RANK-SUM TEST FOR TWO-SAMPLE LOCATION PROBLEM UNDER ORDER RESTRICTED RANDOMIZED DESIGN

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Yiping Sun, M.S.

* * * * *

The Ohio State University

2007

Dissertation Committee:

Professor Omer Ozturk, Adviser
Professor Steven N. MacEachern
Professor Haikady Nagaraja
Professor Douglas A. Wolfe

Approved by

Adviser
Department of Statistics
ABSTRACT

There are many experimental settings, where experimental units have abundance of information. This information is usually available in two forms, either in formal measurements or in informal observations. While the formal measurements are successfully used in traditional analyses as covariates, the informal observations are usually ignored. The order restricted randomized design (ORRD) exploits the use of these informal observations (subjective information) to design an experiment. Sets of experimental units are recruited from a population along with subjective information that they may have. This subjective information is then used to create artificial covariates through judgment ranking of the experimental units. Artificial covariates, with restricted randomization of treatment regimes to experimental units, induce a positive correlation structure among within-set response measurements. This positive correlation structure then acts as a variance reduction technique in the inference of a contrast parameter in an ORRD. This dissertation develops statistical inference based on ORRD for the location shift between two populations.

Chapter 1 provides a review for existing designs in the literature that are closely connected to ORRD. Chapter 2 introduces a new nonparametric test based on the ORRD for the location shift between two populations. Sections 2.1 and 2.2 develop an asymptotic theory for the null distribution of the test statistic. Section 2.3 constructs an optimal design that maximizes the asymptotic Pitman efficacy of the
proposed test. Section 2.4 shows that the size of the test is inflated if the design has some judgment ranking error. Section 2.5 develops point and interval estimates for the location shift parameter.

Chapter 3 develops an asymptotic theory under imperfect ranking and provides a calibration technique to reduce the impact of ranking error. It is shown that the test performs quite well even under imperfect ranking with this calibration. Chapter 4 provides simulation results for the empirical power of the test. Chapter 5 applies the proposed procedure to a clinical trial to draw inference on the difference between control and treatment regimes. Finally Chapter 6 provides some concluding remarks and discusses some open problems for future work.
This work is dedicated to my parents, Quan and Wenrong, and my husband, Xin.
ACKNOWLEDGMENTS

I thank my adviser, Professor Omer Ozturk, for his advice and encouragement. Without his help, I could not have finished this work.

I also thank Professors Steve N. MacEachern, Douglas Wolfe and Haikady Nagaraja for serving on my dissertation committee and their great comments and suggestions.

I thank all of my professors, fellow students, and department staff for their kind support.
VITA

1994 ........................................ Bachelor in Clinical Medicine, Nanjing Medical University
1994-2000 ................................. Physician, Guangzhou Blood Center
2003 .......................... M.S. in Statistics, The Ohio State University
2002-present ......................... Graduate Teaching and Research Associate, The Ohio State University.

PUBLICATIONS


Sun, Y. and Tian, Z. “Therapeutic Plasma Collection”. The National Symposium of Latest Techniques in Transfusion Medicine, 1996

FIELDS OF STUDY

Major Field: Statistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Ranked set sampling</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Order restricted randomized designs</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1</td>
<td>Description for ORRD</td>
<td>11</td>
</tr>
<tr>
<td>1.2.2</td>
<td>Properties of ORRD</td>
<td>14</td>
</tr>
<tr>
<td>2.</td>
<td>RANK-SUM TEST UNDER PERFECT RANKINGS</td>
<td>20</td>
</tr>
<tr>
<td>2.1</td>
<td>Asymptotic null distribution of test statistic $T$</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Asymptotic properties of $T$ under perfect ranking</td>
<td>38</td>
</tr>
<tr>
<td>2.3</td>
<td>Optimal designs</td>
<td>51</td>
</tr>
<tr>
<td>2.4</td>
<td>Simulation study for the ORRD</td>
<td>66</td>
</tr>
<tr>
<td>2.5</td>
<td>The point and interval estimates for location shift parameter $\Delta$</td>
<td>70</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Balanced rank set sample when $H = 4$ and $n = 2$</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Unbalanced ranked set sample when $H = 4$, $n_1 = 1$, $n_2 = 3$, $n_3 = 3$ and $n_4 = 1$</td>
<td>4</td>
</tr>
<tr>
<td>1.3 One ORRD sample with replicate size $n$, set size $H$, sets $\alpha$ and $\beta$</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Brief comparisons among RSS, ORRD and CRD</td>
<td>18</td>
</tr>
<tr>
<td>2.1 The asymptotic relative efficiency (ARE)</td>
<td>65</td>
</tr>
<tr>
<td>2.2 Examples for ARE</td>
<td>65</td>
</tr>
<tr>
<td>2.3 The estimated type I error rates under perfect ranking. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.</td>
<td>69</td>
</tr>
<tr>
<td>2.4 Type I error rates under imperfect ranking when $n = 5$ and underlying distribution is standard normal $N(0,1)$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.</td>
<td>70</td>
</tr>
<tr>
<td>3.1 The estimated Type I error rates when $\rho = 1.0$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.</td>
<td>86</td>
</tr>
<tr>
<td>3.2 The estimated Type I error rates when $\rho = 0.9$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.</td>
<td>87</td>
</tr>
<tr>
<td>3.3 The estimated Type I error rates when $\rho = 0.75$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.</td>
<td>88</td>
</tr>
</tbody>
</table>
3.4 The estimated Type I error rates when $\rho = 0.5$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates. 89

5.1 The sampled ORRD data for week 4 CD4 cell counts . . . . . . . . . . 116
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size (H) and replicate size ($n$). Data sets are generated from N(0,1). Perfect ranking is used and set sizes (H) are marked on each curve.</td>
<td>95</td>
</tr>
<tr>
<td>4.2</td>
<td>Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size (H) and replicate size ($n$). Data sets are generated from t(3). Perfect ranking is used and set sizes (H) are marked on each curve.</td>
<td>96</td>
</tr>
<tr>
<td>4.3</td>
<td>Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size (H) and replicate size ($n$). Data sets are generated from LN(0,1). Perfect ranking is used and set sizes (H) are marked on each curve.</td>
<td>97</td>
</tr>
<tr>
<td>4.4</td>
<td>Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (N(0,1)).</td>
<td>98</td>
</tr>
<tr>
<td>4.5</td>
<td>Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (N(0,1)).</td>
<td>99</td>
</tr>
<tr>
<td>4.6</td>
<td>Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 10 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (N(0,1)).</td>
<td>100</td>
</tr>
<tr>
<td>4.7</td>
<td>Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 10 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (N(0,1)).</td>
<td>101</td>
</tr>
</tbody>
</table>
4.8 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size $(n) = 5$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$). . . . . . . . . 102

4.9 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size $(n) = 5$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$). . . . . . . . . 103

4.10 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size $(n) = 10$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$). . . . . . . . . 104

