Robustness of Wilcoxon Signed-Rank Test Against the Assumption of Symmetry

by

Jutharath Voraprateep

A thesis submitted to
The University of Birmingham
for the degree of
Master of Research (Statistics)

School of Mathematics
The University of Birmingham
October 2013

Abstract

Wilcoxon signed-rank test is one of nonparametric tests which is used to test whether median equals some value in one sample case. The test is based on signed-rank of observations that are drawn from a symmetric continuous distribution population with unknown median. When the assumption about symmetric distribution fails, it can affect the power of test. Our interest in this thesis is to study robustness of the Wilcoxon signed-rank test against the assumption of symmetry. The aim of this study is to investigate changes in the power of Wilcoxon signed-rank test when data sets come from symmetric and more asymmetric distributions through simulations.

Simulations using Mixtures of Normal distributions find that when the distribution changes from symmetry to asymmetry, the power of Wilcoxon signed-rank test increases. That is, the Wilcoxon signed-rank test is not good and applicable under the asymmetry distribution. Therefore, the second objective is to study the inverse transformation method which is a technique in statistics to make observations from an arbitrary distribution to be a symmetric distribution. Moreover, the effect of the inverse transformation method to the Wilcoxon signed-rank test is also studied to answer whether or not the Wilcoxon signed-rank test is still good and applicable after we apply the inverse transformation method to the test.

Acknowledgements

I would like to deeply thank my lead supervisor, Dr. Prakash N. Patil, for providing me with the topic and for his patience and help during the work undertaken herein. I would like to thank my family and all my friends for their encouragement. I am also grateful to the University of Birmingham for the opportunity to take the course of lectures that formed the basis of the thesis. Moreover, I am indebted to the Royal Thai Government for financial support during my study at the University of Birmingham.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Almost every statistical inferential procedure is based on a certain set of assumptions. Thus for the validity of any statistical inferential procedure one has to make sure that the necessary assumptions are met. For example consider a standard t-test to test the hypothesis that the population mean is equal to some hypothesized value. Here, one can always carry out the t-test procedure based on a sample from the corresponding population. But for the outcome of the test procedure to be meaningful or valid one would like to have - sample data used to carry out the test to constitute a 'random sample' and the population from which the sample is selected is to be normally distributed. Now at this stage, among other, one faces following possibilities:

- one may be able to check whether or not these assumptions are met but the procedures for checking/testing for these assumptions will have its own assumptions and problems associated with them,
- there may not be sufficiently reliable procedures to check/test these assumptions or
- the assumptions are not met but it might be clear that the setting may not be in strong violation of the assumptions, e.g. in the application of the t-test, the

population may not be thought to be precisely normally distributed but it may not be 'too far' from being normal.

To find out the way out of the above circumstances, one of the approaches is to ask the question, 'How valid is the inferential procedure under consideration if the assumptions it demands are not fully met?'

Answer to this question, loosely speaking, calibrates the reliability of the statistical procedure under consideration. E.g. it may mean a 95% confidence interval obtained under the standard set of assumption may carry less confidence if the assumptions are not met; that is, there is a slight reduction in the reliability which may be acceptable under the circumstances. Or it may mean a 95% confidence interval obtained under the standard set of assumption may result into completely meaningless if the assumptions are not met and hence not acceptable. A procedure reflecting the character suggested in the former is referred to as a Robust procedure and the latter non-Robust.

Clearly, one of the possible and most commonly preferred and recommended ways out of the situations where standard assumptions are not met is still to use the same procedure if it is robust enough. In fact, this has led to viewing robustness as one of the important desirable properties of a statistical procedure. Further, in statistics literature generally one notices that if there a statistical procedure which is known to be non-robust, then there is likely to be a robust alternative procedure or efforts towards constructing one.

In the next section we give a short literature review, mainly pertaining to understanding and the evolution of the concept of robustness (in statistics). In section 1.3, we consider a particular example of one sample normal measurement model to describe couple of theories or criteria of robust estimation. In section 1.4, we evaluate/discuss the sign test and Wilcoxon signed-rank test as robust alternatives to standard test of t- or Z-test for population mean. Overview of the dissertation is given in the last section.

1.2 Evolution and Development of the Robustness Concept

Robustness have been studied in parametric procedures for many years. One of the earliest work on the robustness is by Box and Andersen (1955) where permutation theory is used in the derivation of robust criteria.

When statistical procedures are applied, the validity of assumptions is always concerned, for example, the normal distribution assumption is satisfied for t- and Z-test when we would like to test a location parameter θ or shift of the distribution say, the population mean. For hypothesis testing, Box and Andersen (1955) pointed out that there are two requirements for a good statistical test that it should be

1. insensitive to changes in extraneous factors not under test,
2. sensitive to change in the specific factors under test.

A test that satisfies the first requirement is said to be **robust test** and a test that satisfies the second is said to be **powerful test**. From two requirements above, we found that parametric tests when the assumptions are true tend to satisfy the second requirement but not necessarily the first, whereas, nonparametric tests tend to satisfy the first requirement but not necessarily the second. Therefore, many studies in statistics were conducted on (1) the robustness of parametric tests and (2) the power of nonparametric tests. Whenever hypothesis testing is used, we should examine whether related assumptions are satisfied. The validity of assumptions is an important thing for statistical hypothesis testing. However, even if the assumption is unsatisfied or departures from assumption occur, one would like the test to be valid.

For example, the test on variances, it is clear that the test depends on the normal distribution assumption. Many statisticians studied the analysis of variance criterion when the distribution is non-normal. They found that the test is remarkably insensitive to general non-normality. Whereas, general non-normality is meant to imply that the observations have the same non-normal parent distribution with possibly different means. Furthermore, when the group sizes are equal, the test is not very sensitive to variance inequalities from group to group.

Permutation test or randomization test is a remarkable new class of tests which was introduced by Fisher in 1935. By Fisher's view, permutation test is concerned with the validity of the test of the null hypothesis. For example, the application of the paired t-test, Fisher showed how the null hypothesis could be tested simply by counting how many of the mean differences obtained by rearranging the pairs exceeded the actual mean difference observed. Also, he showed that the null probability given by the permutation test and the t-test are almost identical. As Pearson's view in 1937, permutation tests are concerned with the power of the test when some alternative hypothesis is true. That is, Pearson emphasized that if the permutation test was to be powerful, the choice of criterion would have to depend on the type of alternative hypothesis which the experimenter had in mind. The inference in the permutation test from two views can be taken. The difference between two views is the conception of population of samples from which the observed sample is supposed to have been drawn. The Fisher's view is confined only to the finite population of samples produced by rearrangement of observations of the experiment. For the Pearson's view, the samples are regarded as being drawn from some hypothetical infinite population in the usual way.

Box and Andersen (1955) applied permutation theory which provides a method for deriving robust criteria, to the problem of comparing variances. Box and Anderson (1955) mentioned that the permutation theory may be employed to provide two results:

1. Robust tests may be formulated by approximation to the permutation test.

2. The effect on standard test procedures of non-normality and certain other departures

from assumption may be evaluated.

Box and Andersen (1955) studied both, approximation to permutation test and the effect of departures from assumption on the null distribution, for one-way classification analysis of variance and randomized blocks. For example, the power and robustness of the standard F-test and modified F-test were investigated for the rectangular, normal and double-exponential parent distributions when comparing two variances. They found that for the rectangular distribution, the modified test corrects almost perfectly, for the double-exponential distribution, the modified test appears to slightly over-correct and for the normal distribution, there is a little bit loss of power.

Lastly, Box and Anderson (1955) proposed that when the validity of the test of the null hypothesis which depends on the normal distribution is not satisfied and the central limit property lacks the criterion which is of much less practical utility, approximating to the appropriate permutation test is one possible way to be an alternative test which have greater robustness since the form of permutation test statistic can be made to depend on the alternative distribution.

Probably the first major review of work on robustness was carried out by Huber (1972), and in his view the robustness defined in Box and Andersen (1955) was vague. Huber (1972) tried to fix robustness concept by considering the problem of estimating a location parameter θ from a large number of independent observations where the distribution function F is not exactly known. Huber (1972) proposed that a robust estimator for a location parameter θ should possess:

1. a high absolute efficiency for all suitably smooth shapes F.

2. a high efficiency relative to the sample mean (and some other selected estimates), and this for all F.

3. a high absolute efficiency over a strategically selected finite set F_i of shapes (parametric family of shapes).

4. a small asymptotic variance over some neighborhood of one shape, in particular of the normal one.

5. the distribution of the estimate should change little under arbitrary small variations of the underlying distribution F.

For Huber's view, the statements in 4 and 5 are the important ones for robustness. Frequently, we have a good idea of the approximate shape of the true underlying distribution, say looking at histograms and probability plots of related previous samples so that it should be enough to consider a neighborhood of only one shape.

In view of Bickel (1976), who carried out detailed review of work on robustness until then, it may have been too late and undesirable to define the robustness narrowly. Bickel (1976) suggested, whenever robustness is to be investigated, one should answer the following three question:

1. Robustness against what? What is the super-model (a new parametric model in particular to enlarge the old one by adding more parameters)?

2. Robustness of what? What kind of procedures are being considered?

3. Robustness in what sense? What are the aims and criterion of performance used?

Also, Bickel et al. (1976) reviewed and discussed many works on robustness against gross errors and new developments. He also gave a brief review of the location problem and adaptation that are presented along with supermodels that correspond to selection of a family $\mathcal{F}$ of possible F's. Bickel focused on the parameter estimation in the normal linear model and considered the important departures from the model in the following senses.

1. Heteroscedasticity

2. Nonlinearity

3. Nonadditivity

The behavior of tests for the above mentioned departures in the normal linear model were studied and presented in many papers. Departures will be difficult for estimation of the parameters in the model. Therefore, the aim of many studies in the past was to adjust standard procedures to be modified procedures that has robustness of validity.

Huber (1981) mentioned robustness in a relatively narrow sense. According to this, robustness is insensitivity small deviations from assumptions. When the shape of the underlying distribution deviates slightly from the assumed model or the standard assumptions of statistics are not satisfied, then how a robust procedure should achieve. Huber (1981) proposed the desirable features for any statistical procedure as follows:

1. It should have a reasonably good efficiency at the assumed model (optimal or nearly optimal).

2. It should be robust in the sense that small deviations from the model assumptions should impair the performance only slightly.

3. Some larger deviations from the model should not cause failure.

Huber (1981) stated that traditionally, robust procedures have been classified together with nonparametric procedure and distribution-free test. The concepts of nonparametric procedure and distribution-free test have a little overlap in the following ideas.

- A procedure is called **nonparametric** if it is supposed to be used for a broad and not-parameterized set of underlying distribution. For example, the sample mean and sample median are the nonparametric estimates of the population mean and median, respectively. Unfortunately, the sample mean is highly sensitive to outliers and is very non-robust.

- A test is called **distribution-free** if the probability of falsely rejecting the null hypothesis is the same for all possible underlying continuous distributions or optimal robustness of validity. Most distribution-free tests happen to have a reasonably stable power and a good robustness of total performance. Anyway, distribution-free test does not imply anything about the behavior of the power function.

In Huber's study, robust methods are much closer to the classical parametric ideas than to nonparametric or distribution-free procedure. Robust methods are destined to work with parametric models. Huber (1981) intended to standardize robust estimates such that they are consistent estimates of the unknown parameters at the idealized model.

Herrendörfer and Feige (1984) stated that it seems virtually impossible to find a definition of robustness that is simultaneously clear and comprehensive. For their study, a combinatorial method in robustness research and two applications, they defined robustness for interval estimations and tests. The robustness investigations of an exact method were presented in their work that deal with known parametric procedures: the u- and t-test statistics in the case of the single sample problem and found out how they behave if the distribution is not the assumed normal distribution under all other conditions are satisfied.

As Posten (1984) stated that there are two directions in robustness research (1) attempt to quantify or measure the degree of robustness inherent in a standard statistical procedure and (2) attempt to develop a new alternative procedure, which is more robust than the standard procedure. The research contributions are still made in both the directions, for example, the study of robustness of the two-sample t-test and the relation between the shape of population distribution and the robustness of four simple test statistics. In recent years, much of the robustness research has been concerned with the development of new procedures. However, the major contributions are in the study of the robustness of standard procedures about the conditions under which the procedure is robust and under which it is non-robust. For example, in 1982, Posten et al. studied robustness of the two-sample t-test under heterogeniety of variance and nonnormality.

In addition, Tiku, Tan and Balakrishnan (1986) noted that robust statistics can provide an alternative approach to classical statistical methods when the observations deviate significantly from the assumptions. Moreover, if the assumptions are only approximately met, the robust statistics will still have a reasonable efficiency and a small bias. Furthermore, Tiku, Tan and Balakrishnan (1986) were interested in the study of robust estimation and hypothesis testing procedures for means and variances when populations are extremely nonnormal symmetric distributions and extremely skew distributions. These new procedures are based on **modified maximum likelihood estimators of location and scale parameters**. The hypothesis testing procedures developed in their study have robustness of validity and robustness of efficiency. Also, Tiku, Tan and Balakrishnan (1986) defined robustness of validity and robustness of efficiency as follows:

- Robustness of validity is the phenomenon that the type I error of a test procedures is stable from distribution to distribution.

- Robustness of efficiency is the phenomenon that the power function of a test procedure is sensitive to underlying distributions and the test is almost as powerful as the classical test for a normal distribution.

From above evolution and development of the robustness concept, many useful features are presented to clarify the concept. In next section, we consider and describe couple of theories or criteria of robust estimation.

1.3 Robustness Criteria/Theories

In 1964, Huber's paper on "Robust estimation of a location parameter" formed the first basis for a theory of robust estimation. It was an important pioneer work that contains a wealth of material for robustness study. In this section, we consider a particular example of one sample normal measurement model to describe theories or criteria of robust estimation. To illustrate first we introduce the one sample normal measurement model. We then present methods of constructing estimators in such models. Finally, we measure the robustness of the estimators.

1.3.1 The One Sample Normal Measurement Model

Let x_1, ..., x_n be random samples sizes n and we represent $x_i = \theta + e_i$, where e_i is measurement error. The measurement errors $e_i s$ are assumed to be independent normal random variables with mean 0 and variance σ^2. The maximum likelihood estimator for θ is then the sample mean $\overline{X}$. It is well known that in this situation $\overline{X}$ is the best linear unbiased, consistent and efficient estimator of θ. However, the sample mean $\overline{X}$ is not robust against the departures from the normality assumption.

Suppose that the measuring instrument, which usually produces normal errors, malfunctions on each observation with probability ε (independent of what the measurement error might have been without malfunction) and produces e_i distributed according to a distribution H. The e_i then have a common distribution $G(x)$, where

$$G(x) = (1 - \varepsilon)\Phi(x/\sigma) + \varepsilon H(x) \tag{1.1}$$

and $F(x) = G(x - \theta)$ is the distribution of X_i. The equation (1.1) is called the **gross error model**. Experience suggests that G has heavier tails than the normal component, for instance, the bad e_i tend to be larger than the good ones in absolute value, and the corresponding X_i tend to be outliers. Kotz, Johnson and Read (1988) pointed out that

- When outliers are present and are large enough, they influence the value of $\overline{X}$ to the large extent which in turn then leads to inaccurate estimates of θ.
- Unless G is symmetric about 0, $\overline{X}$ will be biased.
- Even if G is symmetric about 0, the variance of $\overline{X}$ may be much higher compared to the case when there are no gross errors, and thus $\overline{X}$ may be highly inefficient.

Although, the sample mean is a good estimator for the population mean when sample data were drawn from the normal distribution but its goodness can affected even for slight deviations from normality. There are simple alternatives for such scenarios. For example, the median and the trimmed mean are useful alternatives to the sample mean when the observations do not satisfy the normal distribution assumption. Moreover, the departure of the shape of the error distributions from normality is a nuisance to be guarded against by using robust estimation procedures.

1.3.2 Methods for Constructing Estimators

Many classical statistical procedures that rely heavily on normality assumption are not robust. Nonrobustness is usually caused by high sensitivity to outliers. The nonrobustness of classical statistical methodologies has led statisticians to make them robust by modification, a process called **robustification**, or to find alternative robust procedures, that is, **robust substitutes**. Methods for constructing such estimates are divided in three types as follows:

1. **Maximum Likelihood Type Estimators (M-estimators)**

   M-estimators, that were introduced by Huber (1981), are generalizations of the maximum likelihood estimator (MLE). Let X_1, ..., X_n be independent, identically distributed random variables with a common density function $f(x, \theta)$, where θ is an unknown parameter. The MLE for θ is obtained by maximizing

$$\sum_{i=1}^{n} log f(x_i, \theta)$$

or solving

$$\sum_{i=1}^{n} \dot{l}(x_i, \theta) = 0$$

where $\dot{l}(x_i, \theta)$ is the gradient of log $f(x, \theta)$ with respect to θ.

M-estimators are obtained by replacing the objective function log $f(x, \theta)$ by another function, say $\rho(x, \theta)$, or replacing $\dot{l}(x, \theta)$ by say $\psi(x, \theta)$.

That is, any estimate T_n, defined by a minimization problem of the form

$$\sum_{i=1}^{n} \rho(x_i; T_n) = min! \tag{1.2}$$

or by an implicit equation

$$\sum_{i=1}^{n} \psi(x_i; T_n) = 0 \tag{1.3}$$

where ρ is an arbitrary function and has a derivative $\psi(x; \theta) = (\partial/\partial\theta)\rho(x; \theta)$. Therefore, estimator T_n defined by (1.2) and (1.3) is called an **M-estimators** or **maximum likelihood type estimators**. Note that the choice $\rho(x; \theta)$ = -log$f(x; \theta)$ gives the ordinary maximum likelihood estimator.

