1-1-2014

Rank-based Two Sample Tests Under a General Alternative

Jamye Curry
University of Mississippi

Rank-based Two Sample Tests Under A General Alternative

Jamye Nichelle Curry

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Doctor of Philosophy
major in Mathematics
with concentration in Statistics

The University of Mississippi

2014

ABSTRACT

The problem of testing whether two samples come from the same or different population is a classical one in statistics. In this dissertation, I first study rank based formulation of univariate two-sample distribution-free tests. One form of the test statistic is the average of between-group distances of ranks. The other form of the test statistic is the difference between the average of between-group distances of ranks and the average of within-group distances of ranks. Although they are different in formulation, they are closely related to the two-sample Cramér-von Mises criterion. The first one is a linear transformation of Cramér-von Mises criterion in the case the two samples are of the same size. The second one is a different form of the Cramér-von Mises criterion. The properties of the two-sample test statistic based on the new formulation are studied. In particular, the Hájek projection and orthogonal decomposition technique in deriving the asymptotics of the test statistic is applied. For the first statistic under the balanced case, its limiting distribution is not normal since the projection on one variable is insufficient to represent the variation of the test statistic. By taking the projection on two variables, it is proved to be a weighted mixture of independent chi-square distributions. An operator in the functional space is defined and its eigenfunctions and eigenvalues are applied to derive the limiting distribution.

Rank-based formulations allow generalizations of the two-sample Cramér-von Mises test to the multivariate case by using different notions of multivariate rank functions. In the multivariate case, the rank tests may lose the distribution-free property under a general alternative. They are, however, usually more robust than the parametric tests. I propose two corresponding new tests based on multivariate spatial ranks. The spatial rank function yields a relative center-outward ranking of a data set. It preserves not only ordering on the

magnitude of vectors but also directional information. It characterizes the distribution. One test statistic is the difference between the average of intra-sample rank distances and the average of inter-sample rank distances. The other one is simply the average of intra-sample rank distances for the balanced samples. Unlike the univariate case, those two statistics are no longer equivalent. Comparing with other tests, the proposed ones can be established by the following desirable properties. (1) They are nonparametric with fewer assumptions, although they are not completely distribution-free. (2) They are invariant with respect to orthogonal linear transformations, which doesn't hold for tests based on the component-wise ranks. (3) They are consistent against all alternatives. The simulation results have illustrated the proposed tests to be promising. The bootstrap and permutation procedures are used for yielding a consistent approximation to the null distribution of the test statistics.

DEDICATION

This dissertation is dedicated to my parents, Larry and Barbara Curry,
and my brother, Bo Curry, who have fully encouraged and supported me
throughout my graduate career.
Without their continuous love, endless support and prayers
it would not have been possible.
This work is also dedicated to Kendrick Savage
who always gives me infinite inspiration.

ACKNOWLEDGMENTS

Special thanks to my advisor, Dr. Xin Dang, who graciously accepted me as her student. She provided unlimited support and detailed guidance throughout the course of conducting my research and the writing of my dissertation. I express my sincere gratitude for her immense knowledge, enthusiasm and patience. Her dedication and mentorship made my reseach study worthwhile. I could not have imagined having a better advisor to conduct research with. It is because of her that I am able to complete my dissertation.

My thanks go to Dr. Hailin Sang, Assistant Professor of Mathematics, Dr. Martial Longla, Assistant Professor of Mathematics, Dr. Gerard Buskes, Professor of Mathematics, and Dr. Walt Mayer, Associate Professor of Economics, for serving as members of the examing committee. Especially thanks to Dr. Sang for his collaborative effort on my research and research paper, to Dr. Longla for his advice and proofreading of my dissertation, to Dr. Buskes for his insightful suggestions, and to Dr. Mayer for providing simulation ideas.

Thanks to Dr. Donald Cole and Dr. Gerard Buskes for providing me with financial support through the Graduate Assistance in Areas of National Need (GAANN) Fellowship. The fellowship is greatly appreciated as it took away the stress of tuition expense and the expense of attending research conferences. The funding has opened many doors for me which, with my own means, I could have never walk through.

Finally, I am grateful to Ms. Jacquline Vinson and Ms. Stephanie Brown, the coordinators of the Increasing Minority Access to Graduate Education (IMAGE) Program, for their support and warm encouragement during my graduate career.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Two-sample Problem

The problem of testing whether two samples come from the same or different population is a classical one in statistics. Two-sample tests have been applied to many fields. For example, in medical science, the investigation of a two sample problem can be used to identify cancer genes by analyzing microarray data. Microarray technology allows researchers to examine samples of gene expression levels under diverse circumstances. A common objective in analyzing data from microarray experiments is to identify which genes are differentially expressed, where the samples are obtained under various conditions. Researchers test for differentially expressed genes when searching for disease-related genes. One can compare data of gene expression levels between cancer tissue samples and normal tissue samples and then select genes related to the cancer under investigation. The cancer genes will be detected if their expression levels between the two samples are significantly different.

In business, one example to use the two-sample test is for benchmarking comparison. Benchmarking is a process done when one desires to compare the practice or performance of one organization or company to another. Facilitators utilize this process to identify how well the company is progressing in reference to measures such as quality, cycle time and cost. Data may be collected via surveys, interviews, publications, etc. Once data is obtained from each organization or company, comparisons are then made to identify which company

performed better regarding the performance under study. This process is designed to assist in the improvement of the company.

The set up of a two sample problem is as follows. Suppose $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ are two independent random samples that are drawn from two populations with continuous distribution functions F and G, respectively. The goal is to test whether the two samples are drawn from identical populations, i.e.,

$$H_0 : F(\boldsymbol{x}) = G(\boldsymbol{x}) \text{ for all } \boldsymbol{x}.$$

1.2 Parametric vs Nonparametric Approach

Let us start with this task in the univariate case. Generally speaking, there are two approaches. One is the parametric approach that makes model assumptions on the forms of the underlying distributions, and assumes that the differences between the two populations lie only with respect to some parameters. In such a way, in Neyman-Pearson framework it becomes possible to derive the best test. For example, if we assume that the populations are normally distributed, two-sample student's t test and F test are the best tests for equality of means and for equality of variance, respectively. However, those and other parametric tests may be sensitive to violations of the underlying assumptions inherent in the derivation and construction of these tests. They are only valid when those assumptions are reasonable to make. If there is a suspicion of a violation of those assumptions, or if there is no sufficient information to justify those assumptions, one prefers the nonparametric approach.

Nonparametric tests assume less. The null hypothesis is formulated as identical populations drawn from a common distribution completely unspecified except that it is continuous. Thus, under H_0, the two random samples can be considered as a single random sample of size $N = m + n$. Then the combined configuration of the N random variables in the sample is one of the $\binom{N}{m, n} = \frac{N!}{m!n!}$ possible equally likely arrangements. The possible arrangement of X's and Y's provides information about the type of difference which

may exist in the populations. Such possible arrangement is the natural ranks of X and Y. Under H_0, the test based on ranks has a distribution-free property, which means that the distribution of rank-based test statistics is independent to F. Depending on the alternative hypothesis, various distribution-free tests have been proposed.

A particular alternative is the difference in location. That is,

$$H_L : F(x) = G(x - \theta) \text{ for all } x \text{ and some } \theta \neq 0.$$

Under this alternative, one sample is stochastically larger than the other, and hence the ranks of one sample tends to be large. The Wilcoxon rank-sum test, also know as the Mann-Whitney U test (Wilcoxon (1945), Mann & Whitney (1947)), is a common practice for this problem. The test statistic is based on the sum of ranks of one sample with respect to the combined sample. A small or large test statistic is an evidence to reject H_0. Other tests include Terry-Hoeffding (Terry (1952), Hoeffding (1951)) and van der Waerden (1952). They are tests based on the sum of some monotonic function of the ranks of one sample. The choice of the monotonic function determines the properties of the test.

Similarly, if the difference in scale is of interest, then the scale alternative is used. That is,

$$H_S : F(x) = G(\theta x) \text{ for all } x \text{ and some } \theta \neq 1.$$

Under this alternative, one sample has a large dispersion, and hence more values of one sample should be larger or smaller than the values of the other sample. The Mood test (Mood (1954)) or the Freund-Ansari-Bradley test (Freund & Ansari (1957), Bradley (1968)) are based on the sum of squared or absolute deviations of one sample ranks from the average combined rank. Some other tests like Siegel-Tukey (Siegel & Tukey (1960)) use a special way to transform the ranks so that the scale problem changes to the location problem and then the Wilcoxon rank-sum test can be applied.

The completely general two-sided alternative is

$$H_A : F(x) \neq G(x) \text{ for some } x.$$

This alternative simply states that there is a difference between the two populations, but does not specify where the difference is and how they are different. This is the most general and least restrictive case. Hence, the test designed for this type of alternative has a wider application than others. The Kolmogorov-Smirnov test (Smirnov (1939)) and the Cramér-von Mises test are commonly used for the general two-sample problem.

In this dissertation, my first goal is to study two rank-based formulations for the two-sample problem under a general alternative. The second goal is to extend the rank tests to the multivariate nonparametric two-sample problem.

1.3 Component-wise vs Multivariate Approach

A generalization from the univariate case to the multivariate case can take a component-wise approach. This approach deals with each variate separately using conventional univariate methods. Such an approach is intuitive and simple. However, it completely ignores the correlation between variables. Component-wise approach doesn't take full information of data and hence it usually has drawbacks of low efficiency. Let us look at the following example.

1.3.1 A Motivating Example

Two samples of size 100 from multivariate t-distribution with 3 degrees of freedom in $\mathbb{R}^2$ are generated. That is, $\boldsymbol{x}_1, ..., \boldsymbol{x}_{100} \sim t_3(\mathbf{0}, \Sigma_1)$ and $\boldsymbol{y}_1, ..., \boldsymbol{y}_{100} \sim t_3(\mathbf{0}, \Sigma_2)$, where

$$\Sigma_1 = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$$

and

$$\Sigma_2 = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}.$$

See the scatter plot of the generated data in Figure 1.1.

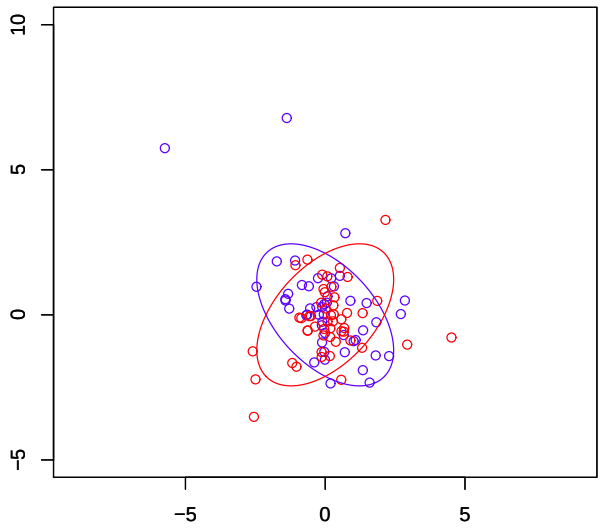


Figure 1.1. Ellipses based on mean 0 and variances Σ_1 and Σ_2 for random samples generated from t-distribution: X-sample in $blue$ and Y-sample in red.

Each dimension of two samples has the same distribution. Thus, any good univariate test should not reject H_0. If we take a component-wise approach, it will fail to reject H_0. However, the two samples are from two distinct distributions. The reason the component-wise approach fails in this case is that it ignores the correlation information between variables. As in this example, the difference in the two distributions only lie in correlations.

Using the component-wise tests Wilcoxon, Cramér-von Mises, Kolmogorov-Smirnov and Mood's test, we can see that the component-wise approach will fail. Each $\boldsymbol{X}$-sample point $\boldsymbol{x}_i$ has two components $(x_{i1}, x_{i2})^T$. The components x_{i1} and x_{i2} are jointly from distribution F, while x_{i1} come from the marginal distribution F_1 and x_{i2} come from the

marginal distribution F_2. Similarly, each $\boldsymbol{Y}$-sample point $\boldsymbol{y}_j$ has two components $(y_{j1}, y_{j2})^T$, where $(y_{j1}, y_{j2})^T \sim G$, with $y_{j1} \sim G_1$ and $y_{j2} \sim G_2$. Thus, in this example, $F_1 = G_1$ and $F_2 = G_2$, but $F \neq G$. Computing the p-value of the random vectors for each test for this example gives the following: For the data $x_{11}, ..., x_{m1}, y_{11}, ..., y_{n1}$, the p-value corresponding to the Wilcoxon test statistic, Cramér-von Mises, Kolmogorov-Smirnov statistic and Mood's test statistic is 0.9862, 0.4998, 0.7166 and 0.0864, respectively. For the data $x_{12}, ..., x_{m2}, y_{12}, ..., y_{n2}$, the p-value corresponding to the Wilcoxon test statistic, Cramér-von Mises, Kolmogorov-Smirnov statistic and Mood's test statistic is 0.2931, 0.1892, 0.1124 and 0.2227, respectively. (See Table 1.1 below). The p-value is the probability of obtaining a test statistic at least as extreme as the one observed, assuming that the null hypothesis is true. Here, the p-value for eacth test is large. Hence, the decision here would be to fail to reject H_0, as expected. However, there is a difference in the two distributions ($F \neq G$). The two samples are drawn from the t-distribution with a difference in correlation. Hence, the component-wise approach ignores correlation and can only test in the direction of the x- and y-plane. Thus, a fully multivariate approach is necessary.

Table 1.1. Motivation example p-values

Test	**p-value:**(x_{i1}, y_{j1})	**p-value:**(x_{i2}, y_{j2})
Wilcoxon rank sum	0.9862	0.2931
Mood	0.0864	0.2227
Kolmogorov-Smirnov	0.7166	0.1124
Cramér-von Mises	0.4998	0.1892

Component-wise approach also leads to methods that are not affine equivariant and are not even invariant or equivariant under orthogonal transformations. Affine equivariance is where any linear translation of the sample observations are paralleled by a similar translation of the location estimator. We say a test statistic T is affine invariant if

$$T(\boldsymbol{Mx_1} + \boldsymbol{c}, ..., \boldsymbol{Mx_m} + \boldsymbol{c}, \boldsymbol{My_1} + \boldsymbol{c}, ..., \boldsymbol{My_n} + \boldsymbol{c}) = T(\boldsymbol{x_1}, ..., \boldsymbol{x_m}, \boldsymbol{y_1}, ..., \boldsymbol{y_n})$$

for every $\boldsymbol{x_1}, ..., \boldsymbol{x_m}, \boldsymbol{y_1}, ..., \boldsymbol{y_n}$, where $\boldsymbol{M}$ is a $d \times d$ nonsingular matrix and $\boldsymbol{c}$ is a $d \times 1$ vector. This property is useful as it assures that the test statistic remains unchanged given any rotations and reflections of the observations about any point. This also includes the changes in scale. However, the component-wise approach does not have this property. For example, the component-wise rank of a point in $\mathbb{R}^2$ might completely change if the data is rotated. Those drawbacks make the component-wise approach less interesting. Thus, a more appealing approach is that of the fully multivariate approach. Therefore, my second goal of the dissertation is to take the multivariate approach and extend the rank tests of the univariate case to the multivariate nonparametric two-sample problem.

In the univariate case, there is a linear ordering and hence the definition of rank is natural and unquestionable. However, in high dimensional space, such natural order no longer exists, which makes ranking conceptually difficult. Hence generalizations of the univariate rank methods to the multivariate case are not always feasible. This makes the multivariate rank based two-sample problem one of permanent interest. Although an ordering can always be defined, a specific ordering is needed. An ordering for multivariate data is defined later in Chapter 3.

1.4 Dissertation Overview

The contributions of the dissertation are given as follows.

- *A new perspective of the Cramér-von Mises test.* The classical Cramér-von Mises test is revisited in the univariate case with a totally different rank-based approach. In addition, a rank formulation that is equivalent to the Cramér-von Mises test is provided. The critical values are provided for small sizes. Hájek projection and orthogonal decomposition techniques are applied in deriving the asymptotics of the test statistics.

- *A new generalization to the multivariate case*. The generalization is based on spatial ranks. Bootstrap and permutation methods are used to determine critical values. Extensive simulation study has been conducted for comparing the proposed test with other existing tests.

- *Broad adaptability*. The proposed test applies to a general two-sample problem, not just for location or scale difference problems. It is nonparametric, requires few assumptions and hence it has a broad adaptability.

This dissertation is organized as follows. Chapter 2 deals with the univariate case. I first review several distribution-free two sample tests for a general alternative with emphasis on the Cramér-von Mises test. Two rank formulations are proposed and their relationship with Cramér-von Mises criterion are discussed. The properties of the test statistic based on the new formulation are studied and the results are consistent with the ones using the classical method. In particular, the Hájek projection and orthogonal decomposition technique in deriving the asymptotics of the test statistic are applied. Chapter 2 ends with simulation procedures and results used to investigate the power performance of the proposed test against other exisiting tests. Chapter 3 extends the rank-based tests to the multivariate case. Popular multivariate rank notions are first introduced and then followed by discussion of the properties of each rank function. Spatial rank is used for the generalization and the properties of the rank test are explored. Unlike in the univariate case, the proposed rank test is no longer distribution-free, although it is nonparametric. Bootstrap and permutation techniques are used for determining critical values. The connection with other tests is discussed. Power performance comparison with those tests is conducted. The final chapter presents the summary, conclusions and future work.

CHAPTER 2

UNIVARIATE RANK FORMULATION TESTS

2.1 Introduction

Let $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$, $m, n \in \mathbb{N}$, be independent random samples from univariate distributions $F(x)$ and $G(x)$, respectively. The empirical distributions of the X and Y samples are defined as

$$F_m(x) = \frac{1}{m} \sum_{i=1}^{m} I(X_i \leq x)$$

and

$$G_n(x) = \frac{1}{n} \sum_{j=1}^{n} I(Y_j \leq x),$$

respectively. Then the empirical distribution of the combined sample $X_1, ...X_m, Y_1, ..., Y_n$ with $N = m + n$ is defined as

$$H_N(x) = \frac{1}{N} \{ \sum_{i=1}^{m} I(X_i \leq x) + \sum_{j=1}^{n} I(Y_j \leq x) \}.$$

For testing the hypothesis

$$H_0 : F = G \quad vs \quad H_a : F \neq G, \tag{2.1}$$

two nonparametric approaches are used. One approach is based on the ranks of observations of the two samples. Under H_0, the two random samples can be considered as a single

random sample of size N drawn from the common continuous, but unspecified distribution F. Then the combined configuration of the m X's and n Y's random variables in the sample is one of the $\binom{N}{m,n} = \frac{N!}{m!n!}$ possible equally likely arrangements. The sample pattern of arrangement (ranks) of X's and Y's provides information about the type of difference which may exist in the populations. This approach is very useful to design efficient rank tests for a particular alternative such as a location difference or scale difference alternative.

The other nonparametric approach is based on some measure of difference between the two empirical functions $F_m(x)$ and $G_n(x)$. A large difference is an evidence to reject the null hypothesis. This approach is used for the Kolmogorov-Smirnov (KS) test, Cramér-von Mises (CM) test and the derivations of the CM test. This approach is especially useful for a general alternative.

This chapter shows the connection of these two approaches. Two new test statistics are proposed and their connection to the KS and CM test are presented. The review of these tests for a general alternative is provided first.

2.1.1 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test is used to measure the difference of two empirical functions $F_m(x)$ and $G_n(y)$ by the sup norm. In other words, the statistic is defined as the maximum distance (difference) between the two empirical distribution functions. That is,

$$K_{m,n} = \sup_x |F_m(x) - G_n(x)|.$$

H_0 is rejected if $K_{m,n}$ is sufficiently large. The rejection region is in the upper tail defined by $K_{m,n} \geq c_\alpha$ where c_α is determined by $P(K_{m,n} \geq c_\alpha | H_0) \leq \alpha$. Under H_0, the distribution of $K_{m,n}$ is independent of F (i.e, the KS test is distribution-free). Hence the exact null distribution can be derived and the table of c_α is available for small m and n. For the asymptotic null distribution when $m, n \to \infty$ and $m/n \to \tau$, where τ is a constant in

$(0, 1)$, Smirnov proved that

$$\lim_{m,n\to\infty} P\left(\sqrt{\frac{mn}{N}}K_{m,n} \leq d\right) = L(d),$$

where $L(d) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2d^2}$. Note that the KS test can be used for a one-sample problem. This is applied to test for the goodness-of-fit, i.e., want to test if the sample comes from a specific distribution. The test compares the empirical distribution function of a random sample with a given hypothesized distribution F_0, which is to test whether the random sample is drawn from the hypothesized distribution.

2.1.2 Cramér-von Mises Test

In 1928, Cramér suggested a goodness-of-fit test using the squared L_2 norm of function difference on R^1. That is,

$$\int_{-\infty}^{\infty}[F_m(x) - F_0(x)]^2\,dx. \tag{2.2}$$

von-Mises independently made an equivalent suggestion in 1933 and developed a distribution-free test by modifying it as

$$\int_{-\infty}^{\infty}[F_m(x) - F_0(x)]^2\,dF_0(x). \tag{2.3}$$

The natural analogue to (2.3) for the two sample problem is

$$\frac{mn}{N}\int_{-\infty}^{\infty}[F_m(x) - G_n(x)]^2 dH_N(x). \tag{2.4}$$

This test statistic and its asymptotics have been studied by Lehmann (1951), Rosenblatt (1952), Fisz (1960), Darling (1957) and Anderson (1962).

Some related works include Pettitt (1976) and Baumgartner *et al.* (1998). They consider Anderson-Darling type of statistics that can be viewed as standardized versions of Cramér-

von Mises statistic, defined as

$$\frac{mn}{N}\int_{-\infty}^{\infty}\frac{[F_m(x)-G_n(x)]^2}{H_N(x)(1-H_N(x))}dH_N(x).$$

Schmid & Trede (1995) utilize L_1 Cramér-von Mises statistic. That is,

$$\frac{mn}{N}\int_{-\infty}^{\infty}|F_m(x)-G_n(x)|dH_N(x). \tag{2.5}$$

A rank-based representation of L_1 Cramér-von Mises statistic (2.5) under a balanced size and its generalizations are studied by Borroni (2001). Next, a rank-based formulation of (2.4) is studied.

2.2 Two Rank Formulations

In this section, the rank formulation of the Cramér-von Mises (CM) criterion is proposed. In addition, a new rank formulation that is a linear transformation of CM is proposed. The formulation of the test statistics are based on ranks in the mixture distribution $H=\tau F+(1-\tau)G$ with $0\leq\tau\leq 1$. The standardized rank of X with respect to a distribution F is defined as $R(X,F)=F(X)$. Hence, the sample version of the standardized rank of $X_i's$ with respect to its empirical distribution F_N is defined as

$$R(X_i,F_N)=F_N(X_i)=\frac{1}{N}\sum_{j=1}^{N}I(X_j\leq X_i)$$

for samples $X_1,X_2,...,X_N$, where the ranks of $X_1,X_2,...,X_N$ are uniform on $\{1/N,2/N,...,N-1/N,1\}$. Let $R(y,H)$ denote the standardized rank of the quantity y with respect to the cumulative distribution function $H(x)$, i.e., $R(y,H)=H(y)$. For testing the hypothesis

(2.1), we propose two forms of test statistics. The first proposed statistic is defined as

$$\begin{aligned} T_1 =& \frac{mn}{N}\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|R(X_i,H_N)-R(Y_j,H_N)| \\ &- \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}|R(X_i,H_N)-R(X_j,H_N)| \\ &- \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|R(Y_i,H_N)-R(Y_j,H_N)|\}. \end{aligned} \tag{2.6}$$

T_1 behaves as the difference of the average of between-group rank differences and the average of within-group rank differences. A large value of T_1 indicates the deviation of two groups. Baringhaus & Franz (2004) studied a version based on the original data. That is,

$$\frac{mn}{N}\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|X_i-Y_j| - \frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}|X_i-X_j| - \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|Y_i-Y_j|\}. \tag{2.7}$$

Although it has a direct generalization to the multivariate case, it is not distribution-free and it requires an assumption on the first moment. It is equivalent to Cramér test (2.2) for a two-sample problem. It is worthwhile to note that the above test (2.7) falls in the unified framework on energy statistics studied by Székely & Rizzo (2013).

The second proposed statistic is defined as

$$T = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|R(X_i,H_N)-R(Y_j,H_N)|. \tag{2.8}$$

T is defined as the average of the absolute difference between the ranks two groups. It has a simpler form than T_1, but as shown later, it is only suitable for balanced samples.

The following lemma gives the connection of the distributions of the standardized ranks with the original distributions and the mixture distribution H.

Lemma 1 *Let X and Y be independent random variables with the distribution functions F*

and G, respectively. Let $H = \tau F + (1-\tau)G$ with $0 \leq \tau \leq 1$. Then $R(X,H)$ and $R(Y,H)$ are independent. Moreover, if J is the distribution of $R(X,H)$ and K is the distribution function of $R(Y,H)$, then $J(x) = F \circ H^{-1}(x)$ and $K(x) = G \circ H^{-1}(x)$ for any $x \in [0,1]$, where $H^{-1}(x) = \inf\{u : H(u) \geq x\}$.

Proof. Since X and Y are independent, then any continuous function of X and any continuous function of Y are also independent. Therefore, H(X) and H(Y) are independent. Since $R(X,H) = H(X)$ and $R(Y,H) = H(Y)$, then $R(X,H)$ and $R(Y,H)$ are independent. For any $x \in [0,1]$,

$$\begin{aligned} J(x) &= P(R(X,H) \leq x) = P(H(X) \leq x) \\ &= P(X \leq H^{-1}(x)) \\ &= F \circ H^{-1}(x). \end{aligned} \tag{2.9}$$

Thus, (2.9) holds, since $P(H^{-1}(x) < X < \sup\{u : H(u) \leq x\}) = 0$ for the mixture distribution H. Similarly, $K(x) = G \circ H^{-1}(x)$ for any $x \in [0,1]$. ■

Remark 2 *The conclusion on $J(x)$ and $K(x)$ in Lemma 1 holds if H is any continuous distribution function. If H is not continuous, then a sufficient condition to ensure (2.9) still holds is that H is to be the mixture of F and G.*

The next lemma gives the expected value of the absolute difference between the standardized ranks of X and Y. The second proposed statistic T is given by the empirical version of (2.10) in the following lemma.

Lemma 3 *Let X and Y be independent random variables from F and G, respectively. Let $H = \tau F + (1-\tau)G$ with $0 \leq \tau \leq 1$, J be the distribution of $R(X,H)$ and K be the distribution function of $R(Y,H)$. Then*

$$\mathbb{E}|R(X,H) - R(Y,H)| = \int_0^1 J(t)(1-K(t))\,dt + \int_0^1 K(t)(1-J(t))\,dt. \tag{2.10}$$

In particular, $\mathbb{E}|R(X,H) - R(Y,H)| = 1/3$ if $F = G$.

Proof. Let $I(A)$ denote the indicator function of A, that is,

$$I(A) = \begin{cases} 1 & , A \text{ is true} \\ 0 & , \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} & |R(X,H) - R(Y,H)| \\ =& I\left[R(Y,H) > R(X,H)\right] \cdot (R(Y,H) - R(X,H)) \\ & + I\left[R(X,H) > R(Y,H)\right] \cdot (R(X,H) - R(Y,H)) \\ =& \int_0^1 \left[I(R(X,H) \le t < R(Y,H))\right] dt + \int_0^1 \left[I(R(Y,H) \le t < R(X,H))\right] dt. \end{aligned} \tag{2.11}$$

Hence, using Lemma 1 and (2.11),

$$\begin{aligned} & \mathbb{E}|R(X,H) - R(Y,H)| \\ =& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1 I(X < H^{-1}(t) < Y)\, dt\, dF(x)\, dG(x) \\ & + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^1 I(Y < H^{-1}(t) < X)\, dt\, dF(x)\, dG(x) \\ =& \int_0^1 F\big(H^{-1}(t)\big)\big(1 - G(H^{-1}(t))\big)\, dt + \int_0^1 G\big(H^{-1}(t)\big)\big(1 - F(H^{-1}(t))\big)\, dt \\ =& \int_0^1 J(t)(1 - K(t))\, dt + \int_0^1 K(t)(1 - J(t))\, dt. \end{aligned}$$

Under the null hypothesis, $F = G$ implies $H = F = G$ and $J(t) = K(t) = tI(0 \le t \le 1)$.

Hence

$$\begin{aligned} \mathbb{E}|R(X,H) - R(Y,H)| =& \int_0^1 t(1-t)\, dt + \int_0^1 t(1-t)\, dt \\ =& 2\int_0^1 t(1-t)\, dt \\ =& 1/3. \end{aligned}$$

■

Theorem 4 *In the case that $\tau = 1/2$, $\mathbb{E}|R(X,H) - R(Y,H)| = 1/3$ if and only if $F = G$, and $\mathbb{E}|R(X,H) - R(Y,H)| > 1/3$ if $F \neq G$.*

Proof. By Lemma 1, we have

$$\mathbb{E}|R(X,H) - R(Y,H)| = \int_0^1 J(t)(1-K(t))\,dt + \int_0^1 K(t)(1-J(t))\,dt \tag{2.12}$$

$$= \int_{-\infty}^{\infty} F(x)(1-G(x))dH(x) + \int_{-\infty}^{\infty} G(x)(1-F(x))dH(x) \tag{2.13}$$

$$= \int_0^1 F(1-G)dH + \int_0^1 G(1-F)dH \tag{2.14}$$

$$= \int_0^1 F(1-G)d\frac{F+G}{2} + \int_0^1 G(1-F)d\frac{F+G}{2} \tag{2.15}$$

$$= \int_0^1 [F - FG + G - FG]\, d\frac{F+G}{2}$$

$$= \int_0^1 [F + G - 2FG]\, d\frac{F+G}{2}$$

$$= \int_0^1 \left[(F+G) - \frac{(F+G)^2}{2} + \frac{(F-G)^2}{2}\right] d\frac{F+G}{2}$$

$$= 2\int_0^1 [(F+G)]\, d[F+G] - \int_0^1 (\frac{(F+G)^2}{2})d\frac{F+G}{2}$$

$$+ \int_0^1 \frac{(F-G)^2}{2} d\frac{F+G}{2}$$

$$= 1 - 2/3 + \frac{1}{2}\int_0^1 (F-G)^2 d\frac{F+G}{2}. \tag{2.16}$$

By a change in variables, (2.14) is obtained from (2.12), and (2.15) is obtained from (2.14) for $\tau = 1/2$. Thus, $\mathbb{E}|R(X,H) - R(Y,H)| = 1/3$ if and only if $F = G$, and $\mathbb{E}|R(X,H) - R(Y,H)| > 1/3$ if $F \neq G$. ■

Remark 5 *The expression (2.16) shows that the test statistic (2.8) is a linear transformation of the classical Cramér-von Mises test statistic in the case $m = n$.*

Corollary 6 *From Lemma 4.1 of Lehmann (1951), $\mathbb{E}|R(X,H) - R(Y,H)|$ may also be interpretated as the probability that a pair of X's lie on the same side of a pair of Y's. That is,*

$P(\max(X_1, X_2) < \min(Y_1, Y_2); \min(X_1, X_2) > \max(Y_1, Y_2))$
$= \frac{1}{3} + 2\int_{-\infty}^{\infty} (F(x) - G(x))^2 d\frac{F(x) + G(x)}{2}$.

Proof. Let $V = \max(X_1, X_2)$ and $W = \min(Y_1, Y_2)$.

Then $P(W > w) = P(\min(Y_1, Y_2) > w) = P(Y_1 > w, Y_2 > w) = [1 - G(w)]^2$ and $P(V < v) = P(\max(X_1, X_2) < v) = P(X_1 < v, X_2 < v) = F^2(v)$. Thus, $P(\max(X_1, X_2) < \min(Y_1, Y_2)) = P(V < W) = P(W > V) = \int_0^1 (1 - G)^2 dF^2$.

Similarly, $P(\max(Y_1, Y_2) < \min(X_1, X_2)) = \int_0^1 (1 - F)^2 dG^2$.

