

Power study of the most commonly used goodness-of-fit tests has been conducted by many researchers including Shapiro, Wilk and Chen (1968), and Aly and Shayib (1992). Some recent research papers concluded that the Shapiro–Wilk's test has the best power for a given significance, followed closely by Anderson-Darling when comparing the Shapiro-Wilk, Kolmogorov–Smirnov, Lilliefors, and Anderson-Darling tests. Although it was mentioned by Steele and Chaseling (2009) that none of the existing test statistics can be regarded as the "best" test statistic, to maximize the power of the test statistic for checking normality is still under explored and modified by many statistics researchers. The purpose of this thesis is to compare these tests.

In the present research, the statistical tests proposed in Chen and Ye (2009) and Xiong (2010) will be adopted and will be extended to test normality for the case that the location parameter and the scale parameter of the normal distribution are both unknown. A table for the critical values of the test statistics is provided using Monte Carlo simulation. The performance of the newly updated statistics are compared with the Shapiro-Wilk's test and the Kolmogorov-Smirnov test.

1.2 Basic Idea

Chen and Ye (2009) proposed a new test statistic (G statistic) for testing uniformity. The test statistic can be used to test if the underlying population distribution is a uniform distribution. On the basis of the probability integral transformation (See F.N. David and N.L. Johnson, 1948), the underlying population distribution can be any distribution. Suppose $x_1, x_2, \dots, x_n$ are the observations of a random sample from a population distribution with distribution function *F(x)*. Suppose also that $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ are the corresponding order statistics. The purpose is to test:

$$H_0 : F(x) = F_0(x).$$

$$H_1 : F(x) \neq F_0(x).$$

It can be seen that $F(x_{(1)}), F(x_{(2)}), \dots, F(x_{(n)})$ are the ordered observations of a random sample from the $U(0,1)$ distribution. The G test statistics can be used to conduct the following test procedure. The test statistics can be defined as

$$G^*(x_{(1)}, x_{(2)}, \dots x_{(n)}) = \frac{(n+1)\sum_{i=1}^{n+1}(F_0(x_{(i)}) - (F_0(x_{(i-1)}) - \frac{1}{n}))^2}{n}. \tag{1}$$

H_0 should be rejected at significance level α if $G^*(x_{(1)}, x_{(2)}, \dots, x_{(n)}) > G^*_\alpha$. Here G^*_α is the upper critical value of the G^* statistic. The value of G^*_α is calculated by the Monte Carlo simulation. For simplicity, $G^*(x_{(1)}, x_{(2)}, \dots, x_{(n)})$ can also be expressed as

$$G^*(x_{(1),} x_{(2)}, \dots x_{(n)}) = \frac{n+1}{n}\sum_{i=1}^{n+1}(F_0(x_{(i)}) - F_0(x_{(i-1)}))^2 - \frac{1}{n}. \tag{2}$$

Expression (2) will be used in the Monte Carlo simulation. The test can be used for testing any hypothesized distribution. The normal distribution is merely a special case.

The range of the function $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ is from 0 to 1 and the mathematical expectation and variance of the test statistic have been given in Chen and Ye (2009).

In Xiong's research, the parameters including the expected value and the standard deviation of the normal distribution are assumed to be known. When the parameters of the distribution are unknown, the test is no longer valid.

To solve this problem, Lilliefors' idea is adopted here to treat the case with unknown parameters. Estimation of the population mean and population variance derived from the sample data is conducted before calculating statistic.

The procedure in this research for simulating the critical values of the G^* test statistic is summarized as follows:

1. Generate a pseudo random sample $x_1, x_2, ..., x_n$ of size n from the standard normal distribution;
2. Calculate the sample mean ($\bar{x}$) and variance (s^2);
3. Find the ordered values $x_{(1)}, x_{(2)}, ..., x_{(n)}$ and define $x_{(0)} = 0$ and $x_{(n+1)} = 1$;
4. Calculate $F(x_{(1)}), F(x_{(2)}), ..., F(x_{(n)})$. Here $F(x)$ is the cumulative distribution function of the distribution;
5. Calculate the value of $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ using Equation (2);
6. Repeat steps 1 to 5 k times (k=1,000,000 in this research);
7. Sort all the values of G^* in ascending order;

8. Find the critical values with α= 0.1, 0.05, 0.01, 0.005, 0.001, that is, to calculate the 90th, 95th, 99th, 99.5th and 99.9th percentiles.

The procedure shown above uses the standard normal distribution. It will not affect the simulation result. In fact, it can be shown that it remains invariant when the parameters of the normal distribution change.

Suppose X has a normal distribution with the mean μ and standard deviation σ.

$$G^*(x_{(1),}x_{(2)},...,x_{(n)}) = \frac{n+1}{n}\sum_{i=1}^{n+1}(F_0(x_{(i)}) - F_0(x_{(i-1)}))^2 - \frac{1}{n}$$

$$= \frac{n+1}{n}(\sum_{i=1}^{n+1}(\int_{-\infty}^{x_{(i)}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt)^2 - \int_{-\infty}^{x_{(i-1)}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt)^2) - \frac{1}{n}.$$

Let $z = \frac{t-\mu}{\sigma}$.

The value of $G^*(x_{(1),}x_{(2)},...,x_{(n)})$ becomes

$$\frac{n+1}{n}(\sum_{i=1}^{n+1}(\int_{-\infty}^{z_{(i)}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \sigma dt)^2 - \int_{-\infty}^{z_{(i-1)}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}} \sigma dt)^2) - \frac{1}{n}$$

$$= \frac{n+1}{n}(\sum_{i=1}^{n+1}(\int_{-\infty}^{z_{(i)}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dt)^2 - \int_{-\infty}^{z_{(i-1)}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dt)^2) - \frac{1}{n}.$$

This is the same as in the case that the standard normal distribution is picked.

The performance of the G^* test statistics is compared with the Shapiro-Wilk's test and the K-S test for testing normality. Since the Shapiro-Wilk's test can only be used for testing normal distribution and lognormal distribution, the normality test for the power comparison is conducted in this research.