Huber(1981) also interested in location estimates. When estimating location in the model $\mathfrak{X} = \mathbb{R}$, $\Theta = \mathbb{R}$, $F_\theta(x) = F(x - \theta)$, then the equation (1.2) and (1.3) can be written as

$$\sum_{i=1}^{n} \rho(x_i - T_n) = min! \tag{1.4}$$

or

$$\sum_{i=1}^{n} \psi(x_i - T_n) = 0. \tag{1.5}$$

For the equation (1.5) can be written equivalently as

$$\sum_{i=1}^{n} w_i(x_i - T_n) = 0 \tag{1.6}$$

with

$$w_i = \frac{\psi(x_i - T_n)}{x_i - T_n} \tag{1.7}$$

Therefore, a formal representation of T_n as a weighted mean is

$$T_n = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{1.8}$$

with weights depending on the sample.

2. **Linear Combinations of Order Statistics (L-estimators)**

Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ be the order sample, then a general L-estimate is of the form

$$\sum_{i=1}^{n} \omega_i X_{(i)}$$

where ω_1, ..., ω_n are fixed weights not depending on the data.

Examples of L-estimators are **trimmed mean** and **Winsorization**. Trimming is the process of removing extreme values from the sample, and Winsorization is the process of changing the extreme values by setting each equal to the values of less extreme observations. For instance, suppose that one has a univariate sample X_1, ..., X_n, let k be a positive integer less than $n/2$, and define $\alpha = k/n$. Then the

α-symmetrically trimmed sample is the original sample after the k smallest and k largest order statistics have been removed. Whereas the α-symmetrically Winsorized sample is obtained by replacing the k smallest and k largest order statistics by $X_{k+1:n}$, $X_{n-k:n}$, respectively.

Sprent (1989) advocated that the trimmed mean be used to estimate the location. This is because it has a number of desirable properties. For example, it is very simple to compute, and it is robust. As α changes from 0 to 1/2, the trimmed mean changes along from the arithmetic mean to the median. Moreover, for samples from a symmetric population, the symmetrically trimmed mean is an unbiased estimator of the population mean.

3. **Estimators Derived from Rank Tests (R-estimators)**

R-estimators are based on the ranks of observations. In one sample case, R-estimators exist for the location problem, and normally, the estimators are derived from one sample rank test but we will consider two samples rank test as follows:

Let X_1, ..., X_m and Y_1, ..., Y_n be two samples with distributions $F(x)$ and $G(x) = F(x+\delta)$, where δ is the unknown location shift. Let R_i be the rank of X_i in the pooled sample of size $N = m + n$. A rank test of $\delta = 0$ against $\delta > 0$ is based on a test statistic

$$S_N = \frac{1}{m}\sum_{i=1}^{m} a_N(R_i). \tag{1.9}$$

Usually, we assume that the weights $a_N(i)$ are generated by some function H as follows:

$$a_N(i) = H\left(\frac{i}{m+n+1}\right). \tag{1.10}$$

There are many other possibilities for deriving weights $a_N(i)$ from H, for example,

$$a_N(i) = H\left(\frac{i-1/2}{m+n}\right) \tag{1.11}$$

or

$$a_N(i) = (m+n)\int_{(i-1)/(m+n)}^{i/(m+n)} H(u)du \tag{1.12}$$

and in fact we prefer to work with the last version. For H and F, all these weights lead to asymptotically equivalent tests. In the case of the Wilcoxon signed-rank test, $H(t) = t - 1/2$, the above three variants create exactly the same tests.

To simplify the presentation, we will assume that m = n. Then, we can write (1.9) as

$$S(F,G) = \int H\left(\frac{1}{2}F(x) + \frac{1}{2}G(x)\right) F(dx) \tag{1.13}$$

or, if we substitute $F(x) = s$,

$$S(F,G) = \int H\left(\frac{1}{2}s + \frac{1}{2}G(F^{-1}(s))\right) ds. \tag{1.14}$$

If F is continuous and strictly monotone, the two equations (1.13) and (1.14) are equivalent. We also assume that

$$\int H(s)ds = 0 \tag{1.15}$$

corresponding to

$$\sum a_i = 0. \tag{1.16}$$

Then the expected value of (1.9) under the null hypothesis is 0.

Let T_n be the sequence of location estimators where $n \geqslant 1$. We can then derive estimators of shift δ from such rank tests:

- In two sample cases, adjust δ such that $S_{n,n} \approx 0$ when computed from (x_1, ..., x_n) and ($y_1 - \delta$, ..., $y_n - \delta$).
- In one sample case, adjust T_n such that $S_{n,n} \approx 0$ when computed from (x_1, ..., x_n) and $(2T_n - x_1, ..., 2T_n - x_n)$.

The idea behind the R-estimator of location in one sample case is the following. From the original sample x_1, ..., x_n, we can construct a mirror image by replacing each x_i by $T_n - (x_1 - T_n)$. We choose the T_n for which the test cannot detect any shift, which means that the test statistic S_N in (1.9) comes close to zero (although it often cannot become exactly zero, being a discontinuous function).

1.3.3 Measures of Robustness

The elementary tools used to describe and measure robustness are the **breakdown point**, the **influence function** and the **robustness measures derived from the influence function**.

- **The Breakdown Point**

  The breakdown point is a quantitative measure of the robustness. It indicates the maximum proportion of gross outliers which the induced estimators $T(F_n)$ can tolerate. For example, the median will tolerate up to 50% gross errors or its breakdown point is 50%. It may be useful to note that the breakdown point of the sample mean is 0%. Therefore, the empirical breakdown point is the smallest fraction of outliers that the estimator can tolerate before being affected by the outliers. Hampel (1986) defined the breakdown point of T_n at F that generalizes an idea of Hodges in 1967 by:

Definition 1.1 *The breakdown point ε^* of the sequence of estimators $\{T_n;\ n \geq 1\}$ at F is defined by*

$\varepsilon^* = sup\Big\{\ \varepsilon \leq 1$*; there is a compact set* $K_\varepsilon \subsetneqq \Theta$ *such that* $\pi(F,G) < \varepsilon$ *implies* $G(\{T_n \in K_\varepsilon\}) \to 1,\ n \to \infty\ \Big\}$

where $\pi(F,G)$ is the Prohorov distance (Prohorov, 1956) of two probability distributions F and G and given by

$$\pi(F,G) = inf\{\varepsilon; F(A) \leqslant G(A^\varepsilon) + \varepsilon \ \text{ for all events A}\} \tag{1.17}$$

where $G(A^\varepsilon)$ is the set of all points whose distance from A is less than ε.

For example, when $\Theta = \mathbb{R}$ we obtain $\varepsilon^* = \sup\{\ \varepsilon \leq 1$; there exists r_ε such that $\pi(F,G) < \varepsilon$ implies $G(\{\mid T_n \mid \leq r_\varepsilon\}) \to 1, n \to \infty\}$. The breakdown point should formally be denoted as $\varepsilon^*(\{T_n; n \geq 1\}, F)$, but it usually does not depend on F. From Definition 1.1 , one can also consider the *gross-error breakdown point* where $\pi(F,G) < \varepsilon$ is replaced by $G \in \Big\{(1-\varepsilon)F + \varepsilon H; H \in \mathfrak{F}(\mathfrak{X})\Big\}$.

Furthermore, there is alternative definition of the breakdown point that is much simpler concept than Definition 1.1 and does not contain probability distribution. A slightly different definition was given by Hampel et al. in 1982.

Definition 1.2 *The finite-sample breakdown point ε_n^* of the estimators T_n at the sample $(x_1, ..., x_n)$ is given by*

$\varepsilon_n^*(T_n; x_1, ..., x_n) = \frac{1}{n} max\Big\{m; max_{i_1,...,i_m} sup_{y_1,...,y_m} \mid T_n(z_1, ..., z_n) \mid < \infty\ \Big\}$

where the sample $(z_1, ..., z_n)$ is obtained by replacing the m data points $x_{i_1}, ..., x_{i_m}$ by the arbitrary values $y_1, ..., y_m$.

Note that this breakdown point usually does not depend on $(x_1, ..., x_n)$, and depends only slightly on the sample size n. In many cases, taking the limit of ε_n^* for $n \to \infty$

yields the asymptotic breakdown point ε^* of Definition 1.1. Hampel et al. (1986) mentioned that in 1983, Donoho and Huber took the smallest m for which the maximal supremum of $| T_n(z_1, ..., z_n) |$ is infinite, so their breakdown point equals $\varepsilon_n^* + 1/n$. For example, we find $\varepsilon_n^* = 0$ for the the arithmetic mean whereas they obtain the value $1/n$.

The breakdown point can be used to investigate rejection rules for outliers in the one dimensional location problem.

- **The Influence Function**

The influence function (IF) was originally referred as *influence curve* (IC). Nowadays, we prefer the more general name *influence function* (IF) in view of the generalization to higher dimensions. The influence function (IF) affects in robust estimation and is an important tool to construct an estimator. The influence function (IF) is defined as follows:

Definition 1.3 *The influence function (IF) of an estimate or test statistic T at F is given by*

$$IF(x; T, F) = \lim_{\varepsilon \to 0} \frac{T((1-\varepsilon)F + \varepsilon\delta_x) - T(F)}{\varepsilon} \tag{1.18}$$

where δ_x denotes the point mass 1 at x (Hample et al.(1986)).

If we replace F by $F_{n-1} \approx F$ and put $\varepsilon = 1/n$, we realize that the IF measures approximately n times the change of T caused by an additional observation in x when T is applied to a large sample of size $n-1$.

The influence function is an useful heuristic tool of robust statistics. It describes the effect of an infinitesimal contamination at the point x on the estimate, standardized by the mass of the contamination. One could say it gives a picture of the infinitesimal behavior of the asymptotic value. Thus, it measures the asymptotic bias caused by contamination in the observations.

If some distribution G is near F, then the first-order that is derived from a Taylor expansion of T at F evaluated in G is given by

$$T(G) = T(F) + \int IF(x; T, F) d(G - F)(x) + remainder. \tag{1.19}$$

Hample et al. (1986) recalled the basic idea of differentiation of statistical functionals. T is a *von Mises functional*, with first kernel function a_1. It is clear that

$$\int a_1(x) dF(x) = 0. \tag{1.20}$$

Now, we consider the important relation between the IF and the asymptotic variance. When the observations X_i are independent identically distributed (i.i.d.) according to F, then the empirical distribution F_n will tend to F by the Glivenko-Cantelli theorem. Therefore, in (1.19) we may replace G by F_n for sufficiently large n. We also assume that $T_n(X_1, ..., X_n) = T_n(F_n)$ may be approximated adequately by $T(F_n)$. By using (1.20), which we can rewrite as $\int IF(x; T, F) dF(x) = 0$, we obtain

$$T_n(F_n) = T(F) + \int IF(x; T, F) dF_n(x) + remainder.$$

Evaluating the integral over F_n and rewriting yields

$$\sqrt{n}(T_n - T(F)) \simeq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i; T, F) + remainder.$$

By the central limit theorem, the leading term on the right-hand side is asymptotically normal with mean 0 , if the X_i are independent with common distribution F. In most cases, the remainder becomes negligible for $n \to \infty$, so T_n itself is asymptotically normal. That is, $\sqrt{n}(T_n - T(F))$ is asymptotically normal with mean 0 and variance

$$V(T,F) = \int IF(x;T,F)^2 dF(x). \tag{1.21}$$

The important thing for (1.21) is that it gives the right answer in all practical cases. Moreover, (1.21) can be used to calculate the asymptotic relative efficiency $ARE_{T,S} = V(S,F)/V(T,F)$ of a pair of estimators $\{T_n; n \geqslant 1\}$ and $\{S_n; n \geqslant 1\}$.

- **Robustness Measures Derived from the Influence Function**

  - **The gross error sensitivity**

    From the previous topic, the influence function, we have seen that the IF describes the standardized effect of an infinitesimal contamination at the point x on the asymptotic value of the estimator. Hampel et al. (1986) also defined the gross error sensitivity of an estimate or test statistic T_n at F by

$$\gamma^*(T,F) = sup_x \mid IF(x;T,F) \mid . \tag{1.22}$$

    The gross error sensitivity is the supremum of the absolute value of the influence function. The supremum being taken over all x where IF$(x;T,F)$ exists. The gross error sensitivity measures the worst (approximate) influence which a small amount of contamination of fixed size can have on the value of the estimator. Thus, it may be regarded as an upper bound on the (standardized) asymptotic bias of the estimator. It is a desirable feature that $\gamma^*(T,F)$ be finite, in which case we say that T is *B-robust* at F (Hampel et al.(1986)). Here, the B comes from "bias".

– **The sensitivity curve**

The sensitivity curve (SC) was proposed by Mosteller and Tukey (Hampel et al.(1986)). In the case of an additional observation one starts with a sample $(x_1, ..., x_{n-1})$ of n - 1 observations and defines the sensitivity curve as

$$SC_n(x) = n(T_n(x_1, ..., x_{n-1}, x) - T_{n-1}(x_1, ..., x_{n-1})). \qquad (1.23)$$

In (1.23), SC is proportional to the change in the estimator when one observation with value x is added to a sample x_1, ..., x_{n-1}. This is simply a translated and rescaled version of the empirical IF. When the estimator is a functional, i.e. when $T_n(x_1, ..., x_n) = T(F_n)$ for any n,any sample $(x_1, ..., x_n)$ and corresponding empirical distribution F_n, then

$$SC_n(x) = \frac{T((1 - 1/n)F_{n-1} + (1/n)\delta_x) - T(F_{n-1})}{1/n} \qquad (1.24)$$

where F_{n-1} is the empirical distribution of $(x_1, ..., x_{n-1})$. This last expression is a special case of (1.18), with F_{n-1} as an approximation for F and with contamination size $t = 1/n$. In many situations, $SC_n(x)$ will converge to IF$(x; T, F)$ when $n \to \infty$.

1.4 Sign and Wilcoxon Signed-Rank Tests as Robust Alternatives to t- and Z-test

In general, parametric methods always depend on crucial population assumptions. If assumptions about the underlying population are questionable or are not satisfied, then nonparametric methods are used instead of their parametric analogues because most nonparametric procedures depend on a minimum of assumptions and do not assume a special distribution function F.

For a location parameter setting, the population mean (μ) is a measure of central tendency. When parametric procedures are suitable, we test the null hypothesis Ho : $\mu = \mu_0$. For example, we use the t-test based on the Student's t distribution in testing hypothesis and constructing confidence intervals for a population mean. When sample sizes are large, the central limit theorem is used to justify the use of the Z-test for both of the procedures (test and confidence interval) about a population mean. When we use the t- or Z- test, we assume that the population from which the sample data have been drawn is normally distributed. If the population distribution assumption is violated, we should find an alternative method of analysis. One of the alternative ways is a nonparametric procedure. Several nonparametric procedures are available for making inferences about a location parameter. Basically, we always refer the **population median** (M) rather than the population mean for the location parameter in nonparametric procedures.

The median is the middle value of a set of measurements arranged in order of magnitude. For a continuous distribution, the median is defined as the point M for which the probability that a value selected at random from the distribution is less than and greater than M, are both equal to 0.5. When random samples are drawn from symmetric population distribution, any conclusion about the median is applicable to the mean, since the mean and the median are coincident in symmetric distribution.

In nonparametric procedures, the well-known tests about the population median (M) are the sign and Wilcoxon signed-rank tests in one sample case. We test the null hypothesis Ho : $M = M_0$. The following are some of the advantages of the sign and Wilcoxon signed-rank tests.

- The tests do not depend on the normal population distribution.
- The computations can be quickly and easily performed.

- The concepts and procedures of tests are easy to understand for researchers with minimum preparation in mathematics and statistics.
- The tests can be applied when the data are measured on a weak measurement scale.

The above advantages plus the fact that the sign and Wilcoxon sign-ranked tests are insensitive when the observations deviate significantly from the normal distribution assumption these tests are very popular. Both of these tests are the analogues of independent one sample t- and Z-test in testing hypothesis for a population median. Thus, the sign and Wilcoxon signed-rank tests are robust alternatives for t- and Z-test.

1. **Sign Test**

   The sign test is the oldest of all nonparametric tests of the location parameter. It is called the sign test because we convert the data for analysis in to a series of plus and minus signs. Therefore, the test statistic consists of either the number of plus signs or the number of minus signs. The sign test does not require the assumption that the population be normally distributed and moreover, it does not require that the population probability distribution be symmetric (Wayne(1990)).

   Let $X_1,\ X_2,\ \ldots\ ,\ X_n$ be a sample of size n from a continuous population with probability density function f with median M. If ξ_p is the *p-th quantile* or the *quantile of order* p, for any number p, where $0 \leqslant p \leqslant 1$, or the distribution of X such that it satisfies (Pratt and Gibbons (1981))

$$P(X \leqslant \xi_p) = F(\xi_p) = p. \tag{1.25}$$

   For example, if $p = 0.5$ then $\xi_{0.5}$ is the 0.5-th quantile or the quantile of order 0.5 or the median M. That is

$$P(X \leqslant \xi_{0.5}) = F(\xi_{0.5}) = F(M) = 0.5.$$

The sign test statistic is

$$S = \sum_{i=1}^{n} I[X_i > M_0] \tag{1.26}$$

or

$$S = \sum_{i=1}^{n} I[X_i < M_0] \tag{1.27}$$

where M_0 is the hypothesized median and I in (1.26) and (1.27) is an indicator function.

Therefore, the sign test statistic in (1.26) and (1.27) are the number of positive or negative observations, respectively. Under the null hypothesis, the sampling distribution of S is the binomial distribution with parameter $p = 0.50$. For samples of size 12 or larger, we use the normal approximation to the binomial. The normal approximation involves approximating a discrete distribution by mean of a continuous distribution and we use a continuity correction factor of 0.5. The sign test statistic when sample sizes are 12 or larger is (Wayne (1990))

$$Z = \frac{(S \pm 0.5) - 0.5n}{0.5\sqrt{n}} \tag{1.28}$$

which we compare with the values of the standard normal distribution for the chosen level of significance.