4.11 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size $(n) = 10$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$). . . . . . . . . 105

4.12 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size $(n) = 5$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($LN(0,1)$). . . . . . . . . 106

4.13 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size $(n) = 5$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($LN(0,1)$). . . . . . . . . 107

4.14 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size $(n) = 10$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($LN(0,1)$). . . . . . . . . 108

4.15 Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size $(n) = 10$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($LN(0,1)$). . . . . . . . . 109
5.1 Scatter plots of the CD4 cell counts at the screening stage (Baseline CD4) versus at week 4. . . . . . . . . . . . . . . . . . . . . . . . . . . 114

5.2 Histograms of the CD4 cell counts at week 4 for each treatment regime and the pooled data. . . . . . . . . . . . . . . . . . . . . . . . . 115
CHAPTER 1

INTRODUCTION

Inference on a two-sample location shift parameter is one of the common problems in many observational and experimental studies. Researchers are particularly interested in drawing inference for the location shift between two populations when their distributions have the same, but arbitrary shape. A lot of parametric and nonparametric inferential procedures are developed for this type of problem. Most of these methods are constructed based on a data set collected through simple random sampling (SRS) from the two populations. This sampling procedure ignores potentially available subjective information in the experimental units.

In general, two types of information may be available in a potential experimental unit (EU). The first type of information may be collected as covariates along with the actual response variables. These covariates are usually precise, well-defined and numerically quantifiable. The use of these covariates has been exploited in regression and analysis of covariance models to improve the statistical inference. If the models are correct, the use of covariates considerably reduces the total error variation in the experiment. The second type of information is not well defined and cannot be converted into covariates easily. It is usually subjective, imprecise, vague and rough
information contained in the EUs. As a result, use of this type of information has been ignored in experiments. Although subjective information may be questionable in quality, it is extremely useful in reducing the total variation in an experiment. As an example, consider a study of efficacy comparison between a new drug and an existing one. In this study, there may be two types of information. The first type may include age, gender, weight, etc. These can be used as covariates in regression and covariance models or as blocking factors in the design of the experiments. The second type of information may include general health status and pre-medical history of patients. This information in general may be incomplete, subjective, imprecise, even biased, and it may not be converted to numerical numbers or categories easily. Nonetheless, the information is very useful to improve statistical inference if it is used properly at the design and analysis stages of the experiment.

1.1 Ranked set sampling

Ranked set sampling (RSS) is a procedure, introduced by McIntyre [26], that uses subjective information at the design stage of an experiment to produce an artificially stratified sample. Due to recent interest, it is republished in McIntyre [27]. This alternative data collection procedure yields improved efficiency over the corresponding SRS procedure.

To create a ranked set sample in its original form, $H$ mutually independent sets of experimental units, each of size $H$, are selected at random from an infinite population. Then within each set, units are ranked from smallest to largest based on available subjective information without full measurements. Finally, from the $h^{th}$ set, the unit with judgment rank $h$ is selected for full measurement ($1 \leq h \leq H$). Thus, in total,
$H$ independent units for full measurement are collected, one from each rank. This procedure is called a cycle. The set size $H$ is a design parameter predetermined by the researcher and is usually kept small ($H \leq 5$) to control possible ranking errors. The above process is repeated for $n$ cycles to obtain $N = nH$ fully measured observations. We illustrate the process with $H = 4$ and $n = 2$ in Table 1.1. Note that in Table 1.1 the bold $X$‘s, $X_{11}^{[1]}, X_{21}^{[1]}, X_{31}^{[3]}, X_{41}^{[1]}, X_{12}^{[1]}, X_{22}^{[2]}, X_{32}^{[3]},$ and $X_{42}^{[4]}$, are selected as a ranked set sample, where $[.]$ indicates the judgment rank of each unit. Since the number of observations in each judgment class population is equal, these fully measured observations are called a balanced ranked set sample (BRSS).

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Set 1</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>$X_{11}^{[1]}$</td>
<td>$X_{21}^{[2]}$</td>
<td>$X_{31}^{[3]}$</td>
<td>$X_{41}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 2</td>
<td>$X_{12}^{[1]}$</td>
<td>$X_{22}^{[2]}$</td>
<td>$X_{32}^{[3]}$</td>
<td>$X_{42}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 3</td>
<td>$X_{13}^{[1]}$</td>
<td>$X_{23}^{[2]}$</td>
<td>$X_{33}^{[3]}$</td>
<td>$X_{43}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 4</td>
<td>$X_{14}^{[1]}$</td>
<td>$X_{24}^{[2]}$</td>
<td>$X_{34}^{[3]}$</td>
<td>$X_{44}^{[4]}$</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>$X_{11}^{[1]}$</td>
<td>$X_{21}^{[2]}$</td>
<td>$X_{31}^{[3]}$</td>
<td>$X_{41}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 2</td>
<td>$X_{12}^{[1]}$</td>
<td>$X_{22}^{[2]}$</td>
<td>$X_{32}^{[3]}$</td>
<td>$X_{42}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 3</td>
<td>$X_{13}^{[1]}$</td>
<td>$X_{23}^{[2]}$</td>
<td>$X_{33}^{[3]}$</td>
<td>$X_{43}^{[4]}$</td>
</tr>
<tr>
<td></td>
<td>Set 4</td>
<td>$X_{14}^{[1]}$</td>
<td>$X_{24}^{[2]}$</td>
<td>$X_{34}^{[3]}$</td>
<td>$X_{44}^{[4]}$</td>
</tr>
</tbody>
</table>

Table 1.1: Balanced rank set sample when $H = 4$ and $n = 2$

If we relax the condition that the numbers of observations in all judgment classes are equal, we obtain an unbalanced ranked set sample. In order to construct an unbalanced ranked set sample, $NH$ experimental units are selected at random from an infinite population and these units are divided randomly into $N$ sets, each of size $H$. Within each set, units are ranked from smallest to largest. Then the $h$-th judgment ranked unit is selected for full measurement in $n_h$ sets for $h = 1, \ldots, H$.
so that $\sum_{h=1}^{H} n_h = N$. If $n_1 = n_2 = \ldots = n_H$, this is equivalent to BRSS. Otherwise, we obtain an unbalanced ranked set sample (URSS). The URSS is illustrated in Table 1.2 when $H = 4$, $n_1 = 1$, $n_2 = 3$, $n_3 = 3$, $n_4 = 1$. The bold $X$’s, $X_{[1]11}$, $X_{[2]21}$, $X_{[2]31}$, $X_{[2]42}$, $X_{[3]41}$, $X_{[3]12}$, $X_{[3]22}$, and $X_{[4]32}$, constitute a ranked set sample.