Let $P = P(\max(X_1, X_2) < \min(Y_1, Y_2); \max(Y_1, Y_2) < \min(X_1, X_2))$. Then

$$
\begin{aligned}
P &= \int_0^1 (1 - G)^2 dF^2 + \int_0^1 (1 - F)^2 dG^2 \\
&= \int_0^1 \left(1 - 2G + G^2\right) dF^2 + \int_0^1 \left(1 - 2F + F^2\right) dG^2 && (2.17) \\
&= 2 + \int_0^1 F^2 dG^2 + \int_0^1 G^2 dF^2 - 4\int_0^1 (FG) dF - 4\int_0^1 (FG) dG && (2.18) \\
&= 2 + \int_0^1 d\left(F^2 G^2\right) - 4\int_0^1 (FG) d\left(F + G\right) && (2.19) \\
&= 2 + \int_0^1 d\left(F^2 G^2\right) - \int_0^1 \left[(F^2 + 2FG + G^2) - (F^2 - 2FG + G^2)\right] d\left(F + G\right) \\
&= 2 + 1 - \int^1 \left[(F + G)^2 - (F - G)^2\right] d(F + G) \\
&= 3 - 2\int^1 \left[(F + G)^2 - (F - G)^2\right] d\frac{F + G}{2} \\
&= 3 - 2\int_0^1 (F + G)^2 d\frac{F + G}{2} + 2\int_0^1 (F - G)^2 d\frac{F + G}{2} \\
&= 3 - 2(\frac{4}{3}) + 2\int_0^1 (F - G)^2 d\frac{F + G}{2} \\
&= \frac{1}{3} + 2\int_0^1 (F - G)^2 d\frac{F + G}{2}.
\end{aligned}
$$

Note that the last two integrals in equality (2.18) are obtained from (2.17) based on the fact that $dF^2 = 2FdF$ and $dG^2 = 2GdG$. The first integral in equality (2.19) is obtained from (2.18) since $d(F^2G^2) = F^2 dG^2 + G^2 dF^2$. ■

Theorem 4 gives the population version of the test statistic T defined in (2.8), and explains the usefulness of T when the two samples have the same size. If the sample sizes are unbalanced, then T cannot be written as a linear combination of the Cramér-von Mises test. However, the following example demonstrates that the conclusion of Theorem 4 may not hold under different assumptions, such that if $\tau \neq 1/2$ or $m \neq n$, and therefore the test statistic (2.8) should not apply.

Example 7 *Suppose that Y has a standard exponential distribution. The cumulative distribution function is $G(x) = (1-e^{-x})I(0,\infty)$ and the density function is $g(x) = e^{-x}I(0,\infty)$. Let X be with distribution function $F(x) = G(x+c) = (1-e^{-(x+c)})I(-c,\infty)$ and density function $f(x) = g(x+c) = e^{-(x+c)}I(-c,\infty)$ for any $c \geq 0$. Suppose $H = \tau F + (1-\tau)G$ with $\tau \neq 1/2$. Then,*

$$
\begin{aligned}
&\mathbb{E}|R(X,H) - R(Y,H)| = \int_0^1 J(t)(1-K(t))dt + \int_0^1 K(t)(1-J(t))dt \\
=&\int_{-\infty}^{\infty} (F(x)+G(x)-2F(x)G(x))d(\tau F(x)+(1-\tau)G(x)) \\
=&\int_{-\infty}^{\infty} F(x)(d\tau F(x)+(1-\tau)G(x)) + \int_{-\infty}^{\infty} G(x)(d\tau F(x)+(1-\tau)G(x)) \\
&-\int_{-\infty}^{\infty} 2F(x)G(x)d(\tau F(x)+(1-\tau)G(x)) \\
=&\int_{-\infty}^{\infty} F(x)(d\tau F(x)) + \int_{-\infty}^{\infty} F(x)(d(1-\tau)G(x)) + \int_{-\infty}^{\infty} G(x)(d\tau F(x)) \\
&+\int_{-\infty}^{\infty} G(x)(d(1-\tau)G(x)) - \int_{-\infty}^{\infty} 2F(x)G(x)d(\tau F(x)+(1-\tau)G(x)) \qquad (2.20) \\
=&\tau\int_{-\infty}^{\infty} F(x)(dF(x)) + (1-\tau)\int_{-\infty}^{\infty} F(x)g(x)dx + \tau\int_{-\infty}^{\infty} G(x)f(x)dx \\
&+(1-\tau)\int_{-\infty}^{\infty} G(x)dG(x) - \int_{-\infty}^{\infty} 2F(x)G(x)d(\tau F(x)+(1-\tau)G(x)) \qquad (2.21) \\
=&\frac{\tau}{2} + (1-\tau)\int_{-\infty}^{\infty} F(x)g(x)dx + \tau\int_{-\infty}^{\infty} G(x)f(x)dx + (1-\tau)\cdot\frac{1}{2} \\
&-\int_{-\infty}^{\infty} 2F(x)G(x)d(\tau F(x)+(1-\tau)G(x)) \\
=&1/2 - 2\int_{-\infty}^{\infty} F(x)G(x)(\tau f(x)+(1-\tau)g(x))dx
\end{aligned}
$$

$$
\begin{aligned}
&+(1-\tau)\int_{-\infty}^{\infty}F(x)g(x)dx+\tau\int_{-\infty}^{\infty}G(x)f(x)dx\\
=&1/2-2\int_{0}^{\infty}(1-e^{-x})(1-e^{-(x+c)})(\tau e^{-(x+c)}+(1-\tau)e^{-x})dx\\
&+(1-\tau)\int_{0}^{\infty}(1-e^{-(x+c)})e^{-x}dx+\tau\int_{0}^{\infty}(1-e^{-x})e^{-(x+c)}dx\\
=&1/2-e^{-c}/6+\tau e^{-2c}/3-\tau e^{-c}/3.
\end{aligned}
$$

Note that $\int a(x)dF(x) = \int a(x)f(x)dx$. *Thus, (2.21) is obtained from (2.20) in the same manner. Now let* $\mathbb{E}|R(X,H)-R(Y,H)| = 1/3$. *This gives two solutions* $c=0$ *and* $c=\ln(2\tau)$. *When* $c=0$, *this corresponds to the null hypothesis* $F=G$. *But because* $\tau \neq 1/2$, $c=\ln(2\tau)\neq 0$ *corresponds to an alternative hypothesis* $F\neq G$. *This contradicts* $\mathbb{E}|R(X,H)-R(Y,H)|=1/3$ *if and only if* $F=G$.

The next theorem establishes the rank based formulation of test statistic T_1 for the Cramér-von Mises test statistic.

Theorem 8 *Let* X, X_1, X_2 *be random variables with distribution* F *and* Y, Y_1, Y_2 *be random variables with distribution* G, *where the* X*'s and* Y*'s are independent. Let* $H = \tau F+(1-\tau)G$ *with* $0\leq\tau\leq 1$ *be the mixture distribution. Then*

$$
\mathbb{E}|R(X,H)-R(Y,H)|-\frac{1}{2}\mathbb{E}|R(X_1,H)-R(X_2,H)|-\frac{1}{2}\mathbb{E}|R(Y_1,H)-R(Y_2,H)|\geq 0. \tag{2.22}
$$

Moreover, the equality holds if and only if $F=G$.

Proof. From Lemma 3 and Lemma 1, we have

$$
\begin{aligned}
&\mathbb{E}|R(X,H)-R(Y,H)|-\frac{1}{2}\mathbb{E}|R(X_1,H)-R(X_2,H)|-\frac{1}{2}\mathbb{E}|R(Y_1,H)-R(Y_2,H)|\\
&=\int_0^1 (J(t)-K(t))^2\,dt
\end{aligned} \tag{2.23}
$$

Hence the inequality (2.22) follows from (2.23). Since $J(t) = K(t)$ implies that $F(t) = G(t)$, then the equality in (2.22) holds. ■

Note that in (2.10) and in Theorem 8, H can be any continuous distribution function. Taking H as the mixture distribution is for the purpose of the rank based tests.

The result of Theorem 8 suggests two corresponding test statistics for testing the hypothesis (2.1). The first test statistic is denoted by T_1 as defined in (2.6). In fact, the test statistic T_1 is the sample plug-in version of the left side of Equation (2.23) multiplied by mn/N with $\tau = m/N$. H_0 is rejected if $T_1 > c_\alpha(m, n)$. The critical value $c_\alpha(m, n)$ is determined by the significance level α and the null distribution of T_1.

Based on (2.23) in the proof of Theorem 8, we may also write T_1 as

$$\frac{mn}{N} \int_0^1 (J_m(t) - K_n(t))^2 \, dt, \tag{2.24}$$

where J_m and K_n are the empirical distributions of standardized ranks of each X's and Y's in the combined sample. Hence the test statistic T_1 is the sample plug-in version of the right side of Equation (2.23) multiplied by mn/N. If a large value is observed, then the null hypothesis is rejected. Moreover, the two corresponding test statistics are two formulations of the Cramér-von Mises statistic since they are the empirical version of Equation (2.23). Since $J_m(x) = F_m \circ H_N^{-1}(x)$ and $K_n(x) = G_n \circ H_N^{-1}(x)$, T_1 also can be written as

$$\begin{aligned} T_1 &= \frac{mn}{N} \int_{-\infty}^{\infty} (F_m(x) - G_n(x))^2 \, dH_N(x) \\ &= \frac{mn}{N^2} \sum_{i=1}^{N} \left[\frac{M_i}{m} - \frac{i - M_i}{n} \right]^2 = \frac{1}{mn} \sum_{i=1}^{N} (M_i - \frac{m}{N} i)^2, \end{aligned} \tag{2.25}$$

where M_i is the number of X's less than or equal to the i^{th} smallest number in the combined sample. The representation of T_1 given in (2.25) facilitates an illuminating interpretation: T_1 is the average of squared distances of natural rank of X and its expected value under H_0 evaluated at each combined natural rank, i.e, the average of deviations to the mean. The

natural rank of each X is distributed uniformly on $\{1, ...N\}$.

Theorem 9 *Under the null hypothesis, T_1 and T are distribution free.*

Proof. Under H_0, $X_1, ..., X_m, Y_1, ..., Y_n$ constitute a random sample of size N from the distribution $F = G = H$. So, assignments of m numbers to $X_1, ..., X_m$ and n numbers to $Y_1, ..., Y_n$ from the set of integers $\{1, 2, ..., N\}$ are equally likely, i.e. has probability $\binom{N}{m, n}^{-1}$. Thus, the distribution is free of F. Using the fact that those number assignments have a one-to-one linear mapping to the standardized ranks, T_1 is distribution free. A similar argument holds for T. ■

The exact null distribution of T_1 can be found by enumerating all possible values of T_1 by considering the $N!/(m!n!)$ orderings of m X's and n Y's. Figures 2.1 and 2.2 provide the null distribution of T_1 for sample sizes $m = n = 7$ and $m = 7, n = 9$. Notice that the number of permutations increases rapidly as m or n increases. In the case of $m = n = 7$, the null distribution appears to be discrete, but in the case of $m = 7$ and $n = 9$, the distribution of T_1 looks continuous especially in $(0, 1)$. This is because a larger n yields much more permutations and those permutations yield values of T_1 to fill in gaps as shown in the Figure 2.1.

Reject H_0 if $T_{1m,n} > c_{m,n}$. The critical value $c_{m,n}$ is determined by the significance level α and the null distribution of $T_{1m,n}$. A large value of $T_{1m,n}$ indicates the deviation of the two groups (samples). Similarly, reject the null hypothesis if a large value of T is observed. Listed in Tables 2.1 and 2.2 are critical values for the statistic $T_{1m,n}$ for small m, n. The critical values of T can be obtained from linear transformation of the critical value of T_1 when $m = n$.

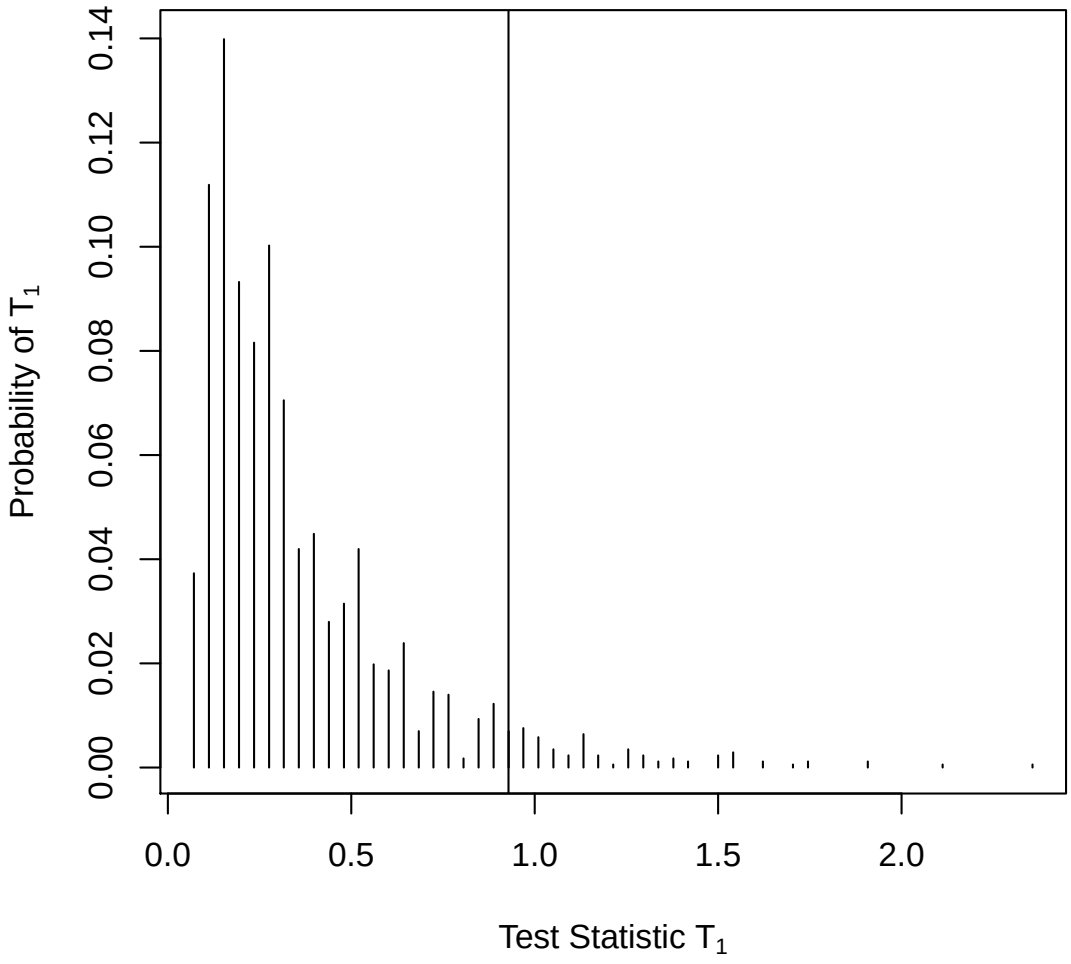


Figure 2.1. The exact null distribution of T_1 for equal sample size $m = 7$, $n = 7$. The vertical line indicates the 5% critical value.

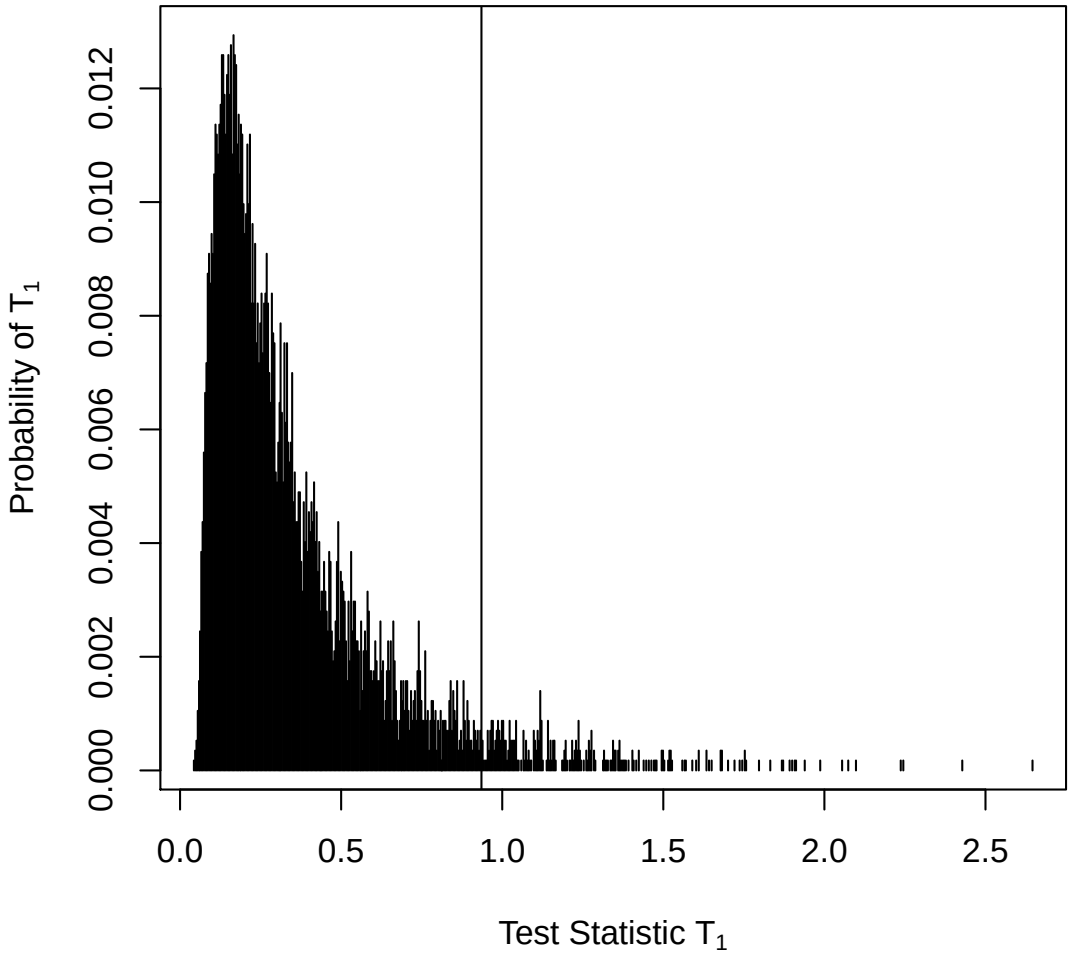


Figure 2.2. The exact null distribution of T_1 for unequal sample sizes $m = 7$, $n = 9$. The vertical line indicates the 5% critical value.

Table 2.1. Critical values for statistic $T_{1\,m,n}$: m $\leq 12, n < 13$; Row 1: $\alpha = 0.05$, Row 2: $\alpha = 0.01$

m/n	3	4	5	6	7	8	9	10	11	12
2	.	0.944	0.629	0.708	0.772	0.825	0.869	0.839	0.874	0.905
		0.944	1.000	1.042	1.074	1.100	1.121	1.139	1.154	1.167
3	.	0.786	0.925	0.852	0.919	0.919	0.898	0.933	0.903	0.933
		1.190	1.292	1.370	1.129	1.205	1.269	1.323	1.240	1.289
4	.	0.750	0.919	0.967	0.883	0.931	0.923	0.900	0.929	0.917
		1.375	1.174	1.317	1.234	1.347	1.346	1.414	1.387	1.375
5	.	.	0.900	0.939	0.934	0.931	0.941	0.924	0.935	0.937
			1.380	1.327	1.344	1.415	1.392	1.378	1.399	1.404
6	.	.	.	0.972	0.912	0.940	0.941	0.929	0.930	0.926
				1.361	1.392	1.411	1.415	1.421	1.440	1.417
7	.	.	.	.	0.929	0.947	0.936	0.943	0.937	0.936
					1.500	1.440	1.424	1.439	1.436	1.444
8	.	.	.	.	.	0.938	0.943	0.940	0.935	0.937
						1.438	1.440	1.443	1.443	1.446
9	.	.	.	.	.	.	0.944	0.938	0.935	0.937
							1.438	1.450	1.450	1.452
10	.	.	.	.	.	.	.	0.930	0.937	0.933
								1.450	1.453	1.456
11	.	.	.	.	.	.	.	.	0.938	0.933
									1.467	1.456
12	.	.	.	.	.	.	.	.	.	0.931
										1.458

Table 2.2. Critical values for statistic $T_{1\,m,n}$: m $\leq 13, n \geq 13$; Row 1: $\alpha = 0.05$, Row 2: $\alpha = 0.01$

m/n	13	14	16	18	20	25	30
2	0.819 0.988	0.848 1.009	0.850 1.044	0.861 1.072	0.859 1.095	0.856 1.105	0.865
3	0.896 1.300	0.908 1.339	0.921 1.279	0.899 1.302	0.917 1.300	0.897 1.299	0.897
4	0.923 1.380	0.914 1.374	0.925 1.369	0.909 1.356	0.911 1.386	0.907 1.379	0.901
5	0.935 1.390	0.932 1.418	0.923 1.401	0.923 1.405	0.924 1.400	0.917 1.397	0.914
6	0.931 1.416	0.931 1.421	0.928 1.420	0.931 1.431	0.924 1.419	0.919 1.419	0.917
7	0.931 1.439	0.930 1.433	0.930 1.436	0.928 1.436	0.926 1.434	0.923	
8	0.932 1.444	0.932 1.443	0.929 1.444	0.929 1.444	0.927	0.925	
9	0.933 1.450	0.934 1.451	0.932 1.452				
10	0.934 1.454	0.934 1.451	0.932 1.455				
11	0.934 1.455	0.933 1.458	0.932 1.458				
12	0.934 1.458	0.933 1.460					
13	0.938 1.459	0.933 1.461					

2.3 Properties of T_1

In this section, the properties of the statistic T_1 are studied. The expectation and variance of T_1 are provided in Anderson (1962) using a different formulation of the statistics with very long and extensive derivations. This section gives a straightforward and simpler derivation. For presentation purposes, let $R(X_i, H_N)$ be denoted as $R(X_i)$. This should not cause any confusion because all ranks are computed with respect to the combined sample. Let the natural rank of X_i be denoted as R_i. It is clear that $R(X_i) = R_i/N$.

Theorem 10 *(Anderson (1962)) Under the null hypothesis,*

$$\mathbb{E}T_1 = \frac{N+1}{6N} = \frac{1}{6} + \frac{1}{6N}$$

and

$$var(T_1) = \frac{N+1}{180N^2}\left[4(N+1) - \frac{3N^2}{mn}\right].$$

The following proof is different from Anderson (1962) as it is based on the new formulation of T_1, and this proof is much simpler.

Proof. Under H_0, $R(X_i)$ and $R(Y_j)$ are sampled from the discrete uniform distribution on $\{1/N, 2/N, ..., (N-1)/N, 1\}$ for all i and j. Hence

$$\begin{aligned}\mathbb{E}[T_1|H_0] =& \frac{mn}{N}\left\{\frac{mn}{mn}\mathbb{E}|R(X_1) - R(Y_1)|\right. \\ & \left. -\frac{m(m-1)}{2m^2}\mathbb{E}|R(X_1) - R(X_2)| - \frac{n(n-1)}{2n^2}\mathbb{E}|R(Y_1) - R(Y_2)|\right\} \\ =& \frac{mn}{N}\{(\frac{1}{2m} + \frac{1}{2n})\mathbb{E}|R(X_1) - R(X_2)|\} = \frac{1}{2N}\mathbb{E}|R_1 - R_2|,\end{aligned}$$

where R_1 and R_2 are natural ranks of X_1 and X_2, respectively. Under H_0, R_1 and R_2 are drawn uniformly from $\{1, 2, ..., N\}$ without replacement. Then,

$$\mathbb{E}|R_1 - R_2| = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}|i-j|$$

$$
\begin{aligned}
=&\frac{2}{N(N-1)}\sum_{i>j}(i-j)\\
=&\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}(i-j)\\
=&\frac{2}{N(N-1)}\sum_{i=2}^{N}\left[i(i-1)-\sum_{j=1}^{i-1}j\right]\\
=&\frac{2}{N(N-1)}\sum_{i=2}^{N}\left[i(i-1)-\frac{i(i-1)}{2}\right]\\
=&\frac{2}{N(N-1)}\sum_{i=2}^{N}\frac{i(i-1)}{2}\\
=&\frac{1}{N(N-1)}\sum_{k=1}^{N-1}k(k+1)\\
=&\frac{1}{N(N-1)}\sum_{k=1}^{N-1}(k^2+k)\\
=&\frac{1}{N(N-1)}\left[\frac{(N-1)N[2(N-1)+1]}{6}+\frac{N(N-1)}{2}\right]\\
=&\frac{1}{N(N-1)}\left[\frac{(N-1)N(2N-1)}{6}+\frac{N(N-1)}{2}\right]\\
=&\frac{2N-1}{6}+\frac{1}{2}\\
=&\frac{2N+2}{6}=\frac{N+1}{3}.
\end{aligned}
\tag{2.26}
$$

Hence $\mathbb{E}T_1=\frac{1}{2N}\left[\frac{N+1}{3}\right]=\frac{N+1}{6N}$.

The derivation of $var(T_1)$ is presented. Let R_1, R_2, R_3, R_4 be natural ranks of X_1, X_2, X_3 and X_4, respectively. We have

$$
\begin{aligned}
&\mathbb{E}(R_1-R_2)^2=\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2\\
=&\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}(i^2-2ij+j^2)\\
=&\frac{1}{N(N-1)}\sum_{i=1}^{N}\left[Ni^2-2i\frac{N(N+1)}{2}+\frac{N(N+1)(2N+1)}{6}\right]
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{N(N-1)}[N\cdot\frac{N(N+1)(2N+1)}{6}-N(N+1)\cdot\frac{N(N+1)}{2}\\
&+\frac{N^2(N+1)(2N+1)}{6}]\\
=&\frac{1}{N(N-1)}\left[\frac{N^2(N+1)(2N+1)}{3}-\frac{N^2(N+1)^2}{2}\right]\\
=&\frac{N^2(N+1)}{N(N-1)}\left[\frac{2N+1}{3}-\frac{N+1}{2}\right]\\
=&\frac{N(N+1)}{N-1}\cdot\frac{4N+2-3N-3}{6}\\
=&\frac{N(N+1)}{6},
\end{aligned}
\tag{2.27}
$$

$$
\begin{aligned}
&\mathbb{E}|R_1-R_2||R_1-R_3| = \frac{1}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{j\neq i}\sum_{k\neq i,j}|i-j||i-k|\\
=&\frac{1}{N(N-1)(N-2)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k\neq j}|i-j||i-k|\\
=&\frac{1}{N(N-1)(N-2)}\sum_{i,j,k=1}^{N}|i-j||i-k|-\frac{1}{N(N-1)(N-2)}\sum_{i,j=1}^{N}(i-j)^2\\
=&\frac{1}{N(N-1)(N-2)}\sum_{i=1}^{N}\left(\sum_{j=1}^{N}|i-j|\right)^2-\frac{N(N+1)}{6(N-2)}\\
=&\frac{1}{N(N-1)(N-2)}\sum_{i=1}^{N}\left(\sum_{j=1}^{i}(i-j)+\sum_{j=i}^{N}(j-i)\right)^2-\frac{N(N+1)}{6(N-2)}\\
=&\frac{1}{4N(N-1)(N-2)}\sum_{i=1}^{N}\left(2i^2-2(N+1)i+N^2+N\right)^2-\frac{N(N+1)}{6(N-2)}\\
=&\frac{1}{4N(N-1)(N-2)}\frac{1}{15}(8N-15N^3+7N^5)-\frac{N(N+1)}{6(N-2)}\\
=&\frac{(N+1)(7N+4)}{60},
\end{aligned}
\tag{2.28}
$$

and

$$\begin{aligned}
&\mathbb{E}|R_1-R_2||R_3-R_4| = \frac{1}{N(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq i,j}^{N}\sum_{l\neq i,j,k}^{N}|i-j||k-l| \\
=&\frac{1}{N(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k\neq i,j}^{N}\sum_{l\neq i,j,k}^{N}|i-j||k-l| \\
=&\frac{1}{N(N-1)(N-2)(N-3)}\sum_{i=1}^{N}\sum_{j=1}^{N}|i-j|\left(\sum_{k=1}^{N}\sum_{l=1}^{N}|k-l|-4\sum_{k=1}^{N}|i-k|+2|i-j|\right) \\
=&\frac{1}{N(N-1)(N-2)(N-3)}[\left(\sum_{i=1}^{N}\sum_{j=1}^{N}|i-j|\right)^2+2\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 \\
&-4\sum_{i,j,k=1}^{N}|i-j||i-k|] \\
=&\frac{N^2(N+1)^2(N-1)^2}{9N(N-1)(N-2)(N-3)}+\frac{N^2(N+1)(N-1)}{3N(N-1)(N-2)(N-3)} \\
&-\frac{N(N-1)(N+1)(7N^2-8)}{15N(N-1)(N-2)(N-3)} \\
=&\frac{(N+1)(5N+4)}{45}.
\end{aligned} \tag{2.29}$$

Using what is known about $\mathbb{E}T_1$ to evaluate $var(T_1)$, since $var(T_1)=\mathbb{E}(T_1^2)-(\mathbb{E}T_1)^2$, the two terms $\mathbb{E}(T_1^2)$ and $(\mathbb{E}T_1)^2$ need to evaluated. Extending T_1^2 and $\mathbb{E}T_1^2$ yields the above three types of expectations $\mathbb{E}(R_1-R_2)^2$, $\mathbb{E}(|R_1-R_2||R_1-R_3|)$ and $\mathbb{E}(|R_1-R_2||R_3-R_4|)$ denoted as E_1, E_2, E_3, respectively. That is,

$$\begin{aligned}
\mathbb{E}T_1^2 = &\frac{m^2n^2}{N^4}\left\{\left[\frac{mn}{m^2n^2}+\frac{2m(m-1)}{4m^2}+\frac{2n(n-1)}{4n^4}\right]E_1\right. \\
&+\left[\frac{mn(m-1)(n-1)}{m^2n^2}+\frac{m(m-1)(m-2)(m-3)}{4m^4}+\frac{n(n-1)(n-2)(n-3)}{4n^4}\right. \\
&\left.-\frac{2mn(m-1)(m-2)}{2m^3n}-\frac{2mn(n-1)(n-2)}{2mn^3}+\frac{2m(m-1)n(n-1)}{4m^2n^2}\right]E_3 \\
&+\left[\frac{m^2n^2-mn-mn(m-1)(n-1)}{m^2n^2}+\frac{4m(m-1)(m-2)}{4m^4}+\frac{4n(n-1)(n-2)}{4n^4}\right. \\
&\left.\left.-\frac{m^2n(m-1)-mn(m-1)(m-2)}{m^3n}-\frac{mn^2(n-1)-mn(n-1)(n-2)}{mn^3}\right]E_2\right\}
\end{aligned}$$

$$= \frac{N+1}{60N^2}\left[3N+3-\frac{N^2}{mn}\right].$$

Hence,

$$\begin{aligned} var(T_1) =& \mathbb{E}T_1^2 - (\mathbb{E}T_1)^2 \\ =& \frac{N+1}{60N^2}\left[3N+3-\frac{N^2}{mn}\right] - \left(\frac{N+1}{6N}\right)^2 \\ =& \frac{3(N+1)\left[3N+3-\frac{N^2}{mn}\right] - 5(N+1)^2}{180N^2} \\ =& \frac{N+1}{180N^2}\left[4(N+1)-\frac{3N^2}{mn}\right]. \end{aligned}$$

■

Remark 11 *If* $m, n \to \infty$, $\mathbb{E}T_1 \to 1/6$ *and* $var(T_1) \to 1/45$.

Lemma 12 *Under* H_0, $\mathbb{E}[|R(X_1) - R(X_2)||X_1] = \frac{1}{2} - \frac{N-2}{N}[F(X_1) - F^2(X_1)]$,
$\mathbb{E}[|R(X_2) - R(X_3)||X_1] = \frac{1}{3} + \frac{2}{N}[F(X_1) - F^2(X_1)]$ *and*
$\mathbb{E}(T_1|X_1) = \frac{1}{6}\left(1+\frac{n}{mN}\right) + \frac{1}{N}\left(1-\frac{n}{m}\right)[F(X_1) - F^2(X_1)]$.