For the power-efficiency of the sign test, Wayne (1990) proposed Walsh's study about the power functions of the sign test with those of the Student's t-test for the case of normal populations. Walsh found that the sign test is approximately 95% efficient for small samples. When sampling from normal populations, he found that the relative efficiency of the sign test decreases as the sample size increases. According to Dixon's study, the power-efficiency of the sign test decreases when sample sizes and level of significance increase.

2. **Wilcoxon Signed-Rank Test**

Wilcoxon signed-rank test is a well-known nonparametric statistical hypothesis test on population location, median. The test is designed to test a hypothesis about the location of a population distribution and does not require the assumption that the population be normally distributed. The test is based on the signed ranks of a random sample from a population which is continuous and symmetric around the median. In many applications, this test is used in place of the one sample t and Z–test when the normality assumption is questionable. The advantage of Wilcoxon signed-rank test is that it does not depend on the shape of the population distribution (Wayne(1990)).

Let $X_1,\ X_2,\ ...\ ,\ X_n$ be a sample of size n from a continuous population with probability density function f, which is symmetric with median M.

The Wilcoxon signed-rank test statistic is

$$T = \sum_{i=1}^{n} R_i \cdot Sign(Z_i) \tag{1.29}$$

where $Z_i = X_i$ - M_0, R_i is rank of $\mid Z_i \mid$, i = 1, 2, 3, ..., n and

$$Sign(Z_i) = \begin{cases} 1 & \text{if } Z_i > 0 \\ -1 & \text{if } Z_i < 0. \end{cases}$$

Note that, the sign test utilizes only the signs of the differences between each observation and the hypothesized median, M_0, but the magnitudes of these observations relative to M_0 are ignored. Assuming that such information is available, a test statistic which takes into account these individual relative magnitudes might be expected to give a better performance. Wilcoxon signed-rank test statistic can provide an alternative test of location which uses by both the magnitudes and signs of these differences. Therefore, one expects Wilcoxon signed-rank test to be more powerful test than the sign test.

The Wilcoxon signed-rank test is a well known rank-based test for a location parameter. For a normal population, the efficiency of Wilcoxon signed-rank test is equal to 0.955 relative to the t-test. For a heavy-tailed distribution, the Wilcoxon signed-rank test can be considered more powerful than the t-test. Moreover, the type I error probability of the Wilcoxon signed-rank test can be computed exactly under the null hypothesis, regardless of what the population distribution may be. However, the type I error rate of the t-test is reasonably stable as the populations deviate from the normal distribution , so the real advantage of the Wilcoxon signed-rank test is robustness of efficiency (Kotz, Johnson and Read (1988)). Wilcoxon signed-rank test will be mainly studied in this thesis.

1.5 Outline of the Thesis

In this thesis, we focus on the Wilcoxon signed-rank test and examine the power of test when the symmetry assumption about the distribution is not satisfied. That is a sample is from asymmetric continuous distribution. In the second chapter, we explain the recently proposed measure of asymmetry. Then we consider the Mixtures of Normal distributions such that their asymmetry coefficient varies from 0.0 to 0.5. The objective of the simulation study is to investigate whether or not the Wilcoxon signed-rank test is robust against the assumption of symmetry. We simulate random samples from the Mixtures of Normal distributions with increasing asymmetry and investigate the power and size of the Wilcoxon signed-rank test as asymmetry changes. Moreover, the simulation results, summarization and discussion are also given.

In the third chapter, we propose to transform the data to achieve symmetry. We then carry out the Wilcoxon signed-rank test on the transformed data. Here first we assume that the actual probability model under the null hypothesis is known. Using this model, data is transformed to uniform distribution. Since having the knowledge of the null distribution in practice is not possible, this proposal is not useful in practice. But we show through simulation that test carried out on such transformed data, irrespective of original data being asymmetric, works fine. We then propose a practical method to carry out the Wilcoxon signed-rank test when data is asymmetric. This involves first estimating the probability density function using kernel method based on the sample, then estimating smooth distribution function. This distribution function can then be used to transform the data to uniform (ie. symmetric) before exploring the Wilcoxon signed-rank test.

Chapter 2

Simulation Study: Power of the Wilcoxon Signed-Rank Test

2.1 Introduction

In Chapter 1, we have introduced and given a short literature review of the evolution and development of the robustness concept. We noted that the robustness have been studied in the parametric procedures for many years. Most of the previous robustness studies were to investigate the robustness of standard procedures in the parametric statistical inference. The aim of those investigations was to check which of the statistical inferential procedures under consideration there are robust and which are not. For example, in the hypothesis testing, to have a meaningful interpretation of the outcome of a test, checking the validity of the necessary assumptions for a test is essential. Then the investigation here will be of the type as explained in the next sentence. If the assumptions are not satisfied, then the question one would like to answer is whether or not the test is still a good and applicable under the circumstance. That is, does the test still has the same size and the power, and if not, are the changes in the size and power are 'small' enough for the test to be still applicable with slight changes in the size and power. Therefore, a test is called **robust test** when the test should impair the performance slightly when there is a small deviation from the assumption.

Furthermore, we have also reviewed robustness criteria/theories which are useful tools for robustness study in Chapter 1, for example, methods for constructing robust estimators and measure of robustness. Lastly, we have mentioned two well-known tests about the population median, the sign test and Wilcoxon signed-rank test. None of these two tests depend on the functional form of the population distribution. Thus they can replace the standard t- and Z-test which are generally used to carry out tests to test hypotheses concerning the population mean when the population distribution is thought to be Normal. But there are differences in the sign and the Wilcoxon signed-rank tests. The test statistic of the sign test uses the signs of the differences between each observation and the hypothesized median and it does provide a good test to test the assertions about population median. However, the sign test's test statistic ignores the magnitudes of those differences for hypothesis test. Whereas, the test statistic of the Wilcoxon signed-rank test uses both the magnitudes and signs of the differences just mentioned. Thus, as one would expect, the Wilcoxon signed-rank test is more powerful test than the sign test, see for example, Wayne (1990), Pratt and Gibbons (1981). There is another important difference between the Wilcoxon signed-rank and the sign test. The former test requires the population to be continuous and symmetric where as the latter test can be valid without such requirement.

So we have the Wilcoxon signed-rank test which itself is reasonably robust against the population distributional assumption but it does require population to be symmetric. Thus the question we are interested in is 'how robust the Wilcoxon signed-ranked test is against the assumption of symmetry?'. A common sense suggest that as the population distribution goes away from symmetry (i.e. becomes more and more asymmetric) the power of the Wilcoxon signed-rank test should decrease.

To answer the question that we raised in the last paragraph, first we will have to answer what does one mean by more and more asymmetric. Since asymmetry is a qualitative feature, the amount of asymmetry in a probability distribution or relative asymmetry of two probability distribution is more likely to be subjective or a judgement call. Thus, to overcome this, one may need to quantify the amount of asymmetry in a probability distribution. In fact, it is likely that the reason mathematical statisticians did not study the robustness of the Wilcoxon signed-rank test until now might be due to the lack of appropriate quantification of the asymmetry. However, recently Patil et al. (2012) have proposed a reasonably satisfactory quantification of asymmetry. This certainly removes one of the major hurdles in taking up the question raised in the last paragraph.

Our interest in this Chapter now is to study robustness of the Wilcoxon signed-rank test against the assumption of symmetry through simulations. For that first we describe the asymmetry measure defined by Patil et al. (2012) in the next section. In section 2.3, we describe our overall plan of the simulation study. It is clear that the Wilcoxon signed-rank test statistic computed for a data from a non-symmetric continuous distribution will not have the standard distribution that one expects when the population is symmetric and continuous. This is illustrated through simulations that are carried out in the section 2.4. However, if one were still to use the Wilcoxon signed-rank test statistics, in section 2.5 we introduce the concept of a "relative power" and then show that as the size of asymmetry increases the relative power of the Wilcoxon signed-rank test decreases.

2.2 Two Problems

One of the aims of the simulation study is to explore whether or not the Wilcoxon signed-rank test is robust against the assumption of symmetry. Thus one of the obvious way to investigate the robustness of the Wilcoxon signed-rank test is first to select samples from a symmetric population, carry out the Wilcoxon signed-rank test and find the empirical power of the test. Then take samples from a population which is ‘slightly’ asymmetric and again carry out the test and find the empirical power. If the test is to be robust, one expect there to be a very small change in the power when the population changes from symmetric to slightly asymmetric. But this raises the another more general question. How does the power of the test behave if one were to apply this test to samples from more and more asymmetric populations? Although it is reasonable to expect that the power of the test to decrease as the populations (from which the samples are selected) becomes more and more asymmetric, one may want to verify this through simulations. Thus, to see how the power of the test changes as the population becomes more and more asymmetric, one can take the above mentioned simulation plan a step further. It will mean repeating the simulations just described by taking samples from the populations with increasing amount of asymmetry and finding the power of the test. Although the aims and the subsequent plans to investigate the robustness of the Wilcoxon-signed rank test or, when it is employed to samples from asymmetric populations, to study its power behavior against the increasing amount of asymmetry in the population distributions seems reasonable, one is likely to face two main difficulties. We describe and address these two problems in the next two subsections respectively.

2.2.1 Measure of Asymmetry

The first problem is what does one mean by a ‘slight’ asymmetric population or ‘increasing amount’ of asymmetry in the population distributions. To provide a meaning to ‘slight’ asymmetric population or ‘increasing amount’ of asymmetric populations, it is necessary to quantify the amount of asymmetry. In fact, at this point one may speculate that because of the lack of a satisfactory quantification of asymmetry in the literature, the relationship between the power of Wilcoxon-signed rank test and the size of asymmetry of the population distribution may not have been explored. However, now there is such quantification available and can be used to study the relation between power and size of asymmetry. We now give a very short review of the attempts of quantifying the asymmetry in a probability density curve and describe a measure which will be used to quantify the size of asymmetry of a distribution in this dissertation.

Symmetry of a probability model is a qualitative characteristics and plays an important role in statistical procedures. Normally, it is useful to know their mathematical quantification. But instead, because of the simple form and easy evaluation, basic skewness measures in statistics, at times are used to assess the symmetry. However, when one wants to compare asymmetries of the two probability density function curves, skewness may not be the right measure. In fact, even otherwise, Li and Morris (1991) illustrate the unreliability of skewness measures when used to make assertions on the symmetry. Thus there are attempts in the literature to quantify the asymmetry, but such discussion is very limited and not very satisfactory. For example see MacGillivray (1986), Li and Morris (1991). Patil et al. (2012) argue that the earlier proposals of quantification neither seem user friendly nor intuitive enough to visualize the amount of asymmetry in a density curve and have suggested a measure which seems to do a reasonable job of quantifying asymmetry. Their proposal quantifies the asymmetry of a continuous probability density function on a scale of -1 to 1, where the value zero means a symmetric density and ± 1 mean positively and negatively most asymmetric densities. We now introduce this measure and for that first recall the definition of symmetry.

Definition 2.1 *A continuous probability density function $f(x)$ with distribution function $F(x)$, $x \in \mathbf{R}$, is said to be symmetric about θ if $F(\theta - x) = 1 - F(\theta + x)$ or equivalently $f(\theta - x) = f(\theta + x)$ for every $x \in \mathbf{R}$.*

A necessary condition used in Patil et al. (2012) to develop a new measure of symmetry is stated in the following lemma.

Lemma 2.1 *Let X be a continuous symmetric random variable with square integrable continuous probability density function f(x) and distribution function F(x) then,*

$$Cov(f(X),F(X)) = 0.$$

Patil et al. (2012) proposed a measure or coefficient, $\eta(X)$, of asymmetry of a random variable X based on the above necessary condition and is defined by

$$\eta(X) = \begin{cases} -Corr(f(X), F(X)) & \text{if } 0 < Var(f(X)) < \infty \\ 0 & \text{if } Var(f(X)) = 0 \end{cases}$$

where $F(X)$ is distribution function of X. Observe that the coefficient of asymmetry $\eta(X)$ is such that -1 < $\eta(X)$ < 1.

For $\eta(X)$ to be defined, one needs Var($f(X)$) < ∞ and that leads to the condition

$$\int_{-\infty}^{\infty} f^3(x)dx < \infty. \tag{2.1}$$

When the values of $\eta(X)$ are closer to zero, it means the density function is close to being a symmetric function and whereas closer to ± 1, it means the density function is close to being the most positively or negatively asymmetric function. For instance, the coefficients of asymmetry of the Cauchy, Normal, Uniform distribution are equal to zero ($\eta(X) = 0$).

The important properties of their measure of asymmetry are

1. For a symmetric random variable X, if (2.1) holds, then $\eta(X) = 0$.

2. If $Y = aX + b$ where $a > 0$ and b any real number, $\eta(X) = \eta(Y)$.

3. If Y = -X, $\eta(X)$ = -$\eta(Y)$.

For various examples illustrating how the above coefficient does an admirable job of quantifying visual impression of the asymmetry of a probability density curve we refer the reader to Patil et al. (2012). However, below we provide some of the examples of asymmetric probability distributions which are used in the simulations of this thesis and their associated asymmetry coefficients.

Let X be a continuous random variable that follows a Mixture of Normal distributions, or $X \sim \alpha\text{N}(\mu_1,\sigma_1^2)$ + (1 - α)N(μ_2,σ_2^2) where 0 <α < 1 is the mixing coefficient.

The probability density function of X is $f(x)$ where

$$f(x) = \alpha \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(x-\mu_1)^2/2\sigma_1^2} + (1-\alpha)\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-(x-\mu_2)^2/2\sigma_2^2}$$

for $-\infty < x < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, σ_1^2 , $\sigma_2^2 > 0, 0 < \alpha < 1$.

We choose the parameters of this Mixture of Normal distribution so that size of asymmetry changes from 0.0 to 0.5. For example, let $\mu_1 = 0$, μ_2 =2, $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$ and let α vary for 0.0 to 0.5. Plots of the Mixtures of normal distributions for different α are given in Figure 2.1.

Using $\eta(X)$ in Patil et al. (2012), we note that the asymmetry coefficient of different mixing coefficient α is given in Table 2.1.

Table 2.1: The mixing coefficient (α_i) and size of asymmetry (η_i) of the Mixtures of Normal distribution when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 4$

α_i	η_i
0.000	0.0
0.101	0.1
0.175	0.2
0.256	0.3
0.382	0.4
0.491	0.5

2.2.2 Relative Power

The second problem is about the null distribution of the Wilcoxon-signed rank test. The sampling distribution of the Wilcoxon-signed rank test statistic is derived under the as-

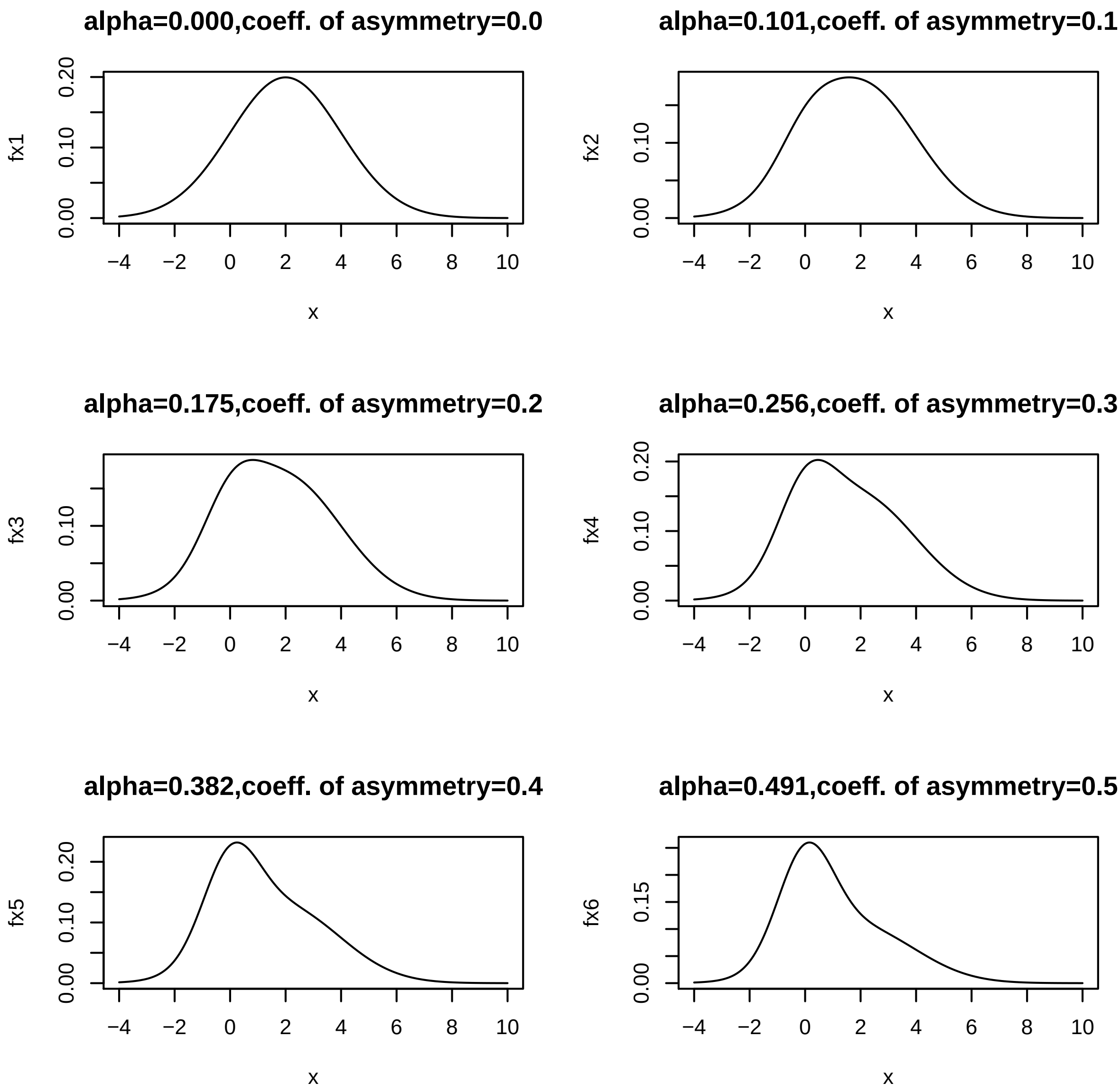

Figure 2.1: Mixtures of Normal density curves when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ with coefficient of asymmetry equal to 0.0, 0.1, 0.2, 0.3, 0.4 and 0.5

sumption that the null population is symmetric. Thus, as soon as one assumes the population under null to be asymmetric, the statistic does not have the standard distribution that one uses to find cut-off points. Therefore carrying out usual test (i.e. using standard cut-off points) will not result in a test of the size desired. Also numerical power obtained using the standard cut-off points will not have much meaning without the precise knowledge of the null distribution. This has been exhibited in Section 2.3. Further, the null distribution of the Wilcoxon-signed rank test statistics when the functional form of the null population is not known, other than that it is asymmetric, remains intractable. In such circumstance to gain some insight into the behavior of the power of the test we introduce and define relative power as follows.