<table>
<thead>
<tr>
<th>Set</th>
<th>Rank 1</th>
<th>Rank 2</th>
<th>Rank 3</th>
<th>Rank 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set 4</td>
<td>$X_{[1]41}$</td>
<td>$X_{[2]41}$</td>
<td>$X_{[3]41}$</td>
<td>$X_{[4]41}$</td>
</tr>
<tr>
<td>Set 8</td>
<td>$X_{[1]42}$</td>
<td>$X_{[2]42}$</td>
<td>$X_{[3]42}$</td>
<td>$X_{[4]42}$</td>
</tr>
</tbody>
</table>

Table 1.2: Unbalanced ranked set sample when $H = 4$, $n_1 = 1$, $n_2 = 3$, $n_3 = 3$ and $n_4 = 1$

RSS is different from SRS in many aspects. One of the differences is that RSS exploits the use of the subjective information on experimental units. For each fully measured observation, it requires $H$ experimental units. This fully measured observation carries the information of the selected unit along with the partial information of the other $H - 1$ units, obtained through judgment ranking. The quality of judgment ranking determines the precision of this information. Since we need to recruit more experimental units in RSS than what is required in SRS for collecting a sample of the same size, it is clear that RSS would be appropriate in situations where measurements are expensive or time consuming while ranking is relatively easy and cheap.
Another difference between RSS and SRS is due to the fact that the RSS creates more structure in the sample than SRS does. Although all observations in a ranked set sample are independent, they do not follow the same distribution except for those sharing the same judgment ranks. In this respect, the RSS is similar to a stratified sample, where strata are constructed based on the perceived values of the judgment ranks of the experimental units. In stratified sampling, the whole population is separated into mutually exclusive strata decided by some predetermined criteria and a simple random sample is selected from each stratum. In RSS, stratification is done “artificially” through judgment ranks after recruiting EUs. Then, the observations are selected from these judgment rank classes. Thus, these judgment rank classes (“artificial strata”), may not be mutually exclusive. Nevertheless, judgment rank classes \((1, \ldots, H)\) can still be considered as \(H\) strata (See, for example, Stokes and Sager [44], and Kaur et al. [21]). The stratification, like in stratified sampling, acts as a variance reduction technique yielding estimators at least as accurate as those estimators from a simple random sample with the same number of fully measured units.

Quality of ranking information plays an important role in RSS. Under perfect ranking, all judgment rank statistics are independent order statistics from the underlying distribution in a sample of size \(H\). Perfect ranking provides the greatest separation between the judgment rank classes and yields the highest efficiency gain for the statistical procedures. Imperfect judgment ranking, however, will introduce ranking errors and separation between judgment rank classes will not be as great as in the perfect ranking case. This usually reduces the efficiency of the statistical procedures. Nevertheless, the efficiency of a balanced ranked set sample procedure
is always at least as good as a SRS procedure. See McIntyre [26], Takahasi [45], Bohn [2], Patil [38] and Ozturk and MacEachern [33] for details.

The efficiency of a RSS procedure can be improved by selecting an unbalanced ranked set sampling design. Balanced RSS is an equal allocation sampling design because all judgment rank classes are treated equally. In contrast, unbalanced RSS is an unequal allocation sampling design, in which researchers intend to select more units from some judgment rank classes than the other classes as done in Table 1.2. If a suitable sampling design for an URSS is selected based on some features of an underlying distribution, such as symmetry and unimodality, efficiency of the statistical procedure can be improved. This issue has been discussed by Kaur et al. [19] and [20]. If the design for an URSS is not suitable, however, its performance can be worse than that of SRS (Patil [37]).

Literature on RSS has expanded rapidly in the last three decades. A few selected applied papers on RSS are Hall and Dell [15], Evans [11], Ridout and Cobby [41], Mode et al. [28] and Barnett [1]. Due to current interest, the original paper of McIntyre [26] was re-published in 2005. Takahasi and Wakimoto [46] provided theoretical foundation for the estimation of the population mean. They showed that the sample mean of a RSS is an unbiased estimator of the population mean and it has a smaller variance than the variance of a sample mean of a simple random sample estimator. Stokes [43] provided an estimator for the population variance based on RSS data and found out that the estimator is asymptotically unbiased and has smaller mean squared error than the variance of SRS data regardless of the quality of judgment rankings. The improvement, however, was not as great as in the case of mean estimator. Stokes and Sager [44] showed that the empirical distribution function based
on RSS data is an unbiased estimator for the population distribution function and superior to SRS estimator. Hettmansperger [17] illustrated that the RSS one-sample sign test is more efficient than the SRS one-sample sign test. For a complete list of references, we refer readers to the research papers Kvam and Samaniego [23], Sinha et al. [42], MacEachern et al. [24], Perron and Sinha [39], Chen [7], the review paper Kaur et al. [22], and the recent monograph Chen, Bai and Sinha [5].

The ranked set sampling procedure can be extended easily to a two sample problem by selecting two independent ranked set samples, one from each population. Ranked set sample for each population can have different set and cycle sizes.

Let \( k \) and \( m \) be the set and cycle sizes for one of the populations, say \( X \)-sample population. We then denote the balanced ranked set sample from the \( X \)-sample population to be \((X_1, \ldots, X_m)\), where \( X_j = (X_{[1]j}, \ldots, X_{[k]j}) \) for \( j = 1, \ldots, m \).

To construct such a sample from the \( X \)-sample population, \( mk^2 \) experimental units are needed and \( mk \) units are fully measured.

Let \( q \) and \( n \) be the set and cycle sizes for the \( Y \)-sample population. Then the balanced ranked set sample from the \( Y \)-sample population is denoted as \((Y_1, \ldots, Y_n)\), where \( Y_t = (Y_{[1]t}, \ldots, Y_{[q]t}) \) for \( t = 1, \ldots, n \).

For this sample, again, \( nq^2 \) experimental units are needed from the \( Y \)-sample population and \( nq \) of them are fully measured.

Unbalanced ranked set sampling designs, similar to the one given in one-sample problems, can also be used for each population in two-sample problems depending on the main feature of each population.

Literature is very rich in two sample ranked set sampling procedures. Bohn and Wolfe [3], developed a nonparametric two-sample inference to estimate and test
the location-shift between two populations based on a balanced RSS under perfect judgment ranking. They introduced RSS analog of the Mann-Whitney-Wilcoxon statistic, $U_{RSS}$, to test the null hypothesis that the two populations are stochastically equivalent against the alternative hypothesis that one population is stochastically larger than the other, where

$$U_{RSS} = \sum_{s=1}^{q} \sum_{t=1}^{n} \sum_{i=1}^{k} \sum_{j=1}^{m} I(X_{[i]j} < Y_{[s]t}).$$

(1.1)

Under perfect judgment ranking, the exact null distribution of $U_{RSS}$ is free of the underlying distribution even though it is computationally intensive. Additionally, the null distribution of $U_{RSS}$ is asymptotically normal and the test based on $U_{RSS}$ has higher asymptotic Pitman efficiency with respect to SRS Mann-Whitney-Wilcoxon test. In a follow up paper, Bohn and Wolfe [4], introduced a judgment ranking model to explain the impact of ranking error on $U_{RSS}$. The construction of their model relies on expected spacing between two order statistics. This model indicates that the probability of incorrect ranking is a decreasing function of the expected spacing between two order statistics. Thus, it is less likely to have perfect ranking when the expected spacing between two units is small. Under imperfect judgment rankings, the exact and asymptotic null distribution of $U_{RSS}$ is not distribution free. It may depend on the underlying distributions and judgment ranking model. Thus, the nominal level of the test is not achieved and level of the test is substantially increased due to the inflation in the null variance of the test statistic.