Proof. Under H_0, $R_1 - 1$ given X_1 has a binomial distribution with parameters $N-1$ and $F(X_1)$:

$$P(R_1 - 1 = u|X_1, H_0) = \binom{N-1}{u} F(X_1)^u [1-F(X_1)]^{N-1-u}, \quad u = 0, 1, ..., N-1.$$

Given R_1, R_2 is uniformly distributed from $\{1, 2, ..., N\}\backslash\{R_1\}$. Since $\mathbb{E}[Z|X] = \mathbb{E}[\mathbb{E}(Z|X, Y)|X]$, we have

$$\begin{aligned} & \mathbb{E}(|R_1 - R_2||X_1) \\ =& \mathbb{E}[\mathbb{E}(|R_1 - R_2||X_1, R_1)|X_1] = \sum_{R_1=1}^{N} \frac{1}{N-1}\left[\sum_{i=1,\neq R_1}^{N} |R_1 - i|\right] P(R_1|X_1) \end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{N-1}\sum_{R_1=1}^{N}\left[\sum_{i=1}^{R_1}(R_1-i)+\sum_{i=R_1}^{N}(i-R_1)\right]P(R_1|X_1)\\
=&\frac{1}{2(N-1)}\sum_{R_1=1}^{N}(N^2+2R_1^2-2NR_1-2R_1+N)P(R_1|X_1)\\
=&\frac{1}{2(N-1)}\sum_{R_1=1}^{N}[(N-1)N-2(N-1)(R_1-1)+2(R_1-1)^2]P(R_1|X_1)\\
=&\frac{1}{2}N-(N-1)F(X_1)+\frac{1}{N-1}[(N-1)F(X_1)(1-F(X_1)+(N-1)^2F^2(X_1)]\\
=&\frac{1}{2}N-(N-2)[F(X_1)-F^2(X_1)].
\end{aligned}
$$

Let R_3 be the natural rank of X_3. Under H_0, we have

$$
\begin{aligned}
\mathbb{E}(|R_2-R_3||X_1)&=\mathbb{E}[\mathbb{E}(|R_2-R_3||X_1,R_1)|X_1]\\
&=\sum_{R_1=1}^{N}\frac{1}{(N-1)(N-2)}\left[\sum_{i=1,i\neq R_1}^{N}\sum_{j=1,j\neq i,R_1}^{N}|i-j|\right]P(R_1|X_1)\\
&=\sum_{R_1=1}^{N}\frac{1}{(N-1)(N-2)}\left[\sum_{i,j=1}^{N}|i-j|-2\sum_{i=1}^{N}|R_1-i|\right]P(R_1|X_1)\\
&=\frac{N(N+1)}{3(N-2)}-\frac{N-2(N-2)[F(X_1)-F^2(X_1)]}{N-2}\\
&=\frac{1}{3}N+2[F(X_1)-F^2(X_1)].
\end{aligned}
$$

Note that $\mathbb{E}(T_1|X_1)$ contains the above conditional expectations $\mathbb{E}[|R(X_1)-R(X_2)||X_1]$ and $\mathbb{E}[|R(X_2)-R(X_3)||X_1]$ denoted as E_1^* and E_2^*, respectively. Then it follows that

$$
\begin{aligned}
&\mathbb{E}(T_1|X_1)\\
&=\frac{mn}{N}\left\{\frac{n}{mn}E_1^*+\frac{mn-n}{mn}E_2^*-\frac{2(m-1)}{2m^2}E_1^*-\frac{(m-1)(m-2)}{2m^2}E_2^*-\frac{n(n-1)}{2n^2}E_2^*\right\}\\
&=\frac{mn}{N}\left\{\frac{1}{m^2}(E_1^*-E_2^*)+\frac{N}{2mn}E_2^*\right\}
\end{aligned}
$$

$$= \frac{1}{6}\left(1+\frac{n}{mN}\right)+\frac{1}{N}\left(1-\frac{n}{m}\right)[F(X_1)-F^2(X_1)].$$

■

Next, let us focus on the limiting distribution. To determine the limiting distribution of T_1, the Hájek projection method is used. Suppose $X_1, X_2, ..., X_n$ are independent random variables, and let $\mathcal{S}_n$ denote a sequence of linear spaces containing all variables of the form $\sum_{i=1}^{n} g_i(X_i)$ with $\mathbb{E}g_i^2(X_i) < \infty$.

Definition 13 *For a statistic $T_n = T(X_1, X_2, ..., X_n)$ with a finite second moment, the Hájek Projection of T_n onto $\mathcal{S}_n$ is $\hat{S}_n = \sum_{i=1}^{n} E[T|X_i] - (n-1)E[T]$.*

Theorem 14 *(Hájek Projection Theorem)*

Let $\hat{S}_n$ denote the Hájek projection of T_n onto $\mathcal{S}_n$. If the linear space $\mathcal{S}_n$ contain constants and $\frac{Var(T_n)}{Var(\hat{S}_n)} \to 1$, then

$$R_n = \frac{T_n - E(T_n)}{\sqrt{Var(T_n)}} - \frac{\hat{S}_n - E(\hat{S}_n)}{\sqrt{Var(\hat{S}_n)}} \xrightarrow{p} 0.$$

Therefore, the Hájek projection of T_1 is defined as

$$\hat{T_1} = \sum_{i=1}^{m} \mathbb{E}[T_1|X_i] + \sum_{j=1}^{n} \mathbb{E}[T_1|Y_j] - (N-1)\mathbb{E}T_1$$

with variance

$$\begin{aligned}
Var\hat{T_1} &= \sum_{i=1}^{m} Var(\mathbb{E}[T_1|X_i]) + \sum_{j=1}^{n} Var(\mathbb{E}[T_1|Y_j]) \\
&= \left[\frac{m(m-n)^2}{m^2N^2} + \frac{n(m-n)^2}{n^2N^2}\right] Var[F(Y_1) - F^2(Y_1)] \\
&= (m-n)^2\left[\frac{m}{m^2N^2} + \frac{n}{n^2N^2}\right]\left(\mathbb{E}\left[\left(F(Y_1) - F^2(Y_1)\right)^2\right] - \mathbb{E}^2\left(F(Y_1) - F^2(Y_1)\right)\right) \\
&= (m-n)^2\left[\frac{1}{mnN}\right]\left(\frac{1}{30} - (\frac{1}{6})^2\right)
\end{aligned}$$

$$= \frac{(m-n)^2}{180mnN}.$$

Note that $Var\hat{T}_1/VarT_1 \to 0$ as $m, n \to \infty$. Therefore the first order Hájek projection does not apply to T_1. Note that Rosenblatt (1952) and Fisz (1960) have derived the limiting distribution of T_1 defined in (2.24) by using the stochastic process method involving random variables depending on a variable parameter. The next section provides the limiting distribution of the test statistic T by applying the projection method.

2.4 Properties of T

The properties of the test statistic (2.8) is studied in this section. Recall that $R(X_i) = R_i/N$. Using (2.26), $\mathbb{E}T = \mathbb{E}|R(X_1) - R(Y_1)| = \mathbb{E}|R_1 - R_2|/N = \frac{N+1}{3N}$ (2.29′). From (2.8),

$$T^2 = \frac{1}{m^2n^2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}\Big[|R(X_i, H_N) - R(Y_j, H_N)||R(X_k, H_N) - R(Y_l, H_N)|\Big].$$

Recall that $E_1 = \mathbb{E}(R_1 - R_2)^2$, $E_2 = \mathbb{E}(|R_1 - R_2||R_1 - R_3|)$ and $E_3 = \mathbb{E}(|R_1 - R_2||R_3 - R_4|)$. Thus, by T^2, (2.27), (2.28) and (2.29),

$$\begin{aligned}
\mathbb{E}T^2 &= \frac{1}{m^2n^2N^2}[mnE_1 + mn(m+n-2)E_2 + mn(m-1)(n-1)E_3] \\
&= \frac{1}{mnN^2}E_1 + \frac{m+n-2}{mnN^2}E_2 + \frac{(m-1)(n-1)}{mnN^2}E_3 \\
&= \frac{1}{mnN^2}\left[\frac{N(N+1)}{6} + \frac{(N-2)(N+1)(7N+4)}{60} + \frac{(mn-N+1)(N+1)(5N+4)}{45}\right] \\
&= \frac{N+1}{180mnN^2}\left[30N + +3(7N+4)(m+n-2) + 4(5N+4)(m-1)(n-1)\right] \\
&= \frac{N+1}{180mnN^2}(20mnN + 16mn + N^2 + 4N - 8).
\end{aligned}$$

Therefore,

$$\begin{aligned} VarT &= \mathbb{E}T^2 - (\mathbb{E}T)^2 \\ &= \frac{N+1}{180mnN^2}(20mnN + 16mn + N^2 + 4N - 8) - \frac{(N+1)^2}{9N^2} \\ &= \frac{(N+1)[20mnN + 16mn + N^2 + 4N - 8 - 20(N+1)mn]}{180mnN^2} \\ &= \frac{(N+1)[-4mn + N^2 + 4N - 8]}{180mnN^2} \\ &= \frac{(N+1)[(n-m)^2 + 4(N-2)]}{180mnN^2}. \end{aligned} \tag{2.30}$$

Studying the projection of T, consider

$$\mathbb{E}[T|X_1] = \frac{1}{m}\mathbb{E}(|R(X_1) - R(X_2)||X_1) + (1 - \frac{1}{m})\mathbb{E}(|R(X_2) - R(X_3)||X_1).$$

By Lemma 12,

$$\begin{aligned} \mathbb{E}[T|X_1] &= \frac{1}{m}\left[\frac{1}{2} - \frac{N-2}{N}[F(X_1) - F^2(X_1)]\right] + (1 - \frac{1}{m})\left[\frac{1}{3} + \frac{2}{N}[F(X_1) - F^2(X_1)]\right] \\ &= \frac{1}{3} + \frac{1}{6m} + \frac{m-n}{mN}[F(X_1) - F^2(X_1)]. \end{aligned} \tag{2.31}$$

Similarly,

$$\mathbb{E}[T|Y_1] = \frac{1}{3} + \frac{1}{6n} + \frac{n-m}{nN}[F(Y_1) - F^2(Y_1)]. \tag{2.32}$$

Then the projection of T is given by $\tilde{T} = \sum_{i=1}^{m}\mathbb{E}[T|X_i] + \sum_{j=1}^{n}\mathbb{E}[T|Y_j] - (N-1)\mathbb{E}T$ with variance

$$\begin{aligned} Var\tilde{T} &= \sum_{i=1}^{m} Var(\mathbb{E}[T|X_i]) + \sum_{j=1}^{n} Var(\mathbb{E}[T|Y_j]) \\ &= mVar\left(\frac{1}{3} + \frac{1}{6m} + \frac{m-n}{mN}[F(X_1) - F^2(X_1)]\right) \end{aligned} \tag{2.33}$$

$$
\begin{aligned}
&+ nVar\left(\frac{1}{3}+\frac{1}{6n}+\frac{n-m}{nN}[F(Y_1)-F^2(Y_1)]\right)\\
=&\left[\frac{m(m-n)^2}{m^2N^2}+\frac{n(m-n)^2}{n^2N^2}\right]Var[F(Y_1)-F^2(Y_1)]\\
=&\frac{(m-n)^2}{180mnN}.
\end{aligned}
$$

The result of (2.33) is due to X_i and Y_j being independent and identically distributed under H_0. Thus, by (2.30),

$$
\begin{aligned}
\frac{Var\tilde{T}}{VarT} &= \frac{(m-n)^2}{180mnN}\cdot\frac{180mnN^2}{(N+1)[(n-m)^2+4(N-2)]}\\
&= \frac{(m-n)^2N}{(N+1)[(n-m)^2+4(N-2)]}\\
&= \frac{(m-n)^2}{(n-m)^2+4(N-2)}.
\end{aligned}
$$

If $(m-n)^2/N\to\infty$ as $N\to\infty$, $\dfrac{Var\tilde{T}}{VarT}\to 1$.

Let L_2 be defined as $L_2(F)=\{g:\int_{-\infty}^{\infty}g^2(x)dF(x)<\infty\}$. Then $g_n\to g$ in $L_2(F)$ if and only if $\int(g_n(x)-g(x))^2dF(x)\overset{n\to\infty}{\to}0$. Convergence in probability F, i.e., $g_n\overset{F}{\to}g$, means that for every $\epsilon>0$, $P_F(|g_n-g|>\epsilon)\overset{n\to\infty}{\to}0$. Notice that convergence in L_2 implies convergence in probability. The following is a lemma useful in deriving the asymptotics of the test statistic T.

Lemma 15 *Let $S_n(X_1,X_2,\cdots,X_n)$ be a function of n independent random variables with decomposition $S_n=M_n+R_n$. If $\mathbb{E}(R_n)=Cov(M_n,R_n)=0$ for all n and $Var(S_n)/Var(M_n)\to 1$ as $n\to\infty$, then $|R_n|/\sqrt{VarS_n}\to 0$ in L_2 and therefore $|R_n|/\sqrt{VarS_n}\to 0$ in probability.*

Proof.

$$
\begin{aligned}
\mathbb{E}[R_n^2/Var(S_n)] &= \frac{\mathbb{E}[(S_n-\mathbb{E}S_n)-(M_n-\mathbb{E}M_n)]^2}{Var(S_n)}\\
&= \frac{Var(S_n)+Var(M_n)-2\mathbb{E}(S_n-\mathbb{E}S_n)(M_n-\mathbb{E}M_n)}{Var(S_n)}
\end{aligned}
$$

$$= \frac{Var(S_n) + Var(M_n) - 2\mathbb{E}(M_n - \mathbb{E}M_n)^2 - 2\mathbb{E}R_n(M_n - \mathbb{E}M_n)}{Var(S_n)}$$
$$= \frac{Var(S_n) - Var(M_n)}{Var(S_n)} = 1 - Var(M_n)/Var(S_n) \to 0.$$

■

Lemma 15 is application of the Hájek projection technique. See Hájek & Šidák (1967) and Hettmansperger & McKean (2010) for details. By Lemma 15, Lindeberg central limit theorem and the fact of $Var\tilde{T}/VarT \to 1$ as $N \to \infty$, the following central limit theorem may be established if $m \neq n$.

Theorem 16 *If* $(m-n)^2/N \to \infty$ *and* $m \neq n$*, then*

$$\frac{\tilde{T} - \frac{N+1}{3N}}{\sqrt{\frac{(m-n)^2}{180mnN}}} \Rightarrow N(0,1). \tag{2.34}$$

Proof. Denote $U = F(X_1) - F^2(X_1)$ and $V = F(Y_1) - F^2(Y_1)$. Without loss of generality, assume that $m > n$. From (2.31) and (2.29′),

$$\begin{aligned}
&\mathbb{E}[T|X_1] - \mathbb{E}(T) \\
&= \frac{1}{6m} - \frac{1}{3N} + \frac{m-n}{mN}U \\
&= \frac{m-n}{6mN}(6U - 1).
\end{aligned}$$

Similarly,

$$\mathbb{E}[T|Y_1] - \mathbb{E}(T) = \frac{m-n}{6nN}(1 - 6V).$$

Let $\mathbb{P} = (\Omega, \mathcal{F}, \mathbb{P})$ be a probablity space, and $X_k : \Omega \to \mathbb{R}$, $k \in \mathbb{N}$, be independent random variables defined on $\mathbb{P}$. Assume $\mathbb{E}[X_k] = \mu_k$ and $Var[X_k] = \sigma_k^2$ exist are finite, and let

$s_n^2 := \sum_{k=1}^{n} \sigma_k^2$. By the Lindeberg central limit theorem, if $\{X_k\}_{k=1}^n$ satisfies the condition:

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbb{E}[(X_k - \mu_k)^2 \cdot I_{\{|X_k-\mu_k|>\epsilon s_n\}}] = 0 \tag{2.35}$$

for all $\epsilon > 0$, then the general central limit theorem holds, i.e., the distribution of the standardized sums $\frac{1}{s_n}\sum_{k=1}^{n}(X_k - \mu_k)$ converges to the standard normal distribution $N(0,1)$. In our case, $E(T|X_k)$ and $E(T|Y_k)$ have the mean $E(T)$ and their second moments exist. Also, $s_n = \sum_{k=1}^{n} Var(E(T|X_k) + \sum_{k=1}^{n} Var(E(T|Y_k) = Var(\tilde{T})$. Since $U = F(X) - F^2(X) \in [0, 1/4]$, for large m and N,

$$\begin{aligned} & I\left\{|\mathbb{E}[T|X_k] - \mathbb{E}(T)| > \epsilon\sqrt{Var\tilde{T}}\right\} \\ &= I\left\{|\frac{m-n}{6mN}(6U-1)| > \frac{\epsilon(m-n)}{\sqrt{180mnN}}\right\} \\ &= I\left\{|6U-1| > \frac{\epsilon\sqrt{mN}}{\sqrt{5n}}\right\} = 0. \end{aligned}$$

By the same argument, since $V = F(Y) - F^2(Y) \in [0, 1/4]$, for large n and N,

$$\begin{aligned} & I\left\{|\mathbb{E}[T|Y_k] - \mathbb{E}(T)| > \epsilon\sqrt{Var\tilde{T}}\right\} \\ &= I\left\{|6V-1| > \frac{\epsilon\sqrt{nN}}{\sqrt{5m}}\right\} = 0. \end{aligned}$$

Then the Lindeberg condition (2.35) is satisfied, hence (2.34) holds. ■

Since $(m-n)^2/N \to \infty$ implies $\dfrac{1}{3N\sqrt{\frac{(m-n)^2}{180mnN}}} \to 0$, the following corollary is established.

Corollary 17 *If $(m-n)^2/N \to \infty$ as $N \to \infty$, then*

$$\frac{\tilde{T} - \frac{1}{3}}{\sqrt{\frac{(m-n)^2}{180mnN}}} \xrightarrow{d} N(0,1).$$

Corollary 18 *If* $(m-n)^2/N \to \infty$ *as* $N \to \infty$*, then*

$$\frac{T-\frac{1}{3}}{\sqrt{\frac{(m-n)^2}{180mnN}}} \xrightarrow{d} N(0,1). \tag{2.36}$$

Proof. By the decomposition presented in a lemma of Efron & Stein (1978), $T - \mathbb{E}(T) = (\tilde{T} - \mathbb{E}(T)) + R_n$ and $cov(\tilde{T}, R_n) = 0$. Under the condition $(m-n)^2/N \to \infty$, $Var\tilde{T}/VarT \to 1$. Then by Corollary 17 and Lemma 15, we have (2.36). ■

However, Example 7 shows that it is not suitable to apply T for a general test if $m \neq n$. For the case $m = n$, the second order projection is necessary since the first order projections are constants [see (2.31) and (2.32)].

Lemma 19 *The second order projection of* T *on one* X *variable and one* Y *variable is*

$$\mathbb{E}[T|X_1, Y_1, H_0] = \frac{2n^3 + n^2 - 2n + 2}{6n^3} + \frac{n-1}{2n^3}[|F(X_1) - F(Y_1)| + F(X_1)(1 - F(X_1)) + F(Y_1)(1 - F(Y_1))]$$

and its variance is $Var\{\mathbb{E}[T|X_1, Y_1, H_0]\} = \frac{(n-1)^2}{90n^6}$.

Proof. Let R_i be the natural rank of X_i, $i = 1, 2, 3, 4$. Under H_0 and given $X_1 < X_2$, (R_1, R_2) has trinomial distribution with parameters $F(X_1)$, $F(X_2) - F(X_1)$ and $1 - F(X_2)$, i.e.,

$$\begin{aligned} &P(R_1 = u, R_2 = v | X_1, X_2, X_1 < X_2, H_0) \\ &= \binom{N-2}{u-1, v-u-1, N-v} [F(X_1)]^{u-1} [F(X_2) - F(X_1)]^{v-u-1} [1 - F(X_2)]^{N-v}. \end{aligned} \tag{2.37}$$

Then

$$\mathbb{E}[R_1 | X_1, X_2, X_1 < X_2, H_0] = (N-2)F(X_1) + 1,$$

and

$$\mathbb{E}[R_2 | X_1, X_2, X_1 < X_2, H_0] = (N-2)F(X_2) + 1.$$

Therefore, the first condition (involving the ranks of the given observations) is given by

$$\mathbb{E}[|R_1 - R_2||X_1, X_2, X_1 < X_2, H_0] = (N-2)[F(X_2) - F(X_1)] + 1. \tag{2.38}$$

Under H_0 and given R_1, R_2, the natural rank of X_3, R_3, has discrete uniform distribution on the set $\{1, 2, \cdots, N\} \setminus \{R_1, R_2\}$, i.e., $P(R_3 = w|R_1, R_2, H_0) = \dfrac{1}{N-2}$ for $1 \leq w \leq N$, $w \neq R_1, R_2$. The second condition (involving the rank of a given obervation and the rank of a different observation) is given by

$$\begin{aligned}
&\mathbb{E}(|R_1 - R_3||X_1, X_2, X_1 < X_2, H_0) \\
&= \mathbb{E}[\mathbb{E}(|R_1 - R_3||R_1, R_2, X_1 < X_2, H_0)|X_1 < X_2, H_0] \\
&= \mathbb{E}\left[\frac{1}{N-2}\left(\sum_{1\leq i<R_1}(R_1 - i) + \sum_{R_1<i\leq N, i\neq R_2}(i - R_1)\right)|X_1 < X_2, H_0\right] \\
&= \mathbb{E}\left(\frac{N^2 + N - 2NR_1 + 2R_1^2 - 2R_2}{2(N-2)}|X_1 < X_2, H_0\right) \\
&= \frac{1}{2(N-2)}\left[N^2 + N - 2N\mathbb{E}(R_1) + 2\mathbb{E}(R_1^2) - 2\mathbb{E}(R_2)|X_1 < X_2, H_0\right] \\
&= \frac{1}{2(N-2)}\Big[N^2 + N - 2N[(N-2)F(X_1) + 1] + \\
&2\big[[(N-2)F(X_1)]^2 + (N-2)F(X_1)(1 - F(X_1) + 2[(N-2)F(X_1) + 1] - 1\big] - \\
&2[(N-2)F(X_2) + 1]\Big] \qquad (2.39) \\
&= \frac{N+1}{2} - (N-3)F(X_1)[1 - F(X_1)] - F(X_2). \qquad (2.40)
\end{aligned}$$

Equality (2.39) is based on $\mathbb{E}[R_1|X_1, X_2, X_1 < X_2, H_0]$, $\mathbb{E}[R_1|X_1, X_2, X_1 < X_2, H_0]$ and the fact that $\mathbb{E}(R_1 - 1)^2 = [\mathbb{E}(R_1 - 1)]^2 + Var(R_1 - 1)$. Note that $\mathbb{E}(R_1)^2 = (N - 2)F(X_1)]^2 + (N-2)F(X_1)(1 - F(X_1) + 2\mathbb{E}(R_1) - 1$. Therefore, $\mathbb{E}[R_1^2|X_1, X_2, X_1 < X_2, H_0]$ is obtained. Hence, we have equality (2.40). Similarly, the third condition is given

by

$$\mathbb{E}(|R_1 - R_3||X_1, X_2, X_1 > X_2, H_0)] = \frac{N-1}{2} - (N-3)F(X_1)[1 - F(X_1)] + F(X_2). \tag{2.41}$$

The final condition (involving the ranks of observations different from the given observations) is given by

$$\begin{aligned}
&\mathbb{E}(|R_3 - R_4||X_1, X_2, X_1 < X_2, H_0) = \mathbb{E}[\mathbb{E}(|R_3 - R_4||R_1, R_2, X_1 < X_2, H_0)|X_1 < X_2, H_0] \\
&= \mathbb{E}\left[\frac{1}{(N-2)(N-3)} \sum_{1 \le i,j \le N, i,j \ne R_1, R_2} |i - j||X_1 < X_2, H_0\right] \\
&= \frac{N-1}{3} + 2F(X_1)[1 - F(X_1)] + 2F(X_2)[1 - F(X_2)].
\end{aligned} \tag{2.42}$$

The last equality (2.42) is from (2.37), which is the conditional distribution of R_1 and R_2. By using (2.8),

$$\begin{aligned}
\mathbb{E}[T|X_1, Y_1, X_1 < Y_1, H_0] = \frac{1}{mn} \{&\mathbb{E}[|R(X_1) - R(Y_1)||X_1 < Y_1, H_0] \\
&+ (n-1)\mathbb{E}(|R(X_1) - R(Y_2)||X_1 < Y_1, H_0) \\
&+ (m-1)\mathbb{E}(|R(X_2) - R(Y_1)||X_1 < Y_1, H_0) \\
&+ (m-1)(n-1)\mathbb{E}(|R(X_2) - R(Y_2)||X_1 < Y_1, H_0)\}.
\end{aligned}$$

In the case that $m = n$, by (2.38), (2.40), (2.41) and (2.42), we have

$$\begin{aligned}
\mathbb{E}[T|X_1, Y_1, X_1 < Y_1, H_0] = \frac{1}{2n^3} \{&(2(n-1)(F(Y_1) - F(X_1)) + 1 \\
&+ (n-1)[\frac{2n+1}{2} - (2n-3)F(X_1)(1 - F(X_1)) - F(Y_1)] \\
&+ (n-1)[\frac{2n-1}{2} - (2n-3)F(Y_1)(1 - F(Y_1)) + F(X_1)] \\
&+ (n-1)^2[\frac{2n-1}{3} + 2F(X_1)(1 - F(X_1)) + 2F(Y_1)(1 - F(Y_1))]\}
\end{aligned}$$

$$=\frac{2n^3+n^2-2n+2}{6n^3}$$
$$+\frac{n-1}{2n^3}[F(Y_1)-F(X_1)+F(X_1)(1-F(X_1))+F(Y_1)(1-F(Y_1))].$$

Similarly, (2.38), (2.40), (2.41) and (2.42) provide

$$\mathbb{E}[T|X_1,Y_1,X_1>Y_1,H_0]=\frac{2n^3+n^2-2n+2}{6n^3}$$
$$+\frac{n-1}{2n^3}[F(X_1)-F(Y_1)+F(X_1)(1-F(X_1))+F(Y_1)(1-F(Y_1))].$$

Note that $\mathbb{E}[T|X_1,Y_1,H_0]$ can be expressed as either

$$\mathbb{E}[T|X_1,Y_1,X_1<Y_1,H_0]$$

or

$$\mathbb{E}[T|X_1,Y_1,X_1>Y_1,H_0].$$

Furthermore, $\mathbb{E}[T|X_1,Y_1,X_1<Y_1,H_0]$ and $\mathbb{E}[T|X_1,Y_1,X_1>Y_1,H_0]$ has the same expression

$$\frac{2n^3+n^2-2n+2}{6n^3}+\frac{n-1}{2n^3}[|F(X_1)-F(Y_1)|+F(X_1)(1-F(X_1))+F(Y_1)(1-F(Y_1))].$$

Hence,

$$\mathbb{E}[T|X_1,Y_1,H_0]=\frac{2n^3+n^2-2n+2}{6n^3}$$
$$+\frac{n-1}{2n^3}[|F(X_1)-F(Y_1)|+F(X_1)(1-F(X_1))+F(Y_1)(1-F(Y_1))].$$

Let $U_1=F(X_1)$ and $U_2=F(Y_1)$ be independent and identically distributed (i.i.d) uniform

random variables on $[0,1]$. Then

$$Var\{\mathbb{E}[T|X_1,Y_1,H_0]\} = \frac{(n-1)^2}{4n^6} Var(|U_1-U_2| + U_1(1-U_1) + U_2(1-U_2)) \\ = \frac{(n-1)^2}{90n^6}.$$

■

Lemma 20 *The second order projection of T on all X variables or all Y variables is*

$$\mathbb{E}[T|X_1,X_2,H_0] = \frac{2n^2+n+2}{6n^2} \\ - \frac{1}{2n^2}[|F(X_1)-F(X_2)| + F(X_1)(1-F(X_1)) + F(X_2)(1-F(X_2))]$$

and

$$\mathbb{E}[T|Y_1,Y_2,H_0] = \frac{2n^2+n+2}{6n^2} \\ - \frac{1}{2n^2}[|F(Y_1)-F(Y_2)| + F(Y_1)(1-F(Y_1)) + F(Y_2)(1-F(Y_2))],$$

and the variances are $Var\{\mathbb{E}[T|X_1,X_2,H_0]\} = Var\{\mathbb{E}[T|Y_1,Y_2,H_0]\} = \frac{1}{90n^4}$.

Proof. By the definition of T as in (2.8),

$$\mathbb{E}[T|X_1,X_2,X_1<X_2,H_0] \\ = \frac{1}{mn}\{n\mathbb{E}(|R(X_1)-R(Y_1)||X_1,X_2,X_1<X_2,H_0) + n\mathbb{E}(|R(X_2)-R(Y_1)||X_1,X_2,X_1<X_2,H_0) \\ + n(m-2)\mathbb{E}(|R(X_3)-R(Y_1)||X_1,X_2,X_1<X_2,H_0)\}.$$

In the case that $m=n$, by (2.40), (2.41) and (2.42),

$$\mathbb{E}[T|X_1,X_2,X_1<X_2,H_0] \\ = \frac{1}{2n^3}\{n[(2n+1)/2 - (2n-3)F(X_1)(1-F(X_1)) - F(X_2)]$$

$$
\begin{aligned}
&+ n[(2n-1)/2 - (2n-3)F(X_2)(1-F(X_2)) + F(X_1)] \\
&+ n(n-2)[(2n-1)/3 + 2F(X_1)(1-F(X_1)) + 2F(X_2)(1-F(X_2))]\} \\
=&\frac{2n^2+n+2}{6n^2} \\
&- \frac{1}{2n^2}[F(X_2) - F(X_1) + F(X_1)(1-F(X_1)) + F(X_2)(1-F(X_2))].
\end{aligned}
$$

By symmetry,

$$
\begin{aligned}
\mathbb{E}[T|X_1, X_2, X_1 > X_2, H_0] = \frac{2n^2+n+2}{6n^2} - \frac{1}{2n^2}[F(X_1) \\
- F(X_2) + F(X_1)(1-F(X_1)) + F(X_2)(1-F(X_2))].
\end{aligned}
$$

$\mathbb{E}[T|X_1, X_2, H_0]$ can be expressed as either

$$\mathbb{E}[T|X_1, X_2, X_1 < X_2, H_0]$$

or

$$\mathbb{E}[T|X_1, X_2, X_1 > X_2, H_0].$$

Furthermore, $\mathbb{E}[T|X_1, X_2, X_1 < X_2, H_0]$ and $\mathbb{E}[T|X_1, X_2, X_1 > X_2, H_0]$ has the same expression

$$\frac{2n^2+n+2}{6n^2} - \frac{1}{2n^2}[|F(X_1) - F(X_2)| + F(X_1)(1-F(X_1)) + F(X_2)(1-F(X_2))].$$

Therefore, given all X variables,

$$
\begin{aligned}
&\mathbb{E}[T|X_1, X_2, H_0] = \frac{2n^2+n+2}{6n^2} \\
&- \frac{1}{2n^2}[|F(X_1) - F(X_2)| + F(X_1)(1-F(X_1)) + F(X_2)(1-F(X_2))]
\end{aligned}
$$

and its variance is

$$\begin{aligned} Var\{\mathbb{E}[T|X_1, X_2, H_0]\} &= \frac{1}{4n^4} Var(|U_1 - U_2| + U_1(1 - U_1) + U_2(1 - U_2)) \\ &= \frac{1}{90n^4}. \end{aligned}$$

The results for the projections on all Y variables are the same due to symmetry of X and Y in T. ■

Notice that

$$\begin{aligned} \mathbb{E}\{\mathbb{E}[T|X_1, Y_1, H_0]\} = \mathbb{E}\{\mathbb{E}[T|X_1, X_2, H_0]\} = \mathbb{E}\{\mathbb{E}[T|Y_1, Y_2, H_0]\} \\ = \frac{2n+1}{6n} = \mathbb{E}T \end{aligned}$$

and $cov(\mathbb{E}[T|Z_1, Z_2, H_0], \mathbb{E}[T|Z_1, Z_3, H_0]) = 0$, where Z_1, Z_2 and Z_3 are three variables from X_i, Y_i, $1 \leq i \leq n$. The following shows that $cov(\mathbb{E}[T|X_1, X_2, H_0]$, $\mathbb{E}[T|X_1, X_3, H_0]) = 0$: Let $C = \frac{2n^2 + n + 2}{6n^2}$ let and $U_1 = F(X_1), U_2 = F(X_2)$ and $U_3 = F(X_3)$ be i.i.d on uniform$[0, 1]$. Then,

$$\begin{aligned} & cov(\mathbb{E}[T|X_1, X_2, H_0], \mathbb{E}[T|X_1, X_3, H_0]) \\ =& cov\bigg(C - \frac{1}{2n^2}[|U_1 - U_2| + U_1(1 - U_1) + U_2(1 - U_2)], \\ & C - \frac{1}{2n^2}[|U_1 - U_3| + U_1(1 - U_1) + U_3(1 - U_3)]\bigg) \\ =& \frac{1}{4n^4}\bigg(cov\big(|U_1 - U_2|, |U_1 - U_3|\big) + 2cov\big(|U_1 - U_2|, U_1(1 - U_1)\big) \\ & + cov\big(U_1(1 - U_1, U_1(1 - U_1)\big)\bigg) \\ =& \frac{1}{4n^4}\bigg(\mathbb{E}(|U_1 - U_2| \cdot |U_1 - U_3|) - \mathbb{E}(|U_1 - U_2|)\mathbb{E}(|U_1 - U_3|) + \\ & 2\left[\mathbb{E}\big(|U_1 - U_2|U_1(1 - U_1)\big) - \mathbb{E}\big(|U_1 - U_2|\big)\mathbb{E}\big(U_1(1 - U_1)\big)\right] \\ & + \mathbb{E}\big(U_1^2(1 - U_1)^2\big) - \mathbb{E}^2\big(U_1(1 - U_1)\big)\bigg) \end{aligned} \tag{2.43}$$

$$=\frac{1}{4n^4}\left(\left[\frac{7}{60}-(\frac{1}{3})(\frac{1}{3})\right]+2\left[\frac{1}{20}-(\frac{1}{3})(\frac{1}{6})\right]+\left[\frac{1}{30}-(\frac{1}{6})(\frac{1}{6})\right]\right)$$
$$=0.$$

Equality (2.43) is zero since U_1, U_2, and U_3 are independent. Thus, $cov(\mathbb{E}[T|X_1, X_2, H_0]$, $\mathbb{E}[T|X_1, X_3, H_0]) = 0$. Similarly, $cov(\mathbb{E}[T|Z_1, Z_2, H_0], \mathbb{E}[T|Z_1, Z_3, H_0]) = 0$.