First we carry out the Wilcoxon-signed rank test as before and then find the empirical "size" say, α, and empirical "power" say, β. But as explained above one does not have the precise knowledge of the null distribution and thus these numbers are not meaningful. However, assuming the linearity, we define the relative power, β^* as (Patil's recommendation)

$$\beta^* = \frac{\beta - \alpha}{\beta}.$$

In Section 2.4 we exhibit through simulation that the relative power of the test decreases as asymmetry increases.

The R codes for the simulations are given in the Appendix.

2.3 Simulation Study: Power of the Wilcoxon Signed-Rank Test

Our main in this section is to investigate, through simulations, whether the Wilcoxon signed-rank test is robust against the assumption of symmetry. For that we carry out the Wilcoxon signed-rank test in an ideal situation and evaluate its empirical power. That is, by having the null population being symmetric, we test the null hypothesis that the median of the population from which a sample is selected, is equal to the median of the symmetric null population. Here samples are taken from the populations which have exactly the same shape (or functional form) as the null population except that their medians are larger than the null median. Then by carrying out Wilcoxn signed-rank tests we find its empirical power. This procedure is then repeated by perturbing the null population so as to make it asymmetric. That is, by having asymmetric population being the null population, we again test the null hypothesis that the median of the population from which we select a sample, is equal to the median of the null population. Here samples are taken from populations which have exactly the same shape (or functional form) as the null population except that their medians are larger than the null median. Then as before by carrying out Wilcoxn signed-rank test we find its empirical power.

To be precise, using the asymmetry quantification described in Section 2.2.1, we will have null populations with increasing asymmetry coefficients. Thus, for the simulations considered in this section we have null populations with asymmetry coefficients equal to 0, 0.1, 0.2, 0.3. 0.4 and 0.5. To illustrate, let X be random variable such that $\eta(X) = 0$, that is, a symmetric population. Without of loss of generality let its median to be zero. Then we select alternative population such that the random variable Y associated with this population is, $Y = X + \theta$. This give us an alternative which has the same shape (which is symmetric here) as the null population except that its median is different. Similarly, if for a random variable associated with null population is such that $\eta(X) > 0$ and median of X is M then we take the alternative population such that the r.v. Y associated with this population is given by $Y = X + \theta$. Again because of the properties of η, $\eta(X) = \eta(Y) > 0$. That is both, the null and alternative populations have the same asymmetric shape and the only difference between them is their median. Thus we find the empirical powers of the Wilcoxon signed-rank test by varying the asymmetry coefficient from zero to 0.5 with an increment of 0.1.

Now we define the probability models that will conform to the requirement that we imposed on the null and alternative populations in the preceding two paragraphs. For that first we recall the mixtures of normal distributions defined in subsection 2.2.1. That is, $X \sim \alpha\mathrm{N}(\mu_1,\sigma_1^2)$ + (1 - α)N(μ_2,σ_2^2) where $0 < \alpha < 1$ is the mixing coefficient. As noted in table 2.1, as α increases from 0 to 0.491 asymmetry coefficient increases from 0 to 0.5. Therefore, clearly if for every fix α we define the null population to be αN(μ_1,σ_1^2) + (1 - α)N(μ_2,σ_2^2) distribution having median M_α and the alternative population to be αN($\mu_1+\delta$,σ_1^2) + (1 - α)N($\mu_2+\delta$,σ_2^2) distribution with median $M_{\alpha\delta}$ then these populations conforms to our requirement.

Let $\mu_1 = 0, \mu_2 = 2, \sigma_1^2 = 1, \sigma_2^2 = 4$ and $\alpha = \alpha_i, i = 1, 2, ..., 6$ where, from Table 2.1, $\alpha_1 = 0, \alpha_2 = 0.101, ..., \alpha_6 = 0.491$. Thus we have six mixed normal populations P_i where P_i is $\alpha_i\, N(\mu_1, \sigma_1^2) + (1-\alpha_i)\, N(\mu_2, \sigma_2^2)$, $i = 1, 2, ..., 6$ with increasing asymmetry. We will take these six populations as our null populations and denote their median by $M_i, i = 1, 2, ..., 6$. Now for every null population we consider 7 alternative populations such that median of the alternative populations is larger than or equal to the median of the null population and also the shape of the alternative population is same as that of the null population. For that define population P_{ij} to be $\alpha_i\, N(\mu_1 + \delta_j, \sigma_1^2) + (1 - \alpha_i)\, N(\mu_2 + \delta_j, \sigma_2^2)$, where $i = 1, 2, ..., 6$ and $j = 1, 2, ..., 7$ with $\delta_1 = 0$, $\delta_2 = 0.01$, $\delta_3 = 0.05$, $\delta_4 = 0.10$, $\delta_5 = 0.15$, $\delta_6 = 0.20$ and $\delta_7 = 0.25$. Let M_{ij} denote the median of the population $P_{ij}, i = 1, 2, ..., 6$ and $j = 1, 2, ..., 7$. Note that $P_{i0} = P_i$ and $M_{i0} = M_i$ for $i = 1, 2, ..., 6$

The medians of all these populations are given in Table 2.2.

Table 2.2: The mixing coefficient (α_i) and medians M_{ij} of of P_{ij} (i.e. medians of $\alpha_i N(0+\delta_j, 1) + (1-\alpha_i)N(2+\delta_j, 4)$)

	Median of M_{ij} of P_{ij}						
α_i	$\delta_1 = 0.00$	$\delta_2 = 0.01$	$\delta_3 = 0.05$	$\delta_4 = 0.10$	$\delta_5 = 0.15$	$\delta_6 = 0.20$	$\delta_7 = 0.25$
0.000	2.000	2.008	2.055	2.102	2.148	2.195	2.242
0.101	1.742	1.758	1.789	1.836	1.898	1.945	1.992
0.175	1.523	1.539	1.586	1.633	1.680	1.727	1.773
0.256	1.289	1.305	1.336	1.398	1.445	1.492	1.539
0.382	0.945	0.945	0.992	1.039	1.086	1.133	1.195
0.491	0.680	0.695	0.742	0.789	0.836	0.883	0.930

Remark:

The medians of P_{ij} (i.e. the medians of the Mixtures of Normal distributions) are obtained from the bisection method.

As explained above first we carry out Wilcoxon signed-rank test in the ideal situation, that is when $\eta = 0$.

- **Case 1: $\eta = 0.0$**

  We now test $H_0 : M = M_1$ against $H_1 : M = M_{1j}$ (or equivalently, $H_0 : X \sim P_1$ against $H_1 : X \sim P_{1j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for K = 10,000 times. For a test of size 0.05, we then record the empirical powers against every M_{1j} and for the sample sizes mentioned above in Table 2.3.

Table 2.3: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.0$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.08	4.83	4.92	4.82	4.75	4.73	4.93	4.95	4.73
0.01	4.53	5.42	4.73	4.96	5.48	5.35	5.11	6.03	5.81
0.05	4.77	6.24	6.81	6.78	6.94	7.31	7.88	9.62	13.64
0.10	6.26	7.56	7.74	8.89	9.12	10.00	12.01	16.61	29.19
0.15	7.10	8.64	9.79	11.38	13.01	14.26	17.68	27.28	49.30
0.20	7.38	11.03	12.32	14.89	15.78	18.65	24.72	39.33	70.62
0.25	8.32	13.75	15.52	18.75	21.41	23.25	33.27	52.56	85.77

- **Case 2: $\eta = 0.1$**

  It means now the null population is sightly asymmetric. Here again we carry out the Wilcoxon signed-rank test by taking the alternative population which has same shape as that of null but its median is shifted to the right. That is, we now test $H_0 : M = M_2$ against $H_1 : M = M_{2j}$ (or equivalently, $H_0 : X \sim P_2$ against $H_1 : X \sim P_{2j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K = 10{,}000$ times. For a test of size 0.05, we then record the empirical powers against every M_{2j} and for the sample sizes mentioned above in Table 2.4.

Table 2.4: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.1$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.63	5.77	5.52	5.68	5.31	6.42	6.94	7.91	9.28
0.01	4.64	5.70	5.92	6.15	6.72	6.40	7.08	8.85	10.97
0.05	5.13	6.83	7.13	7.75	8.23	8.84	9.83	13.86	21.78
0.10	5.59	8.22	9.17	10.02	11.03	11.79	15.04	22.61	40.86
0.15	6.92	9.80	11.22	12.90	15.10	16.92	20.81	33.35	63.26
0.20	7.84	12.29	13.32	16.48	18.42	21.42	28.36	45.95	79.65
0.25	9.26	13.59	16.51	20.47	24.43	26.57	37.77	60.11	91.51

- **Case 3: $\eta = 0.2$**

  It means now the null population is more asymmetric compared to the last case. But here again we carry out the Wilcoxon signed-rank test by taking the alternative population which has same shape as that of null but its median is shifted to the right. That is, we now test $H_0 : M = M_3$ against $H_1 : M = M_{3j}$ (or equivalently, $H_0 : X \sim P_3$ against $H_1 : X \sim P_{3j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K = 10{,}000$ times. For a test of size 0.05, we then record the empirical powers against every M_{3j} and for the sample sizes mentioned above in Table 2.5.

- **Case 4: $\eta = 0.3$**

  It means now the null population is more asymmetric compared to the last case. But here again we carry out the Wilcoxon signed-rank test by taking the alternative population which has same shape as that of null but its median is shifted to the

Table 2.5: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.2$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	5.06	6.77	7.46	8.13	8.28	9.22	10.77	13.41	22.46
0.01	5.12	7.28	7.40	7.87	8.46	9.51	11.08	15.46	25.09
0.05	5.83	7.98	8.96	10.31	11.08	12.01	14.89	23.04	40.58
0.10	6.92	9.78	11.70	13.14	15.16	15.81	21.73	33.26	63.12
0.15	7.36	11.42	13.79	16.57	19.75	20.20	29.48	46.41	79.73
0.20	8.82	13.69	16.95	21.25	24.38	26.72	39.20	60.34	91.85
0.25	10.28	16.27	20.38	25.29	29.48	33.54	47.46	72.91	97.17

right. That is, we now test $H_0 : M = M_4$ against $H_1 : M = M_{4j}$ (or equivalently, $H_0 : X \sim P_4$ against $H_1 : X \sim P_{4j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for K =10,000 times. For a test of size 0.05, we then record the empirical powers against every M_{4j} and for the sample sizes mentioned above in Table 2.6.

Table 2.6: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.3$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	5.94	7.94	8.73	10.05	11.65	12.01	15.55	22.09	41.83
0.01	5.99	8.28	9.83	9.97	12.02	13.33	16.57	25.21	45.65
0.05	6.57	9.80	11.08	13.48	14.48	15.91	22.07	34.15	62.67
0.10	7.03	11.19	13.71	17.05	18.81	20.77	29.01	47.58	80.58
0.15	8.47	14.09	17.63	20.38	24.18	26.73	39.37	60.17	96.74
0.20	10.22	15.96	20.71	25.50	28.71	33.66	48.52	73.43	97.38
0.25	10.67	19.50	24.19	30.37	35.97	39.84	57.68	84.34	99.30

- **Case 5: $\eta = 0.4$**

  It means now the null population is more asymmetric compared to the last case. But here again we carry out the Wilcoxon signed-rank test by taking the alternative population which has same shape as that of null but its median is shifted to the right. That is, we now test $H_0 : M = M_5$ against $H_1 : M = M_{5j}$ (or equivalently, $H_0 : X \sim P_5$ against $H_1 : X \sim P_{5j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for K = 10,000 times. For a test of size 0.05, we then record the empirical powers against every M_{5j} and for the sample sizes mentioned above in Table 2.7.

Table 2.7: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.4$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	6.30	9.28	11.85	12.87	15.70	16.58	22.56	37.10	66.19
0.01	6.49	10.13	13.08	14.06	15.95	17.70	24.81	38.67	70.62
0.05	7.71	11.86	14.35	17.92	19.94	21.63	31.51	49.93	83.67
0.10	8.52	14.37	17.53	21.31	25.11	28.46	39.86	63.64	94.18
0.15	9.74	16.96	21.39	26.56	31.53	34.49	50.81	76.96	98.29
0.20	10.44	18.84	25.58	31.94	38.21	42.52	60.60	86.07	99.64
0.25	12.25	23.33	30.22	37.62	44.83	50.15	69.96	92.54	99.94

- **Case 6: $\eta = 0.5$**

  It means now the null population is more asymmetric compared to the last case. But here again we carry out the Wilcoxon signed-rank test by taking the alternative population which has same shape as that of null but its median is shifted to the right. That is, we now test $H_0 : M = M_6$ against $H_1 : M = M_{6j}$ (or equivalently, $H_0 : X \sim P_6$ against $H_1 : X \sim P_{6j}$) for every $j = 1, 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for K = 10,000 times. For a test of size 0.05, we then record the empirical powers against every M_{6j} and for the sample sizes mentioned above in Table 2.8.

Conclusions: Clearly, from Table 2.3 when the distribution is symmetry or $\eta = 0.0$, we found that the percentages of empirical power of Wilcoxon signed-rank test increase

Table 2.8: The percentages of empirical power of Wilcoxon signed-rank test when $\eta = 0.5$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	6.76	10.65	13.15	15.05	17.87	19.48	27.90	43.96	75.91
0.01	7.37	11.76	13.88	16.02	19.01	20.64	28.40	47.24	79.37
0.05	8.09	13.60	16.35	19.50	23.55	26.26	36.00	58.42	90.04
0.10	8.98	15.71	20.75	24.39	28.79	33.63	46.25	72.44	96.80
0.15	10.51	19.00	24.80	30.15	36.16	40.85	58.65	83.53	99.29
0.20	12.62	22.75	29.76	36.67	43.52	49.05	68.74	91.25	99.93
0.25	14.09	26.65	35.10	43.37	51.50	57.33	77.93	95.96	100.00

when a small constant (δ) and sample size (n) increase. It means that the figures in Table 2.3 are correct results which according to symmetric distribution assumption of the Wilcoxon signed-rank test. As for the figures from Table 2.4 - 2.8, when the distribution is asymmetric and more asymmetry, the percentages of empirical power of Wilcoxon signed-rank test tend to increase when a small constant (δ) and sample size (n) increase. They are not useful to show robustness of the Wilcoxon signed-rank test against the assumption of symmetry because the cut-off points of each of the test is obtained assuming that the null population is symmetric, whereas except for the case of $\eta = 0$ all null populations considered above are asymmetric.

The next section, we consider the relative power of test which is another value to examine the robustness of Wilcoxon signed-rank test against the assumption of symmetry.

2.4 The Relative Power of Test

For the robustness study of Wilcoxon signed-rank test, we expect, if symmetry assumption is at the center of rationale behind this test, then the power of test should decrease when the distribution changes from symmetry to asymmetry. From the simulation results in section 2.3, we found that the empirical power of test increases when the measure of asymmetry changes from 0.0 to 0.5. But as noted earlier those numbers are meaningless because the null population was not symmetric, under null, the statistic computed did not follow the standard distribution that is associated with Wilcoxon signed-rank test statistic. In this section, we take the power and size of test from the simulation study to compute and examine the relative power of test. The relative power of test (β^*), as described in section 2.2.2, is derived by subtracting the size of test from the power of test and then dividing the difference by the power of test. The quantity β^* represents the difference between the power of test and the size of test in units of the power of test.

$$\beta^* = \frac{\beta - \alpha}{\beta} \tag{2.2}$$

where β and α are power and size of test, respectively.

The empirical relative power of test is classified by the measure of asymmetry from 0.0 to 0.5 as follows:

- **Case 1: $\eta = 0.0$**

  First we test $H_0 : M = M_1$ against $H_1 : M = M_{1j}$ (or equivalently, $H_0 : X \sim P_1$ against $H_1 : X \sim P_{1j}$) for every $j = 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{1j} and for the sample sizes mentioned above in Table 2.9.

- **Case 2: $\eta = 0.1$**

  Now we test $H_0 : M = M_2$ against $H_1 : M = M_{2j}$ (or equivalently, $H_0 : X \sim P_2$ against $H_1 : X \sim P_{2j}$) for every $j = 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{2j} and for the sample sizes mentioned above in Table 2.10.