Ozturk [31] constructed a test for testing the equality of the population medians for two-sample inference based on one-sample ranked set sample sign statistics. The new test first constructs confidence intervals for the medians of each population. If these confidence intervals are disjoint, then the test rejects the null hypothesis. Since
the new test is constructed based on median confidence intervals, it is distribution free
and has higher efficiency than its SRS analog. However, under imperfect ranking,
its exact and asymptotic null distributions are not distribution-free.

Ozturk and Wolfe [35] introduced an unbalanced ranked set sampling design to
improve the efficiency of Bohn-Wolfe rank-sum test. An information function is de-
defined to measure the information content of a two-sample ranked set sample. This
information function is numerically maximized to find appropriate ranks of the ob-
servations to be quantified in each cycle from the two populations. The test based
on this design has larger Pitman efficacy than the Pitman efficacy of the test based
on $U_{RSS}$. When the underlying distributions are symmetric, optimal design selects
observations at the mode of each distribution. Similar results are also given for
asymmetric distributions.

Selection of the optimal design is one of the active research fields in RSS. The idea
is to find an allocation scheme to maximize the efficiency of a particular statistical
procedure. Designs that maximize Pitman efficacies of one and two-sample nonpara-
metric tests are developed in Ozturk [30], and Ozturk and Wolfe [36]. Selection of
optimal allocation has also been discussed by Kaur et al. [19], Chen [8] and Nahhas
et al. [29].

Fligner and MacEachern [12] recently developed a nonparametric test to test the
location shift between two populations. The test statistic is constructed as a linear
combination of rank-sum statistic of judgment classes and is distribution-free regard-
less of the quality of judgment ranking. The proposed test can be used with both
balanced and unbalanced RSS.
Even though the procedures based on RSS have improved efficiency over the procedures based on SRS, there are some concerns in its use in the design of experiments, where available experimental units are limited. In a ranked set sample, for each fully measured observation, $H - 1$ additional experimental units are needed for ranking purposes. This would be a concern in the design of experiments, where available units are either limited or expensive. For example, recruiting subjects in a clinical trial can take a long time and be expensive, thus the use of ranked set sampling may be limited for this type of experiments. Another concern in using RSS design to draw inference for the treatment contrast is that the role of randomization is not clearly defined. Almost all available research in design of experiments based on ranked set samples considers observational studies without addressing the role of randomization. In order to address these concerns, a new design, order restricted randomized design, is introduced.

1.2 Order restricted randomized designs

The order restricted randomized design (ORRD) was proposed by Ozturk and MacEachern [33] to overcome the shortcomings of RSS designs discussed in the previous section. It was also discussed in a different context by Chen [6]. The main idea in ORRD is similar to the one presented in RSS design. It combines the judgment ranking principle in RSS and the randomization principle in the design of experiment without any waste of the recruited experimental units. Subjective information creates positive correlations within set measurements through judgment ranking process of the experimental units, but unlike RSS design it uses all available units in a set. In this respect, it is superior to a RSS design. These positive correlations are turned into
negative correlations for the contrast parameter with an appropriate randomization. In the drug efficacy comparison example at the beginning of this chapter, subjective information, such as general health status and pre-medical history of patients, can be exploited to create judgment ranks for participating subjects. Consequently, these judgment ranked subjects are presumably positively correlated even though the ranking process may not be perfect. If these positively correlated subjects are assigned to treatment and control groups, the positive correlations are changed into negative ones when the efficacy difference between the two drugs is estimated. The variance of the difference estimator is expected to be smaller due to the negative correlation, yielding a better inference for the drug efficacy comparison.

1.2.1 Description for ORRD

We use the following model to describe the underlying structure between the responses and control-treatment effects.

\[ Z_{ij} = \mu_i + \gamma_{ij} \quad \text{for } i = 1, 2 \text{ and } j = 1, ..., n , \]  

(1.2)

where \( Z_{ij} \) is the response of the experimental unit \( j \) with treatment \( i \); \( \mu_i \) is the median effect of treatment \( i \); \( \gamma_{ij} \)'s are i.i.d. random errors from a distribution \( F \) with finite Fisher information.

This model indicates that the error term \( \gamma_{ij} \) is the property of an experimental unit. Thus, the heterogeneity among experimental units is explained by these random error terms. The premise of ORRD is to use this heterogeneity by ranking the residual terms in model (1.2) at the design stage. The order restricted randomized design can be constructed through a two step procedure, ranking and randomization.
Step I (Ranking): First fix the set size $H$. Then select $2H$ experimental units at random from an infinite population. These units are divided into 2 sets, each of size $H$. The residuals, $\gamma_{ij}$'s, in each set are judgment ranked with available subjective auxiliary information. Let $R_1, \ldots, R_H$ denote the subjective ranks of the experimental units in each set. Note that ranking is performed pre-experimentally based on inherent variations among the experimental units. This ranking process is similar to the construction of blocks in a randomized block design. If the ranking is perfect, then the residuals that belong to the experimental units within each set are the order statistics from a sample of size $H$. These order statistics are positively correlated. Under mild conditions, we expect that these correlations are still positive even when we have imperfect ranking. Under model (1.2), it is clear that response measurements from ranked units are also positively correlated.

Step II (Randomization): In order to randomize the treatment regimes to experimental units, we separate the ranks $(R_1, \ldots, R_H)$ into two disjoint non-empty subsets Set $\alpha$ and Set $\beta$, where $\alpha = (\alpha_1, \ldots, \alpha_u)$ and $\beta = (\beta_1, \ldots, \beta_{H-u})$. In the first set, we then perform a randomization to decide whether the control or treatment regimes are assigned to units that have ranks in Set $\alpha$ or Set $\beta$. In the second set, opposite allocation is performed without randomization so that each treatment regime is applied to all ranks. Consider the previous example of the drug efficacy comparison experiment. Suppose the set size $H = 4$, $\alpha = \{1, 3\}$ and $\beta = \{2, 4\}$. Eight subjects are recruited and separated randomly into two sets with size 4. Then we randomly assign the subjects that have ranks in $\alpha$ to the control regime and subjects that have ranks in $\beta$ to the treatment regime. The opposite assignment is performed in the
second set. In this example, each treatment regime is applied to four subjects having judgment ranks 1, 2, 3 and 4.