Chapter 2 outlines the method for calculating these three test statistics. The power study results of these three tests are analyzed in Chapter 3. Chapter 4 concludes the performance comparisons of the performance comparisons with the G test, the Shapiro-Wilk's test and the Kolmogorov-Smirnov test.

Monte Carlo simulation was used to conduct power study for these three tests. The computer programming languages used in this research are SAS/IML and SAS/Base.

CHAPTER II METHODOLOGY

The performance of a test statistic can be evaluated by a power study. To evaluate the performance of test statistic $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ proposed in this thesis, various alternative distributions are used to find the power of the test statistic. The G^* test power is compared with the Shapiro-Wilk test and the Kolmogorov-Smirnov test in the present research. The null hypothesis assumes that the underlying distribution is a normal distribution, while the alternative hypothesis assumes a distribution that is not a normal distribution. The alternative distributions used here include the triangle distributions, V-shaped triangle distributions, and Beta distributions.

2.1 Methodology of Three Tests

2.1.1 G test

Suppose $x_1, x_2, ..., x_n$ are the observations of a random sample from a population distribution with a distribution function $F(x)$. Suppose also that $x_{(1)}, x_{(2)}, ..., x_{(n)}$ are the corresponding order statistics. To test whether or not the underlying distribution is a normal distribution, the null and alternative hypotheses are

H_0: The population distribution is a normal distribution,

H_α: The population distribution is not a normal distribution.

As discussed in Chapter 1, when the test statistics $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ is used, H_0 should be rejected at significant level α if $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)}) > G^*_\alpha$, where G^*_α is the critical value of the G^* test statistic. The value of $G^*(x_{(1)}, x_{(2)},..., x_{(n)})$ can be calculated using equation (2) for convenience.

2.1.2 Shapiro-Wilk Test

The Shapiro-Wilk test utilizes the null hypothesis principle to determine whether a sample $x_1, x_2,..., x_n$ come from a normally distributed population. The W test statistic is the ratio of the best estimator of the variance (derived from the square of a linear combination of the order statistics) to the usual corrected sum of squares estimator of the variance (Shapiro and Wilk; 1965). When n is greater than three, the coefficients to compute the linear combination of the order statistics can be approximated by the method of Royston (1992). The statistic W is always greater than zero and less than or equal to one ($0 < W < 1$). Small values of W lead to the rejection of the null hypothesis of normality. The distribution of W is highly skewed. Seemingly large values of W (such as 0.90) may be considered small will result in rejecting the null hypothesis. Research papers show that the Shapiro-Wilk test has the better performance compared with the Anderson-Darling test and the Kolmogorov-Smirmov test.

The test statistic is

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n}$$

is the sample mean; $x_{(i)}$ is the $i\ th$ order statistic; the constants $a_i^{'}s$ are given by

$$(a_1, a_2, ..., a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{\frac{1}{2}}}.$$

Here $m_1, m_2, .., m_n$ are the expected values of the order statistics of independent and identically distributed random variables sampled from the standard normal distribution and V is the covariance matrix of those order statistics. Reject H_0 if W is too small.

To compute the value of test statistic W for a given complete random sample $x_1, x_2, ..., x_n$, the procedure proposed in Shapiro &Wilk (1965) is as follows:

(1) Order the observations to obtain an ordered sample $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$.

(2) Compute $\sum_{i=1}^{n} (x_{(i)} - \bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$, where $\bar{x}$ is the sample mean.

(3) (a) If n is even, $n = 2m$, compute $b = \sum_{i=1}^{m} a_{n-i+1} (x_{(n-i+1)} - x_{(i)})$, where the values of a_{a-i+1} are given in Shapiro and Wilk (1965).

(b) If n is odd, $n = 2m + 1$ the computation is the same as the one in (3)(a) since $a_{m+1} = 0$. Thus $b = a_n (x_{(n)} - x_{(1)}) + ... + a_{m+2} (x_{(m+2)} - x_{(m)})$, where the value of $x_{(m+1)}$, the sample median, does not enter the computation of b.

(4) Compute $W = b^2 / (\sum_{i=1}^{n} (x_i - \bar{x})^2)$.

(5) Compare with the critical values from quantiles of the Shapiro-Wilk test for

normality table. If the calculated value of the test statistic W is smaller than W_α , H_0 is rejected at the significance level α .

2.1.3 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test statistic is defined as

$$D_n = \sup_x \left| F_n^*(x) - F(x) \right| = \max(D_n^+, D_n^-),$$

$$D_n^+ = \sup_x \left[F_n^*(x) - F(x) \right],$$

$$D_n^- = \sup_x \left[F\ (x) - F_n^*(x) \right].$$

Here F is the cumulative distribution function specified by the null hypothesis. Let $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ be the order statistic of $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. Then the empirical distribution function is

$$F_n^*(x) = \frac{i}{n} \text{ for } x_{(i)} \leq x < x_{(i+1)} \quad (i = 1,2,\ldots,n).$$

$$D_n^+ = \max_{0 \leq i \leq n} \sup_{x_{(i)} \leq x < x_{(i+1)}} \left\{ \frac{i}{n} - F(x) \right\} = \max_i \left\{ \frac{i}{n} - \inf_{x_{(i)} \leq x < x_{(i+1)}} F(x) \right\}$$
$$= \max_{0 \leq i \leq n} \left\{ \frac{i}{n} - F(x_{(i)}) \right\} = \max \left\{ \max_{1 \leq i \leq n} \left[\frac{i}{n} - F(x_{(i)}) \right], 0 \right\}$$

Similarly,

$$D_n^- = \max \left\{ \max_{1 \leq i \leq n} \left[F(x_{(i)} - \frac{i-1}{n}) \right], 0 \right\}$$

If the calculated value of the test statistic is greater than D_α , H_0 is rejected at the stated significance level. Here D_α is the critical value of the Kolmogorov-Smirrnov test statistic.

The popularity of the Kolmogorov-Smirnov test relates to the fact that the test does not depend on the underlying cumulative distribution function being tested. However, its disadvantages also limit its application sometimes because it only applies to continuous distributions. The test statistic becomes more sensitive near the center of the distribution than at the both tails. All the parameters such as location, scale and shape of a distribution must be fully specified. If not, the K-S test is no longer valid. It must be estimated by simulation.