Table 2.9: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.0$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.099	0.109	-0.040	0.028	0.133	0.116	0.035	0.179	0.186
0.05	0.145	0.226	0.278	0.289	0.316	0.353	0.374	0.485	0.653
0.10	0.348	0.361	0.364	0.458	0.479	0.527	0.590	0.702	0.838
0.15	0.425	0.441	0.497	0.576	0.635	0.668	0.721	0.819	0.904
0.20	0.447	0.562	0.601	0.676	0.699	0.746	0.801	0.874	0.933
0.25	0.510	0.649	0.683	0.743	0.778	0.797	0.852	0.906	0.945

Table 2.10: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.1$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.002	-0.012	0.068	0.076	0.210	-0.003	0.020	0.106	0.154
0.05	0.097	0.155	0.226	0.267	0.355	0.274	0.294	0.429	0.574
0.10	0.172	0.298	0.398	0.433	0.519	0.455	0.539	0.650	0.773
0.15	0.331	0.411	0.508	0.560	0.648	0.621	0.667	0.763	0.853
0.20	0.409	0.531	0.586	0.655	0.712	0.700	0.755	0.828	0.883
0.25	0.500	0.575	0.666	0.723	0.783	0.758	0.816	0.868	0.899

- **Case 3:** $\eta = \mathbf{0.2}$

  Here we test $H_0 : M = M_3$ against $H_1 : M = M_{3j}$ (or equivalently, $H_0 : X \sim P_3$ against $H_1 : X \sim P_{3j}$) for every $j = 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{3j} and for the sample sizes mentioned above in Table 2.11.

- **Case 4:** $\eta = \mathbf{0.3}$

  Here we test $H_0 : M = M_4$ against $H_1 : M = M_{4j}$ (or equivalently, $H_0 : X \sim P_4$ against $H_1 : X \sim P_{4j}$) for every $j = 2, ..., 7$ and each of this test is carried out for

Table 2.11: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.2$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.012	0.070	-0.008	-0.033	0.021	0.030	0.028	0.133	0.105
0.05	0.132	0.152	0.167	0.211	0.253	0.232	0.277	0.418	0.447
0.10	0.269	0.308	0.362	0.381	0.454	0.417	0.504	0.597	0.644
0.15	0.313	0.407	0.459	0.509	0.581	0.544	0.635	0.711	0.718
0.20	0.426	0.505	0.560	0.617	0.660	0.655	0.725	0.778	0.755
0.25	0.508	0.584	0.634	0.679	0.719	0.725	0.773	0.816	0.769

samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{4j} and for the sample sizes mentioned above in Table 2.12.

Table 2.12: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.3$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.008	0.041	0.112	-0.008	0.031	0.099	0.062	0.124	0.084
0.05	0.096	0.190	0.212	0.254	0.195	0.245	0.295	0.353	0.333
0.10	0.155	0.290	0.363	0.411	0.381	0.422	0.464	0.536	0.481
0.15	0.299	0.436	0.505	0.507	0.518	0.551	0.605	0.633	0.568
0.20	0.419	0.503	0.578	0.606	0.594	0.643	0.680	0.699	0.570
0.25	0.443	0.593	0.639	0.669	0.676	0.699	0.730	0.738	0.579

- **Case 5: $\eta = 0.4$**

  Here we test $H_0 : M = M_5$ against $H_1 : M = M_{5j}$ (or equivalently, $H_0 : X \sim P_5$ against $H_1 : X \sim P_{5j}$) for every $j = 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{5j} and for the sample sizes mentioned above in Table 2.13.

Table 2.13: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.4$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.029	0.084	0.094	0.085	0.016	0.063	0.091	0.041	0.063
0.05	0.183	0.218	0.174	0.282	0.213	0.233	0.284	0.257	0.209
0.10	0.261	0.354	0.324	0.396	0.375	0.417	0.434	0.417	0.297
0.15	0.353	0.453	0.446	0.515	0.502	0.519	0.556	0.518	0.327
0.20	0.397	0.507	0.537	0.597	0.589	0.610	0.628	0.569	0.336
0.25	0.486	0.602	0.608	0.658	0.650	0.669	0.678	0.599	0.338

- **Case 6: $\eta = 0.5$**

  Here we test $H_0 : M = M_6$ against $H_1 : M = M_{6j}$ (or equivalently, $H_0 : X \sim P_6$ against $H_1 : X \sim P_{6j}$) for every $j = 2, ..., 7$ and each of this test is carried out for samples of sizes 10, 20, 30, 40, 50, 60, 100, 200 and 500. This is repeated for $K =$ 10,000 times. For a test of size 0.05, we then record the empirical relative powers against every M_{6j} and for the sample sizes mentioned above in Table 2.14.

Table 2.14: The empirical relative power of Wilcoxon signed-rank test when $\eta = 0.5$ and δ equals to 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the empirical relative power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.01	0.083	0.094	0.053	0.061	0.060	0.056	0.018	0.069	0.044
0.05	0.164	0.217	0.196	0.228	0.241	0.258	0.225	0.248	0.157
0.10	0.247	0.322	0.366	0.383	0.379	0.421	0.397	0.393	0.216
0.15	0.357	0.439	0.470	0.501	0.506	0.523	0.524	0.474	0.235
0.20	0.464	0.532	0.558	0.590	0.589	0.603	0.594	0.518	0.240
0.25	0.520	0.600	0.625	0.653	0.653	0.660	0.642	0.542	0.241

Conclusions: When the distribution becomes from symmetry ($\eta = 0.0$) to more asymmetry ($\eta = 0.5$), we found that

1. When the distribution is symmetric ($\eta = 0.0$), as sample size (n) and small constant

(δ) increase, the empirical relative power of Wilcoxon signed-rank test increases. This is exactly as one would have expected.

2. When the distribution is less symmetric ($\eta = 0.1$, 0.2 and 0.3) we will concentrate only when the sample size is large. For larger sample sizes, for example $n = 200$ or 500 there is evidence that as asymmetry increases, the empirical relative power of the test decreases.

3. There are some figures with negative relative power, for example, Table 2.9 and Table 2.10. It means that in some cases, the power of test is less than the size of test in units of the power of test. Normally, we expect the power of Wilcoxon signed-rank test should decrease when the distribution changes from symmetry to asymmetry when symmetry assumption is valid.

In the next section, we will present summarization and discussion from the section 2.3 - 2.4.

2.5 Summarization and Discussion

For the simulation study in section 2.3, we investigated changes in the power and size of Wilcoxon signed-rank test when the distribution changes from symmetry to asymmetry. The Mixtures of Normal distributions when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ and the size of asymmetry equals to 0.0 to 0.5 are chosen to study in the simulation study. The aims of the simulation study is to explore whether the Wilcoxon signed-rank test is robust test against the assumption of symmetry or not. The simulation will study the behavior of size and power of test when model distribution becomes symmetry to more asymmetric. We expect the power of the test to decrease when the distribution changes from symmetry to asymmetry.

From the simulation study results are given in Table 2.3 - 2.8, we found that when the size of asymmetry increases from 0.0 to 0.5, the power of Wilcoxon signed-rank test tends to increase. The necessary assumption for Wilcoxon signed-rank test is data sets come from a symmetric population distribution and we expect that the test is still robust if small deviations from the symmetric assumption should impair the power of test slightly. All of simulation study results reverse our expectancy. That is due the fact that the population is not symmetric, the null distribution does not follow the standard probability distribution associated with Wilcoxon signed-rank test statistic. Therefore, this simulations make the point that to study the behavior of the test under consideration we either have to: derive the distribution of the Wilcoxon signed-rank test statistic which is dependent on the size of asymmetry and then use this to test the hypothesis or find an alternative method to its power behavior as asymmetry increases.

Although it may be extremely difficult, in principle, deriving the distribution of the Wilcoxon signed-rank test dependent of the size of symmetry may be possible. However, we did not be follow this route. Instead, as an alternative methodology to study the power behavior, in section 2.4 we computed and considered the relative power of Wilcoxon signed-rank test. It seems the relative power of test is meaningful. We found that the relative power of test tends to decrease when the distribution changes from symmetry (η = 0.0) to asymmetry (η = 0.5) at least when sample size is large. So the overall lesson is this: although we may not be able to precisely evaluate the effect of the asymmetry on the power the Wilcoxon signed-rank test (unless we take a very difficult task of evaluating the Wilcoxon signed-rank test statistic's distribution dependent on the size of asymmetry), one clearly needs to be careful when using the Wilcoxon signed-rank test when the population is asymmetric.

In the next Chapter we provide a methodology that can be used to carry out the Wilcoxon signed-rank test when faced with asymmetric population.

Chapter 3

Inverse Transformation Method

3.1 Introduction

In Chapter 2, we have studied robustness of the Wilcoxon signed-rank test against the assumption of symmetry through simulations. From the simulation results, we investigated and found that the Wilcoxon signed-rank test does get effected if the symmetry assumption is violated. Thus our main aim here is to provide way to make the use Wilcoxon signed-rank test possible when null the population is not symmetric.

For that first in the next section we define and explain probability integral transform or inverse transform method. This method will be used to transform a given sample from a population with distribution function $F(x)$ to a sample from Uniform population. In section 3.3 we propose a methodology which, in slightly unrealistic (that is, idealistic) setting, allows one to use Wilcoxon signed-rank test when null the population is asymmetric. The idealistic setting here refers to having a very precise knowledge of the probability distribution under the null hypothesis. There we also propose a practical approach to overcome the requirement of the precise knowledge of the null population so as to make use of the Wilcoxon signed-rank test possible whenever the assumption of symmetry is doubtful. However, the practical implementation or theoretical analysis of the proposed practical approach is beyond the scope of this dissertation. In section 3.4 we carry out the simulation to show that the proposed methodology of carrying out Wilcoxon signed-rank test under the idealistic assumption works.

3.2 The Inverse Transformation

In this section, we propose the inverse transformation method which transform any continuous random variable to a continuous and symmetric uniform (0,1) random variable.

This method is used to generate of random variables from any non-uniform distribution, usually by applying a transformation to uniformly distributed random variables. Thus the uniform distribution is useful for sampling from any distributions. The inverse transformation method relates to the cumulative distribution function (cdf.) of the target random variable which has a uniform (0,1) distribution. This method is very useful in theoretical work and is a basic method for pseudo-random number sampling, i.e. for generating sample numbers at random from any probability distribution given its cumulative distribution function (cdf.). According to the restriction that the distribution is continuous, this method is applicable and can be computationally efficient if the cumulative distribution function (cdf.) can be analytically inverted. But for some probability distributions, it may be too difficult in practice.

Definition 3.1 *If X is a random variable with a continuous cumulative distribution function $F(x) = P(X \leq x)$, then the random variable*

$$U = F(X)$$

has a uniform (0,1) distribution. This fact provides a very simple relationship between a uniform random variable U and a random variable X with cumulative distribution function F:

$$X = F^{-1}(U).$$

This transformation is called **the method of inversion** or **the inverse transformation method** (Devroye (1986)).

That is, suppose that a random variable X has a continuous distribution with the cumulative distribution function F, then the random variable U is defined as

$$U = F_X(X)$$

has a uniform (0,1) distribution.

It is clear that we will get the new random variable $U = F(X)$ which is uniformly distributed on [0,1] from the inverse transformation method. Since the standard uniform distribution is a continuous and symmetric distribution on interval (0,1) and it satisfies essential assumptions for the Wilcoxon signed-rank test. Thus, we will make use the inverse transformation method in our proposal to the use of the Wilcoxon signed-rank test when the null population is not symmetric.

3.3 Methodology: Wilcoxon Signed-Rank Test When the Null Population is Asymmetric

We begin by describing the methodology that we will use when the null population is asymmetric but it is completely known. For that let $X_1, X_2, ..., X_n$ be a given random sample and one would like test,

$$H_0 : M = M_0 \text{ against } H_1 : M = M_1$$

where M_0 and M_1 are the medians of the null and alternative populations respectively. For simplicity assume that $M_0 < M_1$. Let the distribution function associated with the null population be $F(x)$. Let us also assume that $F(x)$ is completely known and it is not symmetric. The population under the alternative hypothesis has the same shape as that of $F(x)$ but it is shifted to the right.

As will be the case with any test hypothesis problem, we will assume that the sample is coming from the null and then find the evidence to conclude otherwise. Thus first construct a sample $U_1, U_2, ..., U_n$ where $U_i = F(X_i), i = 1, 2, ..., n$. Also define $\theta_0 = F(M_0)$ and $\theta_1 = F(M_1)$. Clearly, under the null hypothesis, $U_1, U_2, ..., U_n$ is a random sample from a uniform distribution which is **symmetric**. Also $\theta_0 = 1/2$. Now, if θ denotes the population median, carry out the usual Wilcoxon signed-rank test based on sample $U_1, U_2, ..., U_n$ to test the hypothesis

$$H_0 : \theta = 1/2 \text{ against } H_1 : \theta = \theta_1.$$

Clearly now the conditions of the Wilcoxon signed-rank test are met. Therefore, under the null hypothesis, the sample is form the null population which symmetric, and hence Wilcoxon signed-rank test statistic will have the standard distribution that is associated with this test. Thus carrying said test now will lead to meaningful and valid inference.

For example if sample is really from the null population, the test statistic will be very less likely to be too large and hence null hypothesis will not be rejected.

However, if the sample is really coming from the alternative population, then the transformed sample will automatically have a median larger than the hypothesized null value. This is because distribution function used to transform the data is shifted to left compared to the true distribution function. This will mean with high probability the test statistic value will be large and hence the null will get rejected.

The above methodology, although interesting, it is based on the unrealistic assumption of the complete knowledge of the distribution function associated with the null population. If it is unknown, one can modify the above methodology as follows.

As before, let $X_1, X_2, ..., X_n$ be the given random sample and one would like test,

$$H_0 : M = M_0 \text{ against } H_1 : M = M_1$$

where M_0 and M_1 are the medians of the null and alternative populations respectively. For simplicity assume that $M_0 < M_1$. Let the distribution function associated with the null population be $F(x)$. We assume that it is asymmetric and that the population under the alternative hypothesis has the same shape as that of $F(x)$ but it is shifted to the right. But most importantly, as opposed the idealistic situation considered above, we assume that $F(x)$ is unknown.

Now use the given sample to estimate the probability density function, $f(x)$ associated with the population from which sample is coming from. That is, let $\widehat{f}_h(x)$ denotes the kernel estimate of $f(x)$. Then define $\widehat{F}(x) = \int_{-\infty}^{x} \widehat{f}_h(u)du$. Now construct a sample $V_1, V_2, ..., V_n$ where $V_i = \widehat{F}(X_i), i = 1, 2, ..., n$.

The most important difference, at this point, between idealistic scenario that is described above and the realistic situation that is considered now is this: If $X_1, X_2, ..., X_n$ is really coming from the null population then $U_1, U_2, ...U_n$ are uniformly distributed over zero to one. But in the current scenario, irrespective of whether the null population or alternative population produces the sample $X_1, X_2, ..., X_n$, $V_1, V_2, ..., V_n$ are expected to have uniform distribution. Also define $\theta_0 = \widehat{F}(M_0)$, $\theta_1 = \widehat{F}(M_1)$ and set $W_i = V_i - \theta_0, i = 1, 2, ..., n$. Now use the signs of W_i and the ranks of $|W_i|$ to compute Wilcoxon signed-rank test and denote it by T.

Observe that if the original sample $X_1, X_2, ..., X_n$ is from the null population, one expects W_is to be distributed symmetric about zero. Thus one expects T to have the standard distribution associated with Wilcoxon signed-rank test statistic when the null is true. Thus, with very low probability it will be large and, as one would like, it will not lead to the rejection of H_0 with high probability.

Now if the original sample $X_1, X_2, ..., X_n$ is from the alternative population, one does not expect expects W_is to be distributed symmetric about zero. In fact, one expects high proportion of of W_is to be greater than zero. This is because populations under the null and the alternative are of the same shape, except that the latter population is shifted to right, one expects $\theta_1 = 1/2 > \theta_0$. This means one expects with a high probability T to take larger values compared to the standard distribution associated with Wilcoxon signed-rank test statistic when T is computed from symmetrically spread W_is. Thus, such values of T will lead to the rejection of H_0 as one would like.

To illustrate that the above methodology does the right thing we carry out the simulation on exactly the same null and alternative populations that we used in the simulations of the last chapter. That is, we still study the Mixtures of Normal distributions from two populations that follow a Normal distribution with mean μ_1, μ_2, variance σ_1^2, σ_2^2, respectively, and a mixing coefficient α where $0 < \alpha < 1$ or $X \sim \alpha\mathrm{N}(\mu_1,\sigma_1^2)$ + (1 - α)N(μ_2,σ_2^2).

For the Mixtures of Normal distributions, let X be a random variable with the Mixtures of Normal distributions $\alpha\mathrm{N}(\mu_1, \sigma_1^2)$ + (1 - α)N(μ_2, σ_2^2). The probability density function of X is $f(x)$ where

$$f(x) = \alpha\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-(x-\mu_1)^2/2\sigma_1^2} + (1-\alpha)\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-(x-\mu_2)^2/2\sigma_2^2}$$

for $-\infty < x < \infty$, $-\infty < \mu_1 < \infty$, $-\infty < \mu_2 < \infty$, σ_1^2 , $\sigma_2^2 > 0, 0 < \alpha < 1$.

Let

$$\begin{aligned} U_1 &= F(X_1) \\ U_2 &= F(X_2) \\ U_3 &= F(X_3) \\ &. \\ &. \\ &. \\ U_n &= F(X_n). \end{aligned}$$

Clearly, $U_1, U_2, U_3, ..., U_n$ are continuous uniformly distributed on [0,1] random variables or $U_1, U_2, U_3, ..., U_n \sim U(0, 1)$.

We generate random samples U and test a hypothesis about the population median by the Wilcoxon signed-rank test. The statistical hypothesis for the Wilcoxon signed-rank test is

$$\text{Ho} : \text{M} = 0.5 \qquad H_1 : \text{M} > 0.5$$

The Monte Carlo study is conducted with extreme care in order to provide quality assurances on the accuracy of the results. We transform random samples from the Mixtures of Normal distributions to the uniform distribution.