These two steps are called a *replicate*. This basic process is repeated $n$ times to increase the sample size. Table 1.3 illustrates the whole process for replicate size $n$, set size $H$, sets $\alpha$ and $\beta$.

<table>
<thead>
<tr>
<th>Replicate Size</th>
<th>Set</th>
<th>Control</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>set1</td>
<td>$X_{[\alpha_1]1}, \ldots, X_{[\alpha_u]1}$</td>
<td>$Y_{[\beta_1]1}, \ldots Y_{[\beta_{H-u}]1}$</td>
</tr>
<tr>
<td></td>
<td>set2</td>
<td>$X_{[\beta_1]1}, \ldots, X_{[\beta_{H-u}]1}$</td>
<td>$Y_{[\alpha_1]1}, \ldots Y_{[\alpha_u]1}$</td>
</tr>
<tr>
<td>2</td>
<td>set1</td>
<td>$X_{[\alpha_1]2}, \ldots, X_{[\alpha_u]2}$</td>
<td>$Y_{[\beta_1]2}, \ldots Y_{[\beta_{H-u}]2}$</td>
</tr>
<tr>
<td></td>
<td>set2</td>
<td>$X_{[\beta_1]2}, \ldots, X_{[\beta_{H-u}]2}$</td>
<td>$Y_{[\alpha_1]2}, \ldots Y_{[\alpha_u]2}$</td>
</tr>
</tbody>
</table>

Table 1.3: One ORRD sample with replicate size $n$, set size $H$, sets $\alpha$ and $\beta$

For any fixed value of $H$, the ORRD is determined by the choices of Set $\alpha$ and Set $\beta$. Thus, the set $(\alpha, \beta)$ is called the design. When $H$ is 2, the design is unique. When $H$ is greater than 2, there are $2^{H-1} - 1$ possible designs depending on $(\alpha, \beta)$. An optimal design can be chosen from all possible designs based on some reasonable criteria.

An alternative construction is possible for ORRD. Suppose that we wish to construct an ORRD with sample size $2nH$, where $n$ is the replicate size and $H$ is the set size.

**Step I:** Recruit $2nH$ experimental units for the study.

**Step II:** Randomly separate these units into $2n$ sets, each of size $H$.

**Step III:** Rank the units in each set from 1 to $H$ in a consistent manner based on available subjective information.
Step IV: Randomly separate the $2n$ sets into two groups, $n$ sets of each. In one of the groups, perform a randomization to decide whether the control or treatment regime is assigned to units that have ranks in Set $\alpha$ or in Set $\beta$. In the other group, perform an opposite allocation without a randomization.

Obviously, if we can recruit $2nH$ experimental units during a very short period of time, the alternative construction for the ORRD is feasible. Otherwise, we use the original construction scheme.

1.2.2 Properties of ORRD

Main features of a data set obtained from an ORRD can be summarized as follows.

- In each control and treatment groups, there are $n$ observations with judgment rank $h$ for $h = 1, \ldots, H$. Thus, the design is balanced.

- Within set observations are not independent. Under some mild assumptions, they are presumably positively correlated.

- Between set observations are independent.

The measurements from an ORRD can be modeled as

$$Z_{i[h][j]}^* = \mu_i + \gamma_{i[h][j]}^* \quad \text{for} \quad i = 1, 2, \ldots, H \quad \text{and} \quad j = 1, \ldots, n,$$

where $Z_{i[h][j]}^*$ is the response measurement from the experimental unit in replicate $j$ with rank $h$ for treatment $i$; $\mu_i$ is the median effect of treatment $i$; $\gamma_{i[h][j]}^*$ is the random error associated with the experimental unit in replicate $j$ with rank $h$ for treatment $i$. In this model we assume that within set residuals are positively correlated.
This specific correlation structure of the data leads to the most important characteristic of the ORRD. Since the ranked observations from the same set are separated into two treatment groups, the within set positive correlations translate into negative correlations for the contrast of the two treatment groups and leads to the variance reduction for the contrast estimator.

There are two main advantages of an ORRD over the RSS design. One of the advantages is that the ORRD requires a smaller number of experimental units than the required number of units in a RSS design in order to get the same number of fully measured observations. For example, to obtain $H$ observations for full measurements, a balanced RSS requires $H^2$ experimental units, $H(H-1)$ of which are used only for judgment ranking, while ORRD requires $H$ experimental units, all of which are used for full measurements. The second advantage of ORRD is that the randomization technique is used to highlight the inference of the contrast parameters. Thus it provides a superiority over RSS where randomization is seldom applied.

Ozturk and MacEachern [33] introduced ORRD for control versus treatment comparison involving one control population and one or more treatment populations. Assuming that the control and treatment populations have the same but an arbitrary shape with unique medians, they developed a multiple comparison procedure to test whether or not all medians of treatment populations were equal to the control population median. The testing procedure is constructed based on one-sample sign statistics and their associated confidence intervals. The confidence coefficients of the intervals are chosen to provide an upper bound for the family-wise error rate. Regardless of the ranking quality, the performance of the new test is better than those
based on RSS and SRS procedures in terms of empirical power and asymptotic relative efficiency.

In another article, Ozturk and MacEachern [34] investigated two-sample problems under the ORRD. Specifically, they were interested in drawing inference for $\Delta = \mu_1 - \mu_2$, the difference between the two treatment means. They constructed an estimator along with associated confidence interval for $\Delta$. The proposed estimator is defined as

$$\hat{\Delta}_{ORR} = \frac{2}{Hn} \sum_{j=1}^{n} [d_{1j} + d_{2j}]$$

where

$$d_{1j} = \sum_{h \in \alpha} X_{[h]j} - \sum_{h \in \beta} Y_{[h]j}$$

$$d_{2j} = \sum_{h \in \beta} X_{[h]j} - \sum_{h \in \alpha} Y_{[h]j}$$

for $j = 1, \ldots, n/2$. It is shown that $\hat{\Delta}_{ORR}$ is unbiased for $\Delta$ and has a limiting normal distribution. The authors provided an analytic expression for the variance of $\hat{\Delta}_{ORR}$. When $H = 2$, the ORRD is unique. However, for a fixed $H > 2$, the design depends on the choice of $\alpha$ and $\beta$, and the number of designs increases with the set size $H$. The optimal design, which minimizes the variance of $\hat{\Delta}_{ORR}$ over all possible designs, is obtained when units with adjacent judgment ranks are assigned to different treatment regimes.