2.1.4 Monte Carlo Simulation Method

Monte Carlo simulation is applied to generate pseudo random samples from a variety of alternative distributions to compare the power of the G^* test , Shapiro-Wilk test and K-S test. Firstly, k pseudo random samples of size n are generated from a specified distribution. In this research, V-shape triangle distribution, triangle distribution and the Beta distribution are chosen. For each pseudo random sample, the observations $x_1, x_2, \ldots, x_n$ are sorted and become order statistics one $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. Then put the ordered statistics into the formulas of the G^* test and Shapiro-Wilk test, and the K-S test. The values of the test statistics can be computed. By comparing the calculated value and the critical values of the G test, Shapiro-Wilk test and the K-S test, the rejection rates can be found.

In the present research, the sample sizes n =5, 10, 20, 30, 40, 50 are selected to conduct Monte Carlo simulation. The number of repetitions is selected to be $k = 1{,}000{,}000$ to ensure the accuracy of the power. The procedure for calculating the power is summarized as follows:

1. Generate a random sample $x_1, x_2, ..., x_n$ from the specified alternative distribution listed above;
2. Find the ordered values $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(n)}$ and define $x_{(0)} = 0$ and $x_{(1)} = 1$;
3. Calculate the corresponding $F_0(x_{(1)}), F_0(x_{(2)}), ..., F_0(x_{(n)})$,where F_0 is the cumulative distribution function of normal here and the expected value and standard deviation of the normal distribution are calculated from the pseudo random sample;
4. Calculate the value of $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ using equation (2);
5. Compare the value of $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ with the indicated critical value at significance level α=0.05, and determine whether H_0 is rejected;
6. Using the method mentioned above to calculate the value of the Shapiro-Wilk test statistic W ;
7. Compare the value of W with W_α at the same significance level as in step 5, and determine whether H_0 is rejected;
8. Using the method mentioned above to calculate the value of the K-S test statistic D ;
9. Compare the value of D with D_α at the same significance level, and determine whether H_0 is rejected;
10. Repeat steps 1 to 9 1,000,000 times;
11. Calculate the rejection rates for the G^* test, the Shapiro-Wilk test and the K-S test.

2.2 Power Study

The power of a statistical test is the probability that it correctly rejects the null hypothesis when the null hypothesis is false. That is, Power = P (reject null hypothesis/ null hypothesis is false) which can be denoted as $\pi = 1 - \beta$ where β is the probability of committing type II error.

The power of the test statistic in this research can be presented as

$P_{H_\alpha}(G^* > G_\alpha) = P(G^* > G_\alpha | H_\alpha)$ for G test;

$P_{H_\alpha}(W < W_\alpha) = P(W < W_\alpha | H_\alpha)$ for Shapiro-Wilk test;

$P_{H_\alpha}(D > D_\alpha) = P(D < D_\alpha | H_\alpha)$ for Kolmogorov-Smirnov test;

The power estimate is high means that the performance of the test is good. Statistical power may depend on a number of factors. Some of these factors may be particularly because of a specific testing situation, but at a minimum, power always depends on the following three factors: sample size, the significance level, and the sensitivity of the data.

The rejection rates of the three test statistics are used as estimates of their power in the thesis. High rejection rates means the power of the test statistic is high. To evaluate the performance of the test statistic $G^*(x_{(1)}, x_{(2)}, ..., x_{(n)})$ proposed in this research, various alternative distributions including V-shape triangle distribution, Beta distributions and triangle distributions are used to study the power of this test statistic.

To find the powers of the G^* test, Shapiro-Wilk test and Kolmogorov-Smirnov test, Monte Carlo simulation was used to generate pseudo random samples from the various alternative distributions. To accomplish this, k pseudo random samples of size n are generated from a specified distribution. For each pseudo random sample, the observation $x_1, x_2, ..., x_n$ are sorted and the sorted observations become $x_{(1)}, x_{(2)}, ..., x_{(n)}$ .Then the values of the test statistics can be computed for all three tests. Finally, the rejection rates or powers for the G^* test, Kolmogorov-Smirnov test and Shapiro-Wilk test are calculated.

In the present research, the sample sizes n = 5 to 50 are selected to conduct the Monte Carlo simulation. In order to ensure the accuracy of the power study, the number of the repetitions is selected to be k= 1,000,000.

CHAPTER III POWER COMPARISON

3.1 Alternative Distributions

3.1.1 V-shaped Triangle Alternative Distributions

The probability density function of the V-shaped triangle distribution is

$$f(x)=\begin{cases} 2-\dfrac{2x}{h} & 0\le x\le h \\ 2-\dfrac{2(1-x)}{1-h} & h<x\le 1 \\ 0 & elsewhere. \end{cases}$$

Here h is a constant between 0 and 1. The following V-shaped triangle distributions are used in this research:

Alternative Distribution 1

Consider h = 0.25. This is a left-skewed V-shaped triangle distribution. The power comparison result under this V-shaped triangle distribution is shown in Table 2. It can be found that the G^* test is performs better than Shapiro-Wilk test when sample size is 5. When sample size increases, the power of these three tests also increases. Compared with the Kolmogorov-Smirnov test, the G^* test outperforms the Kolmogorov-Smirnov test in all cases. The Shapiro-Wilk test performs better than the other two tests when the sample size becomes large.

Alternative Distribution 2

Consider h = 0.5. This is a symmetric V-shaped triangle distribution. The power comparison result under this V-shaped triangle distribution is shown in Table 3. The result is similar to the previous case. The G^* test performs better than Shapiro-Wilk test when sample size is 5 and is more powerful than Kolmogorov-Smirnov test for all sample sizes. When sample size increases, the power of Shapiro-Wilk test increase faster than the other two test statistics.

Alternative Distribution 3

Consider h = 0.75. This is a right-skewed V-shaped triangle distribution. The power comparison result under this V-shaped triangle distribution is shown in Table 4. It shows that the G^* test is performs better than Shapiro-Wilk test also when sample size is 5. Shapiro-Wilk test performs very well when sample size increases. The G^* test still outperforms the Kolmogorov-Smirnov test in all cases.