That is, if

$$X \sim \alpha N(\mu_1, \sigma_1^2) + (1 - \alpha)N(\mu_2, \sigma_2^2)$$

then the null distribution is

$$U = F(x) = \alpha\Phi\left(\frac{x - \mu_1}{\sigma_1}\right) + (1 - \alpha)\Phi\left(\frac{x - \mu_2}{\sigma_2}\right)$$

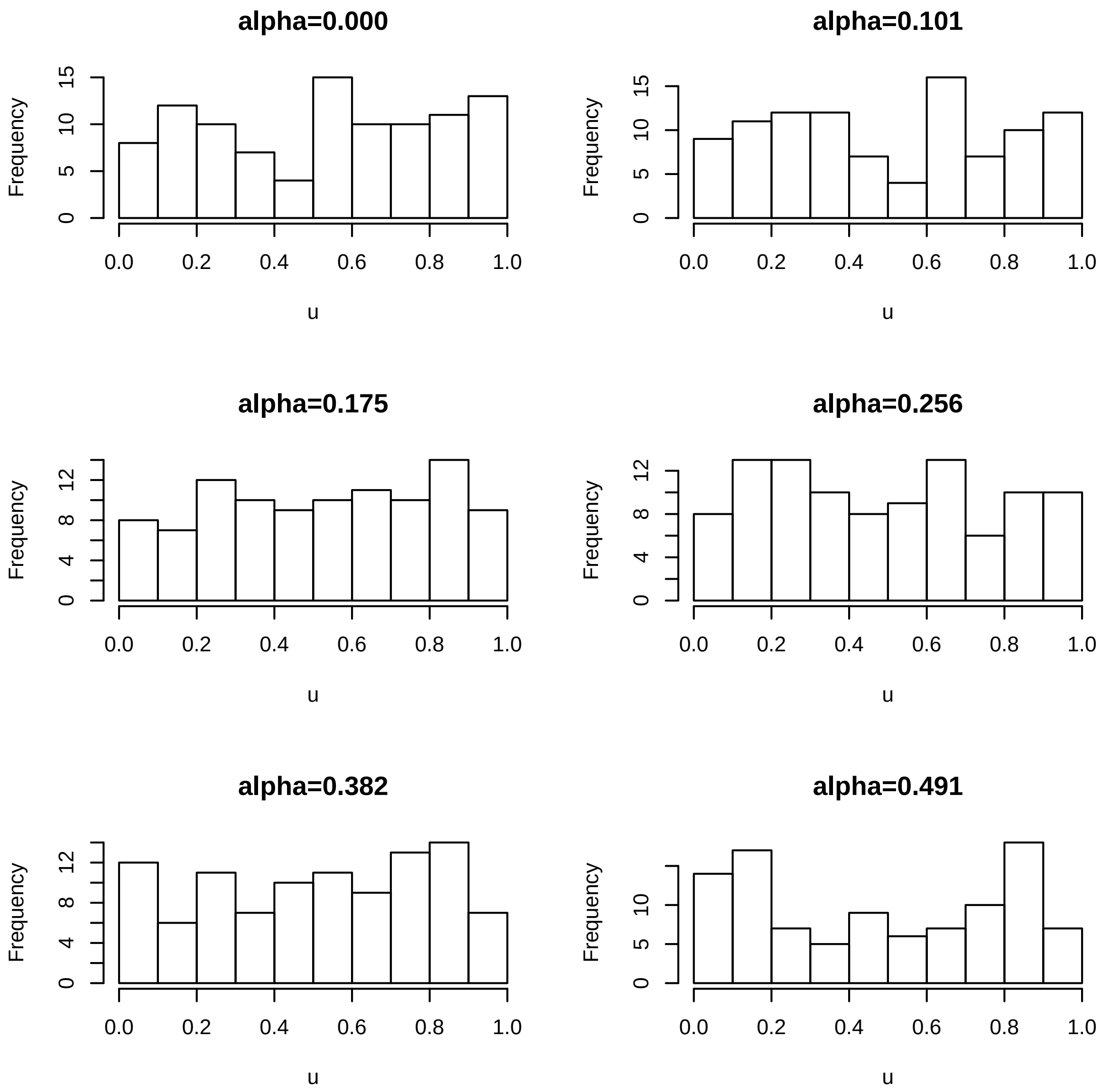

Figure 3.1: Histograms of $U = F(x) = \alpha\Phi(\frac{x-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{x-\mu_2}{\sigma_2})$ when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

where F is the distribution function under the null hypothesis.

Histograms of $U = F(x) = \alpha\Phi(\frac{x-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{x-\mu_2}{\sigma_2})$ when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491 are shown in Figure 3.1.

We generate random samples: Y from the Mixtures of Normal distributions when $\mu_1 = 0 + \delta_i$, $\mu_2 = 2 + \delta_i$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ with sample sizes 10, 20, 30, 40, 50, 60, 100, 200, 500 iteration 10,000 times and δ_i is a small constant that shifts the hypothesized median between 0.00 to 0.25. After that, we transform random samples Y_i to U_i.

That is, if

$$Y \sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$$

then

$$U = F(y) = \alpha\Phi\left(\frac{y-\mu_1}{\sigma_1}\right) + (1-\alpha)\Phi\left(\frac{y-\mu_2}{\sigma_2}\right)$$

where F is the distribution function under the null hypothesis.

Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1)}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2)}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$ when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.01, 0.05, 0.10, 0.15, 0.20, 0.25$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491 are shown in Figure 3.2 - 3.7.

From Figure 3.1 - 3.7, we found that a random variable U is continuous uniformly distributed on [0,1] when we transform a random variable Y to U by using the cumulative distribution function of the null population as the transforming function.

The investigation of the stability of the size and power of the Wilcoxon signed-rank test after applying the inverse transformation method is carried out in the next section. The R codes for the simulations are given in the Appendix.

3.4 Simulation Study

In this section, through simulation we examine whether or not the use of the inverse transformation makes Wilcoxon signed-rank test applicable when the assumption of symmetry is violated. For that we consider the mixtures of two Normal distributions with $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$. By varying the mixing coefficient α as we did in the last Chapter we construct populations of various sizes of asymmetry with size varying from 0.0 to 0.5. Then as per the procedure described in the last section we carry out the Wilcoxon sign-rank test on the samples which are transformed using the cumulative distribution function of the null population as the transforming function. There we record the empirical powers of the test and the results are classified by the coefficient of asymmetry (η) of the Mixture of Normal distributions.

- **Case 1:** $\eta = 0.0$

  The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.0$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.1.

Table 3.1: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.0$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.35	4.88	4.95	5.03	4.82	4.39	4.88	5.11	5.02
0.01	4.42	5.08	5.27	4.89	5.35	5.34	5.47	5.57	6.40
0.05	5.08	5.70	5.77	6.49	6.82	7.37	8.22	9.01	13.45
0.10	5.83	7.15	7.73	9.00	9.22	10.07	11.38	17.26	29.45
0.15	6.79	8.18	10.19	11.19	12.69	14.11	17.57	26.81	49.46
0.20	7.46	10.10	12.92	14.42	16.63	18.78	25.04	39.13	70.56
0.25	8.41	13.49	15.36	18.77	21.69	23.78	33.02	53.26	86.73

- **Case 2:** $\eta = 0.1$

  The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.1$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.2.

- **Case 3:** $\eta = 0.2$

Table 3.2: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.1$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.20	4.82	4.89	5.12	4.79	4.83	4.68	4.86	4.90
0.01	4.18	4.52	5.11	5.34	5.11	5.60	5.60	5.45	6.53
0.05	4.83	6.01	6.53	6.42	7.30	7.58	7.96	9.19	13.26
0.10	5.58	7.38	7.46	8.87	8.90	9.61	11.88	17.00	28.33
0.15	6.29	8.91	10.15	10.89	12.67	14.06	17.84	27.15	48.81
0.20	7.00	10.38	12.85	14.02	16.12	18.00	24.62	38.74	69.71
0.25	8.46	12.45	15.76	17.76	20.71	22.82	32.27	51.43	84.83

The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.2$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.3.

Table 3.3: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.2$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.06	4.32	5.11	5.00	4.93	4.61	4.69	4.86	5.05
0.01	4.14	5.12	4.61	4.79	4.99	5.20	5.21	5.69	5.87
0.05	4.75	5.91	6.01	6.56	6.51	6.99	8.46	9.55	13.11
0.10	5.72	7.54	7.86	8.20	9.64	10.13	11.61	16.67	28.84
0.15	6.00	8.89	9.90	11.27	13.26	13.92	18.38	26.32	48.88
0.20	7.66	10.87	11.70	13.86	16.64	18.03	24.30	38.95	69.81
0.25	8.36	13.50	14.90	17.90	21.01	22.65	32.41	53.74	85.77

- **Case 4: $\eta = 0.3$**

  The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.3$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.4.

Table 3.4: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.3$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.66	5.19	4.52	4.58	4.88	4.88	5.18	4.94	5.00
0.01	4.25	4.82	5.25	5.02	5.32	5.34	5.42	5.84	6.47
0.05	4.93	6.01	6.55	6.29	6.44	6.65	7.91	9.28	13.42
0.10	5.32	7.65	7.95	9.28	9.33	10.68	12.32	17.22	28.80
0.15	6.70	9.22	10.83	11.27	12.85	13.62	18.94	27.45	50.44
0.20	7.42	11.24	13.11	14.73	17.35	19.14	25.20	40.54	70.53
0.25	8.29	12.44	15.88	19.09	21.20	24.49	33.86	54.67	86.79

- **Case 5: $\eta = 0.4$**

  The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.4$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.5.

Table 3.5: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.4$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.24	4.76	4.87	5.24	4.99	4.91	4.62	4.49	4.78
0.01	4.19	5.03	4.88	5.47	5.53	5.24	5.21	5.98	6.57
0.05	4.33	6.20	6.68	6.76	7.17	7.43	8.05	10.16	14.73
0.10	5.43	7.48	8.73	9.08	9.55	10.83	13.56	19.03	32.02
0.15	6.83	9.72	10.13	12.08	13.74	13.97	20.41	30.82	55.42
0.20	7.38	11.50	14.09	16.35	18.26	20.21	27.78	44.24	76.85
0.25	8.93	13.56	16.83	20.79	23.53	26.13	38.22	59.99	91.19

- **Case 6: $\eta = 0.5$**

  The empirical powers of Wilcoxon signed-rank test are obtained when the inverse transformation method is applied in case of $\eta = 0.5$ with constant (δ) being the distance between the null and alternative median and are recorded in Table 3.6.

Table 3.6: The percentages of empirical power of Wilcoxon signed-rank test when the inverse transformation is applied in case $\eta = 0.5$ and δ equals to 0.00, 0.01, 0.05, 0.10, 0.15, 0.20 and 0.25

	the percentages of empirical power of test								
δ_i	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 100	n = 200	n = 500
0.00	4.20	4.69	4.47	4.64	4.44	4.70	4.85	4.75	4.50
0.01	4.27	5.11	5.17	5.06	5.21	5.41	5.55	5.95	6.59
0.05	5.01	6.41	6.84	7.02	7.32	7.37	8.36	10.67	15.57
0.10	6.36	7.96	9.19	9.88	10.40	11.56	13.95	20.07	36.55
0.15	6.82	10.45	11.27	13.12	15.09	15.79	22.03	34.45	61.33
0.20	8.23	12.63	14.77	17.22	19.71	22.00	30.43	49.62	82.69
0.25	9.88	14.94	19.02	21.80	26.07	29.75	42.46	65.64	94.61

Conclusions: When we use the transformation method before computing the Wilcoxon signed-rank test statistic, the samples from the Mixtures of Normal distributions with various sizes of asymmetry (i.e. asymmetry size 0.0 to 0.5) are transformed using the distribution function associated with the null population. Thus whenever a sample is from the null population, after transformation it will be a sample from a uniform (0,1) distribution which is a continuous and symmetric distribution. We expect that the size and power of the Wilcoxon signed-rank test should be stable for every samples which are transformed. From the simulation results, we found that

1. The percentages of empirical size of test (when $\delta = 0.00$) are quite similar for every sizes of asymmetry (η). The percentages are between 4.06% (when $\eta = 0.2$, n = 10) and 5.24% (when $\eta = 0.4$, n = 40).

2. For every sizes of asymmetry (η), the percentages of empirical power of test are quite similar, that is, the power increases when sample size (n) and the small constant (δ) increase. The percentages are between is 4.14% (when $\eta = 0.2$, n = 10, $\delta = 0.01$) and 94.61% (when $\eta = 0.5$, n = 500, $\delta = 0.25$).

That means that when we transform the samples by using the cumulative distribution function of the null population as the transforming function, we observed that there are the stability of the size and power of Wilcoxon signed-rank test.

3.5 Summarization and Discussion

From the simulation study in Chapter 2, we found that the Wilcoxon signed-rank test is not robust against the assumption of symmetry. The Wilcoxon signed-rank test does require the population to be continuous and symmetric. In this chapter, we study the inverse transformation method which transform a sample from any distribution to a standard uniform distribution if the transforming function is the cumulative distribution of the function from which the original sample is selected. We apply the inverse transformation to a given sample before computing the Wilcoxon signed-rank test statistic and examine whether or not the test does the right thing.

The simulation study was carried out to investigate the power of the Wilcoxon signed-rank test after we apply the inverse transformation method to the test. The Mixtures of Normal distributions with $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$ with size of asymmetry varying from 0.0 to 0.5 were chosen to study. From the simulation results, we found that the empirical size and power of the Wilcoxon signed-rank test are very similar irrespective of the size of asymmetry. The empirical size of test is between 4.06% and 5.24% and as sample size (n) and the constant (δ difference between the alternative and null median) increase, the empirical power of test increases, and that increase is between 4.14% and 94.61%.

Clearly, the Wilcoxon signed-rank test is not robust against the assumption of symmetry and the continuous and symmetric distribution are essential for the Wilcoxon signed-rank test. In this chapter, we studied and applied the inverse transformation method to the Wilcoxon signed-rank test to change any distribution which may or may not be a symmetric distribution to symmetry, a uniform (0,1) distribution. Moreover, we also investigated the stability of size and power of the Wilcoxon signed-rank test and found that the test is still good and applicable after applying the inverse transformation method. Thus, one may say that the inverse transformation method makes it possible to use Wilcoxon signed-rank test when assumption of symmetry is not be valid.

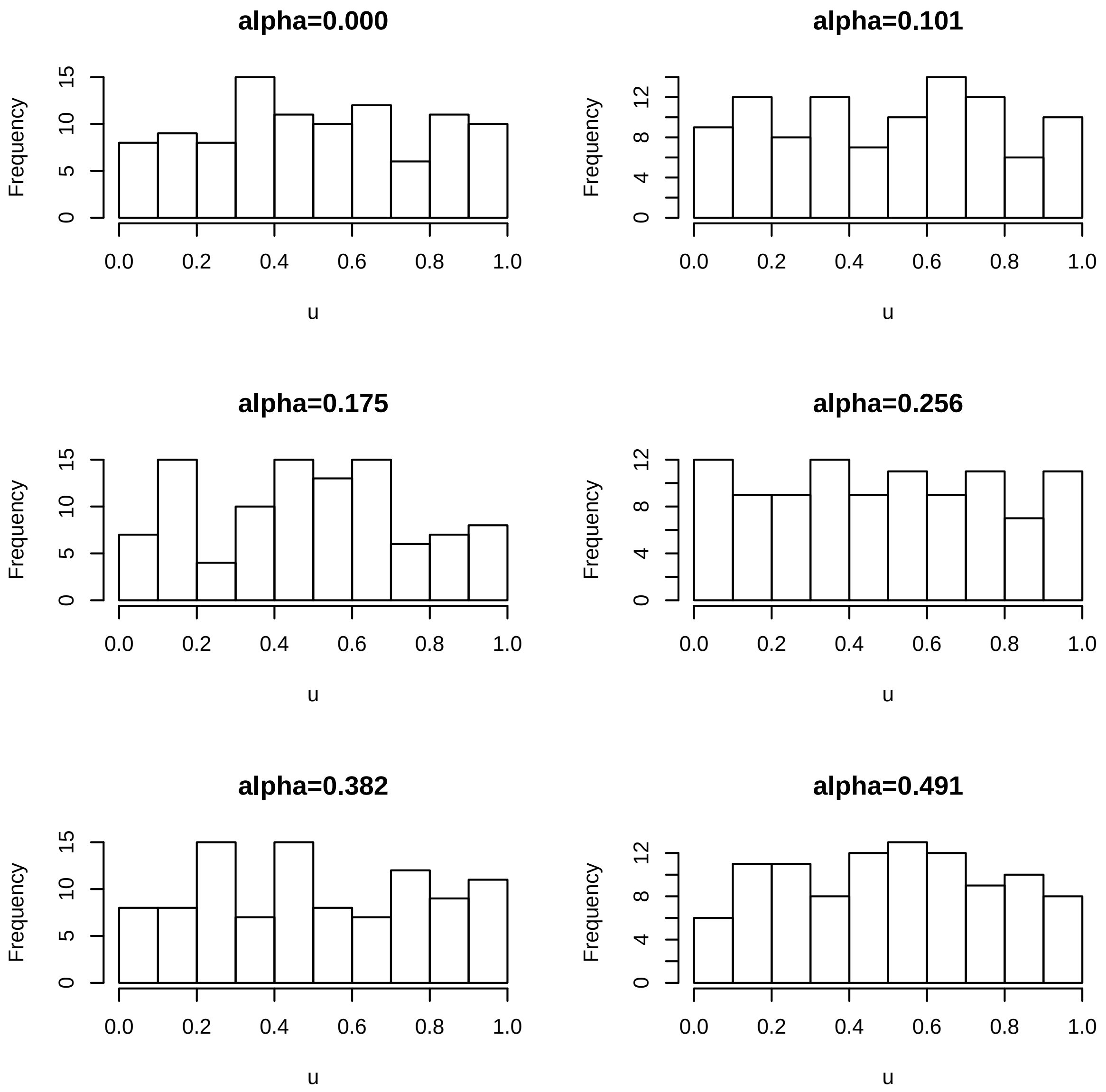

Figure 3.2: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.01$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