Ozturk and MacEachern [34] introduced a test to test the null hypothesis $H_0 : \Delta = \Delta_0$ against the alternative hypothesis $H_a : \Delta \neq \Delta_0$. They used the test statistic

$$T_n = \frac{\hat{\Delta}_{ORR} - \Delta_0}{\sqrt{\hat{V}(\hat{\Delta}_{ORR})}}$$
to reject $H_0$ for large values of $T_n$, where $\hat{V}(\hat{\Delta}_{ORR})$ is a consistent estimator of the variance of $\hat{\Delta}_{ORR}$. Under $H_0$, $T_n$ converges to a standard normal distribution as $n \to \infty$. On the other hand, for small sample sizes, the Student’s $t$-distribution provides a better approximation for the distribution of $T_n$. Under different judgment quality of rankings, the authors performed a simulation study to investigate the estimated Type I error rates based on a normal distribution approximation, a $t$-distribution (with $n-2$ degrees of freedom) approximation and the Satterthwaite’s approximation. When sample size is large, all three approximations provide Type I error rates that are close to the nominal rate. When the sample size is small, the Student’s $t$-distribution provides a reasonably good estimates for the Type I error rates, especially for moderately heavy tailed underlying distributions. The authors also provided a $(1-\gamma) 100\%$ confidence interval for $\Delta$ by converting the test statistic $T_n$.

Ozturk and MacEachern [34] also discussed two competitor estimators for the contrast parameter $\Delta$, SRS estimator $\hat{\Delta}_{SRS} = \bar{X} - \bar{Y}$ and RSS estimator $\hat{\Delta}_{RSS} = \bar{X}_[.] - \bar{Y}_[.]$, where $\bar{X}_[.] = \frac{1}{Hn} \sum_{h=1}^{H} \sum_{j=1}^{n} X_{[h]j}$ and $\bar{Y}_[.] = \frac{1}{Hn} \sum_{h=1}^{H} \sum_{j=1}^{n} Y_{[h]j}$. The authors established some relationships among variances of these estimators and showed analytically that the variance of $\hat{\Delta}_{ORR}$ under the optimal design was smaller than the variances of the other two estimators.

Ozturk [32] developed a two-sample Mood median test under an ORRD. He showed that the new test is superior to the corresponding tests under RSS and completely randomized designs.

When the set size $H$ is relatively large, it may not be easy to recruit $H$ patients during a short period of time, especially for some rare diseases. Under this situation, we suggest to use a smaller set size.
Brief comparisons among RSS, ORRD and CRD (Completely Randomized design) are shown in Table 1.4.

<table>
<thead>
<tr>
<th>Property</th>
<th>CRD</th>
<th>RSS</th>
<th>ORRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measured observations</td>
<td>$nH$</td>
<td>$nH$</td>
<td>$nH$</td>
</tr>
<tr>
<td>Number of recruited units</td>
<td>$nH$</td>
<td>$nH^2$</td>
<td>$nH$</td>
</tr>
<tr>
<td>Independence</td>
<td>Yes</td>
<td>Yes</td>
<td>Those from different sets</td>
</tr>
<tr>
<td>Identical distribution in same treatments</td>
<td>Yes</td>
<td>Those with same ranks</td>
<td>Those with same ranks</td>
</tr>
<tr>
<td>Judgment rankings</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Randomization</td>
<td>Yes</td>
<td>Not clear</td>
<td>With order restrictions</td>
</tr>
</tbody>
</table>

Table 1.4: Brief comparisons among RSS, ORRD and CRD

In the following chapters, we develop a nonparametric inference for a two-sample location shift problem. A new rank-sum test statistic is constructed based on the ORRD data. It is shown that the asymptotic null distribution of the proposed test statistic is normal and the test outperforms two competitor tests in the literature with respect to Pitman asymptotic relative efficacies. We also show that the Type I error rate of the test is close to the nominal test level even under imperfect ranking.

The rest of this dissertation is organized as follows. In Chapter 2, we introduce the test statistic and develop its asymptotic properties under perfect rankings. The asymptotic properties under imperfect ranking are derived in Chapter 3. In Chapter 4, we perform simulation studies for the proposed test under both perfect and imperfect rankings to demonstrate its superiority over its competitors proposed in the literature. We apply the ORRD and the proposed test to an existing data set
from the AIDS clinical trials group protocol 320 study. We provide some concluding remarks and discussion for some open research problems in Chapter 6.
CHAPTER 2

RANK-SUM TEST UNDER PERFECT RANKINGS

Let $F(x)$ and $G(y) = F(y - \Delta)$ be the c.d.f.s of the control and treatment populations, respectively. The parameter, $\Delta = \theta_Y - \theta_X$, represents a location shift between these two distributions, where $\theta_X$ and $\theta_Y$ are the medians of $G$ and $F$. For this contrast parameter, $\Delta$, the ORRD generates two samples from $G$ and $F$. Let 

$$\{X_{[i]}, i = 1, \ldots, H; j = 1, \ldots, n\} \text{ and } \{Y_{[k]}, k = 1, \ldots, H; t = 1, \ldots, n\}$$

be the control and treatment samples, respectively. These data can also be written as

$$(X_{[\alpha_1]}1, \ldots, X_{[\alpha_u]}1, X_{[\beta_1]}1, \ldots, X_{[\beta_{H-u}]}1, \ldots, X_{[\alpha_1]}n, \ldots, X_{[\alpha_u]}n, X_{[\beta_1]}n, \ldots, X_{[\beta_{H-u}]}n), \quad (2.1)$$

and

$$(Y_{[\alpha_1]}1, \ldots, Y_{[\alpha_u]}1, Y_{[\beta_1]}1, \ldots, Y_{[\beta_{H-u}]}1, \ldots, Y_{[\alpha_1]}n, \ldots, Y_{[\alpha_u]}n, Y_{[\beta_1]}n, \ldots, Y_{[\beta_{H-u}]}n), \quad (2.2)$$

where sets $\alpha = (\alpha_1, \ldots, \alpha_u)$ and $\beta = (\beta_1, \ldots, \beta_{H-u})$ are the two pre-determined, disjoint, nonempty subsets of $(1, \ldots, H)$. Let $F_{[i]}$ and $G_{[i]}$ be the c.d.f.s for $X_{[i]}$ and $Y_{[i]}$, respectively. In this notation, the square bracket indicates the quality of ranking information. If ranking is perfect, we replace square brackets with standard notation (round bracket) and judgment rank statistics become order statistics from a sample of size $H$. 

20
We aim to develop a nonparametric test for the hypothesis

\[ H_0 : \Delta = 0 \]  \hspace{1cm} (2.3)

\[ H_1 : \Delta \neq 0. \]

Let

\[ T = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{[i][j]} \leq Y_{[k][t]}) . \]  \hspace{1cm} (2.4)

We reject the null hypothesis for extreme values of \( T \). The proposed test statistic \( T \) is similar to Bohn-Wolfe rank sum test statistic \( U_{RSS} \), both of which count the number of \( X \)- observations that are less than the \( Y \)- observations. On the other hand, these two test statistics possess different properties because they use data sets generated by different designs.