Since there is no function call of V-shape triangle distribution in SAS, the following proposition is needed.

Proposition:

Suppose U is a random variable with uniform distribution on interval (0, 1). Then $H(U)$ has a V-shaped triangle distribution with parameter h. Here $H(u)$ is defined as

$$H(u) = \begin{cases} h - \sqrt{h^2 - hu} & 0 \le u < h \\ h + \sqrt{h^2 - h + (1-h)u} & h \le u \le 1. \end{cases}$$

Let $X = H(U)$. Then the cumulative distribution of X is

$$F_X(x) = P(X \le x) = P(H(U) \le x) = \begin{cases} P(h - \sqrt{h^2 - hu} \le x) & 0 \le \mu < h \\ P(h + \sqrt{h^2 - h + (1-h)U} \le x) & h \le x \le 1 \end{cases}$$

$$= \begin{cases} 0 & x < 0 \\ P(U \le 2x - \dfrac{x^2}{h}) & 0 \le x < h \\ P(U \le \dfrac{1}{1-h}x^2 + \dfrac{2h}{h-1}x + \dfrac{h}{1-h}) & h \le x \le 1 \\ 1 & x > 1. \end{cases}$$

That is,

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 2x - \dfrac{x^2}{h} & 0 \le x < h \\ \dfrac{1}{1-h}x^2 + \dfrac{2h}{h-1}x + \dfrac{h}{1-h} & h \le x \le 1 \\ 1 & x > 1. \end{cases}$$

Then the pdf of X (denoted as $f(x)$) is:

$$f(x) = \begin{cases} 2 - \dfrac{2x}{h} & 0 \le x < h \\ 2 - 2\dfrac{(1-x)}{1-h} & h \le x \le 1 \\ 0 & elsewhere, \end{cases}$$

which is the probability density function of V-shape triangle distribution with parameter h.

3.1.2 Beta Alternative Distributions

The probability density function of the Beta distribution is

$$f(x) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1 \\ 0 & elsewhere \end{cases} \qquad (\alpha > 0, \beta > 0).$$

The following special cases of the beta distributions are used in the power study:

Alternative Distribution 4

$B(4,2)$ distribution. This is a left-skewed Beta distribution. The power comparison result under this Beta alternative distribution is shown in Table 5. It can be found that the G^* test performs better than Shapiro-Wilk test under small sample size case when n = 5. When sample size increases, the power of these three tests increases too. The Shapiro-Wilk test increases more than G^* test. The Shapiro-Wilk test performs still well in most cases. However, the G^* test is better than Shapiro-Wilk test for small sample size when $n = 5$. Kolmogorov-Smirnov test outperforms the G^* test under this distribution.

Alternative Distribution 5

$B(0.5,0.5)$ distribution. This is a symmetric bathtub-shaped Beta distribution. The power comparison result under this Beta alternative distribution is shown in Table 6. It can be found from the table that G^* test performs better than Shapiro-Wilk test under small sample size when n =5. The G^* test is still more powerful than the Kolmogorov-Smirnov test in all cases.

Alternative Distribution 6

$B(2,4)$ distribution. This is a right-skewed Beta distribution. The power comparison result under this Beta alternative distribution is shown in Table 7. The power comparison result shows that G^* test performs better than Shapiro-Wilk test under small sample size when n = 5. G^* test is more powerful than the Kolmogorov-Smirnov test when sample size is 10,20,30 ,40,50 except the case of sample size 5

Alternative Distribution 7

$B(1,1)$ distribution. This is actually a uniform distribution. The power comparison result under this Beta alternative distribution is shown in Table 8. It can be seen from the table that G^* test performs better than Shapiro-Wilk test under small sample size case including n = 5 and 10. The G^* test is also more powerful than the Kolmogorov-Smirnov test when sample size is 10, 20, 30, 40, 50.

3.1.3 Triangle Alternative Distribution

The probability density function of the triangle distribution is

$$f(x) = \begin{cases} \dfrac{2x}{h} & 0 \le x \le h \\ \dfrac{2(1-x)}{1-h} & h < x \le 1 \\ 0 & elsewhere. \end{cases}$$

Here h is a constant between 0 and 1. The following triangle distributions are used in the power study.

Alternative Distribution 8

Consider h=0.75. This is a left-skewed triangle distribution. The power comparison result of this alternative distribution is shown in Table 9. It can be found form the figure that the G^* test performs better than Shapiro-Wilk test when sample size n =5 and 10. The Shapiro-Wilk size performs well when the sample size increases. Kolmogorov-Smirnov test performs better than the G^* test in this case.

Alternative Distribution 9

Consider h=0.5. This is a symmetric triangle distribution. The power comparison result of this alternative distribution is shown in Table 10. The G^* test statistic performs the best in this case. G^* test is more powerful than the Shapiro-Wilk test when sample size is n = 5,10,20,30.

Alternative Distribution 10

Consider h=0.25. This is a right-skewed triangle distribution. The power comparison result of this alternative distribution is shown in Table 11. It can be found form the figure that the G^* test performs better than Shapiro-Wilk test when sample size n =5. The Kolmogorov-Smirnov test performs better than the G^* test in this case.

3.2 Summary of Power Comparison

From the above analysis, we can conclude the following:

- For all the above alternative distributions, the G^* test statistics performs better than the Shapiro-Wilk test for small sample size;
- For all the V-shaped alternative distributions, including three V-shaped triangle distributions and the left-skewed, bathtub shaped, right-skewed Beta distribution, the G^* test statistics performs better than the Shapiro-Wilk test for small sample size.
- For all the V-shaped alternative distributions and bathtub shaped Beta distributions, the G^* test statistics performs better than the Kolmogorov-Smirnov test for all cases;

- For the symmetric triangle alternative distribution, the G^* test statistics performs the best among all these cases conducted. It performs better than the Shapiro-Wilk test does when sample sizes are 5,10,20,30 40;
- For the left-skewed and right-skewed triangle alternative distributions, the G^* test performs better than the Kolmogorov-Smirnov test for small sample sizes;
- For the uniform alternative distribution, the G^* test statistics performs better than the Kolmogorov-Smirnov test and shows similar power to the Kolmogorov-Smirnov test when sample size increases;
- For the uniform alternative distributions, the G^* test statistics performs better than the Shapiro-Wilk test when sample size is less than 10.