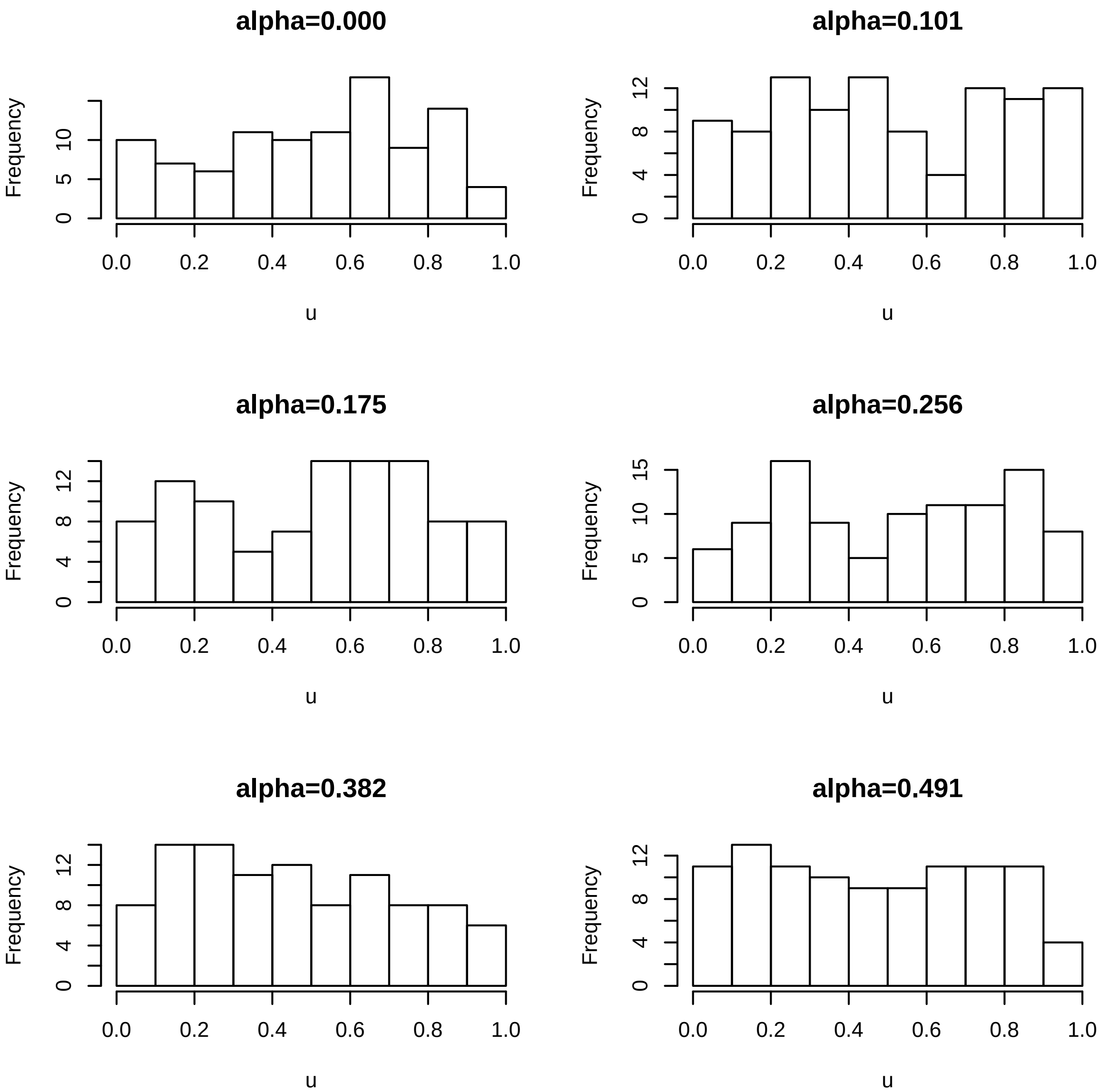

Figure 3.3: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.05$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

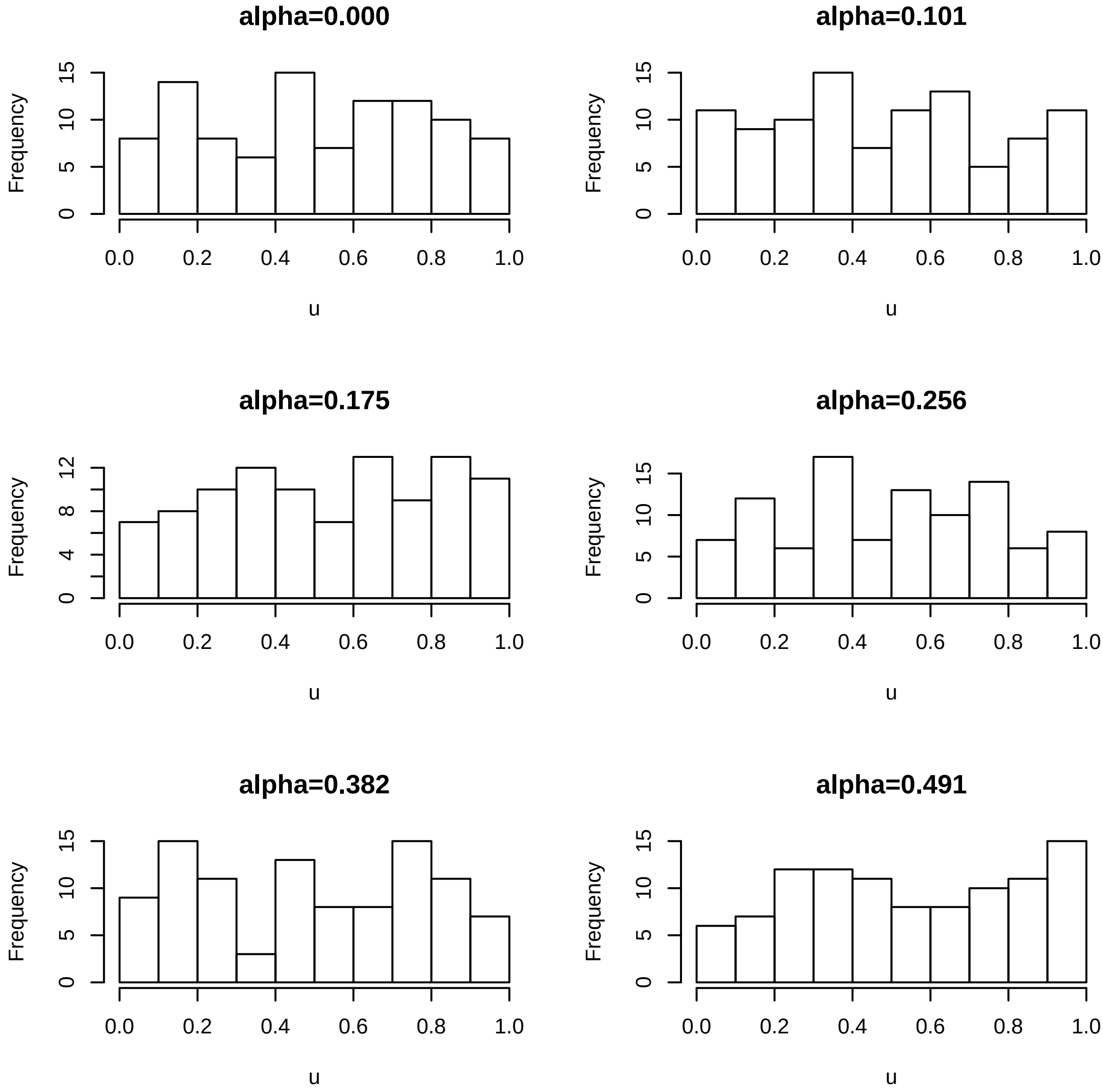

Figure 3.4: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.10$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

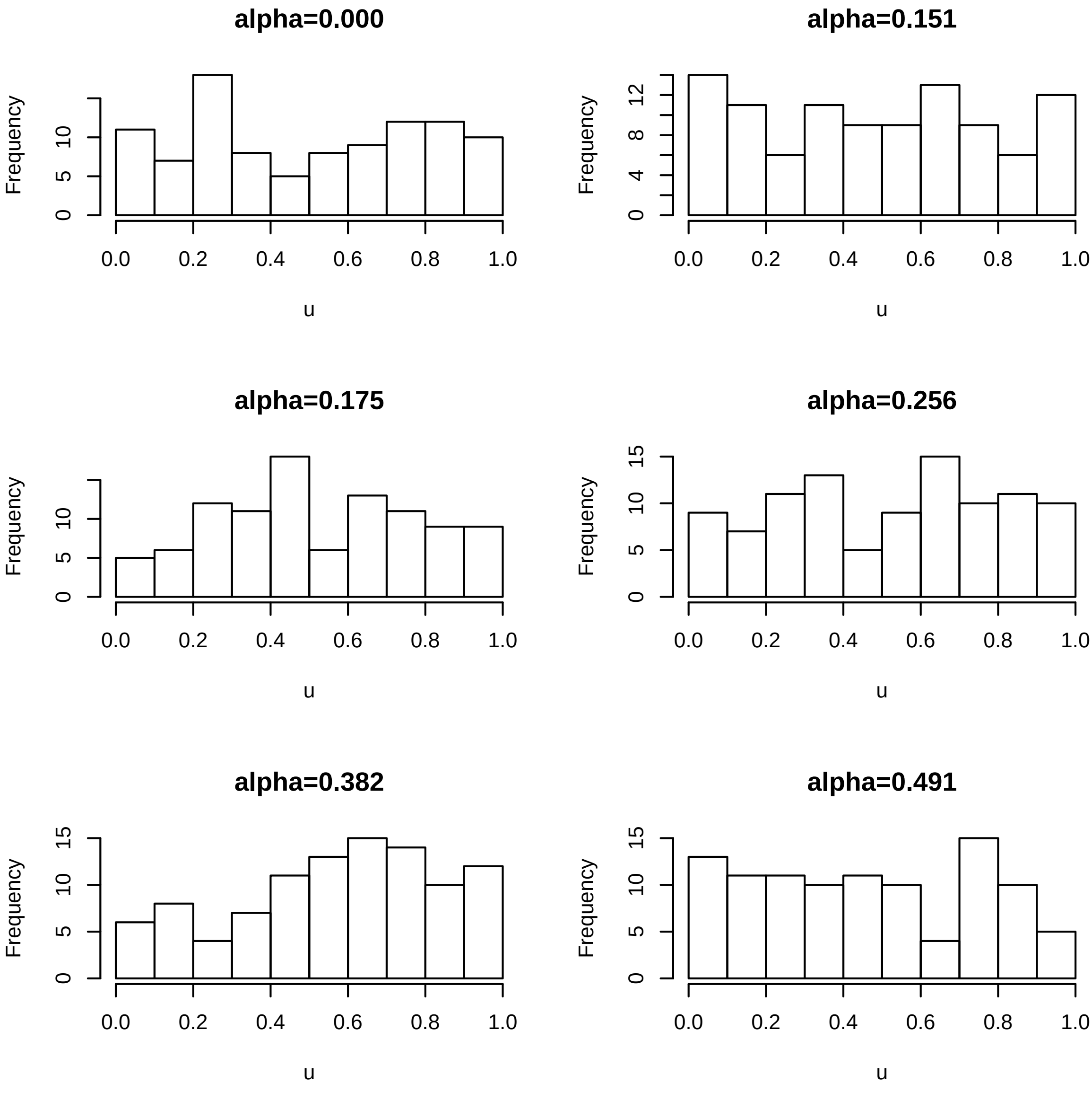

Figure 3.5: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.15$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

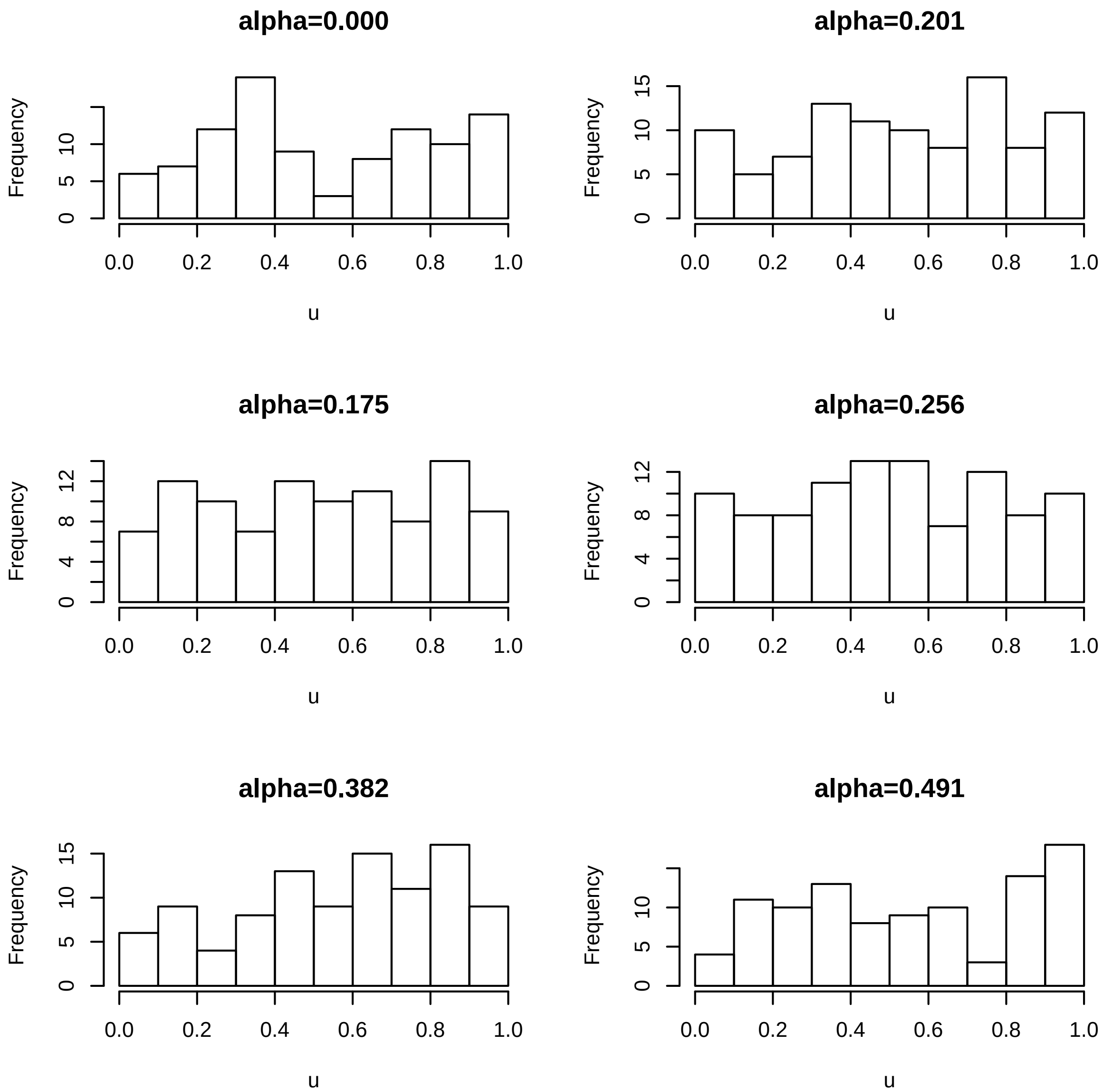

Figure 3.6: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.20$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

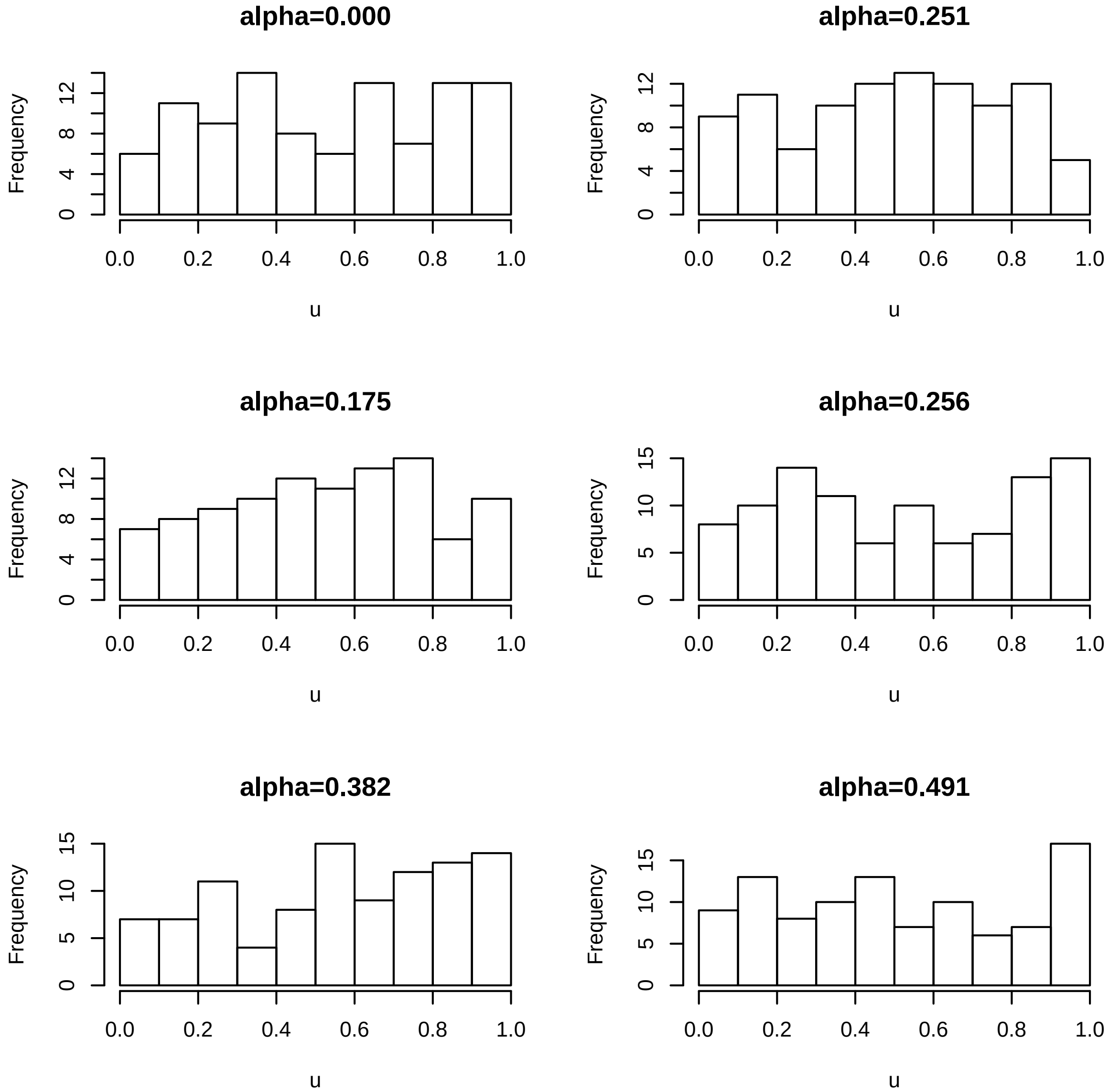

Figure 3.7: Histograms of $U = F(y) = \alpha\Phi(\frac{y-\mu_1}{\sigma_1}) + (1-\alpha)\Phi(\frac{y-\mu_2}{\sigma_2})$ where Y $\sim \alpha N(\mu_1 + \delta, \sigma_1^2) + (1-\alpha)N(\mu_2 + \delta, \sigma_2^2)$, when $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, $\delta = 0.25$ with sample sizes = 100 and the mixing coefficients (α) equal to 0.000, 0.101, 0.175, 0.256, 0.382 and 0.491

Chapter 4

Conclusion and Future Work

4.1 Introduction

Clearly, the necessary assumptions are important for the validity of every statistical inferential procedure. In this dissertation, we particularly focus on the Wilcoxon signed-rank test which is the well-known nonparametric test about the population median in one sample case. When observations are not drawn from the normal distribution, the Wilcoxon signed-rank test is the analogue of independent one sample t- and Z- test. The continuous and symmetric distribution are essential assumptions for the Wilcoxon signed-rank test. Our interest was to study robustness of the Wilcoxon signed-rank test against the assumption of symmetry through simulations.