The exact null distribution of \( T \), under perfect ranking, can be constructed through exhaustive enumeration of the probability contents of all possible permutations of \( X \)- and \( Y \)- observations. Unlike SRS, these permuted sequences are not equally likely. It seems that there does not exist a feasible algorithm, especially for large sample sizes, to evaluate the probability contents of these sequences under imperfect ranking. In addition to computational difficulty, there is another challenge to construct the null distribution of \( T \). Under imperfect ranking, probability content of each permuted sequence of \( X \)- and \( Y \)- observations can not be evaluated without a reasonable model that explains the judgment ranking scheme. In this dissertation, we therefore develop asymptotic null distribution of \( T \) under an arbitrary but consistent judgment ranking scheme. We later illustrate that this asymptotic approximation works reasonably well for moderately small sample sizes. We first provide an assumption that is used throughout this dissertation.
**Assumption:** The same ranking scheme is used in all sets. Under the null hypothesis, this implies that $F_{[i]}(.) = G_{[i]}(.)$, $i = 1, \cdots, H$. This ranking scheme will be called as consistent ranking mechanism in this dissertation.

For computational convenience, we center the test statistic $T$ and write

$$T^* = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[i]j} \leq Y_{[k]t}) - \tau_{ik}]$$

$$= \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{[i]j} \leq Y_{[k]t}) - \frac{n^2 H^2}{2},$$

where

$$\tau_{ik} = E[I(X_{[i]j} \leq Y_{[k]t})].$$

It is clear that $E[T^*] = 0$. For notational convenience, we also define $\overline{T}^* = \frac{T^*}{(nH)^2}$ and $\overline{T} = \frac{T}{(nH)^2}$.

In Section 2.1 we develop the asymptotic null distribution of $T$ under an arbitrary but consistent judgment ranking scheme. Section 2.2 investigates the asymptotic properties of the test under perfect judgment ranking. Asymptotic Pitman efficacy is derived in this section. Section 2.3 discusses an optimal design constructed by maximizing the Pitman efficacy of $T$. Section 2.4 presents simulation results to evaluate convergence rate of the null distribution of the test statistic. Section 2.5 provides the point estimator and the confidence intervals for the location shift parameter.

## 2.1 Asymptotic null distribution of test statistic $T$

Standard asymptotic theory of a statistic relies on the assumption that it is written as an average of independent random variable so that some version of central limit
theorem provides the desired limiting distribution. It is clear that our test statistic $T$ is not a sum of independent random variables. Thus, the standard asymptotic theory does not apply. In this case, one of the common approaches is to project the test statistic onto a space of linear functions of independent variables. Then the central limit theorem is applied to this independent sum to obtain the limiting distribution. Hajek and Sidak [14] provided a detailed discussion on the use of the projection theorem. We state the main theorem without a proof.

**Theorem 2.1.** Let $T = T(Z_1, \ldots, Z_n)$ be a random variable based on a random sample $Z_1, \ldots, Z_n$ such that $E[T] = 0$. Let

$$p_k^*(x) = E[T|Z_k = x] \quad k = 1, \ldots, n.$$  

We say the projection of $T$ onto the space of linear functions of $Z_1, \ldots, Z_n$ is

$$T_p = \sum_{k=1}^{n} p_k^*(Z_k).$$

Let $W = \sum_{i=1}^{n} p_i(Z_i)$ where $p_i$ is a function of $Z_i$. Then $E[(T - W)^2]$ is minimized by taking $p_i(x) = p_i^*(x)$. Furthermore, $E[(T - T_p)^2] = Var[T] - Var[T_p]$. If the asymptotic distribution of $T_p$ exists and $Var[T] - Var[T_p] \to 0$ as $n \to \infty$, then the asymptotic distribution of $T$ is the same as that of $T_p$.

**Proof.** See Theorem 2.4.6 in Hettmansperger and McKean [18].

We rearrange the data in the ORRD as follows

$$Z_{1s} = (X_{[\alpha_1]s}, \ldots, X_{[\alpha_u]s}, Y_{[\beta_1]s}, \ldots, Y_{[\beta_{H-1}u]s}) \quad \text{for} \quad s = 1, \ldots, n$$  

$$Z_{2s} = (X_{[\beta_1]s}, \ldots, X_{[\beta_{H-1}u]s}, Y_{[\alpha_1]s}, \ldots, Y_{[\alpha_{u-1}]s}) \quad \text{for} \quad s = 1, \ldots, n$$  

$$Z = (Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n}), \quad (2.10)$$
where \( Z_{1s} \) and \( Z_{2s} \) indicate the Set 1 and Set 2 observations in the \( s \)-th replication, respectively. Then \( Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n} \) are mutually independent since the sets are randomly selected in ORRD.

The test statistic \( T^* \) in (2.5) can be considered as a random variable based on \( Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n} \), such that \( \mathbb{E}[T^*] = 0 \). Thus, Theorem 2.1 can be applied to \( T^* \) to find its asymptotic distribution.

We construct the projection of \( T^*, V_p \), onto the space of linear functions of \( Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n} \) in Lemma 2.2. The asymptotic distribution of \( V_p \) is shown in Corollary 2.3. Lemma 2.4 proves \( \frac{1}{(nH)^3} \text{Var}(T^*) - \frac{1}{(nH)^3} \text{Var}(V_p) \to 0 \) as \( n \to \infty \). The apparent conclusion of the asymptotic null distribution of \( T^* \) is given in Theorem 2.5.

**Lemma 2.2.** Let \( F(x) \) and \( G(y) = F(y - \Delta) \) be the c.d.f.s of the control and treatment populations, respectively. Let \( Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n} \) be random variables defined in (2.8), (2.9) and (2.10) and \( T^* \) be the test statistic in (2.5). Then, under the null hypothesis in (2.3) and an arbitrary but consistent ranking scheme, the projection of \( T^* \) onto the space of linear functions of \( Z_{11}, \ldots, Z_{1n}, Z_{21}, \ldots, Z_{2n} \) is given by

\[
V_p = \sum_{s=1}^{n} \left[ E(T^* | Z_{1s}) + E(T^* | Z_{2s}) \right]
\]

\[
= nH \sum_{j=1}^{n} \left\{ \sum_{i=1}^{u} \left[ 1 - F(X_{[\alpha_i,j]}) - \bar{\tau}_{\alpha_i} \right] + \sum_{k=1}^{H-u} \left[ F(Y_{[\beta_k,j]}) - \bar{\tau}_{\beta_k} \right] \right\} + \sum_{j=1}^{n} \left\{ \sum_{i=1}^{H-u} \left[ 1 - F(X_{[\beta_i,j]}) - \bar{\tau}_{\beta_i} \right] + \sum_{k=1}^{n} \left[ F(Y_{[\alpha_k,j]}) - \bar{\tau}_{\alpha_k} \right] \right\} + o(n),
\]