CHAPTER IV CONCLUSION AND DISCUSSION

The goodness-of-fit test is a statistical procedure to measure the discrepancy between observed values and the values expected under a specific distribution. The goal of the goodness-of-test is to check whether the underlying probability distribution differs from a hypothesized distribution. There are many existing test statistics including some commonly used goodness-of-fit tests such as the Shapiro-Wilk test, Kolmogorov-Smirnov test, Anderson-Darling test and Cramer-Von Mises test. All these commonly used statistical tests can be used to test normality. Among them, an alternative statistical test G test for uniformity was proposed by Chen and Ye (2009). The test was used by Xiong (2010) to test normality for the case that both location parameter and the scale parameter of the normal distribution are known. The purpose of this thesis is to extend the result to the case that the parameters are unknown.

Power study is conducted to compare the performance of this proposed test with the Shapiro-Wilk test and the Kolmogorov-Smirnov test. The result of the Monte Carlo simulation shows that the G^* test performs better than the Shapiro-Wilk test for small sample cases under all the alternative distributions used in this research. The G^* test also outperforms the Kolmogorov-Smirnov test in most of cases. It can also be found that the Shapiro-Wilk test performs better when the sample size increases. Since the computation of the G^* test statistic is less complicated than the Shapiro-Wilk test. Therefore, the G^* test statistics in this thesis is worth being recommended to be an alternative approach for testing normality, especially when the sample size is small. However, when sample size increases, its power does not increase as fast as the Shapiro-Wilk test does.

Since the Kolmogorov-Smirnov test and the G^* test can be used to test any hypothesized distribution, while Shapiro-Wilk test can be only used for checking normality and lognormality, the power comparison between them of normality test among the three tests is conducted in this research.

Extending the usage of the G^* test for the case that the parameters of the distribution are unknown is useful. When normality is tested, the mean and the variance of the distribution are usually unknown. For testing whether or not the underlying distribution of a data set belongs to a specified distribution family such as the normal distribution family, the exponential distribution family and so on, G^* test can still be used even if the parameters of the distribution are unknown.

Table 1 Critical Value of G test Statistic

n	$G^*_{0.100}$	$G^*_{0.050}$	$G^*_{0.010}$	$G^*_{0.005}$	$G^*_{0.001}$
5	0.1865	0.2190	0.2813	0.3026	0.3414
6	0.1664	0.1958	0.2548	0.2769	0.3190
7	0.1492	0.1757	0.2316	0.2524	0.2955
8	0.1349	0.1583	0.2100	0.2307	0.2743
9	0.1227	0.1440	0.1919	0.2115	0.2508
10	0.1127	0.1320	0.1757	0.1941	0.2340
11	0.1040	0.1215	0.1616	0.1784	0.2164
12	0.0965	0.1125	0.1495	0.1655	0.2024
13	0.0900	0.1047	0.1393	0.1542	0.1883
14	0.0841	0.0976	0.1296	0.1433	0.1759
15	0.0791	0.0917	0.1210	0.1340	0.1649
16	0.0705	0.0812	0.1065	0.1178	0.1445
17	0.0705	0.0812	0.1065	0.1178	0.1445
18	0.0669	0.0770	0.1008	0.1114	0.1370
19	0.0635	0.0729	0.0953	0.1052	0.1295
20	0.0605	0.0692	0.0904	0.1000	0.1225
21	0.0577	0.0660	0.0858	0.0948	0.1159
22	0.0552	0.0631	0.0817	0.0902	0.1109
23	0.0529	0.0603	0.0781	0.0860	0.1049
24	0.0507	0.0578	0.0745	0.0822	0.1004
25	0.0487	0.0554	0.0712	0.0784	0.0961
26	0.0469	0.0532	0.0682	0.0750	0.0916
27	0.0453	0.0516	0.0657	0.0721	0.0878
28	0.0437	0.0494	0.0631	0.0693	0.0845

29	0.0421	0.0476	0.0606	0.0664	0.0817
30	0.0408	0.0460	0.0586	0.0642	0.0783
31	0.0395	0.0445	0.0564	0.0617	0.0745
32	0.0382	0.0430	0.0544	0.0595	0.0718
33	0.0371	0.0416	0.0525	0.0575	0.0694
34	0.0360	0.0404	0.0509	0.0555	0.0673
35	0.0349	0.0392	0.0493	0.0537	0.0651
36	0.0340	0.0392	0.0493	0.0537	0.0624
37	0.0330	0.0370	0.0463	0.0506	0.0606
38	0.0322	0.0360	0.0450	0.0490	0.0591
39	0.0314	0.0350	0.0435	0.0472	0.0573
40	0.0305	0.0341	0.0424	0.0461	0.0552
41	0.0300	0.0332	0.0412	0.0448	0.0540
42	0.0291	0.0323	0.0401	0.0435	0.0521
43	0.0284	0.0315	0.0390	0.0424	0.0508
44	0.0277	0.0308	0.0380	0.0413	0.0490
45	0.0271	0.0301	0.03702	0.0401	0.0478
46	0.0265	0.0294	0.0362	0.0392	0.0465
47	0.0259	0.0287	0.0353	0.0382	0.0456
48	0.0254	0.0281	0.0344	0.0372	0.0446
49	0.0248	0.0275	0.0336	0.0363	0.0431
50	0.0243	0.0269	0.0329	0.0355	0.0422

Table 2 Power Comparison : V-Shape triangle (h=0.25)

n	G-Test	SW-Test	KS-Test
5	0.1680	0.1245	0.1224
10	0.2882	0.3954	0.2371
20	0.5390	0.8692	0.5330
30	0.7502	0.9884	0.7425
40	0.8915	0.9996	0.8811
50	0.9635	1	0.9554