The second order projection of T is given by

$$\hat{T} = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[T|X_i, Y_j] + \sum_{i<j}\mathbb{E}[T|X_i, X_j] + \sum_{i<j}\mathbb{E}[T|Y_i, Y_j] - [n^2 + n(n-1)]\mathbb{E}T.$$

Since $\mathbb{E}T$ and the expectation of the second order projection are constants, $\mathbb{E}\hat{T} = 0$ and

$$\begin{aligned} Var\hat{T} =& \sum_{i=1}^{n}\sum_{j=1}^{n}Var(\mathbb{E}[T|X_i, Y_j]) + \sum_{i<j}Var(\mathbb{E}[T|X_i, X_j]) + \sum_{i<j}Var(\mathbb{E}[T|Y_i, Y_j]) \\ =& \frac{n^2(n-1)^2}{90n^6} + 2\times\frac{n(n-1)}{2}\times\frac{1}{90n^4} \\ =& \frac{1}{45n^2} + o(n^{-2}). \end{aligned}$$

By (2.30), in the case $m = n$,

$$\begin{aligned} VarT &= \frac{(N+1)(4(N-2))}{180n^2(2n)^2} \\ &= \frac{4(2n+1)(2n-2)}{720n^4} \\ &= \frac{8(2n+1)(n-1)}{720n^4} \\ &= \frac{1}{45n^2} + o(n^{-2}). \end{aligned} \tag{2.44}$$

Therefore, $Var\hat{T}/VarT \to 1$ as $n \to \infty$. Efron & Stein (1978) discussed a general orthogonal decomposition of a statistic. Notice that T is decomposed as $\tilde{T} + \hat{T} + R_n$, where $\tilde{T}$ is the constant $\mathbb{E}T$ and R_n is a negligible term. Hence, the limiting distribution of

T is determined by the limiting distribution of $\hat{T}$, which corresponds to the results obtained using the Hájek projection method.

Let $h(x,y) = |F(x)-F(y)|+F(x)[1-F(x)]+F(y)[1-F(y)]-2/3$. Then $h(x,y)$ is a degenerate kernel function since $h(x,y)$ is symmetric and $\mathbb{E}h(X,y) = 0$. The following shows that $\mathbb{E}h(X,y) = 0$.

$$\begin{aligned}
&\mathbb{E}_X(h(X,y)) = \mathbb{E}_X|F(X) - F(y)| + F(X)(1-F(X)) + F(y)(1-F(y)) - 2/3 \\
&= \mathbb{E}_U|U - F(y)| + \mathbb{E}U(1-U) + F(y)(1-F(y)) - 2/3 \\
&= \int_0^{F(y)} (F(y)-u)du + \int_{F(y)}^1 (u-F(y))du + \int_0^1 u(1-u)du + F(y)(1-F(y)) - 2/3 \\
&= F^2(y) - \frac{F^2(y)}{2} + \frac{1}{2} - \frac{F^2(y)}{2} - F(y) + F^2(y) + \frac{1}{2} - \frac{1}{3} + F(y) - F^2(y) - \frac{2}{3} \\
&= 0.
\end{aligned}$$

Thus, $\mathbb{E}h(X,y) = 0$. By Lemma 19 and Lemma 20, $\hat{T}$ can be written as

$$\hat{T} = \frac{n-1}{2n^3}\sum_{i=1}^n\sum_{j=1}^n h(X_i,Y_j) - \frac{1}{2n^2}\sum_{i<j} h(X_i,X_j) - \frac{1}{2n^2}\sum_{i<j} h(Y_i,Y_j).$$

Then,

$$Var[h(Z_1,Z_2)] = 2/45 \tag{2.45}$$

and $Cov(h(Z_1,Z_2),h(Z_1,Z_3)) = 0$, i.e., $h(Z_1,Z_2)$ and $h(Z_1,Z_3)$ are orthogonal, where Z_1, Z_2 and Z_3 are three different variables from X_i, Y_i, $1 \le i \le n$. In this way, $Var\hat{T}$ can also be obtained.

Now let A be an operator on $L_2(F)$ defined by

$$Ag(x) = \int_{-\infty}^{\infty} h(x,y)g(y)dF(y), \quad x \in \mathbb{R},\ g \in L_2.$$

Let $\lambda = \lambda_1, \lambda_2 \cdots$ be the real eigenvalues (not necessarily distinct) corresponding to the

solutions of the equation $Ag = \lambda g$. Let $T_n = \hat{T}/\sqrt{VarT}$.

Theorem 21 *Let $\chi_1^2, \chi_2^2 \cdots$ be independent χ^2 variables and T_n be a statistic that converges to the sum of weighted chi-square variables, i.e., $T_n \Rightarrow -\frac{\sqrt{45}}{2}\sum_{j=1}^{\infty}\lambda_j(\chi_j^2 - 1)$. Then $(T - \mathbb{E}T)/\sqrt{VarT} \Rightarrow -\frac{\sqrt{45}}{2}\sum_{j=1}^{\infty}\lambda_j(\chi_{1j}^2 - 1)$ since $(T - \mathbb{E}T - \hat{T})/\sqrt{VarT} \to 0$ in probability.*

Proof. Let $h(x, y) = \sum_{k=1}^{\infty}\lambda_k\phi_k(x)\phi_k(y)$, where $\{\phi_j(\cdot)\}$ are the orthonormal eigenfunctions corresponding to the eigenvalues $\{\lambda_j\}$, see Dunford & Schwartz (1963) and Serfling (1980). Since $\mathbb{E}h(X, y) = 0$, it follows that

$$0 = Var\{\mathbb{E}[h(X, Y)|Y]\} = \sum_{k=1}^{\infty}\lambda_k^2(\mathbb{E}\phi_k(X))^2 Var(\phi_k(Y)).$$

The λ_k^2 and $Var(\phi_k(Y)$ terms are positive, therefore, $\mathbb{E}(\phi_k(X)) = 0$ for all $k \geq 1$ and $Var[h(X, Y)] = \mathbb{E}h^2(X, Y) = \sum_{k=1}^{\infty}\lambda_k^2$. By (2.45),

$$\sum_{k=1}^{\infty}\lambda_k^2 = Var[h(X, Y)] = 2/45.$$

Define $T_{n,K}$ by

$$\begin{aligned}
(\sqrt{VarT})T_{n,K} =& \frac{n-1}{2n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K}\lambda_k\phi_k(X_i)\phi_k(Y_j) \\
&- \frac{1}{2n^2}\sum_{i<j}\sum_{k=1}^{K}\lambda_k\phi_k(X_i)\phi_k(X_j) - \frac{1}{2n^2}\sum_{i<j}\sum_{k=1}^{K}\lambda_k\phi_k(Y_i)\phi_k(Y_j) \\
=& \frac{n-1}{2n^2}\sum_{k=1}^{K}\lambda_k\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_k(X_i)\right]\left[\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\phi_k(Y_j)\right] \\
&- \frac{1}{2n}\sum_{k=1}^{K}\frac{\lambda_k}{2}\left\{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_k(X_i)\right]^2 - \frac{1}{n}\sum_{i=1}^{n}\phi_k^2(X_i)\right\} \\
&- \frac{1}{2n}\sum_{k=1}^{K}\frac{\lambda_k}{2}\left\{\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_k(Y_i)\right]^2 - \frac{1}{n}\sum_{i=1}^{n}\phi_k^2(Y_i)\right\}. \qquad (2.46)
\end{aligned}$$

From Serfling (1980), page 197, $|\mathbb{E}e^{ixT_n} - \mathbb{E}e^{ixT_{n,K}}| \leq |x|[\mathbb{E}(T_n - T_{n,K})^2]^{1/2}]$. Then

$$\mathbb{E}(T_n - T_{n,K})^2 = \left(\frac{n^2(n-1)^2}{4n^6} + \frac{n(n-1)}{4n^4}\right)\frac{1}{VarT}\sum_{k=K+1}^{\infty}\lambda_k^2$$
$$= [\frac{45}{2} + o(1)]\sum_{k=K+1}^{\infty}\lambda_k^2 \leq 23\sum_{k=K+1}^{\infty}\lambda_k^2.$$

For a given $\epsilon > 0$ and any $x \in \mathbb{R}$, there exists a $K \in \mathbb{N}$ such that

$$|x|(23\sum_{k=K+1}^{\infty}\lambda_k^2)^{1/2} < \epsilon,$$

because $\sum_{k=1}^{\infty}\lambda_k^2$ converges. Thus, for all x$\in \mathbb{R}$ and $\epsilon > 0$, there exists a $K \in \mathbb{R}$ such that $|\mathbb{E}e^{ixT_n} - \mathbb{E}e^{ixT_{n,K}}| < \epsilon$ for all n. Let $W_{n,k}(X) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_k(X_i)$, and $Z_{n,k}(X) = \frac{1}{n}\sum_{i=1}^{n}\phi_k^2(X_i)$. By (2.46) and (2.44), $T_{n,K}$ can be written as

$$T_{n,K} = \frac{\sqrt{45}}{2}\sum_{k=1}^{K}\lambda_k(W_{Xnk}W_{Ynk} - W_{Xnk}^2/2 - W_{Ynk}^2/2 + Z_{Xnk}/2 + Z_{Ynk}/2) + R_n$$
$$= \frac{\sqrt{45}}{2}\sum_{k=1}^{K}\lambda_k(-(W_{Xnk} - W_{Ynk})^2/2 + Z_{Xnk}/2 + Z_{Ynk}/2) + R_n.$$

where $R_n \to 0$. The term R_n is negligible as it comes from $VarT$ in (2.44). When (2.46) is divided by $\sqrt{VarT}$, this gives $T_{n,K}$. Let $\chi_1^2, \cdots, \chi_k^2$ be i.i.d. χ^2 random variables. Denote

$$Y_K = \frac{\sqrt{45}}{2}\sum_{k=1}^{K}\lambda_k(-\chi_{1k}^2 + 1).$$

Since $\mathbb{E}W_{Xnk} = \mathbb{E}W_{Ynk} = 0$, and W_{Xnk} and W_{Ynk} are orthonormal, $(W_{Xnk} - W_{Ynk})/\sqrt{2} \Rightarrow N(0, I_{K\times K})$ by the Linderberg-Levý central limit theorem, where $Z_{Xnk} \to 1$ and $Z_{Ynk} \to 1$ for $1 \leq k \leq K$ by the strong law of large number. Hence $T_{nK} \Rightarrow Y_K$ as $n \to \infty$ and $\mathbb{E}(e^{ixT_{nK}}) - \mathbb{E}(e^{ixY_K})| < \epsilon$ for all n. By the same argument as in Serfling (1980), $\mathbb{E}(e^{ixY_K}) -$

$\mathbb{E}(e^{ixY})| < \epsilon$ for all n. Then together with $|\mathbb{E}e^{ixT_n} - \mathbb{E}e^{ixT_{nK}}| < \epsilon$, $|\mathbb{E}e^{ixT_n} - \mathbb{E}e^{ixY}| \to 0$ as $n \to \infty$. Therefore, $T_n \Rightarrow -\frac{\sqrt{45}}{2}\sum_{j=1}^{\infty}\lambda_j(\chi^2_{1j} - 1)$. Since $Var\hat{T}/VarT \to 1$, it follows that $(T - \mathbb{E}T - \hat{T})/\sqrt{VarT} \to 0$ in probability. Then by Lemma 15,

$$(T - \mathbb{E}T)/\sqrt{VarT} \Rightarrow -\frac{\sqrt{45}}{2}\sum_{j=1}^{\infty}\lambda_j(\chi^2_{1j} - 1). \tag{2.47}$$

■

This limiting distribution agrees with the one obtained via stochastic process methods in Rosenblatt (1952) and Fisz (1960). The projection approach applied here is typically useful in U-statistic theory, but it shall be emphasized that neither T nor $\hat{T}$ is U-statistic.

Although the limiting distribution of T and T_1 is known, it is rarely useful in practice. There are two reasons. One is that F is usually not known, hence $h(x, y)$ is unknown. Thus, the eigenvalues are not available. The second reason is that even if F is known, the computation of λ is extremely difficult. So, for moderate or large sample sizes, a random permutation (bootstrap) to obtain the critical values of T and T_1 is used. The procedure is described below.

1. Generate a random sample (from a specific distribution) without (with) replacement from $1 : N$
2. The first m values of the sample are natural ranks of X and the remaining n are natural ranks of Y
3. Compute the test statistic T or T_1
4. Repeat steps 1-3 M times to obtain M values T or T_1
5. $(1 - \alpha)$ quantile of T or T_1 is the $100\alpha\%$ critical value.

Sampling without replacement corresponds to the permutation procedure and sampling with replacement is the bootstrap procedure. In the next section, we will investigate the

power performance of the test T_1 (or, similarly T when $m = n$) using those two methods and compare with other tests.

2.5 Simulation

This section consists of the discussion of results that were obtained in investigating the power performance of the proposed test and its comparison with other tests. The traditional bootstrap procedure and the permutation procedure were used to study the performances of the tests in the one-dimensional case. The bootstrap procedure was used to compare the performances of the Kolmogorov-Smirnov test, Wilcoxon rank sum test, Baringhaus and Franz's Cramér test, Fernándes *et al.* test, and the proposed rank-based test. The permutation procedure was used only for Baringhaus and Franz's Cramér test, Fernándes *et al.* test, and the proposed spatial rank test. Refer to the Kolmogorov-Smirnov test as KS, Wilcoxon rank sum test as W, Baringhaus and Franz's Cramér test as CT, Fernándes *et al.* test as DT, and the proposed test as RCT, i.e. T_1. The notations CTP and CTB are used to denote the CT test under the permutation method and the bootstrap method, respectively. The same holds for the DTP and DTB tests, and the proposed tests $RCTP$ and $RCTB$.

2.5.1 Simulation Results for Location Alternatives

First, is the discussion of the results in the case of location alternatives. For the case $d = 1$, two independent samples were generated with equal sizes ($n = m = 35$ and $n = m = 50$) and unequal sizes ($n = 50$ and $m = 20$) from the normal distribution, t_3-distribution, t_1-distribution, exponential distribution, Pareto distribution, and Poisson distribution at the chosen significance level $\alpha = 0.05$. For each distribution, 1000 iterations were computed for both the bootstrap and permutation methods to determine the estimated powers by calculating the fraction of p-values less than or equal to 0.05. Since the results for the bootstrap method behave quite similar to the permutation method, only the results

obtained from the bootstrap method are reported. However, the results of both methods are displayed in each figure below for all considered distributions. Also, each test corresponds similarly to each other under each distribution when the samples are generated with equal sizes of either $n = m = 35$ or $n = m = 50$. The power peformance tends to be stronger when the samples are of size $n = m = 50$, where the difference in performance can be as large as 10%. For simplicity, only the samples of size $n = m = 50$ will be discussed, however, the results for size $n = m = 35$ are displayed in Figures 2.13-2.18 for the Normal, t and Pareto distributions.

For the normal distribution, $X_1,\ldots,X_n \sim$ N(0, 1) and $Y_1,\ldots,Y_m \sim$ N(Δ, 1) are generated, with $\Delta = 0, \ldots, 1$ in steps of 0.10. Figure 2.3 shows the power performance for each test under the normal distribution. For equal sample sizes, note that the statistical power of the $RCTB$ test compares favorably to the W test and the CTB test. The statistical power of the $RCTB$ test is higher than that of the DTB test, and is significantly higher than that of the KS test where the difference in powers can be a large as 15%. Similarly, for unequal sample sizes, the statistical power of the $RCTB$ test is comparable to the W test and the CTB test, while it is significantly higher than that of the KS test and the DTB test.

The experiment is repeated for the t-distribution to studied the performance when $df = 3$ and when $df = 1$, where df denotes the degrees of freedom. Then $X_1,\ldots,X_n \sim$ t(0, 1) and $Y_1,\ldots,Y_m \sim$ t(Δ, 1) were generated, with $\Delta = 0, \ldots, 1$ in steps of 0.10. Looking at Figure 2.5 we see the power performance for each test under the t_3-distribution. Note that the statistical power of the $RCTB$ test outperforms each test for all considered alternatives for both equal and unequal sample sizes. In the case of the t_1-distribution, for equal sample sizes, Figure 2.7 indicates that the statistical power of the $RCTB$ test is comparable to the KS test and is slightly higher than that of the DTB test for alternatives between 0.45 and 0.80. For unequal sample sizes, the $RCTB$ test is comparable to the DTB test for alternatives between 0 and 0.7, while the DTB test is slightly higher than that of the $RCTB$ test for alternatives between 0.7 and 1. Also, for unequal sample sizes, the $RCTB$ test is

slightly higher than the KS test for all alternatives considered. For both equal and unequal sample sizes, the $RCTB$ test outperforms the CTB test for all considered alternatives. (Note that the CTP test performs better than the CTB test for both equal and unequal sample sizes in Figure 2.5.

To study the power for the exponential distribution, simulations for exponential variates $X_1,\ldots,X_n \sim$ E(1) and $Y_1,\ldots,Y_m \sim$ E(Δ) are conducted, with $\Delta = 0.5,\ldots,1.5$ in steps of 0.10. Figure 2.9 displays the difference in the power performance for each test under exponential distribution in which each test remains consistent in relation to each other for both equal and unequal sample sizes. The CTB test outperforms all considered tests, where the KS test has the weakest performance. Notice that the $RCTB$ test is comparable to the W test and outperforms the DTB test for alternatives between 0.5 and 0.85.

In the case of Poisson samples $X_1,\ldots,X_n \sim$ P(2) and $Y_1,\ldots,Y_m \sim$ P(Δ), with $\Delta = 1,\ldots,3$ in steps of 0.10, for equal sample sizes, Figure 2.10 displays the obtained results of the statistical power for each test. The power performance of the $RCTB$ test is comparable to the W test and the CTB test, and has a higher power performance than that of the DTB test. However, the power performance of the $RCTB$ test is significantly higher than the KS test where the difference in powers can be as large as 35%. Similarly, for unequal sample sizes, the same results hold as the $RCTB$ test compares fairly well to the W test and the CTB test and outperforms the DTB test. The difference in the power performances for the $RCTB$ test and the KS test can be as large as 37%.

Shown in Figure 2.11 is the power performance for the Pareto distribution, where $X_1,\ldots,X_n \sim$ Pa(2, 2) and $Y_1,\ldots,Y_m \sim$ Pa(2+Δ, 2) are generated, with $\Delta = 0,\ldots,1$ in steps of 0.10. The $RCTB$ test is comparable to the KS test and outperforms all other considered tests. The results are the same for both equal and unequal sample sizes. The power difference between the $RCTB$ test and that of the CTB test can be as large as 43% for equal sample sizes and can be as large as 40% for unequal sample sizes.

2.5.2 Simulation Results for Scale Alternatives

Here is the discussion of the results found in the case of scale alternatives, in which similarly procedures were performed to that which was conducted in the case of location alternatives. Two independent samples were generated for both equal ($n = m = 35$ and $n = m = 50$) and unequal sizes ($n = 50$ and $m = 20$) from a normal distribution, t_3-distribution, t_1-distribution, and Pareto distribution all for the case $d = 1$. All considered tests for scale alternatives as in the experiment for location alternatives were used with the exception of the Wilcoxon (W) test, as this test is a test for location. Instead, the scale test known as Mood's test is used, which is refer to as the M test. Each test corresponds similarly to each other under each distribution when the samples are generated with equal sizes of either $n = m = 35$ or $n = m = 50$. The power peformance tends to be stronger when the samples are of size $n = m = 50$, where the difference in performance can be as large as 20%. Here, only the samples of size $n = m = 50$ will be discussed, however, the results for size $n = m = 35$ are displayed in Figures 2.13-2.18 for the Normal, t and Pareto distributions.

In the case of the normal distribution, $X_1,\ldots,X_n \sim$ N(0, 1) and $Y_1,\ldots,Y_m \sim$ N(0, Δ) were generated, where $\Delta = 1,\ldots,3$ in steps of 0.10. Figure 2.4 displays the results obtained and in this case, for equal sample sizes the $RCTB$ test does not compare favorably to all considered tests other than the KS test, in which it peforms significantly better than the KS test, while the M test outperforms all tests. The same results hold for unequal sample sizes, however in this case, the $RCTB$ test performs similar to that of the KS test.

For the t_3-distribution, $X_1,\ldots,X_n \sim t_3(0, 1)$ and $Y_1,\ldots,Y_m \sim t_3(0, \Delta)$ were generated, with $\Delta = 1,\ldots,3$ in steps of 0.10. Also, similar samples for the t_1-distribution were generated. The results obtained for the t_3-distribution are displayed in Figure 2.6 with similar results which are displayed in Figure 2.4 for the normal distribution (for both equal and unequal sample sizes). In the case of the t_1-distribution, in Figure 2.8, the $RCTB$ test is not as powerful as the other considered tests for equal sample sizes. For unequal sample

sizes, the $RCTB$ test is comparable to the KS test and outperforms the CTB test (only for the bootstrap method) for all considered alternatives, however, it is not as powerful as the remaining tests. It is necessary to note that the permutation method of the CT test performs better than the bootstrap method of the CT test. The M test remains the most powerful of all tests for equal and unequal sample sizes.

For the scale alternatives, the proposed test does not compare as favorably to its non-parametric competitors as this is seen in the case of location alternatives for the normal distribution, t_3-distribution and t_1-distribution. However, in the case of the Pareto distribution, the proposed test is more favorable in the case of unequal sample sizes. First, for the equal sample sizes, looking at Figure 2.12 we see for Pareto samples $X_1,\ldots,X_n \sim$ Pa(2, 2) and $Y_1,\ldots,Y_m \sim$ Pa(2, 2Δ), with $\Delta = 1, \ldots, 3$ in steps of 0.10, the power performance of the CTB test is higher than the $RCTB$ test. But the $RCTB$ test is slightly higher than the DTB test and it outperforms the KS test. One recognizes that all considered tests outperform the M test by a great margin. In the case of unequal sample sizes, the $RCTB$ test outperforms all considered tests and the difference in the power performance compared to that of the CTB test can be as large as 14%.

2.5.3 Simulation Results for Local Power

The local power performance of each test is displayed in Tables 2.3 - 2.5 for the Normal distribution, t-distribution and Pareto distribution for the chosen significance level $\alpha = 0.05$. The local power is analyzed as an extension of the location altervative, where the local power is based on sequences such that the generated data approaches the null hypothesis and not necessarily satisfying the alternative hypothesis. Let $\Delta = \delta/\sqrt{n}$. Then the X-sample is from distribution $F(x)$ and the Y-sample is from distribution $G(x) = F(x + \Delta) = F(x + \frac{\delta}{\sqrt{n}})$, with $\delta = 1$ and 3.5, and with a sequence of $n = 2^4, 2^5, 2^6, 2^7$ and 2^8. Tables 2.3 - 2.5 shows that the local power decreases as n increases and as the change in $\frac{\delta}{\sqrt{n}}$ approaches zero. The proposed test $RCTB$ outperfoms all other tests in the case of

the t-distribution when $\delta = 3.5$. It is worth noting that the local power of the KS test is < 0.05 for $n > 2^5$ for the Normal distribution when $\delta = 1$. This implies that the KS has low performance and hence will fail in this case.

Table 2.3. Local Power for Normal Distribution (Top: $\delta = 1$, Bottom: $\delta = 3.5$)

$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	0.072	0.049	0.040	0.042	0.036
	W	0.125	0.076	0.067	0.056	0.058
	CTP	0.117	0.076	0.058	0.062	0.061
	DTP	0.111	0.076	0.062	0.048	0.058
	RCTP	0.122	0.076	0.063	0.060	0.054
	CTB	0.118	0.077	0.056	0.062	0.059
	DTB	0.108	0.071	0.066	0.053	0.054
	RCTB	0.122	0.078	0.064	0.066	0.062

$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	0.946	0.707	0.418	0.252	0.140
	W	0.977	0.837	0.560	0.329	0.193
	CTP	0.979	0.831	0.557	0.334	0.197
	DTP	0.964	0.75	0.482	0.284	0.171
	RCTP	0.974	0.817	0.537	0.330	0.188
	CTB	0.977	0.833	0.556	0.337	0.195
	DTB	0.964	0.747	0.485	0.287	0.171
	RCTB	0.974	0.811	0.542	0.332	0.196

Table 2.4. Local Power for t Distribution (Top: $\delta = 1$, Bottom: $\delta = 3.5$)

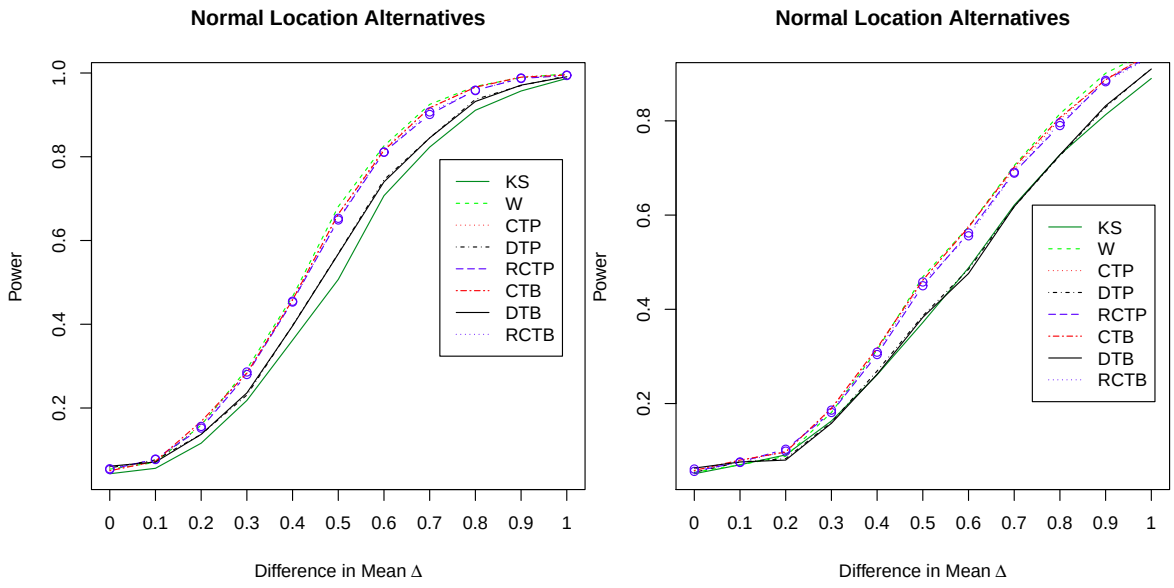
$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	0.111	0.098	0.054	0.042	0.043
	W	0.146	0.113	0.073	0.069	0.059
	CTP	0.149	0.104	0.074	0.072	0.060
	DTP	0.138	0.108	0.073	0.055	0.062
	RCTP	0.155	0.127	0.073	0.065	0.051
	CTB	0.142	0.107	0.065	0.067	0.057
	DTB	0.137	0.106	0.074	0.055	0.062
	RCTB	0.149	0.124	0.076	0.068	0.055

$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	0.891	0.603	0.341	0.207	0.099
	W	0.919	0.657	0.397	0.243	0.117
	CTP	0.92	0.641	0.373	0.246	0.113
	DTP	0.889	0.605	0.359	0.214	0.121
	RCTP	0.930	0.682	0.422	0.257	0.127
	CTB	0.913	0.626	0.368	0.235	0.113
	DTB	0.895	0.608	0.367	0.212	0.118
	RCTB	0.933	0.680	0.420	0.258	0.124

Table 2.5. Local Power for Pareto Distribution (Top: $\delta = 1$, Bottom: $\delta = 3.5$)

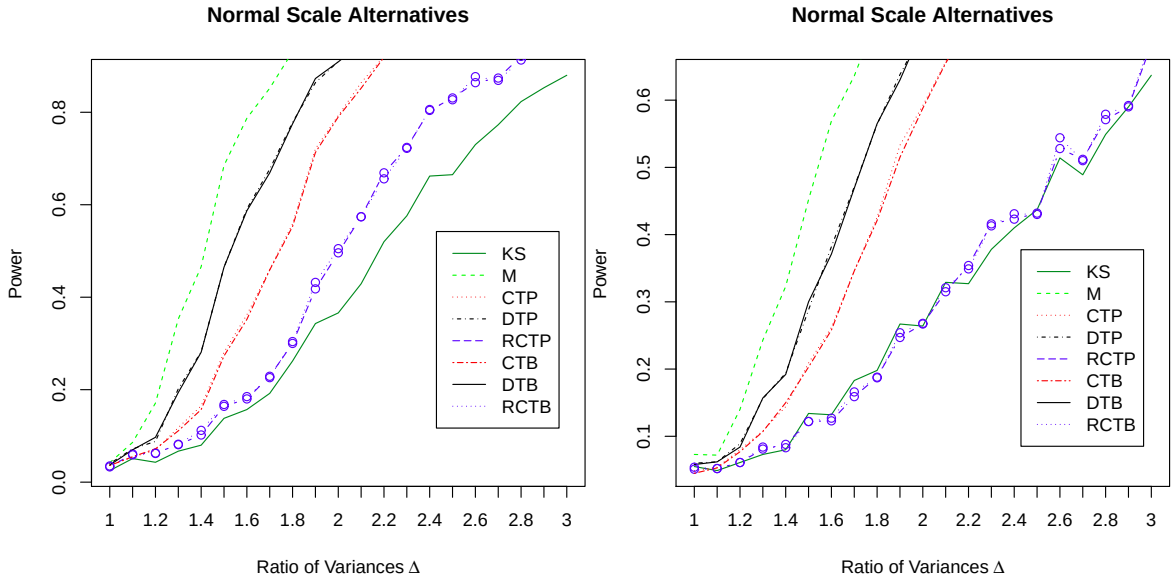
$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	0.402	0.194	0.110	0.074	0.045
	W	0.427	0.251	0.153	0.133	0.074
	CTP	0.192	0.116	0.074	0.082	0.056
	DTP	0.284	0.158	0.093	0.094	0.059
	RCTP	0.465	0.259	0.148	0.125	0.067
	CTB	0.164	0.088	0.056	0.074	0.045
	DTB	0.280	0.159	0.089	0.094	0.060
	RCTB	0.465	0.261	0.152	0.122	0.068

$\Delta = \delta/\sqrt{n}$	Test	$n = 2^4$	$n = 2^5$	$n = 2^6$	$n = 2^7$	$n = 2^8$
	KS	1	0.992	0.922	0.619	0.303
	W	0.999	0.956	0.821	0.600	0.340
	CTP	0.981	0.835	0.522	0.309	0.150
	DTP	1	0.953	0.720	0.442	0.239
	RCTP	1	0.981	0.888	0.647	0.356
	CTB	0.957	0.769	0.458	0.246	0.125
	DTB	1	0.950	0.723	0.429	0.229
	RCTB	1	0.982	0.891	0.647	0.356



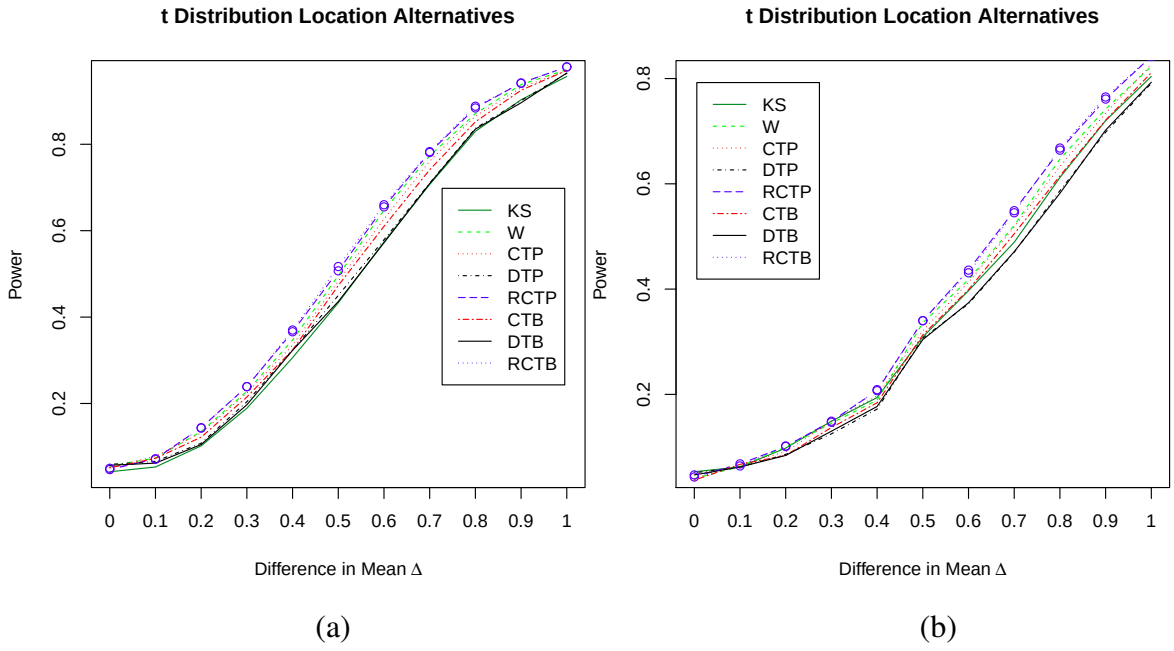
(a) (b)