Simulations were carried out to investigate the stability of the size and power of the Wilcoxon signed-rank test when the population distribution changes from symmetry to more and more asymmetry. The Mixtures of Normal distributions are some of the examples of asymmetric probability distributions which are used in the simulations and their associated asymmetry coefficients. In the simulations, we examined the empirical size and power of the Wilcoxon signed-rank test when the size of asymmetry changes from 0.0 to 0.5. Furthermore, we proposed the inverse transformation method to make the Wilcoxon signed-rank test applicable when the assumption of symmetry is not met.

4.2 Conclusion

For the simulations in this dissertation, the Mixtures of Normal distributions with $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_1^2 = 1$, $\sigma_2^2 = 4$, mixing coefficients (α) vary for 0.000 to 0.491 and its size of asymmetry changes from 0.0 to 0.5 were chosen to study the robustness of Wilcoxon signed-rank test.

In Chapter 2, we investigated the percentages of empirical size and power of the Wilcoxon signed-rank test. We found that when the distribution is symmetric ($\eta = 0.0$), the size of test is between 4.08% and 4.95%, and as a small constant (δ) and sample size (n) increase, the power of test increases as well, the ranges of the power are from 4.53% to 85.77% (see Table 2.3). When the distribution becomes more and more asymmetric (η = 0.1 - 0.5), the power of test increases when a small constant (δ) and sample size (n) increase (see Table 2.4 - 2.8). However, clearly these numbers were meaningless. When we considered the relative power of the Wilcoxon signed-rank test (β^*), we found that the relative power increases when the distribution is symmetric, a small constant (δ) and sample size (n) increase (see Table 2.9). When the distribution is less symmetric ($\eta = 0.1$ - 0.3), the relative power is small increase when a small constant (δ) and sample size (n) increase (see Table 2.10 - 2.12). When the distribution is more symmetric ($\eta = 0.4$ and 0.5), the relative power tend to decrease when a small constant (δ) is fixed and sample size (n) increases (see Table 2.13 - 2.14). This would mean that the Wilcoxon signed-rank test is not robust against the symmetric assumption.

In Chapter 3, we proposed the inverse transformation method which is a technique to transform any distribution to a continuous and symmetric distribution, a uniform (0,1) distribution. Also, in simulation study we applied the inverse transformation method to samples before employing Wilcoxon signed-rank test. In these simulations we used the known population distribution, the Mixtures of Normal distributions, to transform to the standard uniform distribution and investigated the percentages of empirical size and power of the Wilcoxon signed-rank test. We found that for every distributions for which size of asymmetry equals 0.0 - 0.5, the size and power of the Wilcoxon signed-rank test are not different. The ranges of the size and power of the test are from 4.06% to 5.24% and 4.14% to 94.61%, respectively (see Table 3.1 - 3.6). That is, the Wilcoxon signed-rank test is applicable after applying the inverse transformation method to the desired distribution. Thus, we could apply the inverse transformation method if the population distribution is not symmetric and then, carry on the Wilcoxon signed-rank test.

4.3 Future Work

From the main results, it is clear that the Wilcoxon signed-rank test is not robust when the population distribution is not symmetric. The essential assumptions for the Wilcoxon signed-rank test are continuous and symmetric population distribution. For the test to be applicable, we suggest the inverse transformation method to transform any arbitrary distribution to a continuous and symmetric distribution, a uniform (0,1) distribution, and apply the method to the Wilcoxon signed-rank test. The simulations in this dissertation, we particularly study the Mixtures of Normal distributions with the size of asymmetry changes from 0.0 to 0.5 to be a known population distribution. However, we can not know about a population distribution in the real situation. Thus, one would like to do, estimate the population distribution before using it to trasform the data.

The suggestion for the future work is to estimate the population distribution by kernel density estimation. The univariate kernel density estimation and its related criteria will be studied and applied for any observations. Then, we will apply the inverse transformation method to the distribution which we get from the kernel density estimation. Moreover, we will investigate the empirical size and power of the Wilcoxon signed-rank test when the population distribution is transformed to be a continuous and symmetric distribution through simulation studies. Kernel density estimation and the inverse transformation method will then be applied in some real data. Finally, if time permits, some other related topics will be studied.

Appendix A

R Codes

The R codes used in the simulation studies are listed in this section. The Mixtures of Normal distribution (α, μ_1, σ_1, μ_2, σ_2) when $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 2$, $\sigma_2 = 2$ is chosen to study for this thesis.

1. The Medians of the Mixtures of Normal Distribution

```
###Mixture of normal distribution (alpha,mu1,sigma1,mu2,sigma2)


mixture<-function(x,x1,x2,x3,x4,x5)
      {x1*pnorm((x-x2)/x3,0,1)+(1-x1)*pnorm((x-x4)/x5,0,1)}


###Bisection method for finding the median when alpha=0


a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0,0,1,2,2)<0.5) a<-c
```

```
        else b<-c
      }
}
      print (c)
###########
###Bisection method for finding the median when alpha=0.101

a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0.101,0,1,2,2)<0.5) a<-c
        else b<-c
      }
}
      print (c)
###########
###Bisection method for finding the median when alpha=0.175

a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0.175,0,1,2,2)<0.5) a<-c
```

```
        else b<-c
      }
}
      print (c)
##############
###Bisection method for finding the median when alpha=0.256

a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0.256,0,1,2,2)<0.5) a<-c
        else b<-c
      }
}
      print (c)
##############
###Bisection method for finding the median when alpha=0.382

a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0.382,0,1,2,2)<0.5) a<-c
```

```
        else b<-c
      }
}
      print (c)
#############
###Bisection method for finding the median when alpha=0.491


a<-0
b<-8
delta<-0.01
while (abs(a-b)>delta){
      c<-(a+b)/2
          {if (mixture(c,0.491,0,1,2,2)<0.5) a<-c
        else b<-c
      }
}
      print (c)
#############
```

2. The Empirical Size of Wilcoxon Signed-Rank Test when n = 10 and α = 0

```
## iteration = 10,000 times
## Mixture normal dist.,alpha=0,1-alpha=1,mu1=0,sigma1=1,mu2=2,sigma2=2
## true median(M)= 2,
## true median + delta=2+0.00=2.00
## Ho:Median=M vs. Ha:Median=M+delta
## n=10
```

```
pval=rep(0,10000)
for (i in 1:10000)
{
   y1<-rnorm(10,0.00,1)
   y2<-rnorm(10,2.00,2)
   x1<-rbinom(10,1,0)
   x<- x1*y1+(1-x1)*y2
   pval[i]=wilcox.test(x,alternative="greater",mu=2,exact=TRUE)$p.value
}
sum(pval<0.05)
```

Remark: For $\alpha = 0.101, 0.175, 0.256, 0.382, 0.491$, we change the value of x1 according to α and n = 20, 30, 40, 50, 60, 100, 200, 500, we change the first value of y1, y2 and x1 according to n.

3. The Empirical Power of Wilcoxon Signed-Rank Test when n = 10, $\delta = 0.01$ and $\alpha = 0$

```
## iteration = 10,000 times
## Mixture normal dist.,alpha=0,1-alpha=1,mu1=0,sigma1=1,mu2=2,sigma2=2
## true median(M)= 2,
## true median + delta=2+0.01=2.01
## Ho:Median=M vs. Ha:Median=M+delta
## n=10
pval=rep(0,10000)
for (i in 1:10000)
{
   y1<-rnorm(10,0.01,1)
   y2<-rnorm(10,2.01,2)
```

```
   x1<-rbinom(10,1,0)
   x<- x1*y1+(1-x1)*y2
   pval[i]=wilcox.test(x,alternative="greater",mu=2,exact=TRUE)$p.value
}
sum(pval<0.05)
```

Remark: For $\alpha = 0.101, 0.175, 0.256, 0.382, 0.491$, we change the value of x1 according to α, for δ is a small constant that shifts the hypothesized median between 0.01 to 0.25 and n = 20, 30, 40, 50, 60, 100, 200, 500, we change the first value of y1, y2 and x1 according to n.

4. The Empirical Size of Wilcoxon Signed-Rank Test When applying the Inverse Transformation Method when n = 10 and $\alpha = 0$

```
## iteration = 10,000 times
## Mixture normal dist.,alpha=0,1-alpha=1,mu1=0,sigma1=1,mu2=2,sigma2=2
## n=10
mixture<-function(x,x1,x2,x3,x4,x5)
      {x1*pnorm((x-x2)/x3,0,1)+(1-x1)*pnorm((x-x4)/x5,0,1)}
pval=rep(0,10000)
for (i in 1:10000)
{
   y1<-rnorm(10,0.00,1)
   y2<-rnorm(10,2.00,2)
   x1<-rbinom(10,1,0)
   x<- x1*y1+(1-x1)*y2
   u<- mixture(x,0,0,1,2,2)
   pval[i]=wilcox.test(u,alternative="greater",mu=0.5,exact=TRUE)$p.value
}
```

```
sum(pval<0.05)
```

Remark: For $\alpha = 0.101, 0.175, 0.256, 0.382, 0.491$, we change the value of x1 according to α and n = 20, 30, 40, 50, 60, 100, 200, 500, we change the first value of y1, y2 and x1 according to n.

5. The Empirical Power of Wilcoxon Signed-Rank Test When Applying the Inverse Transformation Method when n = 10, $\delta = 0.01$ and $\alpha = 0$

```
## iteration = 10,000 times
## Mixture normal dist.,alpha=0,1-alpha=1,mu=0,sigma1=1,mu2=2,sigma2=2,
## n=10
mixture<-function(x,x1,x2,x3,x4,x5)
      {x1*pnorm((x-x2)/x3,0,1)+(1-x1)*pnorm((x-x4)/x5,0,1)}
pval=rep(0,10000)
for (i in 1:10000)
{
   y1<-rnorm(10,0.01,1)
   y2<-rnorm(10,2.01,2)
   x1<-rbinom(10,1,0)
   x<- x1*y1+(1-x1)*y2
   u<- mixture(x,0,0,1,2,2)
   pval[i]=wilcox.test(u,alternative="greater",mu=0.5,exact=TRUE)$p.value
}
sum(pval<0.05)
```

Remark: For $\alpha = 0.101, 0.175, 0.256, 0.382, 0.491$, we change the value of x1 according to α, for δ is a small constant that shifts between 0.01 to 0.25, we change the value of u1 according to δ and n = 20, 30, 40, 50, 60, 100, 200, 500, we change the first value of y1, y2 and x1 according to n.

References

[1] Bickel, P.J. (1976) Another look at robustness: A review of reviews and some new developments. Scandinavian Journal of Statistics 3, 145-168.

[2] Blair, R.C. and Higgins, J.J. (1985) Comparison of the power of the paired samples t test of that of Wilcoxon's signed-ranks test under various population shapes. Psychological Bulletin 97, 119-128.

[3] Box, G.E.P. and Andersen, S.L. (1955) Permutation theory in the derivation of robust criteria and the study of departures from assumption. Journal of the Royal Statistical Society. Series B (Methodological) 17, 1-34.

[4] Devroye, L. (1986) Non-uniform random variate generation. Springer-Verlag New York Inc., New York.

[5] Erceg-Hurn, D.M. and Mirosevich, V.M. (2008) Modern robust statistical methods. The American Psychologist 63, 591-601.

[6] Gibbons, J.D. (1971) Nonparametric statistical inference. McGraw-Hill, New York.

[7] Good, P. (1994) Permutation tests: a practical guide to resampling methods for testing hypotheses. Springer-Verlag New York, Inc., New York.

[8] Hampel, R.F., Ronchetti, M.E., Rousseeuw J.P. and Stahel A.W. (1986) Robust statistics: the approach based on influence functions. John Wiley&Sons, Inc., New York.

[9] Herrendörfer, G. and Feige, K. (1984) A combinatorial method in robustness research and two applications, Robustness of statistical methods and nonparametric statistics (edited by Dieter Rasch and Moti Lal Tiku). D. Reidel Publishing Company, German.

[10] Hettmansperger, T.P., McKean, J.W. and Sheather S.J. (2000) Robust nonparametric methods. Journal of the American Statistical Association 95, 1308-1312.

[11] Hollander, M. and Wolfe, D.A. (1973) Nonparametric statistical methods. John Wiley&Sons, Inc., New York.

[12] Huber, P.J. (1972) The 1972 Wald lecture robust statistics: a review. The Annals of Mathematical Statistics 43, 1041-1067.

[13] Huber, P.J. (1981) Robust statistics. John Wiley&Sons, Inc., New York.

[14] Jurečková, J. and Sen, P.K. (1996) Robust statistical procedures. John Wiley&Sons, Inc., New York.

[15] Kotz, S., Johnson, L.N. and Read, B.C. (1988) Encyclopedia of statistical sciences volume 8. John Wiley&Sons, Inc., New York.

[16] Levine, D.W. and Dunlap, W.P. (1982) Power of the F test with skewed data: should one transform or not? Psychological Bulletin 92, 272-280.

[17] Muller, H-G. and Zhou, H. (1991) Transformations in density estimation: Comment. Journal of the American Statistical Association 86, 356-358.

[18] Patil, P.N., Patil, P.P. and Bagkavos, D. (2012) A measure of asymmetry. Statistical Papers 53, 971-985.

[19] Portnoy, S. and He, H. (2000) A robust journey in the New Millennium. Journal of the American Statistical Association 95, 1331-1335.

[20] Posten, H.O. (1984) Robustness of the two-sample t-test, Robustness of statistical methods and nonparametric statistics (edited by Dieter Rasch and Moti Lal Tiku). D. Reidel Publishing Company, German.

[21] Pratt, J.W. and Gibbons, J.D. (1981) Concepts of nonparametric theory. Springer-Verlag, New York.

[22] Randles, R.H. and Wolfe, D.A. (1979) Introduction to the theory of nonparametric statistics. New York Chichester (etc.): Wiley, New York.

[23] Randles, R.H., Fligner, M.A., Policello, G.E. and Wolfe, D.A. (1980) An asymptotically distribution-free test for symmetry versus asymmetry. Journal of the American Statistical Association 75, 168-172.

[24] Rudemo, M. (1991) Transformations in density estimation: Comment. Journal of the American Statistical Association 86, 353-354.

[25] Ruppert, D. and Cline, D.B.H. (1994) Bias reduction in Kernel density estimation by smoothed empirical transformations. The Annals of Statistic 22, 185-210.

[26] Sen, P.K. (1968) On a further robustness property of the test and estimator based on Wilcoxon's signed rank statistic. The Annals of Mathematical Statistics 39, 282-285.

[27] Silverman, B.W. (1986) Density estimation for statistics and data analysis (Monographs on statistics and applied probability). Chapman and Hall, London.

[28] Simonoff, J.S. (1996) Smoothing methods in statistics. Springer-Verlag New York, Inc., New York.

[29] Sprent, P. (1989) Applied nonparametric statistical methods. Chapman and Hall, London.

[30] Staudte, R.G. and Sheather, S.J. (1990) Robust estimation and testing. John Wiley&Sons, Inc., New York.

[31] Tiku, M.L., Tan, W.Y. and Balakrishmn, N. (1986) Robust inference. Marcel Dekker, Inc., New York.

[32] Wand, M.P., Marron, J.S. and Ruppert, D. (1991) Transformations in density estimation. Journal of the American Statistical Association 86, 343-353.

[33] Wand, M.P. and Jones, M.C. (1995) Kernel smoothing. Chapman and Hall, London.

[34] Wayne, W.D. (1978) Applied nonparametric statistics. Houghton Miffin Company, Boston.

[35] Wayne, W.D. (1990) Applied nonparametric statistics (2nd ed.). PWS-KENT Publishing Company, Boston.