Table 3 Power Comparison: V-Shape triangle (h=0.5)

n	G-Test	SW-Test	KS-Test
5	0.2462	0.1681	0.1576
10	0.5004	0.5307	0.3429
20	0.7751	0.9586	0.3443
30	0.9173	0.9992	0.7393
40	0.9770	1	0.9829
50	0.9956	1	0.9974

Table 4 Power Comparison: V-triangle (h=0.75)

n	G-Test	SW-Test	KS-Test
5	0.1690	0.1246	0.1209
10	0.2873	0.3947	0.2366
20	0.5378	0.8697	0.5335
30	0.7504	0.9884	0.7413
40	0.8916	0.9996	0.8804
50	0.9632	1	0.9556

Table 5 Power Comparison: Beta (α= 4, β=2)

n	G-Test	SW-Test	KS-Test
5	0.0575	0.0408	0.0528
10	0.0639	0.0650	0.0632
20	0.0718	0.1222	0.0979
30	0.07831	0.1856	0.1208
40	0.0869	0.2902	0.1512
50	0.0980	0.3935	0.1884

Table 6 Power Comparison: Beta (α= 0.5, β=0.5)

n	G-Test	SW-Test	KS-Test
5	0.1436	0.1045	0.0964
10	0.2135	0.2819	0.1526
20	0.3466	0.7301	0.3376
30	0.5043	0.9519	0.5026
40	0.6766	0.9969	0.6643
50	0.8271	0.9999	0.7971

Table 7 Power Comparison: Beta (α= 2, β=4)

n	G-Test	SW-Test	KS-Test
5	0.0577	0.0409	0.0527
10	0.0635	0.0650	0.0636
20	0.0714	0.1222	0.0981
30	0.0780	0.1848	0.1206
40	0.0869	0.2909	0.1519
50	0.0975	0.3924	0.1889

Table 8 Power Comparison: Beta (α= 1, β=1)

n	G-Test	SW-Test	KS-Test
5	0.0721	0.0451	0.0523
10	0.0921	0.0763	0.0626
20	0.1161	0.2033	0.1072
30	0.1394	0.4122	0.1434
40	0.1713	0.6869	0.1945
50	0.2111	0.8601	0.2574

Table 9 Power Comparison: Triangle (h=0.25)

n	G-Test	SW-Test	KS-Test
5	0.0614	0.0433	0.0552
10	0.0716	0.0712	0.0710
20	0.0837	0.1372	0.1220
30	0.0923	0.2147	0.1608
40	0.1026	0.3417	0.2090
50	0.1143	0.4604	0.2663

Table 10 Power Comparison: Triangle (h=0.5)

n	G-Test	SW-Test	KS-Test
5	0.0495	0.0334	0.0451
10	0.0512	0.0342	0.0417
20	0.0537	0.0336	0.0441
30	0.0538	0.0364	0.0401
40	0.0539	0.0559	0.0402
50	0.0550	0.0753	0.0416

Table 11 Power Comparison: Triangle (h=0.75)

n	G-Test	SW-Test	KS-Test
5	0.0612	0.0433	0.0556
10	0.0713	0.0708	0.0707
20	0.0833	0.1376	0.1219
30	0.0931	0.2154	0.1604
40	0.1030	0.3409	0.2097
50	0.1146	0.4598	0.2655

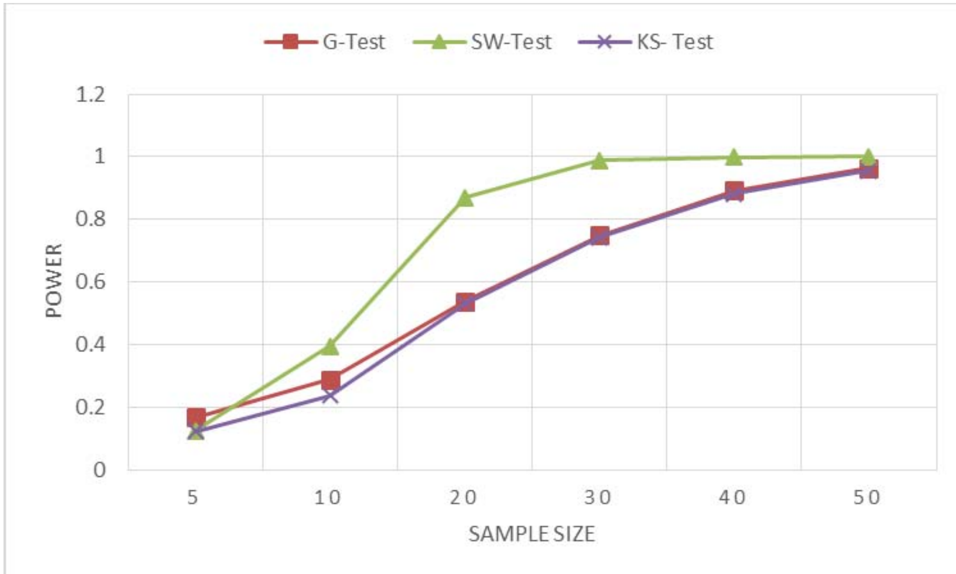

Figure 1 Power Comparison: V-Shape triangle (h=0.25)

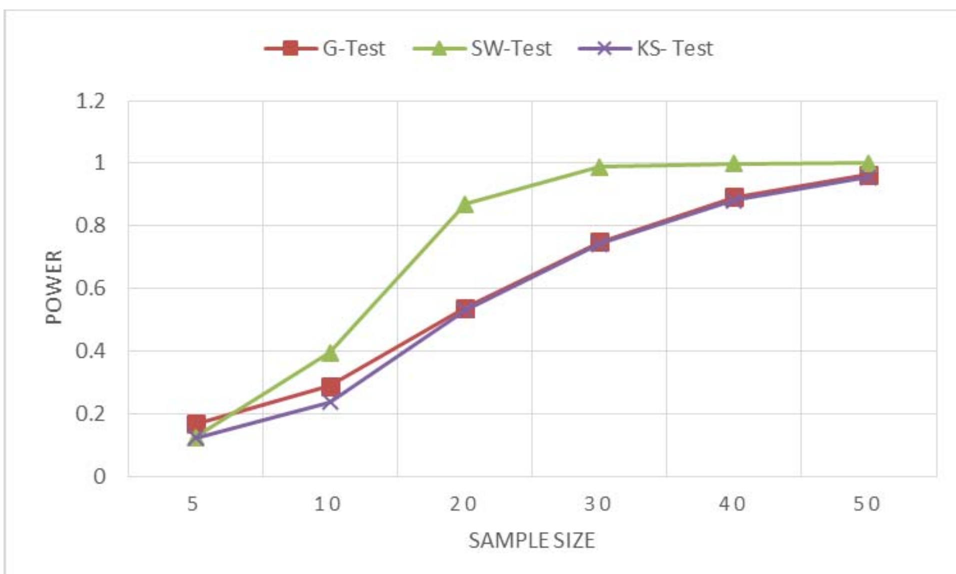

Figure 2 Power Comparison: V-Shape triangle (h=0.5)