Figure 2.3. Power performance for Normal distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



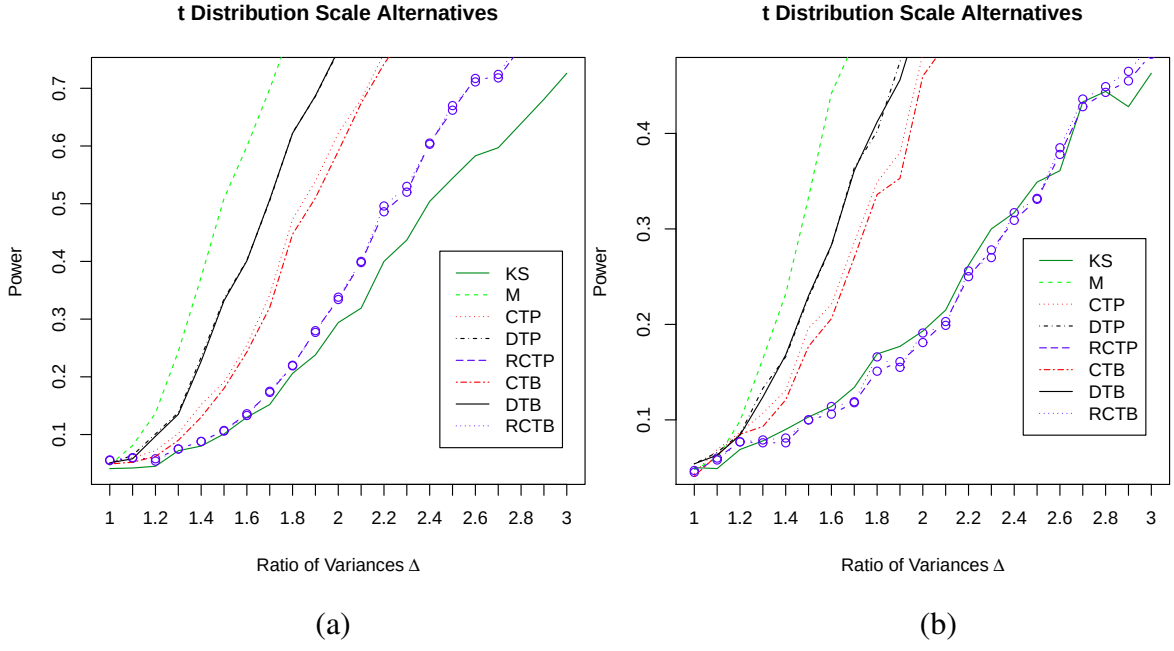
(a) (b)

Figure 2.4. Power performance for Normal distribution scale alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



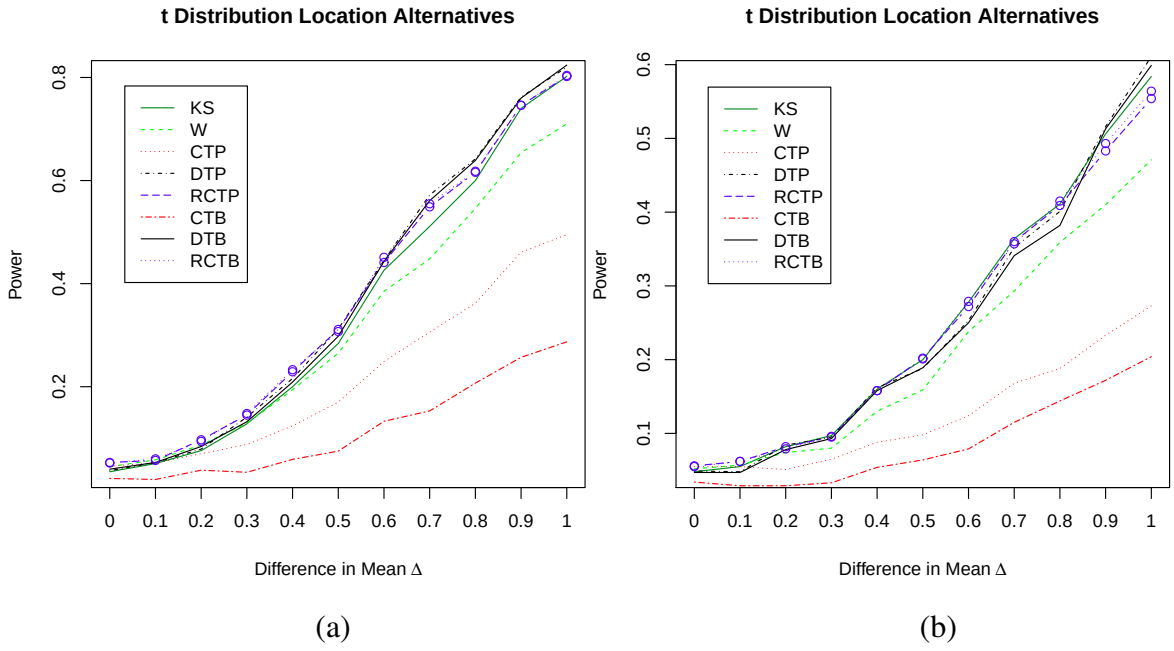
(a) (b)

Figure 2.5. Power performance for t-distribution location alternatives, $df = 3$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



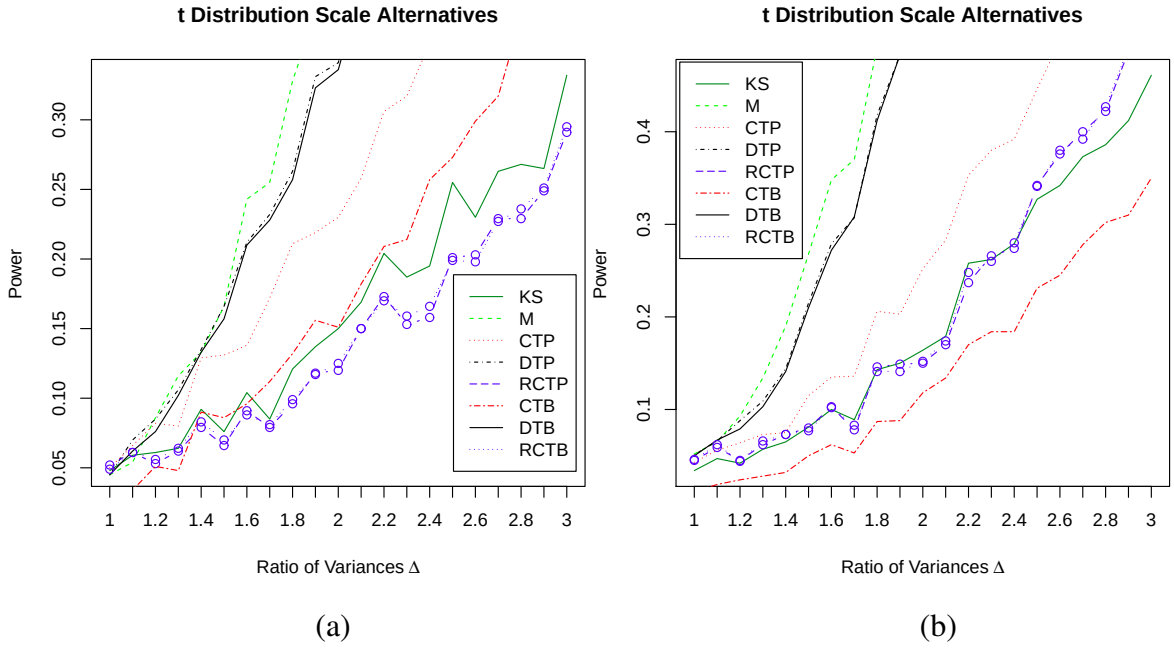
(a) (b)

Figure 2.6. Power performance for t-distribution scale alternatives, $df = 3$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



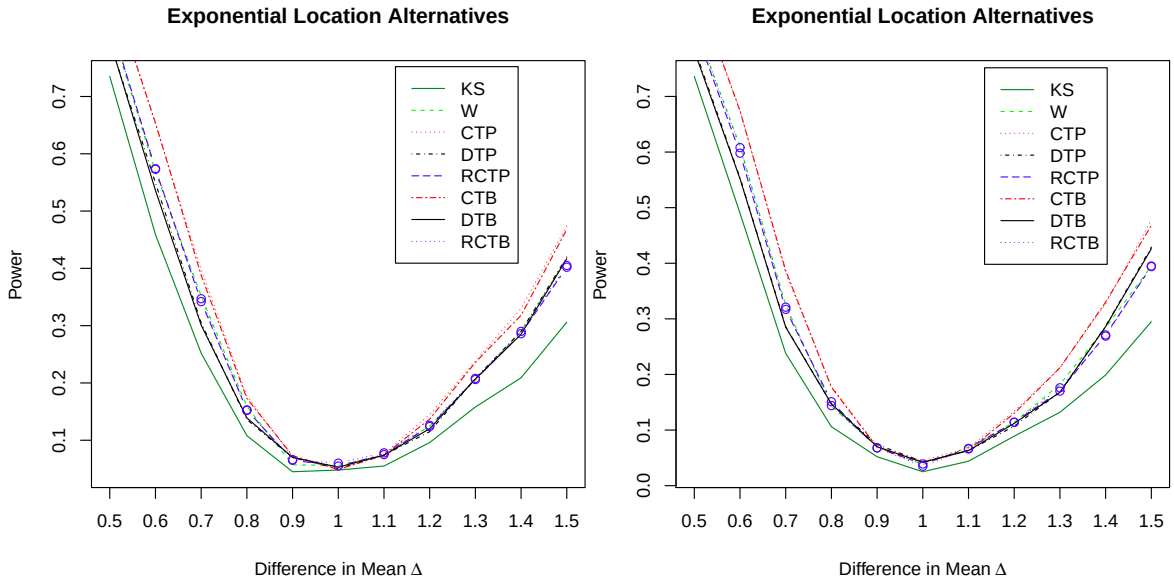
(a) (b)

Figure 2.7. Power performance for t-distribution location alternatives, $df = 1$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



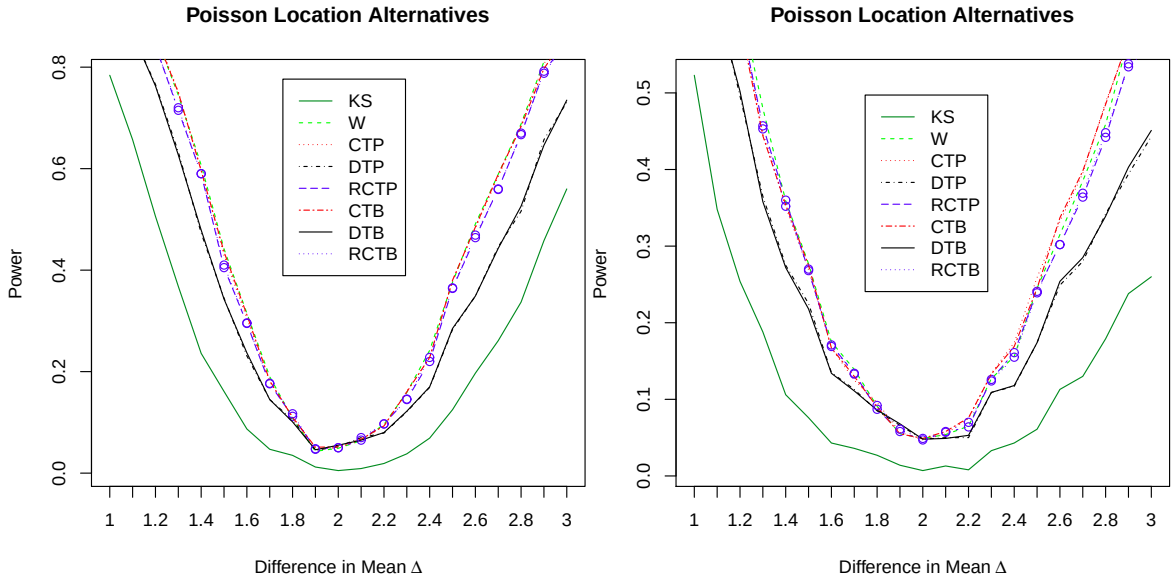
(a) (b)

Figure 2.8. Power performance for t-distribution scale alternatives, $df = 1$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



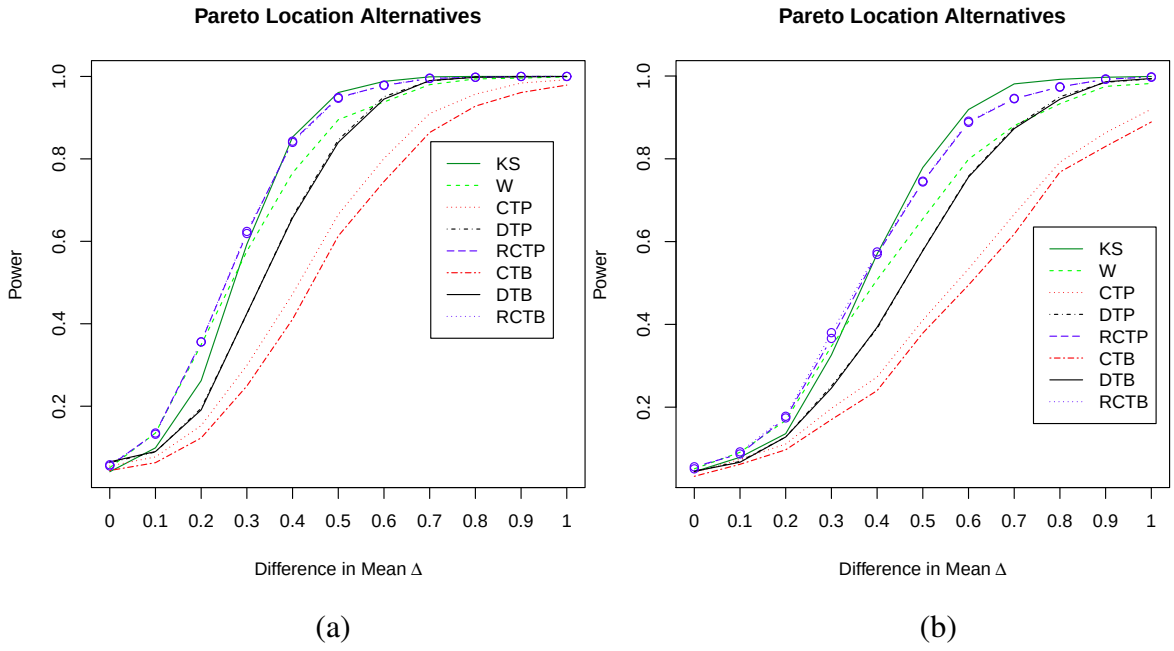
(a) (b)

Figure 2.9. Power performance for Exponential distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



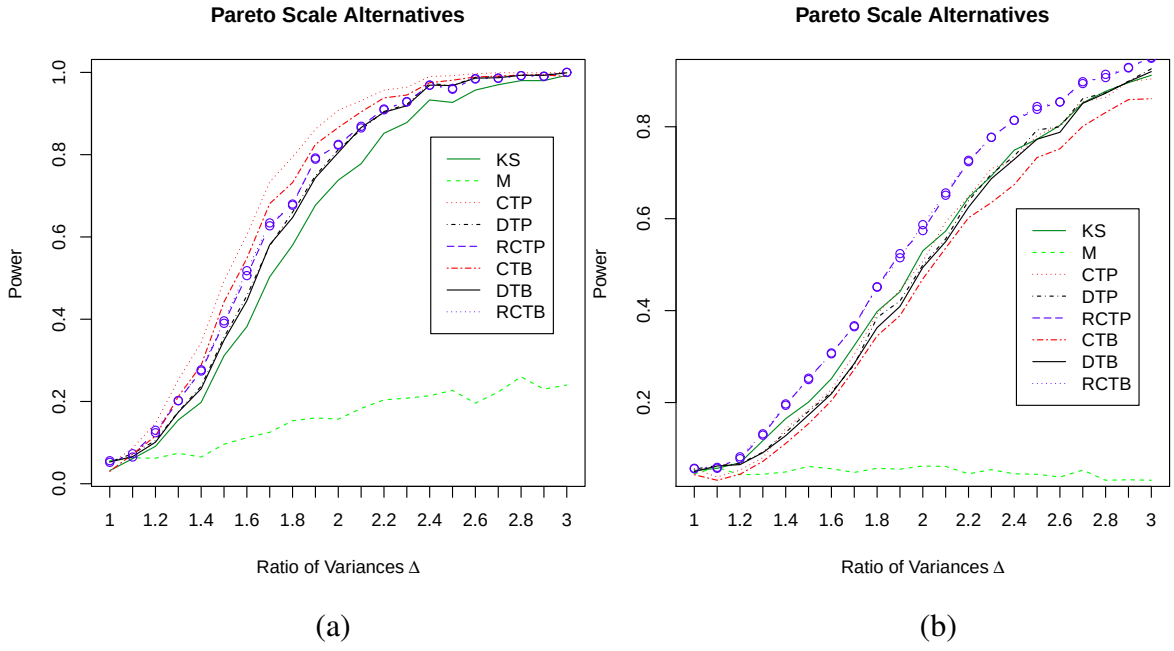
(a) (b)

Figure 2.10. Power performance for Poisson distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



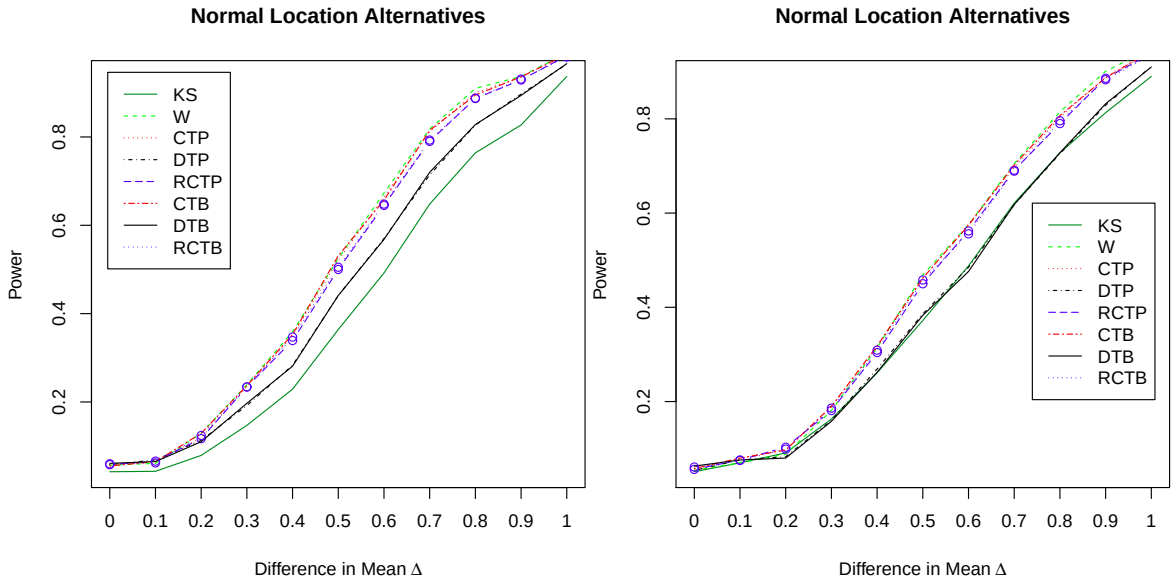
(a) (b)

Figure 2.11. Power performance for Pareto distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



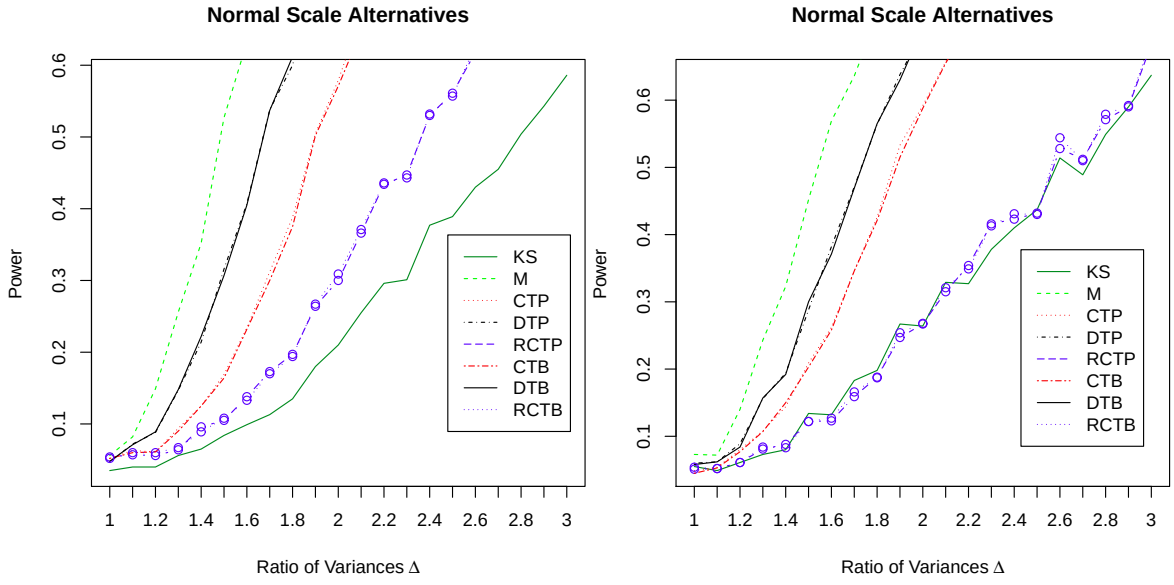
(a) (b)

Figure 2.12. Power performance for Pareto distribution scale alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



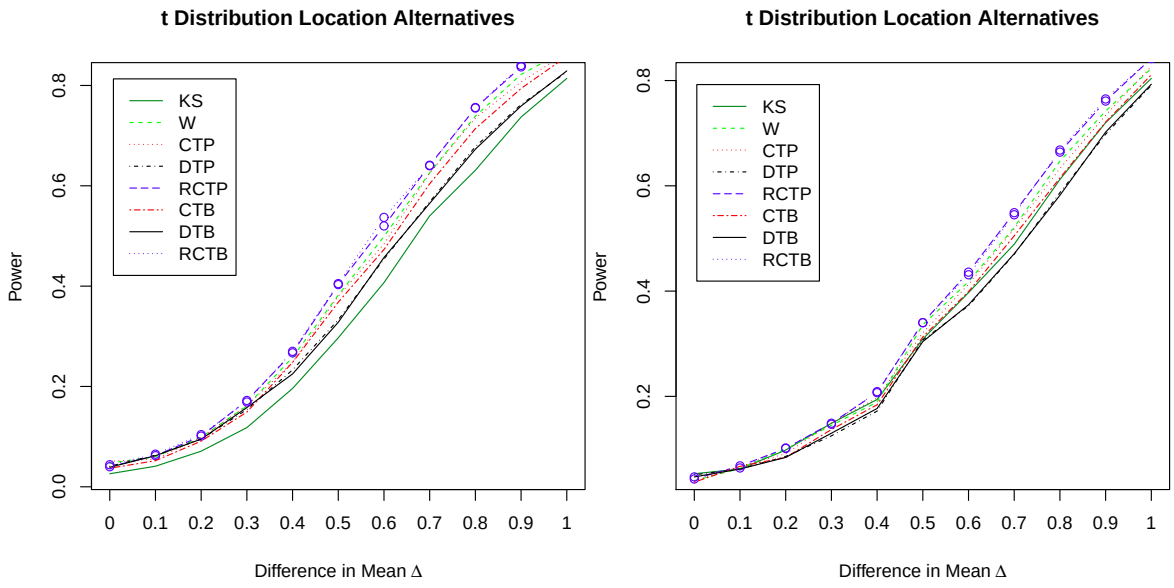
(a) (b)

Figure 2.13. Power performance for Normal distribution location alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



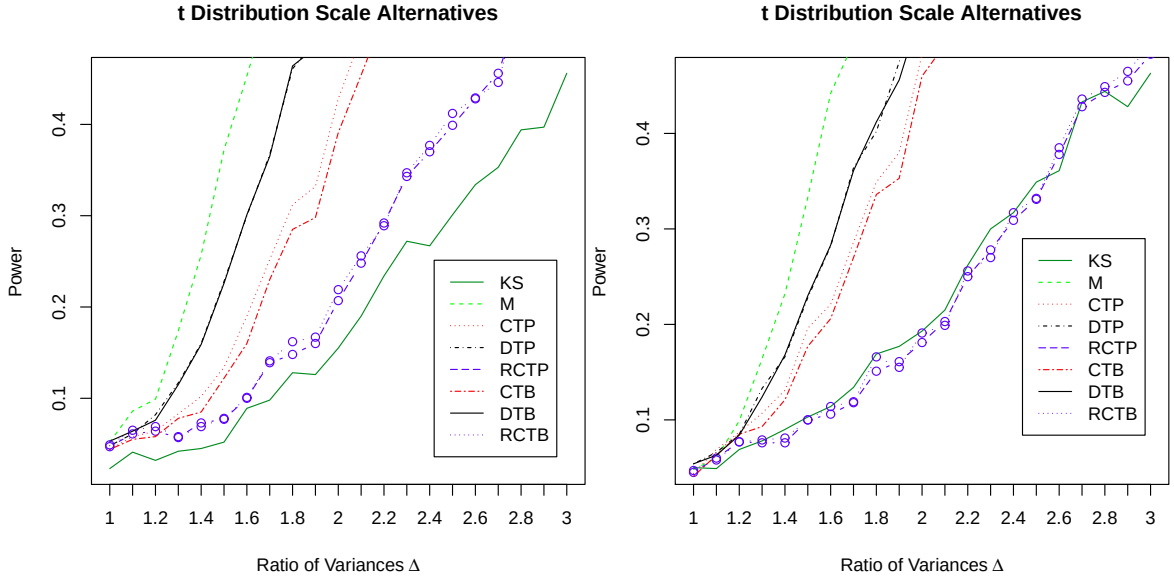
(a) (b)

Figure 2.14. Power performance for Normal distribution scale alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



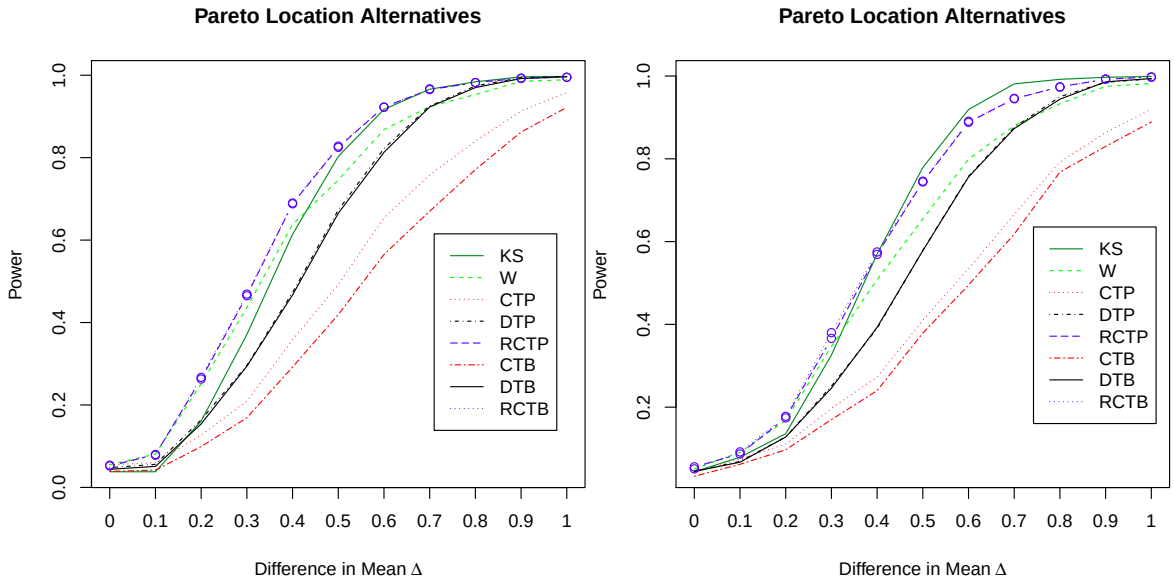
(a) (b)

Figure 2.15. Power performance for t distribution location alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



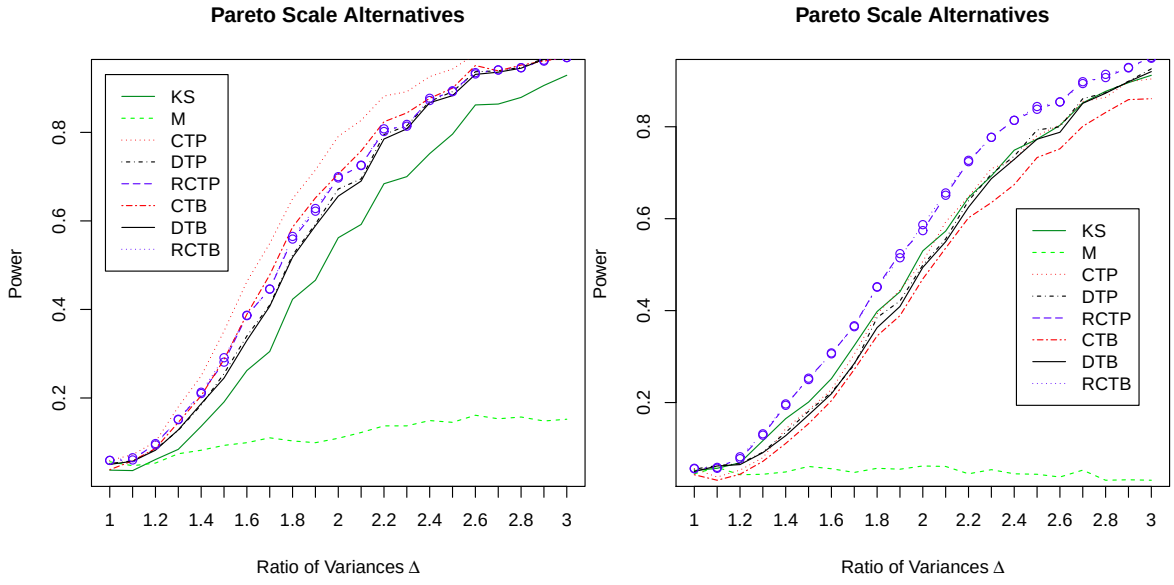
(a) (b)

Figure 2.16. Power performance for t distribution scale alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



(a) (b)

Figure 2.17. Power performance for Pareto distribution location alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



(a) (b)

Figure 2.18. Power performance for Pareto distribution scale alternatives. (a)**: equal sample size,** $n = m = 35$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**

2.6 Summary and Discussion

The nonparametric Cramér-von Mises test is revisited with a totally different rank-based approach, providing a different perspective and insight. Two formulations have been studied. Their properties have been derived. The limiting null distribution was explored through techniques of Hájek projection and orthogonal decomposition. For the statistic T under the balanced case, the limiting distribution is not normal since the projection on one variable is insufficient to represent the variation of the test statistic. By taking the projection on two variables, the limiting distribution was proved to be a weighted mixture of independent chi-square distributions. An operator in the functional space was defined and its eigenfunctions and eigenvalues were used to derive the limiting distribution. Also, statistic T is equivalent to statistic T_1 for the balanced case, as in this case, T is written as a linear transformation of T_1.

Rank-based formulations allow generalizations of two-sample Cramér-von Mises test to the multivariate case straightforward by using different notions of multivariate rank functions. In the multivariate case, the rank tests may lose the distribution-free property under a general alternative. They are, however, usually more robust than the parametric tests. The next chapter will explore these generalizations and make suggestions for different notions of rank functions.

CHAPTER 3

NEW MULTIVARIATE RANK TESTS

3.1 Introduction

A generalization of rank-based tests to the multivariate case requires a multivariate rank concept. In the univariate case, there is a linear ordering and hence the definition of rank is natural and unquestionable. However, in high dimensional space, such natural order no longer exists, which makes rank conceptually difficult. To compensate for the lack of the linear ordering, the ordering is now oriented to a center and center-outward ordering.