Figure 3 Power Comparison: V-triangle (h=0.75)

Figure 4 Power Comparison: Beta ($\alpha = 4, \beta = 2$)

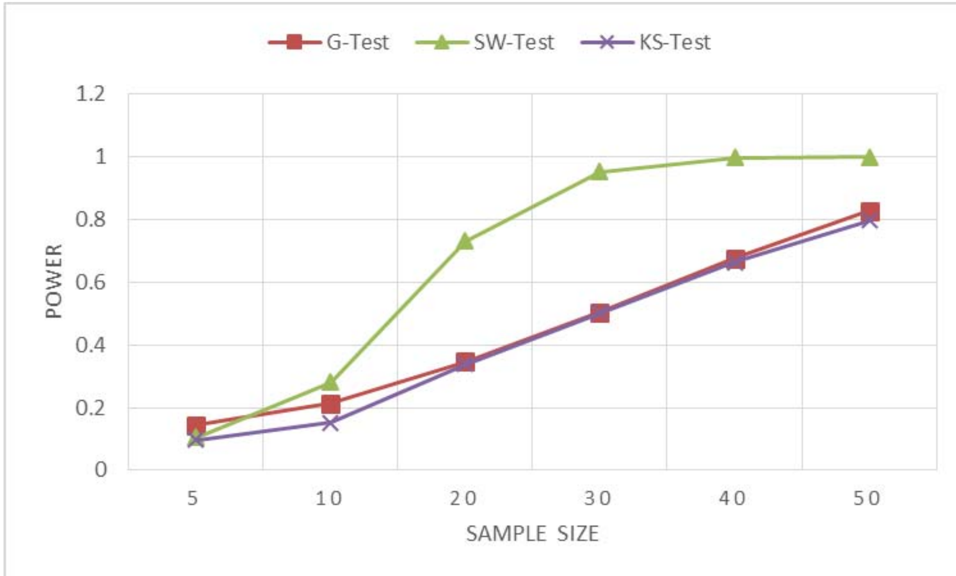

Figure 5 Power Comparison: Beta ($\alpha = 0.5, \beta = 0.5$)

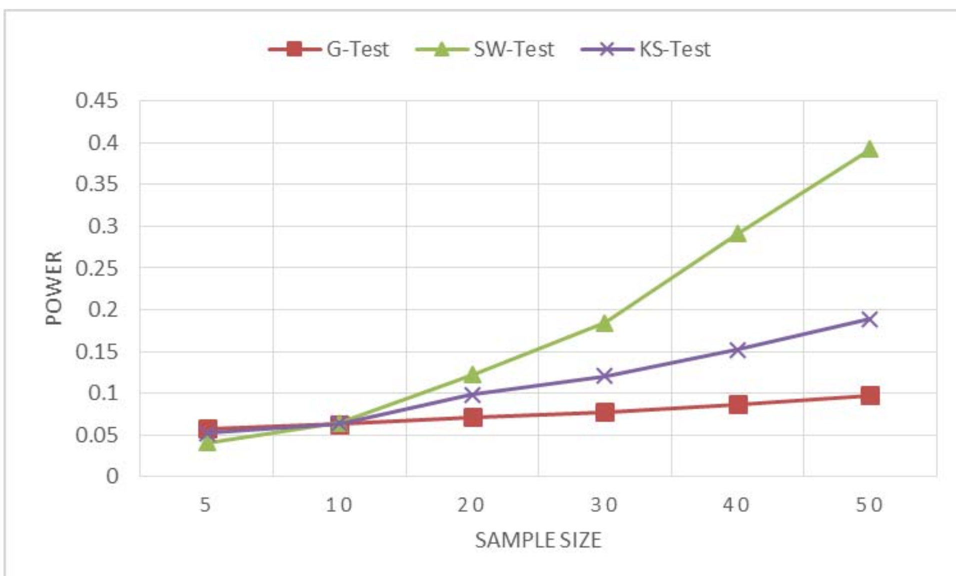

Figure 6 Power Comparison: Beta ($\alpha = 2, \beta = 4$)

Figure 7 Power Comparison: Beta ($\alpha = 1, \beta = 1$)

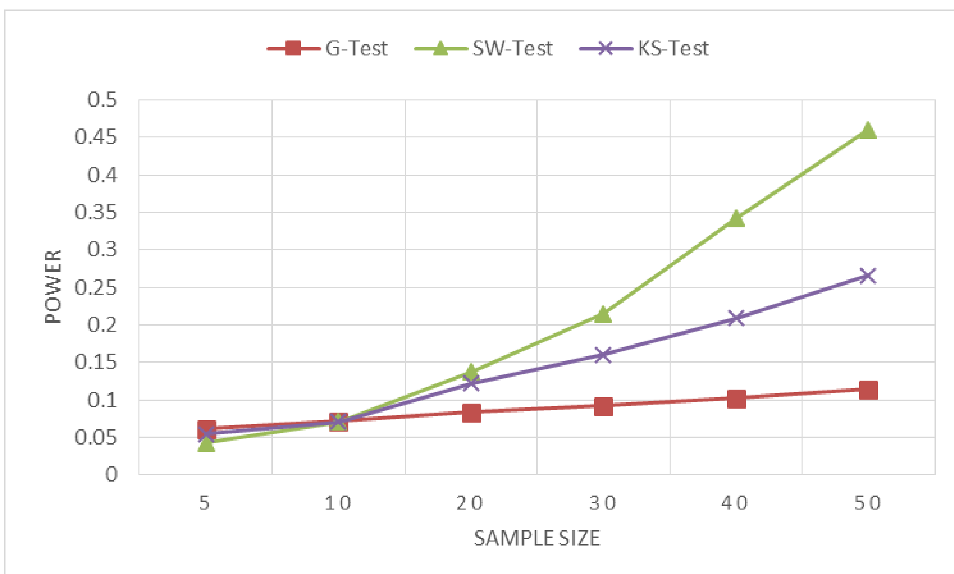

Figure 8 Power Comparison: Triangle (h=0.25)

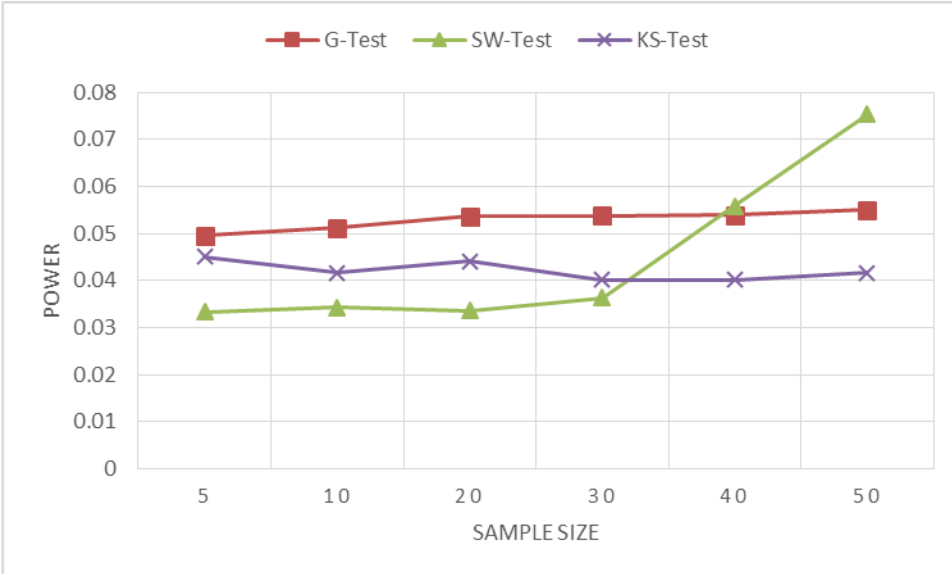

Figure 9 Power Comparison: Triangle (h=0.5)

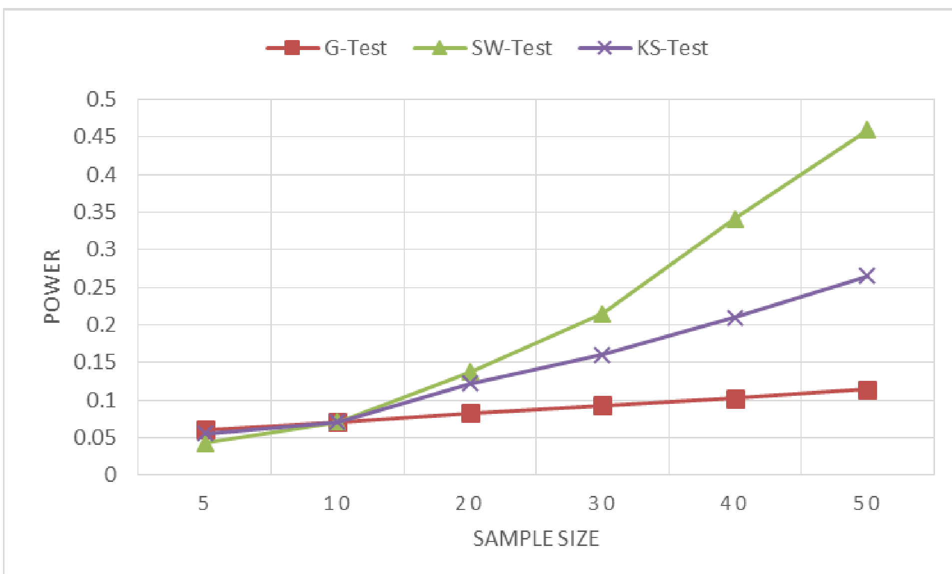

Figure 10 Power Comparison: Triangle (h=0.75)

REFERENCES

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

Anderson, T. W. and Darling, D.A. (1952) Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Process, *Annals of Mathematical statistics*, 23:193-212.

Kolmogorov, A.N. (1933) Sulla Determinazione Empirica di Una Legge di Distribuzione*, Giornale dell' Istituto degli attuari*, 4:83-91

Steele,M,N.Smart, C.Hurst and J. Chaseling (2009) Evaluating the Statistical power of goodness-of-fit tests for health and medicine survey data, *18th IMACS world congress MODSIM 09: International congress on modelling and simulation: Interfacing modelling and simulation with mathematical and computational sciences. Cairns, Qld,Australia.*

F.N. David ad N.L. Johnson (1948) The probability integral transformation when parameters are estimated from the sample, *Biometrika 06/1948; 35(Pts 1-2):182-90.*

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

Lilliefors, W.H. (1969) On the Kolmogorov-Smirnov test for the exponential distribution with unknown mean, *Journal of the American Statistics Association*, 64:387-389.

Shapiro, S. S. and Francia, R. S. (1972) Approximate Analysis of Variance Tests for Normality, *Journal of the American Statistics Association*, 67:215-216.

Shapiro, S. S. and Wilk, M. B. (1965) An Analysis of avariance Test for Normality (complete Samples), *Biometrika*, 52:591-611.

Shapiro, S. S., Wilk, M. B. and Chen, H. (1968) A Comparative Study of Various Tests for Normality, *Journal of the American Statistics Association*, 63:1343-1372.
Jun, X. (2010). *An alternative statistic for testing normality.* Unpublished master's thesis, Florida International University, Miami, Florida.