In this chapter, popular multivariate rank notions are introduced followed by discussion of properties of each rank function and explanation of the reason why spatial rank is used for the generalization. Then the properties of the rank test are explored. Unlike in the univariate case, the proposed rank test is no longer distribution-free, although it is nonparametric. Bootstrap and permutation techniques are used for determining critical values, and the connection with other tests is discussed. Finally, an extensive simulation study on power comparison with those tests is conducted under various scenarios.

3.2 Multivariate Rank Functions

3.2.1 Marginal Sign and Rank

The marginal sign function in high dimensions can be componentwisely generalized from the univariate case. In the case $x \in \mathbb{R}$, the sign function $sign(x)$ takes value $1, 0$ or

-1 as $x > 0$, $x = 0$ or $x < 0$. The (sample) sign and rank functions associated with a random sample $\{x_1, ..., x_n\}$ are defined by

$$S(x, F_n) = sign(x - Med(x_1, ..., x_n)),$$

and

$$R(x, F_n) = ave\{S(x - x_i)\} = \frac{1}{n}\sum_{i=1}^{n} sign(x - x_i).$$

Note that $R(x)$ is in fact the derivative of criterion function $ave\{|x - x_i|\}$. The notation ave is the average taking on the index i. Here, it is equivalent to $\frac{1}{n}\sum_{i=1}^{n}$.

For the d-variate data set $\mathbb{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$, where $\boldsymbol{x}_i = [x_{i1}, ..., x_{id}]^T \in \mathbb{R}^d$, consider the objective functions,

$$H_1(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = |x_1| + , ..., + |x_d|,$$

and

$$D_1(\boldsymbol{x}) = ave\{\|\boldsymbol{x} - \boldsymbol{x_i}\|_1\}.$$

One can define the marginal sign function $\boldsymbol{S}_1(\boldsymbol{x})$ and marginal rank function $\boldsymbol{R}_1(\boldsymbol{x})$ as the gradient of the above objective functions,

$$\boldsymbol{S}_1(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} H_1(\boldsymbol{x}),$$

$$\boldsymbol{R}_1(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} D_1(\boldsymbol{x}).$$

Thus $\boldsymbol{S}_1(\boldsymbol{x}) = [sign(x_1), ..., sign(x_d)]^T$. The vector of the marginal rank of $\boldsymbol{x}$ is $\boldsymbol{R}_1(\boldsymbol{x}) = ave\{\boldsymbol{S}_1(\boldsymbol{x} - \boldsymbol{x}_i)\}$. The marginal sign of $\boldsymbol{x}$ is $\boldsymbol{S}_1(\boldsymbol{x} - \boldsymbol{M}_1(\mathbb{X}))$, where $\boldsymbol{M}_1(\mathbb{X})$ is the marginal median (also called the component-wise median) which minimizes the criterion function

$D_1(\boldsymbol{x})$. The marginal median also satisfies the equality

$$\boldsymbol{R}_1\left(\boldsymbol{M}_1(\mathbb{X})\right) = \mathbf{0}.$$

Using the marginal rank leads to a component-wise approach. As discussed in Chapter 1, a component-wise approach is not appealing since it completely ignores correlation information between variables.

3.2.2 Oja Sign and Oja Rank

The volume of d-variate simplex determined by $\boldsymbol{x}$ and d observations with indices $i_1 < i_2 < ... < i_d$ is

$$\frac{1}{d!} abs\left[det\begin{pmatrix} 1 & ... & 1 & 1 \\ \boldsymbol{x}_{i_1} & ... & \boldsymbol{x}_{i_d} & \boldsymbol{x} \end{pmatrix}\right].$$

Consider the objective functions,

$$H_2(\boldsymbol{x}) = ave\left[abs\left[det\begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ \mathbf{0} & \boldsymbol{x}_{i_1} & ... & \boldsymbol{x}_{i_{d-1}} & \boldsymbol{x} \end{pmatrix}\right]\right],$$

and

$$D_2(\boldsymbol{x}) = ave\left[abs\left[det\begin{pmatrix} 1 & ... & 1 & 1 \\ \boldsymbol{x}_{i_1} & ... & \boldsymbol{x}_{i_d} & \boldsymbol{x} \end{pmatrix}\right]\right].$$

The Oja sign and rank functions, $\boldsymbol{S}_2(\boldsymbol{x})$ and $\boldsymbol{R}_2(\boldsymbol{x})$, are defined as the gradient functions as follows,

$$\boldsymbol{S}_2(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} H_2(\boldsymbol{x}),$$

$$\boldsymbol{R}_2(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} D_2(\boldsymbol{x}).$$

The solution of $\boldsymbol{R}_2\left(\boldsymbol{M}_2(\mathbb{X})\right) = \mathbf{0}$ is called Oja median. Oja rank takes information of multivariate data and hence it leads to procedures that have nice properties such as affine

equivariance and high efficiency in heavy-tailed distributions. However, its computation expense $O(n^d)$ is prohibitive. Oja rank is not feasible in applications to data of large sample size in high dimension.

3.2.3 Spatial Sign and Spatial Rank

Different from the simple way as the marginal sign function taking the component-wise sign or the complex way as the Oja sign based on the notion of simplex, consider the following two objective functions,

$$H(\boldsymbol{x}) = \|\boldsymbol{x}\| = \sqrt{x_1^2 + ... + x_d^2},$$

and

$$D(\boldsymbol{x}) = ave\{\|\boldsymbol{x} - \boldsymbol{x}_i\|\}.$$

The spatial sign function and the spatial rank function are defined as the gradient of them,

$$\boldsymbol{S}(x) = \nabla_{\boldsymbol{x}} H(\boldsymbol{x}),$$

$$\boldsymbol{R}(x) = \nabla_{\boldsymbol{x}} D(\boldsymbol{x}).$$

So the spatial sign function is given by $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\|$ ($\boldsymbol{S}(\boldsymbol{0}) = \boldsymbol{0}$). In fact, the spatial sign can be viewed as the unit vector in the direction of $\boldsymbol{x}$. The spatial sign of $\boldsymbol{x}$ with respect to (w.r.t.) a random sample $\mathbb{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_n\}$ is $\boldsymbol{S}(\boldsymbol{x}, F_n) = \boldsymbol{S}(\boldsymbol{x} - M(\mathbb{X}))$, where $M(\mathbb{X})$ is the spatial median. The (sample) spatial rank is thus derived accordingly:

$$\boldsymbol{R}(\boldsymbol{x}, F_n) = ave\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{x}_i)\} = \frac{1}{n}\sum_{i=1}^{n} \frac{\boldsymbol{x} - \boldsymbol{x}_i}{\|\boldsymbol{x} - \boldsymbol{x}_i\|}.$$

The population version of the spatial rank function with respect to a distribution F in $\mathbb{R}^d$ is

$$R(\boldsymbol{x}, F) = \mathbb{E}\frac{\boldsymbol{x} - \boldsymbol{X}}{\|x - \boldsymbol{X}\|}.$$

In order to be more convenient in developing the theoretical result, the population version of the spatial rank function will be given as well. If $\boldsymbol{X} \in \mathbb{R}^d$ is a random variable from a distribution with cumulative distribution function F, the expected Euclidean distance from $\boldsymbol{x}$ to $\boldsymbol{X}$ is $D(\boldsymbol{x}, F) = E_F\|\boldsymbol{x} - \boldsymbol{X}\|$. The spatial median of F minimizes the criterion function D w.r.t. $\boldsymbol{x}$. The multivariate centered spatial rank function is defined as the gradient of D:

$$\boldsymbol{R}(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} D(\boldsymbol{x}, F) = E_F \frac{\boldsymbol{x} - \boldsymbol{X}}{\|\boldsymbol{x} - \boldsymbol{X}\|} = E_F\{\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{X})\}.$$

The spatial rank function of $\boldsymbol{x}$ is the *expected direction* to $\boldsymbol{x}$ from $\boldsymbol{X}$. It is called centered because the rank of a random vector from the same distribution F has expected value at $\mathbf{0}$, that is, $E_F\boldsymbol{R}(\boldsymbol{X}, F) = \mathbf{0}$. It is interesting to see, the three objective functions D_1, D_2 (see subsections 3.2.1 and 3.2.2) and D would degenerate to the same absolute value function in the univariate case. If the population version is considered, then the marginal rank, Oja Rank and spatial rank objective functions all equal to $D(x, F) = E_F|x - X|$. Their gradient functions (rank functions) lead to the univariate rank function $R(x, F) = E_F\text{sign}(x - X) = 2F(x) - 1 \in [-1, 1]$. Thus, the spatial rank in the univariate case is a linear transformation of the standardized rank. All results of Chapter 2 will hold if the standardized rank is replaced with the spatial rank.

3.3 New Rank-based Tests

For the multivariate setting, the spatial rank function of vector $\boldsymbol{x}$ with respect to distribution $H = \tau F + (1 - \tau)G$ with $0 \leq \tau \leq 1$ is defined as

$$R(\boldsymbol{x}, H) = \tau E_F \frac{\boldsymbol{x} - \boldsymbol{X}}{\|\boldsymbol{x} - \boldsymbol{X}\|} + (1 - \tau) E_G \frac{\boldsymbol{x} - \boldsymbol{Y}}{\|\boldsymbol{x} - \boldsymbol{Y}\|} = \tau R(\boldsymbol{x}, F) + (1 - \tau) R(\boldsymbol{x}, G).$$

Let $R(\boldsymbol{X}_i, H_N)$ and $R(\boldsymbol{Y}_j, H_N)$ denote the spatial rank of $\boldsymbol{X}_i$ and $\boldsymbol{Y}_j$, respectively, from the mixture distribution H_N among $\boldsymbol{X}_1, ..., \boldsymbol{X}_m, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$. Then the proposed multivariate two-sample spatial rank statistic, denoted by T_{M1}, is defined as

$$\begin{aligned} T_{M1} &= \frac{mn}{N}\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|R(\boldsymbol{X}_i, H_N) - R(\boldsymbol{Y}_j, H_N)\| \\ &\quad -\frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|R(\boldsymbol{X}_i, H_N) - R(\boldsymbol{X}_j, H_N)\| \\ &\quad -\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|R(\boldsymbol{Y}_i, H_N) - R(\boldsymbol{Y}_j, H_N)\|\}. \end{aligned} \tag{3.1}$$

The test statistic T_{M1} can be interpreted as the Euclidean distances of the average of the intra-group rank differences and the average of the inter-group rank differences. Recall that $\|\boldsymbol{x}\|$ is the Euclidean distance of $\boldsymbol{x}$ from 0. The null hypothesis is rejected for large values of T_{M1}, and the critical values of the statistic may be computed via simulation. The critical value $c_{m,n}$ is determined by the significance level α and the null distribution of T_{M1}. A large value of T_{M1} indicates the deviation of the two groups.

When it comes to the multivariate case, the correspondences of the transformed observations become of interest. Due to using the spatial rank function, the distances of the ranked observations are not as intuitive as in the univariate case. Similar to the univariate setting, the following lemmas and theorems serve as motivation for the statistic T_{M1} in the multivariate setting.

Lemma 22 *For each $\boldsymbol{x} \in \mathbb{R}^d$, let μ be the uniform distribution on the surface of the unit ball $S^{d-1} = \{\boldsymbol{a} \in \mathbb{R}^d : \|\boldsymbol{a}\| = 1\}$, then*

$$\|\boldsymbol{x}\| = \gamma_d \int_{S^{d-1}} |\boldsymbol{a}^T\boldsymbol{x}|\, d\mu(\boldsymbol{a}),$$

where

$$\gamma_d = \begin{cases} \frac{(d-1)\sqrt{\pi}\Gamma((d-1)/2)}{2\Gamma(d/2)} & , d \geq 2 \\ 1 & , d = 1 \end{cases}.$$

Proof. The result of $d = 1$ is trivial and the result is trivial for $\boldsymbol{x} = \mathbf{0}$. Let $\boldsymbol{x} \neq \mathbf{0}$. Because the uniform distribution on S^{d-1} is invariant with respect to orthogonal transformations,

$$\int_{S^{d-1}} |\boldsymbol{a}^T \boldsymbol{x}| \, d\mu(\boldsymbol{a}) = \|\boldsymbol{x}\| \int_{S^{d-1}} |\boldsymbol{a}^T \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}| \, d\mu(\boldsymbol{a}) = \|\boldsymbol{x}\| \int_{S^{d-1}} |\boldsymbol{a}^T \boldsymbol{e}_1| \, d\mu(\boldsymbol{a}),$$

where $\boldsymbol{e}_1 = (1, 0, ..., 0)^T \in \mathbb{R}^d$. It needs to be shown that $\gamma_d^{-1} = \int_{S^{d-1}} |\boldsymbol{a}^T \boldsymbol{e}_1| \, d\mu(\boldsymbol{a})$. For $\boldsymbol{a} \in S^{d-1}$,

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} cos(\phi_1) \\ sin(\phi_1)cos(\phi_2) \\ \vdots \\ sin(\phi_1) \cdots sin(\phi_{d-2})sin(\phi_{d-1}) \end{pmatrix}.$$

By the Jacobian of the spherical coordinates transformation,

$$d\boldsymbol{a} = sin^{d-2}(\phi_1)sin^{d-3}(\phi_2) \cdots sin(\phi_{d-2})d\phi_1 d\phi_2 \cdots d\phi_{d-1} = Jd\phi_1 d\phi_2 \cdots d\phi_{d-1}.$$

The uniform distribution μ on S^{d-1} is given by

$$\begin{aligned} d\mu(\boldsymbol{a}) =& \frac{1}{\text{surface area}(S^{d-1})} d\boldsymbol{a} \\ =& \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} d\boldsymbol{a}. \end{aligned}$$

Then,

$$\begin{aligned} &\int_{S^{d-1}} |\boldsymbol{a}^T \boldsymbol{e}_1| \, d\mu(\boldsymbol{a}) = \int_0^{2\pi} \cdots \int_0^{\pi} \int_0^{\pi} |cos(\phi_1)| \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} Jd\phi_1 d\phi_2 \cdots d\phi_{d-1} \\ &= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \left[2\pi \int_0^{\pi/2} 2|cos(\phi_1)|sin^{d-2}(\phi_1)d\phi_1 \int_0^{\pi/2} 2sin^{d-3}(\phi_2)d\phi_2 \cdots \int_0^{\pi/2} 2sin(\phi_{d-2})d\phi_{d-2} \right] \end{aligned}$$

$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \left[2\pi \int_0^{\pi/2} 2sin^{d-2}(\phi_1) dsin(\phi_1) \int_0^{\pi/2} 2sin^{d-3}(\phi_2) d\phi_2 \cdots \int_0^{\pi/2} 2sin(\phi_{d-2}) d\phi_{d-2} \right]$$
$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \left[2\pi \left[2\frac{sin^{d-1}(\phi_1)}{d-1} \right]_0^{\pi/2} \int_0^{\pi/2} 2sin^{d-3}(\phi_2) d\phi_2 \cdots \int_0^{\pi/2} 2sin(\phi_{d-2}) d\phi_{d-2} \right]$$
$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \left[\frac{4\pi}{d-1} \int_0^{\pi/2} 2sin^{d-3}(\phi_2) d\phi_2 \int_0^{\pi/2} 2sin^{d-4}(\phi_3) d\phi_3 \cdots \int_0^{\pi/2} 2sin(\phi_{d-2}) d\phi_{d-2} \right].$$

Note that the beta function is equal to

$$B(\alpha, \beta) = 2 \int_0^{\pi/2} sin^{2\alpha-1}(\phi) cos^{2\beta-1}(\phi) d\phi, \quad \text{for} \quad Re(\alpha) > 0, Re(\beta) > 0.$$

Thus,

$$\frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \left[\frac{4\pi}{d-1} \int_0^{\pi/2} 2sin^{d-3}(\phi_2) d\phi_2 \int_0^{\pi/2} 2sin^{d-4}(\phi_3) d\phi_3 \cdots \int_0^{\pi/2} 2sin(\phi_{d-2}) d\phi_{d-2} \right]$$
$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \frac{4\pi}{d-1} \left[B(\frac{d-2}{2}, \frac{1}{2}) \right] \left[B(\frac{d-3}{2}, \frac{1}{2}) \right] \cdots \left[B(1, \frac{1}{2}) \right].$$

The beta function can also be written as

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Therefore,

$$\frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \frac{4\pi}{d-1} \left[B(\frac{d-2}{2}, \frac{1}{2}) \right] \left[B(\frac{d-3}{2}, \frac{1}{2}) \right] \cdots \left[B(1, \frac{1}{2}) \right]$$
$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \frac{4\pi}{d-1} \left[\frac{\Gamma(\frac{d-2}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{d-1}{2})} \right] \left[\frac{\Gamma(\frac{d-3}{2})\Gamma(\frac{1}{2})}{\Gamma(\frac{d-2}{2})} \right] \cdots \left[\frac{\Gamma(1)\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{2})} \right]$$
$$= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}} \frac{4\pi}{d-1} \left(\Gamma(\frac{1}{2}) \right)^{d-3} \left[\frac{\Gamma(\frac{d-2}{2})}{\Gamma(\frac{d-1}{2})} \right] \left[\frac{\Gamma(\frac{d-3}{2})}{\Gamma(\frac{d-2}{2})} \right] \cdots \left[\frac{\Gamma(1)}{\Gamma(\frac{3}{2})} \right].$$

For $\Gamma\left(\frac{1}{2}\right) = (\pi)^{1/2}$ and by cancellation,

$$\begin{aligned}
&\frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\frac{4\pi}{d-1}\left(\Gamma(\frac{1}{2})\right)^{d-3}\left[\frac{\Gamma(\frac{d-2}{2})}{\Gamma(\frac{d-1}{2})}\right]\left[\frac{\Gamma(\frac{d-3}{2})}{\Gamma(\frac{d-2}{2})}\right]\cdots\left[\frac{\Gamma(1)}{\Gamma(\frac{3}{2})}\right] \\
&= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\frac{4\pi}{d-1}\left((\pi)^{1/2}\right)^{d-3}\frac{1}{\Gamma(\frac{d-1}{2})} \\
&= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\frac{4\pi\pi^{\frac{d-3}{2}}}{(d-1)\Gamma(\frac{d-1}{2})} \\
&= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\frac{4\pi^{\frac{d}{2}}\pi^{\frac{-1}{2}}}{(d-1)\Gamma(\frac{d-1}{2})} \\
&= \frac{\Gamma(\frac{d}{2})}{2\pi^{d/2}}\frac{4\pi^{d/2}}{\sqrt{\pi}(d-1)\Gamma(\frac{d-1}{2})} \qquad (3.2) \\
&= \frac{2\Gamma(\frac{d}{2})}{\sqrt{\pi}(d-1)\Gamma(\frac{d-1}{2})}. \qquad (3.3)
\end{aligned}$$

Through a spherical coordinate transformation, the result of γ_d is obtained. ■

Lemma 23 *Let $\boldsymbol{X}$ be d-variate random vectors with distribution F. Denote F^a as the univariate function for $\boldsymbol{a}^T\boldsymbol{X}$. Then*

$$R(\boldsymbol{a}^T\boldsymbol{x}, F^a) = \boldsymbol{a}^T\boldsymbol{R}(\boldsymbol{x}, F) = 2F^a(\boldsymbol{a}^T\boldsymbol{x}) - 1,$$

where $\boldsymbol{a} \in S^{d-1}$.

Proof.

$$R(\boldsymbol{a}^T\boldsymbol{x}, F^a) = \mathbb{E}\frac{\boldsymbol{a}^T\boldsymbol{x} - \boldsymbol{a}^T\boldsymbol{X}}{\|\boldsymbol{a}^T\boldsymbol{x} - \boldsymbol{a}^T\boldsymbol{X}\|} = \boldsymbol{a}^T\mathbb{E}\frac{\boldsymbol{x} - \boldsymbol{X}}{\|\boldsymbol{x} - \boldsymbol{X}\|} = \boldsymbol{a}^T\boldsymbol{R}(\boldsymbol{x}, F).$$

On the other hand, $\boldsymbol{R}(\boldsymbol{a}^T\boldsymbol{x}, F_{\boldsymbol{a}})$ is the spatial rank of $\boldsymbol{a}^T\boldsymbol{x}$ on the one dimensional space, which is equal to $2F_{\boldsymbol{a}}(a^T\boldsymbol{x}) - 1$. ■

Theorem 24 *Let $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2$ and $\boldsymbol{Y}, \boldsymbol{Y}_1, \boldsymbol{Y}_2$ be independent random vectors distributed*

from F and G, respectively. Let $H = \tau F + (1-\tau)G$ with $0 \le \tau \le 1$. Then

$$E\|\boldsymbol{R}(\boldsymbol{X}, H) - \boldsymbol{R}(\boldsymbol{Y}, H)\| - \frac{1}{2}E\|\boldsymbol{R}(\boldsymbol{X}_1, H) - \boldsymbol{R}(\boldsymbol{X}_2, H)\| - \frac{1}{2}E\|\boldsymbol{R}(\boldsymbol{Y}_1, H) - \boldsymbol{R}(\boldsymbol{Y}_2, H)\| \ge 0, \tag{3.4}$$

where equality holds if and only if $F = G$.

Proof. From Lemma 23 and Theorem 8 of Chapter 2,

$$\begin{aligned}&\mathbb{E}|R(\boldsymbol{a}^T\boldsymbol{X}, H^a) - R(\boldsymbol{a}^T\boldsymbol{Y}, H^a)| - \frac{1}{2}\mathbb{E}|R(\boldsymbol{a}^T\boldsymbol{X}_1, H^a) - R(\boldsymbol{a}^T\boldsymbol{X}_2, H^a)| \\ &- \frac{1}{2}\mathbb{E}|R(\boldsymbol{a}^T\boldsymbol{Y}_1, H^a) - R(\boldsymbol{a}^T\boldsymbol{Y}_2, H^a)| \ge 0\end{aligned}$$

for each $\boldsymbol{a} \in S^{d-1}$ where $H^a = \tau F^a + (1-\tau)G^a$. Integration of $\boldsymbol{a}$ with respect to μ obtains (3.4). Equality holds if and only if for μ-almost all $\boldsymbol{a} \in S^{d-1}$ the distributions of $\boldsymbol{a}^T\boldsymbol{X}$ and $\boldsymbol{a}^T\boldsymbol{Y}$ coincide. For each $t \in \mathbb{R}$ the functions $\mathbb{E}\exp(it\boldsymbol{a}^T\boldsymbol{X})$ and $\mathbb{E}\exp(it\boldsymbol{a}^T\boldsymbol{Y})$ with $\boldsymbol{a} \in S^{d-1}$ are continuous. Thus, equality in (3.4) holds if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same Fourier transformation, hence have the same distribution. ■

Note that the statistic T_{M1} is the sample plug-in version of the left side of (3.4), which is a generalization of T_1.

Similarly, we can have a multivariate version of T, denoted as T_M, when $m = n$. That is,

$$T_M = \frac{mn}{N}\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|\boldsymbol{R}(\boldsymbol{X}_i, H_N) - \boldsymbol{R}(\boldsymbol{Y}_j, H_N)\|\}. \tag{3.5}$$

The population version of test statistics T_M has the following corresponding theorem stated below.

Theorem 25 *In the case that $\tau = 1/2$, $\mathbb{E}\|\boldsymbol{R}(\boldsymbol{X}, H) - \boldsymbol{R}(\boldsymbol{Y}, H)\| = 2\gamma_d/3$ if and only if $F = G$, and $\mathbb{E}\|\boldsymbol{R}(\boldsymbol{X}, H) - \boldsymbol{R}(\boldsymbol{Y}, H)\| > 2\gamma_d/3$ if $F \ne G$.*

Proof. By Lemma 22, Lemma 23 and Theorem 4,

$$\mathbb{E}\|\boldsymbol{R}(\boldsymbol{X}, H) - \boldsymbol{R}(\boldsymbol{Y}, H)\| \;=\; \mathbb{E}\gamma_d \int_{S^{d-1}} |\boldsymbol{a}^T[\boldsymbol{R}(\boldsymbol{X}, H) - \boldsymbol{R}(\boldsymbol{Y}, H)]|\, d\mu(\boldsymbol{a})$$

$$
\begin{aligned}
&= \gamma_d \int_{S^{d-1}} \mathbb{E}|\boldsymbol{a}^T\boldsymbol{R}(\boldsymbol{X},H) - \boldsymbol{a}^T\boldsymbol{R}(\boldsymbol{Y},H)|\, d\mu(\boldsymbol{a}) \\
&= \gamma_d \int_{S^{d-1}} \mathbb{E}|R(\boldsymbol{a}^T\boldsymbol{X},H^a) - R(\boldsymbol{a}^T\boldsymbol{Y},H^a)|\, d\mu(\boldsymbol{a}) \\
&\geq \gamma_d \int_{S^{d-1}} \frac{2}{3}\, d\mu(\boldsymbol{a}) \qquad (3.6) \\
&= \frac{2\gamma_d}{3}
\end{aligned}
$$

Equality holds only when $F = G$. The inequality in (3.6) is because of Theorem 4 using the spatial (centered) rather than the univariate standardized rank. ■

Theorem 25 provides a basis to reject H_0 if T_M is sufficiently large.

Now, both T_{M1} and T_M have the orthogonal invariant property. That is, for every $\boldsymbol{X}_1, ..., \boldsymbol{X}_m, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_n$, orthogonal $d \times d$ matrix $O(O^T = O^{-1})$, d-vector $\boldsymbol{c}$ and scaler $b > 0$,

$$
\begin{aligned}
&T_{M1}(bO\boldsymbol{X}_1 + c, ..., bO\boldsymbol{X}_m + \boldsymbol{c}, bO\boldsymbol{Y}_1 + \boldsymbol{c}, ..., bO\boldsymbol{Y}_n + \boldsymbol{c}) = T_{M1}(\boldsymbol{X}_1, ..., \boldsymbol{X}_m, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_n) \\
&T_M(bO\boldsymbol{X}_1 + c, ..., bO\boldsymbol{X}_m + \boldsymbol{c}, bO\boldsymbol{Y}_1 + \boldsymbol{c}, ..., bO\boldsymbol{Y}_n + \boldsymbol{c}) = T_M(\boldsymbol{X}_1, ..., \boldsymbol{X}_m, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_n).
\end{aligned}
$$

The proof follows from the fact that the spatial rank is orthogonal invariant. That is, denoting $F_{bO\boldsymbol{X}+\boldsymbol{c}}$ as the distribution of $bO\boldsymbol{X} + \boldsymbol{c}$,

$$
\boldsymbol{R}(bO\boldsymbol{x} + \boldsymbol{c}, F_{bO\boldsymbol{X}+\boldsymbol{c}}) = \mathbb{E}_F \frac{bO\boldsymbol{x} + \boldsymbol{c} - (bO\boldsymbol{X} + \boldsymbol{c})}{\|bO\boldsymbol{x} + \boldsymbol{c} - (bO\boldsymbol{X} + \boldsymbol{c})\|} = \mathbb{E}_F \frac{O(\boldsymbol{x} - \boldsymbol{X})}{\|\boldsymbol{x} - \boldsymbol{X}\|} = O\boldsymbol{R}(\boldsymbol{x}, F).
$$

Orthogonal invariance ensures that the tests are invariant under rotation, translation and homogeneous scale change. But they do not allow heterogeneous scale changes. The above equations do not hold for a general $d \times d$ nonsingular matrix A. In other words, they are not fully affine invariant. Hence the tests based on T_{M1} and T_M may not be suitable for data whose scale of coordinates is widely different. They are, however, desirable for data with isometric variables such as preprepared gene data.

3.4 Bootstrap and Permutation Approximation

Unlike the distribution-free property of the test statistics in the univariate case, in the multivariate cases T_M and T_{M1} do depend on the distribution F. Since neither the null distribution nor the asymptotic null distribution of test statistics are known, bootstrap and permutation methods provide attractive approaches to determine a critical point for the tests, or equivalently, to approximate the p-value of the observed value of the test statistics.

For the bootstrap method, bootstrap samples are drawn from the empirical distribution function H_N of the pooled sample $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_m, \boldsymbol{Y}_1, ..., \boldsymbol{Y}_n\}$. Let $\boldsymbol{X}_1^*, \boldsymbol{X}_2^*, ..., \boldsymbol{X}_m^*$ and $\boldsymbol{Y}_1^*, \boldsymbol{Y}_2^*, ..., \boldsymbol{Y}_n^*$ be two independent random samples from H_N. Let T_{M1}^* denote the bootstrap version of T_{M1}, that is,

$$\begin{aligned} T_{M1}^* &= \frac{mn}{N}\{\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|R(\boldsymbol{X}_i^*, H_N) - R(\boldsymbol{Y}_j^*, H_N)\| \\ &\quad -\frac{1}{2m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|R(\boldsymbol{X}_i^*, H_N) - R(\boldsymbol{X}_j^*, H_N)\| \\ &\quad -\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|R(\boldsymbol{Y}_i^*, H_N) - R(\boldsymbol{Y}_j^*, H_N)\|\}. \end{aligned} \tag{3.7}$$

According to Arcones & Giné (1992), we get that T_{M1} and T_{M1}^* both converge in law to the same limit under H_0. That is, T_{M1} has the same asymptotic null distribution as T_{M1}^*. This provides a way to approximate the bootstrap critical point $t_{M1}^*(\alpha)$ and the bootstrap p-value, p^*, as follows:

1. Calculate $T_{M1,obs}$, the value of T_{M1} for the original samples $\{\boldsymbol{X}_i\}_{1\le i\le m}$ and $\{\boldsymbol{Y}_j\}_{1\le j\le n}$.

2. Generate B bootstrap samples $(\boldsymbol{X}_1^{*b}, \boldsymbol{X}_2^{*b}, ..., \boldsymbol{X}_m^{*b}, \boldsymbol{Y}_1^{*b}, \boldsymbol{Y}_2^{*b}, ..., \boldsymbol{Y}_n^{*b}), b = 1, ..., B$ from H_N.

3. Calculate T_{M1} for each bootstrap sample and denote it by $T_{M1}^{*b}, b = 1, ..., B$.

4. Approximate the p-value by means of the expression

$$\hat{p}^* = \frac{\#\{b : T^{*b}_{M1} \geq T_{M1,obs}\}}{B},$$

or approximate the critical value by $T^{*(a)}_{M1}$, where $a = [(1-\alpha)B] + 1$, $[x]$ is the greatest integer less than x, and $T^{*(1)}_{M1} \leq T^{*(2)}_{M1} \leq ... \leq T^{*(B)}_{M1}$ are the order statistic of $T^{*b}_{M1}, b = 1, ..., B$.

Next, is another method used to approximate the null distribution of the test statistics: the permutation distribution. To simplify notation, let $\{\boldsymbol{Z}_1, \boldsymbol{Z}_2, ..., \boldsymbol{Z}_m, \boldsymbol{Z}_{m+1}..., \boldsymbol{Z}_N\}$ denote the pooled sample (where the observed values of both samples are denoted with the same letter, with $\boldsymbol{Z}_i = \boldsymbol{X}_i$, $1 \leq i \leq m$, and $\boldsymbol{Z}_{m+i} = \boldsymbol{Y}_i$, $1 \leq i \leq n$). Let $\sigma = (\sigma(1), \sigma(2), ..., \sigma(N))$ be a permutation of $(1, 2, ..., N)$. Let $\boldsymbol{X}'_i = \boldsymbol{Z}_{\sigma(i)}$, $1 \leq i \leq m$, and $\boldsymbol{Y}'_j = \boldsymbol{Z}_{\sigma(m+j)}$, $1 \leq j \leq m$. The permutation version of T_{M1} is denoted as T'_{M1}. Again by Arcones & Giné (1992), T'_{M1} has the asymptotic distribution as T_{M1}. As usual, the permutation p-value, p', or the permutation critical point, T'_{M1} is approximated as follows.

1. Calculate $T_{M1,obs}$, the value of T_{M1} for the original samples $\{\boldsymbol{X}_i\}_{1\leq i\leq m}$ and $\{\boldsymbol{Y}_j\}_{1\leq j\leq n}$.

2. Generate B bootstrap samples $(\boldsymbol{X}'^b_1, \boldsymbol{X}'^b_2, ..., \boldsymbol{X}'^b_m, \boldsymbol{Y}'^b_1, \boldsymbol{Y}'^b_2, ..., \boldsymbol{Y}'^b_n), b = 1, ..., B$ from H_N.

3. Calculate T_{M1} for each bootstrap sample and denote it by $T'^b_{M1}, b = 1, ..., B$.

4. Approximate the p-value by means of the expression

$$\hat{p}' = \frac{\#\{b : T'^b_{M1} \geq T_{M1,obs}\}}{B},$$

or approximate the critical value by $T'^{(a)}_{M1}$, where $a = [(1-\alpha)B] + 1$, $[x]$ is the greatest integer less than x, and $T'^{(1)}_{M1} \leq T'^{(2)}_{M1} \leq ... \leq T'^{(B)}_{M1}$ are the order statistic of $T'^b_{M1}, b = 1, ..., B$.

Note that the bootstrap approximation and permutation approximation are asymptotically equivalent, in the sense that the corresponding distribution estimators converge to the same law. For small samples, it is expected that the power of tests based on permutation approximation is little bit higher than that of tests based on bootstrap approximation since a permutation sample takes all observation values, while a bootstrap sample may not.

Similarly, bootstrap and permutation methods can be applied to approximate the p-value or critical value for tests based on T_M.

As stated in Chapter 2, test statistics T and T_1 are equivalent in the univariate setting for balanced samples. In this case, T is written as a linear transformation of T_1. However, in the multivariate setting, they are no longer equivalent. The absolute value is replaced by the Euclidean distance and the spatial (centered) rank function is used rather than the univariate standardized rank. T_M has a simplier form than T_{M1} and hence T_{M1} provides more information than T_M. Thus, in the multivariate case, T_{M1} is more powerful than T_M. The results are shown later in Section 3.6.

3.5 Connection with Other Test Statistics

Baringhaus & Franz (2004) considered a T_{M1} type of test based on the original data, which is a direct generalization from the univariate case (2.7) by changing the absolute value $|\cdot|$ in one dimension to the Euclidean distance $\|\cdot\|$ in high dimension. The extension is natural, however, the test needs the assumption on the finite first moment. If the sample data X and/or Y come from distributions without the first moment, then their test is not feasible. The proposed test is based on rank transformation of data that always exist, hence it is suitable for all types of data.

A kernel method is considered by Gretton *et al.* (2008). Their test statistics are T_k or $T_k^{1/2}$ with the form of

$$T_k = \frac{1}{m^2}\sum_{i,j=1}^{m}\kappa(\boldsymbol{x}_i,\boldsymbol{x}_j) - \frac{2}{mn}\sum_{i,j=1}^{m,n}\kappa(\boldsymbol{x}_i,\boldsymbol{y}_j) + \frac{1}{n^2}\sum_{i,j=1}^{n}\kappa(\boldsymbol{y}_i,\boldsymbol{y}_j), \tag{3.8}$$

where $\kappa(\boldsymbol{x}, \boldsymbol{y})$ is a positive definite kernel function, which implicitly defines feature mapping $\phi(\boldsymbol{x})$ and $\kappa(\boldsymbol{x}, \boldsymbol{y}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{y}) \rangle$, the inner product in the induced feature space. A common used kernel is the Gaussian RBF kernel, that is, $\kappa(\boldsymbol{x}, \boldsymbol{y}) = e^{-\|x-y\|/\sigma}$, where σ is the kernel parameter.

Although their tests have a nice interpretation of the maximum mean discrepancy, the performance of the tests heavily depends on the choice of the kernel or kernel parameter and in practice there is no guide on how to choose them efficiently.

Another similar test developed by Fernández *et al.* has a form

$$T_c = \frac{1}{m^2} \sum_{i,j=1}^{m} e^{-\|x_i - x_j\|^2/2} - \frac{2}{mn} \sum_{i,j=1}^{m,n} e^{-\|x_i - y_j\|^2/2} + \frac{1}{n^2} \sum_{i,j=1}^{n} e^{-\|y_i - y_j\|^2/2}. \tag{3.9}$$

The motivation of this test comes from the empirical counterpart of $\int |C_F(\boldsymbol{t}) - C_G(\boldsymbol{t})|^2 \, dW(\boldsymbol{t})$, where $C_F(\boldsymbol{t})$ and $C_G(\boldsymbol{t})$ are the characteristic function of the distribution F and G, respectively, and W is some measure on $\mathbb{R}^d$. The above test statistic T_c takes W to be the normal distribution, $N(\mathbf{0}, I)$, where I is the identity matrix. Again the choice of W is very important for the performance of the test and the choice of the normal distribution is more for mathematical convenience.

Upon observation, many of these tests take some transformation on the pair differences and the choice of transformation is more heuristic. The proposed test take the rank transformation before taking pair differences among ranks. The spatial rank transformation has a nice interpretation and it has nice properties such that it characterizes the distribution. In other words, if the rank function $\boldsymbol{R}(\boldsymbol{x}, F)$ is known, then the distribution F is known (Oja (2010)). Hence, the proposed test is more principled than the above mentioned ones.

Oja & Randles (2004) studied a multivariate test based on the spatial ranks. Their test, called the spatial rank test, is a Hotelling's T^2 test based on ranks. Hotelling test is a multivariate generalization of the t -test under the normality assumption. The spatial rank test is nonparametric and is asymptotically distribution-free, however, it only applies to the

location alternatives. For a general alternative, it is not suitable.

The next section gives the comparison of power performance of these tests along with the proposed tests.

3.6 Simulation

In this section, results that were conducted to investigate the power performance of the proposed multivariate rank test and its comparison with other multivariate tests are discussed. Similar to the simulation conducted for univariate data, the traditional bootstrap procedure and the permutation procedure were used to study the performances of the tests for the high-dimensional setting. The bootstrap procedure was used to compare the performances of the multivariate Kolmogorov-Smirnov test, Hotelling T^2 test, Spatial Rank test, multivariate Cramér two-sample test, Fernándes *et al.* test, and the proposed multivariate spatial rank test. The permutation procedure was used only for Baringhaus and Franz's Cramér test, Fernándes *et al.* test, and the proposed multivariate test. Denote the Kolmogorov-Smirnov test as KS, Hotelling's test as $T2$, Spatial Rank test as SR, Cramér two-sample test as CT, Fernándes *et al.* test as DT, and the proposed test as RCT. The notations CTP and CTB are used to denote the CT test under the permutation method and the bootstrap method, respectively. The same holds for the DTP and DTB tests, and the proposed $RCTP$ and $RCTB$ tests. All computations were conducted using programs written in the R language.

3.6.1 Simulation Results for Location Alternatives

In this section, the power for the multivariate two-sample problem is under investigation. To conduct simulations, two multivariate independent samples were generated with equal sizes ($n = m = 50$) and unequal sizes ($n = 50$ and $m = 20$) from the multivariate normal distribution, t-distribution, exponential distribution, Pareto distribution, and Poisson distribution at the chosen significance level $\alpha = 0.05$. For each distribution, 1000

iterations were computed for the bootstrap and permutation methods for each sample to compute the estimated powers by calculating the fraction of p-values less than or equal to 0.05. Since the results for the bootstrap method behave quite similar to the permutation method, only the results obtained for the bootstrap method are reported. However, the results of both methods are displayed in each figure below for all considered distributions.

For the multivariate normal distribution, multivariate samples $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim N_d(\mathbf{0}, I_d)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim N_d(\boldsymbol{\mu}, I_d)$ were generated, with $\boldsymbol{\mu} = (\Delta, 0, \ldots, 0)$, $\Delta = 0, \ldots, 1$ in steps of 0.10. Figure 3.1 shows the power performance for each test under multivariate normal distribution. For equal sample sizes, note that the statistical power of the $RCTB$ test compares favorably to the $T2$ test, SR test and the CTB test. The statistical power of the $RCTB$ test is higher than that of the KS test, and is significantly higher than that of the DTB test where the difference in powers can be a large as 14%. Similarly, for unequal sample sizes, the statistical power of the $RCTB$ test is comparable to the $T2$ test, SR test and the CTB test, while it is significantly higher than that of the DTB test.

The experiment was repeated for the multivariate t_d-distribution in which the performance was observed when $df = 3$ and when $df = 1$, where df denotes the degrees of freedom. Multivariate samples $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim t_d(\mathbf{0}, I_d)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim t_d(\boldsymbol{\mu}, I_d)$ were generated, with $\boldsymbol{\mu} = (\Delta, 0, \ldots, 0)$, $\Delta = 0, \ldots, 1$ in steps of 0.10. Looking at Figure 3.3 the power performance for each test under the t_d-distribution with $df = 3$ is shown. For equal samples sizes, the statistical power of the $RCTB$ test compares favorably to the KS test and is slightly higher than the SR test, while outperforming the other tests for all considered alternatives. For unequal sample sizes, the KS test outperforms each of the considered tests. The $RCTB$ test is slightly higher than the SR test, but it performs better than the CTB and DTB tests. The difference in powers between the $RCTB$ test and the $T2$ test can be a large as 20%. In the case of the multivariate t_d-distribution when $df = 1$, for equal sample sizes, Figure 3.5 indicates that the statistical power of the KS test performs the best. The $RCTB$ test is slightly higher than that of the DTB test for alternatives

between 0 and 0.60. However, the $RCTB$ test outperforms the SR test by as much as 11% while outperforming the CTB and $T2$ test by a large margin. For unequal sample sizes, the $RCTB$ test outperforms the SR, CTB and $T2$ tests. Also note that the $RCTB$ test is comparable to the DTB test for alternatives between 0.6 and 1, while the DTB test is slightly higher than that of the $RCTB$ test for alternatives between 0 and 0.6. For unequal sample sizes, the KS test outperfoms each test for all alternatives considered. (Note that the CTP test performs better than the CTB test for both equal and unequal sample sizes in Figure 3.5).

To study the power for the multivariate exponential distribution, the simulations were repeated for exponential d-variates $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ E(1) and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ E(Δ), with $\Delta = 0.5,\ldots,1.5$ in steps of 0.10 (here $d = 2$). First, two univariate X-samples, $X_{11},\ldots,X_{1n} \sim$ E(1) and $X_{21},\ldots,X_{2n} \sim$ E(1) were generated, then the two components were combined to form the vector sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ E(1). Similarly, two univariate Y-samples, $Y_{11},\ldots,Y_{1m} \sim$ E(Δ) and $Y_{21},\ldots,Y_{2m} \sim$ E(Δ) were generated to form the vector sample $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ E(Δ). Figure 3.9 displays the difference in the power performance for each test under multivariate exponential distribution in which each test remains consistent in relation to each other for equal sample sizes. The CTB and $T2$ tests outperforms all considered tests, where the KS test has the weakest performance. Note that the $RCTB$ test is comparable to the SR test and the DTB test. For unequal sample sizes, the $T2$ test and CTB test performs the best only for alternatives 0.5 to 0.9. Then $T2$ performs slightly worse than the other considered test for alternative 1.1 to 1.5. The $RCTB$ test is comparable to the SR, KS and DTB tests.

Now, observe the case of Poisson multivariate samples $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ P(2) and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ P(Δ), with $\Delta = 1,\ldots,3$ in steps of 0.10. Similar to the exponential case, two univariate X-samples $X_{11},\ldots,X_{1n} \sim$ P(2) and $X_{21},\ldots,X_{2n} \sim$ P(2) were generated to form the vector sample $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ P(2) . Similarly, the vector sample $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ P(Δ) is constructed by combining two univariate Y-samples $Y_{11},\ldots,Y_{1m} \sim$ P(Δ) and $Y_{21},\ldots,Y_{2m}$

$\sim$ P(Δ). Figure 3.10 displays the obtained results of the statistical power for each test. It is shown that the power performance of the $RCTB$ test seem to have the lowest performance of all considered tests and the SR test has the highest power. This is the case for both equal and unequal sample sizes.

Shown in Figure 3.11 is the power performance for the multivariate Pareto distribution, where $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ Pa(2, 2) and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ Pa(2+Δ, 2) were generated, with $\Delta = 0,\ldots,1$ in steps of 0.10. The vector samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ are formed by generating univariate samples $X_{11},\ldots,X_{1n} \sim$ Pa(2, 2) and $X_{21},\ldots,X_{2n} \sim$ Pa(2, 2), and $Y_{11},\ldots,Y_{1m} \sim$ Pa(2+Δ, 2) and $Y_{21},\ldots,Y_{2m} \sim$ Pa(2+Δ, 2), respectively. The KS test performs the best, however, the $RCTB$ test and outperforms the remaining considered tests. The results are the same for both equal and unequal sample sizes. The power difference between the $RCTB$ test and that of the $T2$ test can be as large as 60% for equal sample sizes and can be as large as 51% for unequal sample sizes.

For the proposed test T, the simulations for all distributions discussed above were repeated. However, the simulations were only conducted for equal sample sizes. The test statistic T is compared to CT and $T1$, i.e. RCT, and the power performances are shown in Figures 3.13(a) - 3.17 For the Normal distribution, CT and $T1$ have the same power performance, where both statistics outperforms that of T. When the distribution is the t-distribution for $df = 1$, the proposed statistics $T1$ and T outperforms CT where the difference in the power performance can be as large as 63%. For the t-distribution when $df = 3$, CT and T performs the same, where $T1$ is slightly higher than both CT and T. In the case of the Pareto distribution, $T1$ outperforms CT and T, however T outperforms CT where the difference in the power performance can be as large as 32%. For the exponential distribution, $T1$ is higher than that of T, while CT is slightly higher than that of $T1$. In the case of the Poisson distribution, CT outperforms both $T1$ and T, while $T1$ performs higher than T.

3.6.2 Simulation Results for Scale Alternatives

Now, results found for scale alternatives are of discussion. Similar procedures that were conducted in the case of location alternatives were performed. Two multivariate independent samples for both equal ($n = m = 50$) and unequal sample sizes ($n = 50$ and $m = 20$) were generated from a multivariate normal distribution, t-distribution, and Pareto distribution. All considered tests were used for scale alternatives as in the experiment for location alternatives with the exception of the Hotelling T^2 test ($T2$) and the Spatial Rank (SR) test, as these tests are location tests. Instead, as in the univariate case, the scale test known as Mood's test is used which will be refer to as the M test.

In the case of the multivariate normal distribution, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim N_d(\mathbf{0}, I_d)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim N(\mathbf{0}, \Delta I_d)$ were generated, where $\Delta = 1,\ldots,3$ in steps of 0.10 and

$$I_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Figure 3.2 displays the results obtained and in this case, for equal sample sizes the $RCTB$ test does not compare favorably to the other considered tests, while the M test outperforms all tests. The same results hold for unequal sample sizes. Also generated was $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim N_d(\mathbf{0}, \Sigma)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim N_d(\mathbf{0}, \Delta\Sigma)$, where $\Delta = 1,\ldots,3$ in steps of 0.10 and

$$\Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 2 \end{bmatrix}.$$

Figure 3.7 displays the results obtained in this case presenting similiar results as Figure 3.2. For both equal and unequal sample sizes the $RCTB$ test performs better than the KS test for alternatives between 1 and 2.

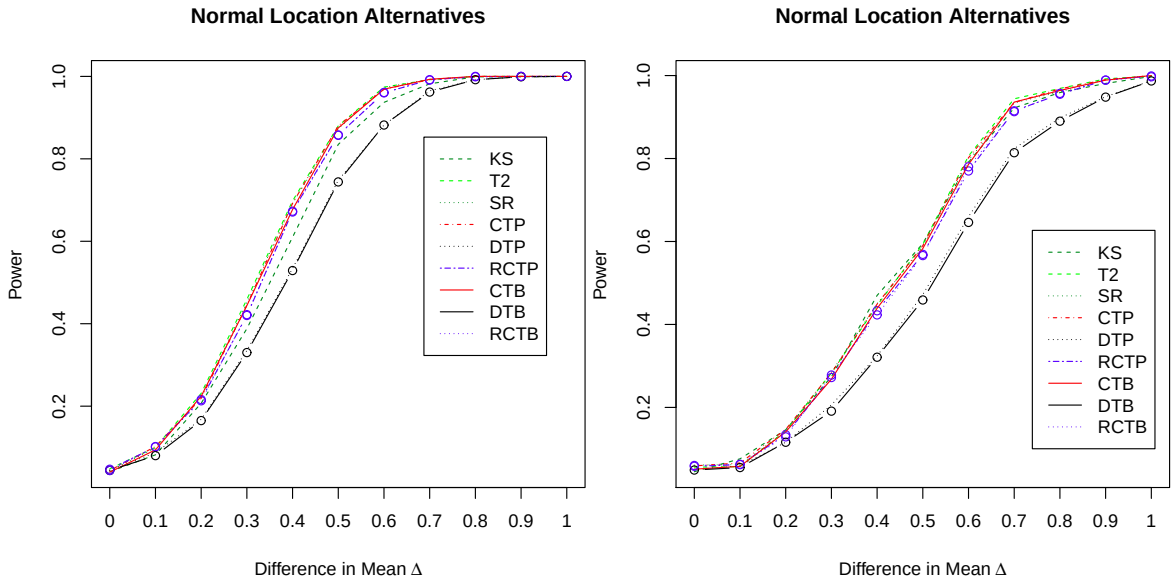
For the multivariate t-distribution with $df = 3$, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim t_d(\mathbf{0}, I_d)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim t_d(\mathbf{0}, \Delta I_d)$ were generated, with $\Delta = 1,\ldots,3$ in steps of 0.10. Also similar samples

when $df = 1$ were generated. The results obtained for $df = 3$ are displayed in Figure 3.4 and it is shown that similar results are displayed in Figure 3.2 for the multivariate normal distribution (for both equal and unequal sample sizes). When $df = 1$, we see in Figure 3.6 that the $RCTB$ test is not as powerful as the other considered tests other than the CTB test for equal sample sizes. It is necessary to note that the permutation method of the CT test performs better than the bootstrap method of the CT test. The M test remains the most powerful of all tests for equal and unequal sample sizes. Similar to the multivariate normal case, $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim t_d(\mathbf{0}, \Sigma)$ and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim t_d(\mathbf{0}, \Delta\Sigma)$ were generated , where $\Delta = 1, \ldots, 3$ in steps of 0.10. Figure 3.8 shows the obtained results. The results are similar to Figure 3.4, but the $RCTB$ test is now comparable to the CTB test and the KS test for unequal sample sizes.

For the scale alternatives, the proposed test does not compare as favorably to the other considered tests as seen in the case of location alternatives for the multivariate normal distribution and t-distribution. However, in the case of the Pareto distribution, the proposed test is more favorable for both equal and unequal sample sizes. As in the case for location, the vector samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ are formed by generating univariate samples $X_{11},\ldots,X_{1n}$ $\sim$ Pa(2, 2) and $X_{21},\ldots,X_{2n} \sim$ Pa(2, 2), and $Y_{11},\ldots,Y_{1m} \sim$ Pa(2, 2*Δ) and $Y_{21},\ldots,Y_{2m}$ $\sim$ Pa(2, 2*Δ), respectively. First, for the equal sample sizes, looking at Figure 3.12, for Pareto samples $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n \sim$ Pa(2, 2) and $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m \sim$ Pa(2, 2*Δ), with $\Delta = 1, \ldots, 3$ in steps of 0.10, the power performance of the $RCTB$ test is slightly higher than the CTP test, while outperforming the remaining considered tests. The difference in the power performance between the $RCTB$ test and the CTB test can be as large as 16%. One recognizes that all considered tests outperform the M test by a large margin. In the case of unequal sample sizes, the $RCTB$ test outperforms all considered tests with the difference in power performance compared to that of the CTB test can be as large as 50%.

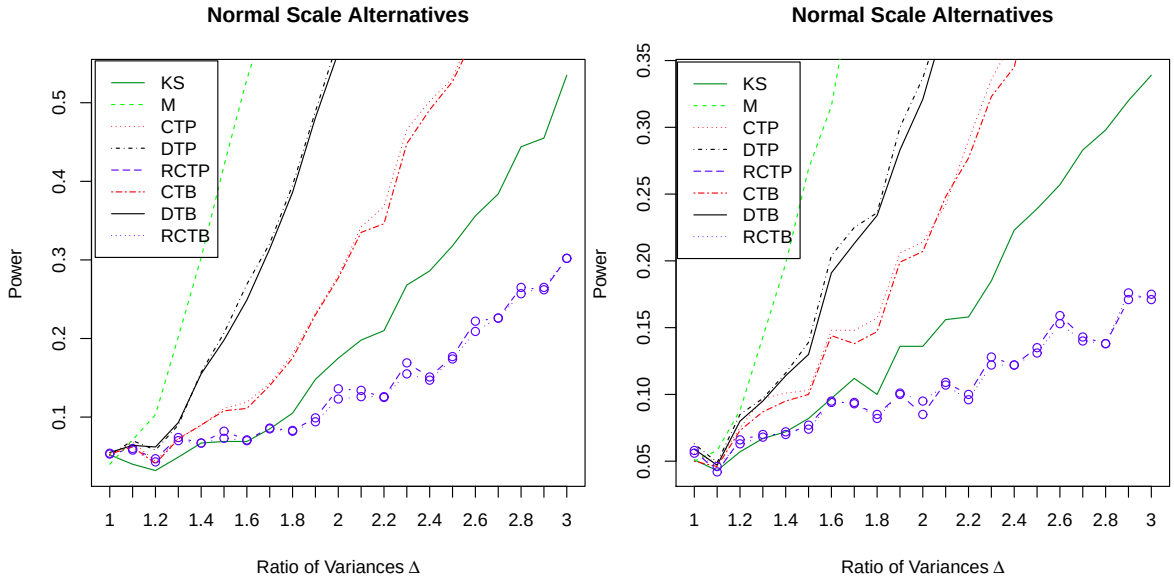
For the proposed test T, the simulations were repeated for all distributions discussed above in the case of scale alternatives. Simulations were only conducted for equal sam-

ple sizes and only the identity matrix is used for the scale change. The test statistic T is compared to CT and $T1$, i.e. RCT, as shown in Figures 3.13(b) - 3.16. For the normal distribution, the power performance of CT is higher than that of the proposed tests, although $T1$ is higher than T. When the distribution is the t-distribution for $df = 1$, the statistics $T1$ and T outperforms CT, however the power performance of $T1$ is the highest of the three. Conversely, when the distribution is the t-distribution for $df = 3$, CT outpeforms both $T1$ and T where the difference in the power performance can be as large as 44% between CT and T. For the Pareto distribution, $T1$ outperforms both CT and T, and T performs higher than CT.



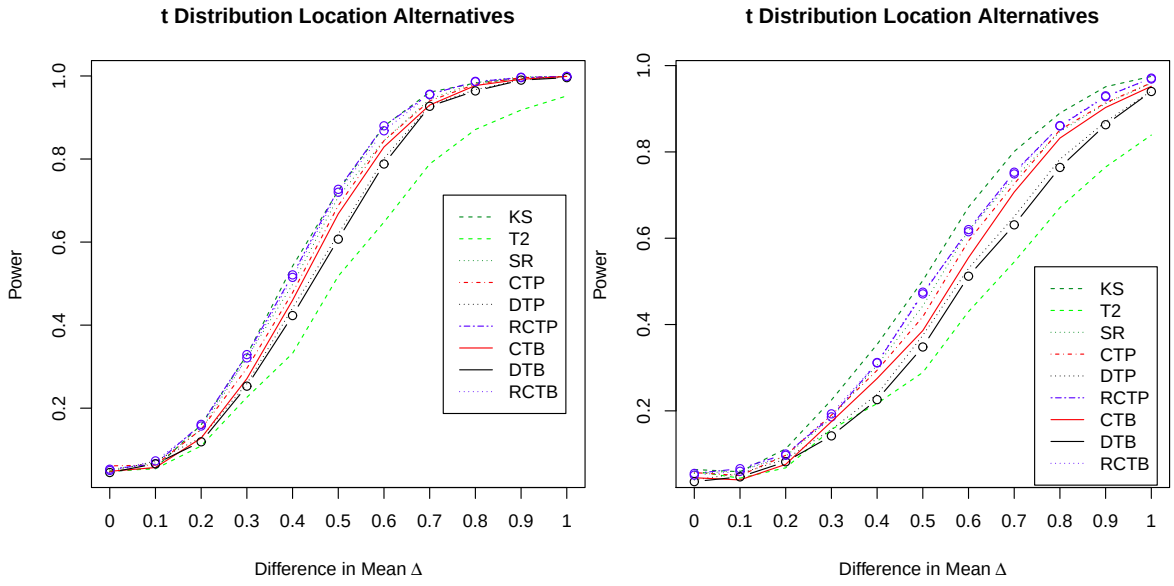
(a) (b)

Figure 3.1. Power performance for multivariate normal distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



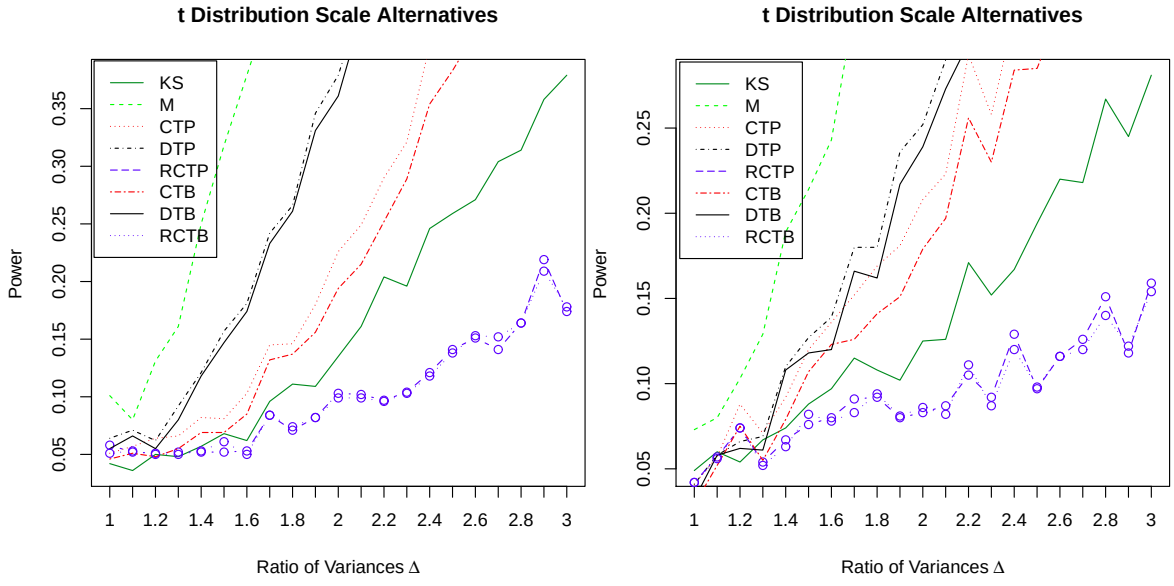
(a) (b)

Figure 3.2. Power performance for multivariate normal distribution scale alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**, (here variance is the identity matrix).**



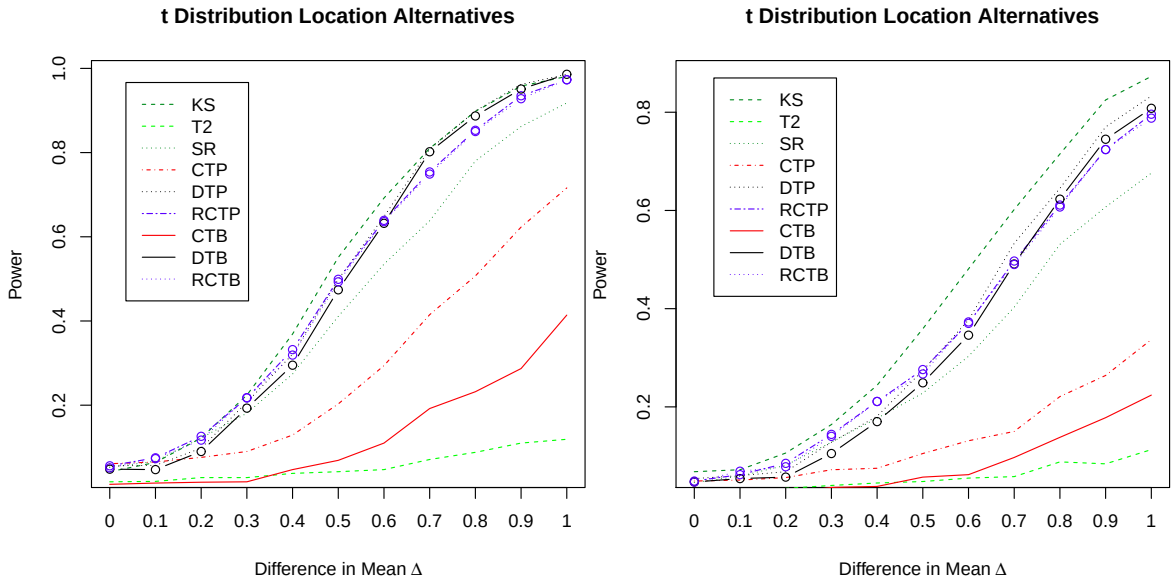
(a) (b)

Figure 3.3. Power performance for multivariate t-distribution location alternatives, $df = 3$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



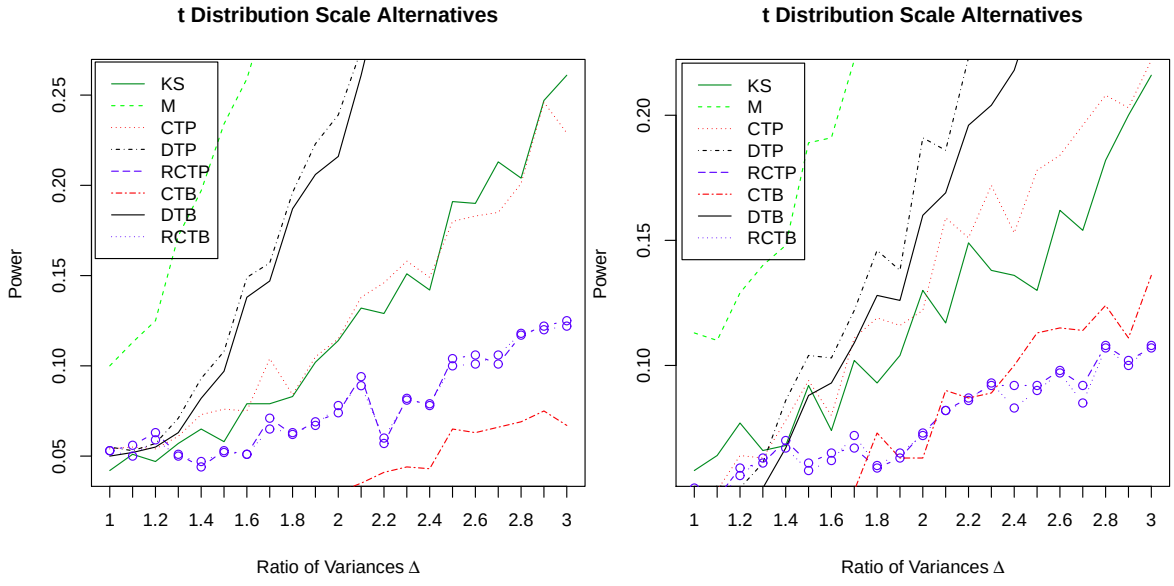
(a) (b)

Figure 3.4. Power performance for multivariate t-distribution scale alternatives, $df = 3$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$, (here variance is the identity matrix).



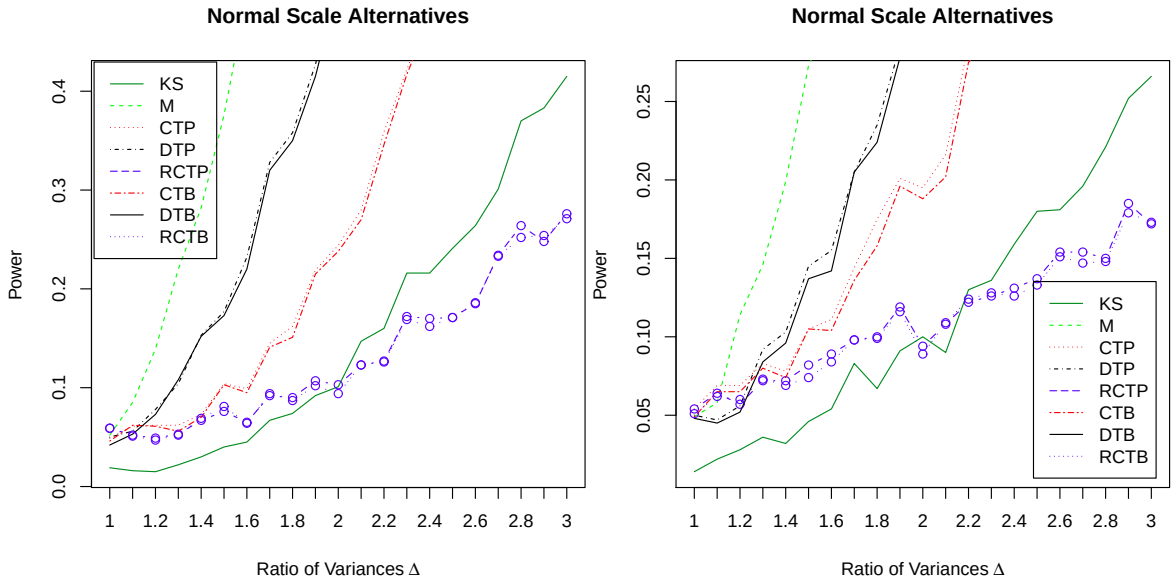
(a) (b)

Figure 3.5. Power performance for multivariate t-distribution location alternatives, $df = 1$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$.



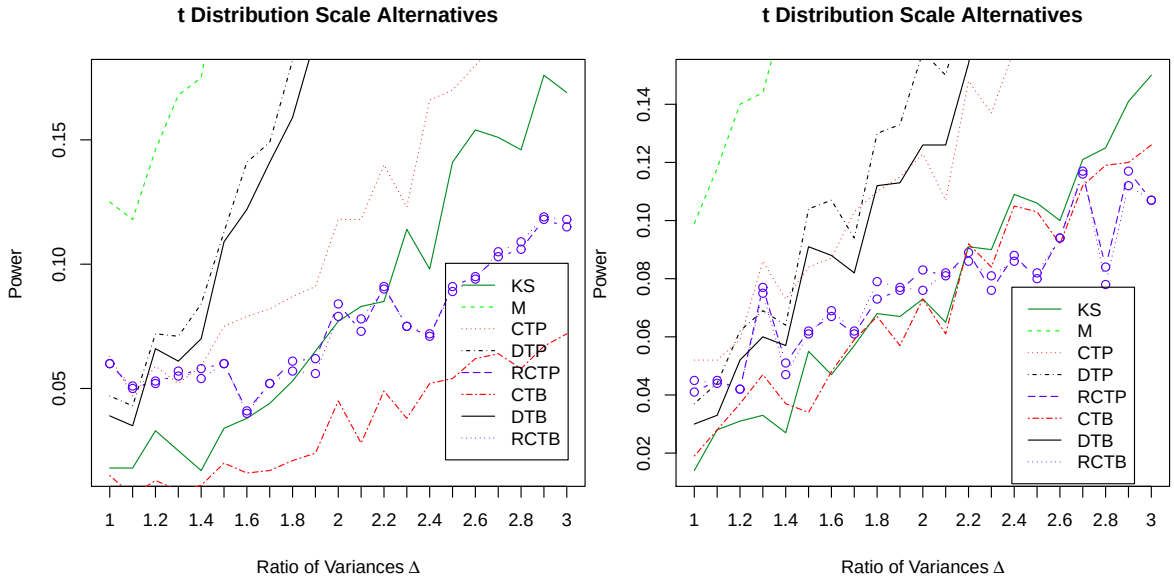
(a) (b)

Figure 3.6. Power performance for multivariate t-distribution scale alternatives, $df = 1$. (a): equal sample size, $n = m = 50$; (b): unequal sample size, $n = 50$ and $m = 20$, (here variance is the identity matrix).



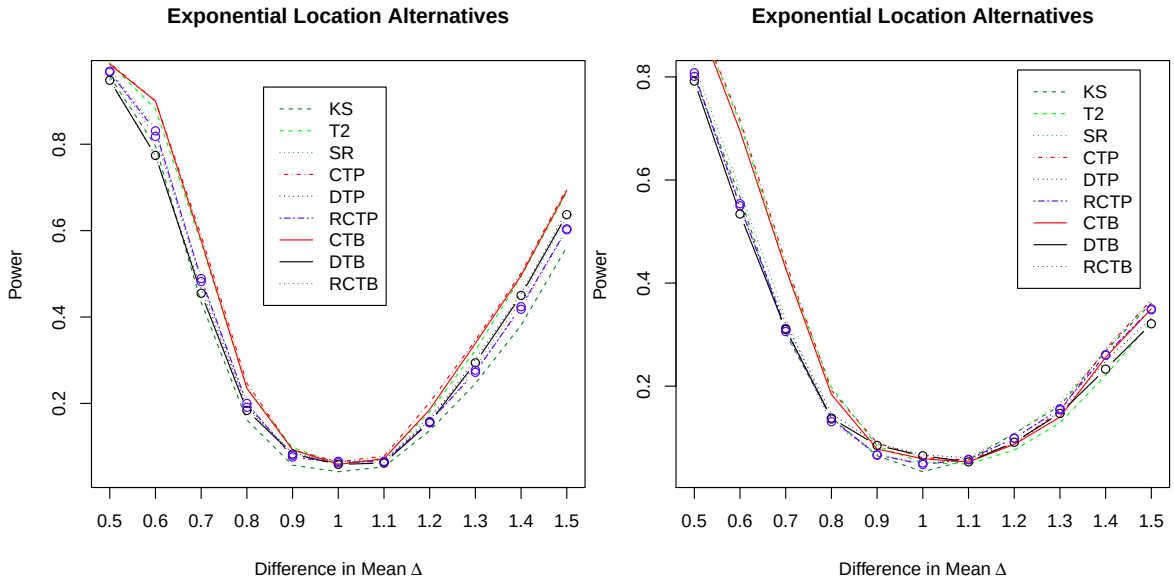
(a) (b)

Figure 3.7. Power performance for multivariate normal distribution scale alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**, (here variance is** Σ **)**



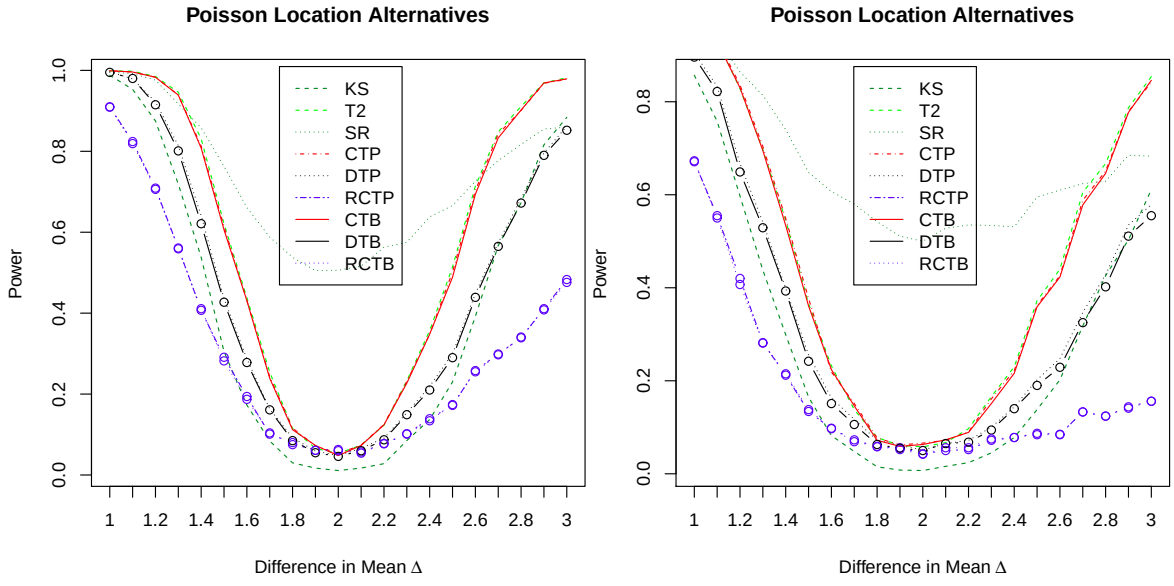
(a) (b)

Figure 3.8. Power performance for multivariate t**-distribution scale alternatives,** $df = 1$**.** (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**, (here variance is** Σ **).**



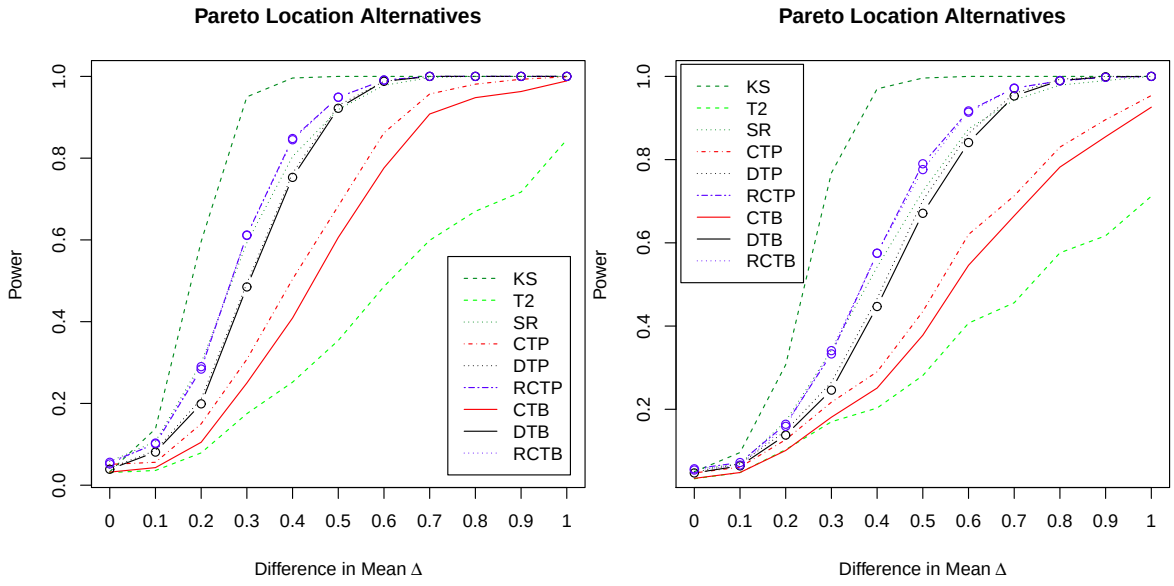
(a) (b)

Figure 3.9. Power performance for multivariate exponential distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



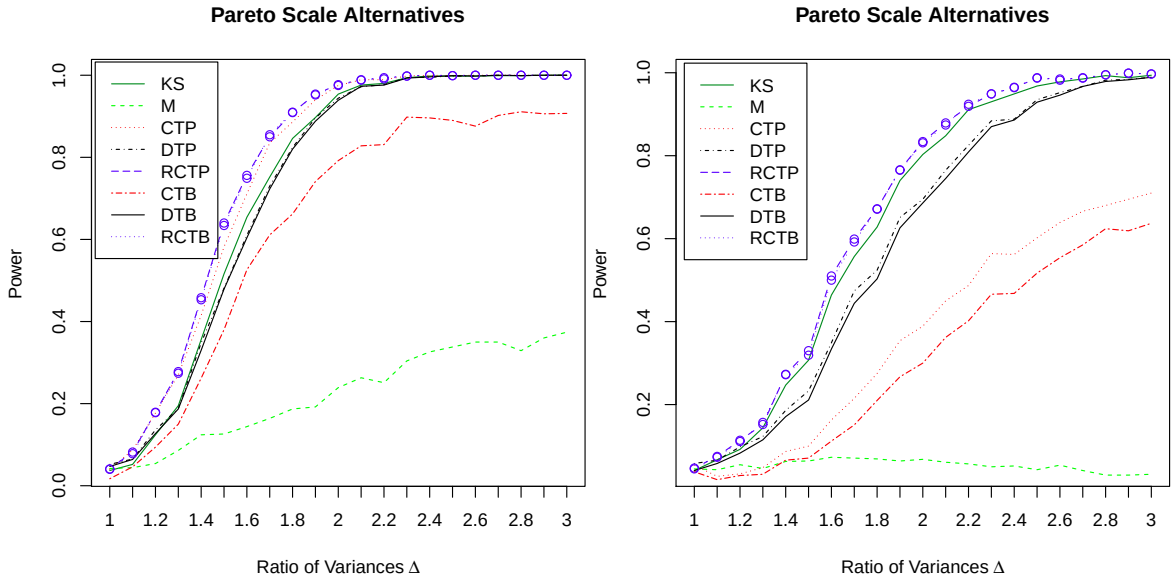
(a) (b)

Figure 3.10. Power performance for multivariate Poisson distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



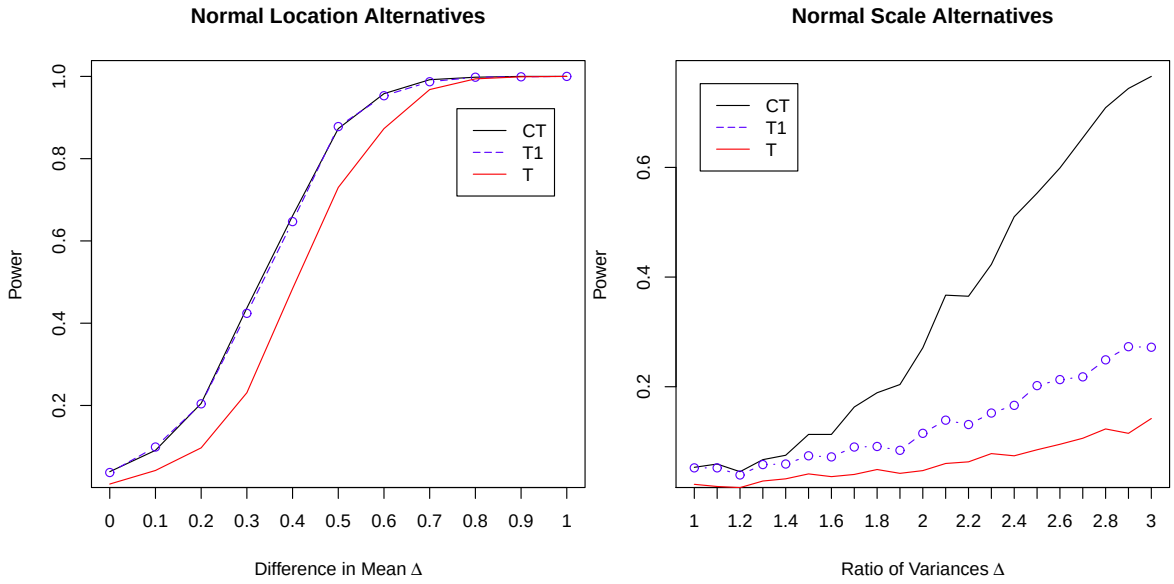
(a) (b)

Figure 3.11. Power performance for multivariate Pareto distribution location alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



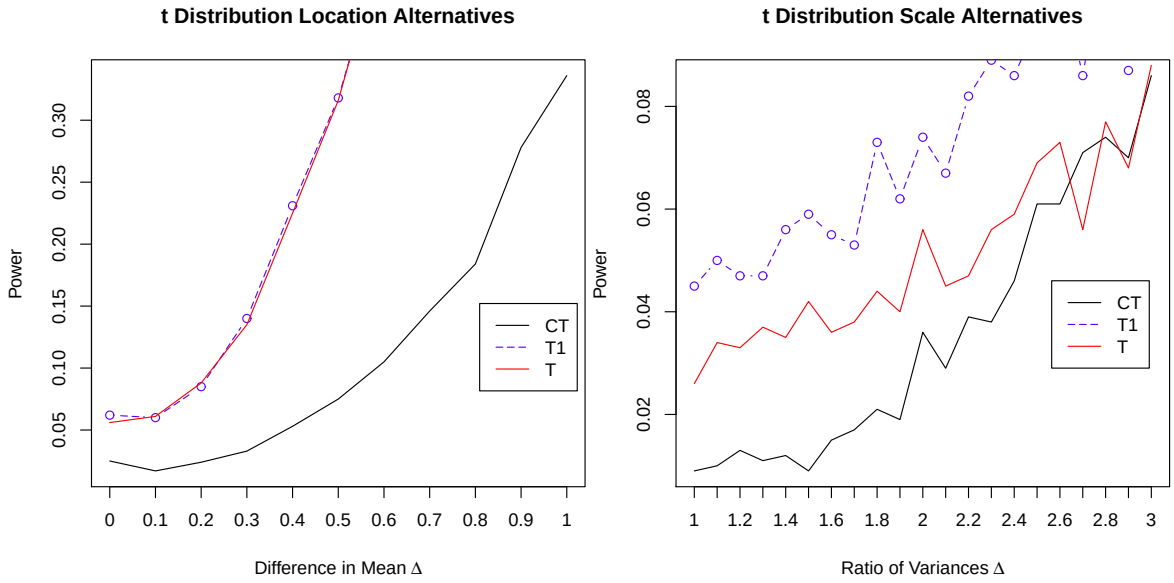
(a) (b)

Figure 3.12. Power performance for multivariate Pareto distribution scale alternatives. (a)**: equal sample size,** $n = m = 50$**;** (b)**: unequal sample size,** $n = 50$ **and** $m = 20$**.**



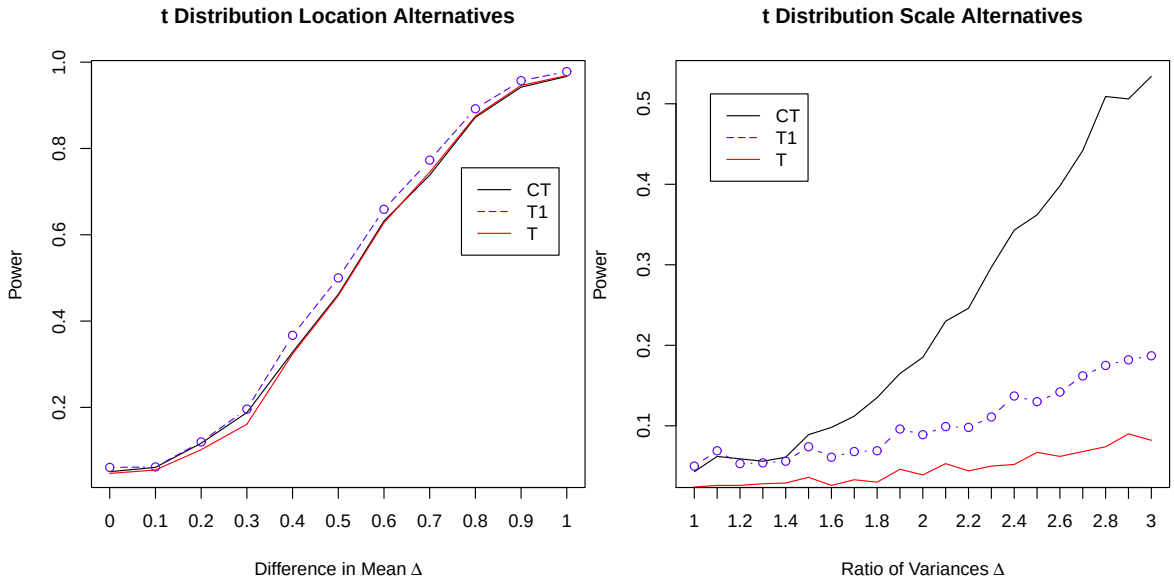
(a) (b)

Figure 3.13. Power performance for Normal distribution for location and scale alternatives with equal sample size, $n = m = 50$. (a): Location alternatives; $(b)t$: Scale alternatives.



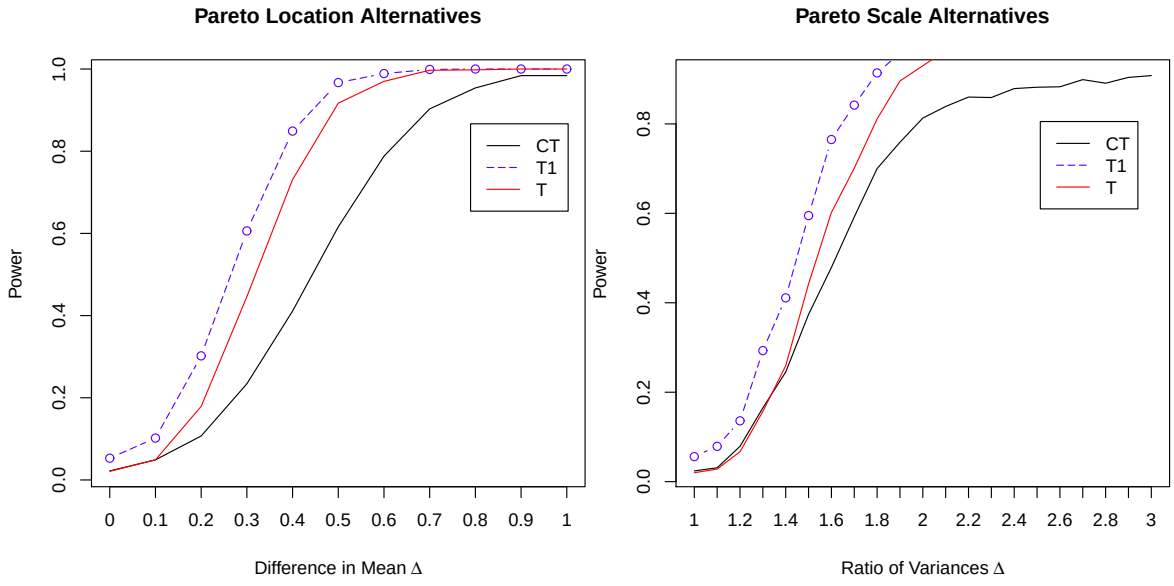
(a) (b)

Figure 3.14. Power performance for t distribution for location and scale alternatives with equal sample size, $n = m = 50$, $df = 1$. (a): Location alternatives; (b): Scale alternatives.



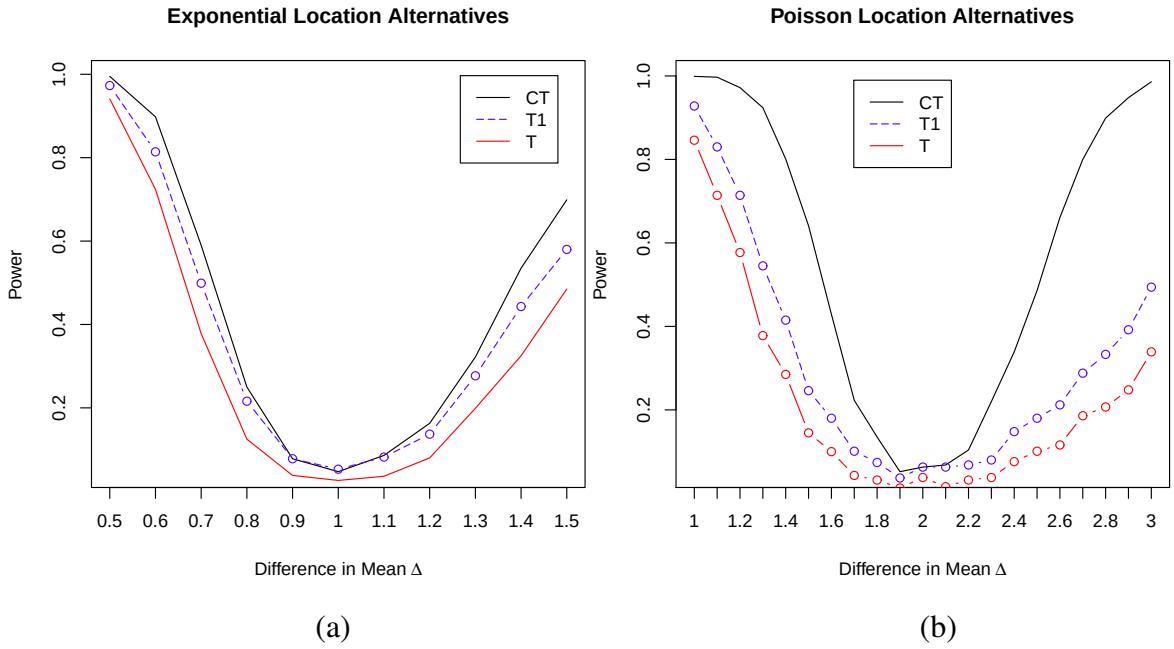
(a) (b)

Figure 3.15. Power performance for t distribution for location and scale alternatives with equal sample size, $n = m = 50$, $df = 3$. (a): Location alternatives; (b): Scale alternatives.



(a) (b)

Figure 3.16. Power performance for Pareto distribution for location and scale alternatives with equal sample size, $n = m = 50$. (a): Location alternatives; (b): Scale alternatives.



(a) (b)

Figure 3.17. Power performance for Exponential and Poisson distribution for location alternatives with equal sample size, $n = m = 50$. (a): Exponential location alternatives; (b): Poisson location alternatives.

CHAPTER 4

SUMMARY AND FUTURE WORK

4.1 Summary and Conclusions

The problem of testing whether two samples come from the same or different population is a classical one in statistics. In this dissertation, I first study rank based formulation of univariate two-sample distribution-free tests. One form of the test statistic is the average of between-group distances of ranks. The other form of the test statistic is the difference between the average of between-group distances of ranks and the average of within-group distances of ranks. Although they are different in formulation, they are closely related to the two-sample Cramér-von Mises criterion. The first one is a linear transformation of Cramér-von Mises criterion in the case the two samples have the same sample size. The second one is a different form of the Cramér-von Mises criterion.

The properties of the two-sample test statistic based on the new formulations are studied. In particular, the Hájek projection and orthogonal decomposition technique are applied in deriving the asymptotics of the test statistic. For the statistic T under the balanced case, its limiting distribution is not normal since the projection on one variable is insufficient to represent the variation of the test statistic. By taking the projection on two variables, it was proved to be a weighted mixture of independent chi-square distributions. An operator in the functional space was defined and its eigenfunctions and eigenvalues were applied to derive the limiting distribution.

Rank-based formulations allow generalizations of two-sample Cramér-von Mises test to the multivariate case by using different notions of multivariate rank functions. In the

multivariate case, the rank tests may lose the distribution-free property under a general alternative. They are, however, usually more robust than the parametric tests. I propose two corresponding new tests based on multivariate spatial ranks. The spatial rank function yields a relative center-outward ranking of a data set. It preserves not only ordering on the magnitude of vectors but also directional information, and it characterizes the distribution. Similar to the univariate case, one test statistic is the difference between the average of intra-sample rank distances and the average of inter-sample rank distances. The other one is simply the average of intra-sample rank distances for the balanced samples. Unlike in the univariate case, those two statistics are no longer equivalent. Comparing with other tests, the proposed tests can be established by the following desirable properties. (1) They are nonparametric with fewer assumptions, although they are not completely distribution-free. (2) They are invariant with respect to orthogonal linear transformations, which doesn't hold for tests based on the component-wise ranks. (3) They are consistent against all alternatives. The simulation results have illustrated the proposed tests promising. The bootstrap and permutation procedures are used for yielding a consistent approximation to the null distribution of the test statistics.

4.2 Future Work

In this study, I extend the rank tests based on the spatial rank because of its statistical efficiency, computational ease. But it is only orthogonally equivariate. To obtain affine equivariate property of the test, transformation and retransformation technique (TR) will be applied. A study of TR spatial rank based tests is a continuation of this work.

One other research direction of this work is to study multiple sample problem. Rather than making decision on whether two samples are from the same population distribution, multiple sample problem deals with more than two samples. A simple extension from a two-sample problem to a multiple sample problem is to consider all combinations of a two-sample problem, and then apply the Bonferroni multiple comparison procedure to control

the type I error. However, more direct and efficient extensions are possible. I will continue to explore rank-based tests for multivariate multiple sample problem.

BIBLIOGRAPHY

BIBLIOGRAPHY

ANDERSON, T. W. (1962) "On the distribution of the two-sample Cramér-von Mises criterion." *Annals of Mathematical Statistics*, Vol. 33(3), pp. 1148–1159.

ARCONES, M. A. & GINÉ, E. (1992) "On the bootstrap of U and V statistics." *Annals of Statistics*, Vol. 20, pp. 655–674.

BARINGHAUS, L. & FRANZ, C. (2004) "On a new multivariate two-sample test." *Journal of Multivariate Analysis*, Vol. 88, pp. 190–206.

BAUMGARTNER, W.; WEISS, P.; & SCHINDLER, H. (1998) "A nonparametric test for the general two sample problem." *Biometrics*, Vol. 54, pp. 1129–1135.

BORRONI, C. G. (2001) "Some notes about the nonparametric tests for the equality of two population." *Test*, Vol. 10(1), pp. 147–159.

BRADLEY, J. V. (1968) *Distribution-free Statistical Tests*. Prentice-Hall, NJ.

DARLING, D. A. (1957) "The Kolomogorov-Smirnov, Cramér-von Mises tests." *Annals of Mathematical Statistics*, Vol. 28(4), pp. 823–838.

DUNFORD, N. & SCHWARTZ, J. T. (1963) *Linear operators Part II: Spectral theory*. Self adjoint operators in Hilbert space.

EFRON, B. & STEIN, C. (1978) "The jackknife estimate of variance." *Technical Report No.40*, Vol. 40.

FISZ, M. (1960) "On a result by M." *Rosenblatt concerning the von Mises - Smirnov Test. Annals of Mathematical Statistics*, Vol. 31(2), pp. 427–429.

FREUND, J. E. & ANSARI, A. R. (1957) "Two-way rank sum tests for variance." *Virgina Polytechnic Institute Technical Report to Office of Ordnance Research abd National Science Foundation*, Vol. 34.

GRETTON, A.; BORGWARDT, K. M.; RASCH, M. J.; SCHÖLKOPF, B.; & SMOLA, A. (2008) "A kernel method for the two-sample problem." *Journal of Machine Learning Research*, Vol. 1, pp. 1–10.

HÁJEK, J. & ŠIDÁK, Z. (1967) *Theory of Rank Tests*. Academic Press.

HETTMANSPERGER, T. P. & MCKEAN, J. W. (2010) *Robust Nonparametric Statistical Methods, 2nd edition*. Chapman & Hall.

HOEFFDING, W. (1951) "Optimum nonparametric tests." *Proceedings of the Second Berkeley Symposium on Mathematical Staistics and Probability, ed. by Neyman, J.*, Vol. pages 83–92.

LEHMANN, E. L. (1951) "Consistency and unbiasedness of certain nonparametric tests." *Annals of Mathematical Statistics*, Vol. 22, pp. 165–179.

MANN, H. B. & WHITNEY, D. R. (1947) "On a test whether one of two random variable is stochastically larger than the other." *Annals of Mathematical Statistics*, Vol. 18, pp. 50–60.

MOOD, A. M. (1954) "On the asymptotic efficiency of certain nonparametric two-sample tests." *Annals of Mathematical Statistics*, Vol. 25, pp. 514–522.

OJA, H. (2010) *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer.

OJA, H. & RANDLES, R. H. (2004) "Multivariate nonparametric tests." *Institute of Mathematical Statistics*, Vol. 19(4), pp. 598–605.

PETTITT, A. N. (1976) "A two-sample Anderson-Darling rank statistic." *Biometrika*, Vol. 63(1), pp. 161–168.

ROSENBLATT, M. (1952) "Limit theorems associated with variants of the von Mises statistic." *Annals of Mathematical Statistics*, Vol. 23, pp. 617–623.

SCHMID, F. & TREDE, M. (1995) "A distribution free test for the two sample problem for general alternatives." *Computational Statistics and Data Analysis*, Vol. 20, pp. 409–419.

SERFLING, R. (1980) *Approximation Theorems of Mathematical Statistics*. Wiley.

SIEGEL, S. & TUKEY, J. W. (1960) "A nonparametric sum of ranks procedure for relative spread in unpaired samples." *Journal of the American Statistical Association*, Vol. 55, pp. 429–445.

SMIRNOV, N. V. (1939) "Estimate of deviation between empirical distribution functions in two independent samples." *Bulletin of Moscow University*, Vol. 2, pp. 3–16.

SZÉKELY, G. J. & RIZZO, M. L. (2013) "Energy statistics A class of statistics based on distances." *Journal of Statistical Planning and Inference*, Vol. 143, pp. 1249–1272.

TERRY, M. (1952) "Some rank order tests which are most powerful against specific parametric alternatives." *Annals of Mathematical Statistics*, Vol. 23, pp. 346–366.

VAN DER WAERDEN, B. (1952) "Order tests for the two-sample problem and their power." *Indagationes Mathematics*, Vol. 14, pp. 453–458.

WILCOXON, F. (1945) "Individual comparisons by ranking methods." *Biometrics*, Vol. 1, pp. 80–83.

VITA

EDUCATION

M.S., Mathematics, Minor in Statistics, Mississippi State University, May 2008

B.S., Mathematics, University of Mississippi, May 2006

TEACHING EXPERIENCE

Graduate Instructor, August 2008 - Present

University of Mississippi

Courses: Calculus I, Business Calculus I and II, Trigonometry, College Algebra, Elementary Statistics

Graduate Teaching Assistant, January 2007 - May 2008

Mississippi State University

Course: College Algebra

FELLOWSHIPS

The Graduate Assistance in Areas of National Need (GAANN) Fellowship, 2009 - Present

The University of Mississippi Minority Fellowship, 2008 - 2009

UNIVERSITY SERVICE

Treasurer, American Mathematical Society, University of Mississippi Chapter, 2013 - Present

Mentor, Increasing Minority Access to Graduate Education Program, 2010 - Present

Graduate Panelist, Alliance for Graduate Education in Mississippi, 2010

Competition Judge, The University of Mississippi Brain Brawl Tournament, 2009

PRESENTATION

Contributed talk ”On a New Distribution-Free Two-Sample Test”, 2013 Joint Statistical Meetings, Montreal, Canada, August 2013