where

\[
\bar{\tau}_{\alpha_i} = \sum_{t=1}^{H} \tau_{\alpha_i,t} / H, \quad \bar{\tau}_{\beta_i} = \sum_{t=1}^{H} \tau_{\beta_i,t} / H, \quad \bar{\tau}_{\alpha_k} = \sum_{t=1}^{H} \tau_{\alpha_k,t} / H, \quad \bar{\tau}_{\beta_k} = \sum_{t=1}^{H} \tau_{\beta_k,t} / H.
\]
Proof. Note that under the null hypothesis, $F(x) = G(y)$. We partition $T^*$ into four different sums

$$T^* = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X[i,j] \leq Y[k,t]) - \tau_{ik}]$$

$$= \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[\alpha_i]}j \leq Y[k,t]) - \tau_{\alpha_i,k}] + \sum_{i=1}^{H-u} \sum_{j=1}^{u} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[\beta_i]}j \leq Y[k,t]) - \tau_{\beta_i,k}]$$

$$= \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[\alpha_i]}j \leq Y_{[\alpha_i]t}) - \tau_{\alpha_i,\alpha_k}] +$$

$$\sum_{i=1}^{H-u} \sum_{j=1}^{u} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[\beta_i]}j \leq Y_{[\beta_i]t}) - \tau_{\alpha_i,\beta_k}] +$$

$$\sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{u} \sum_{t=1}^{n} [I(X_{[\beta_i]}j \leq Y_{[\alpha_i]t}) - \tau_{\beta_i,\alpha_k}] +$$

$$\sum_{i=1}^{H-u} \sum_{j=1}^{H-u} \sum_{k=1}^{n} \sum_{t=1}^{n} [I(X_{[\beta_i]}j \leq Y_{[\beta_i]t}) - \tau_{\beta_i,\beta_k}]$$

$$= T_1 + T_2 + T_3 + T_4.$$ 

Since $Z_{11}, ..., Z_{1n}, Z_{21}, ..., Z_{2n}$ are i.i.d. random vectors under the null hypothesis, the projection of $T^*$ can be computed from the conditional expectations,

$$V_P = \sum_{s=1}^{n} [E(T^*[Z_{1s}) + E(T^*[Z_{2s})].$$

The conditional expectation of $T^*$ given $Z_{1s}$ is

$$E[T^*[Z_{1s}] = \begin{cases} 
\sum_{i=1}^{u} \sum_{k=1}^{H-u} [1 - F_{[\alpha_i]}(X_{[\alpha_i]s}) - \tau_{\alpha_i,\alpha_k}] + \\
\sum_{i=1}^{H-u} [F_{[\beta_i]}(Y_{[\beta_i]s}) - \tau_{\beta_i,\beta_k}], & \text{if } s = j = t \\
(n-1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} [1 - F_{[\alpha_i]}(X_{[\alpha_i]s}) - \tau_{\alpha_i,\alpha_k}] + \\
\sum_{i=1}^{H-u} [F_{[\beta_i]}(X_{[\alpha_i]s}) - \tau_{\alpha_i,\beta_k}], & \text{if } s = j \neq t \\
(n-1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} [F_{[\alpha_i]}(Y_{[\beta_i]s}) - \tau_{\alpha_i,\beta_k}] + \\
\sum_{i=1}^{H-u} [F_{[\beta_i]}(Y_{[\beta_i]s}) - \tau_{\beta_i,\beta_k}], & \text{if } s = t \neq j 
\end{cases}.$$ 

25
In a similar fashion, the second expectation can be computed as

$$E[T^*|Z_{2s}] = \begin{cases} 
\sum_{i=1}^{n} \sum_{k=1}^{u} [F_{[\alpha_i]}(Y_{[\alpha_k]}|s) - \tau_{\alpha_i\alpha_k}] + \\
\sum_{i=1}^{H-u} \sum_{k=1}^{H-u} [1 - F_{[\beta_k]}(X_{[\beta_i]}|s) - \tau_{\beta_i\beta_k}], & \text{if } s = j = t \\
(n-1)\{\sum_{i=1}^{H-u} \sum_{k=1}^{u} [1 - F_{[\alpha_k]}(X_{[\alpha_i]}|s) - \tau_{\alpha_i\alpha_k}] + \\
\sum_{i=1}^{H-u} \sum_{k=1}^{H-u} [1 - F_{[\beta_k]}(X_{[\beta_i]}|s) - \tau_{\beta_i\beta_k}]\}, & \text{if } s = j \neq t
\end{cases}$$

By putting these two terms together, we obtain

$$V_P = \sum_{s=1}^{n} [E(T^*|Z_{1s}) + E(T^*|Z_{2s})]$$

$$= \sum_{s=1}^{n} \{ (n-1) \sum_{i=1}^{u} \sum_{k=1}^{u} [1 - F_{[\alpha_k]}(X_{[\alpha_i]}|s) - \tau_{\alpha_i\alpha_k}] +$$

$$= (n-1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} [1 - F_{[\beta_k]}(X_{[\alpha_i]}|s) - \tau_{\alpha_i\beta_k}] +$$

$$= (n-1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} [F_{[\alpha_i]}(Y_{[\beta_k]}|s) - \tau_{\alpha_i\beta_k}] +$$

$$= (n-1) \sum_{i=1}^{H-u} \sum_{k=1}^{u} [1 - F_{[\alpha_k]}(X_{[\beta_i]}|s) - \tau_{\beta_i\alpha_k}] +$$

$$= (n-1) \sum_{i=1}^{H-u} \sum_{k=1}^{H-u} [1 - F_{[\beta_k]}(X_{[\beta_i]}|s) - \tau_{\beta_i\beta_k}] +$$

$$= (n-1) \sum_{i=1}^{u} \sum_{k=1}^{u} [F_{[\beta_i]}(Y_{[\alpha_k]}|s) - \tau_{\beta_i\alpha_k}] +$$

$$= (n-1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} [F_{[\beta_i]}(Y_{[\alpha_k]}|s) - \tau_{\beta_i\alpha_k}] + o(n)\}.$$ 

Under an arbitrary but consistent ranking scheme, Dell and Clutter [10] showed that

$$\sum_{i=1}^{H} F_{[i]}(t) = HF(t). \quad (2.11)$$
By using this equation, the projection of \( T^* \) reduces to
\[
V_P = nH \sum_{j=1}^{n} \left\{ \sum_{i=1}^{u} [1 - F(X_{[\alpha_i]}j) - \bar{\tau}_{\alpha_i} + \sum_{k=1}^{H-u} [F(Y_{[\beta_k]}j) - \bar{\tau}_{\beta_k}]] + \sum_{i=1}^{u} [1 - F(X_{[\alpha_i]}j) - \bar{\tau}_{\alpha_i} + \sum_{k=1}^{H-u} [F(Y_{[\beta_k]}j) - \bar{\tau}_{\beta_k}]] + nH \sum_{j=1}^{n} [1 - F(X_{[\beta_i]}j) - \bar{\tau}_{\beta_i}] + \sum_{k=1}^{u} [F(Y_{[\alpha_i]}j) - \bar{\tau}_{\alpha_i}] + o(n),
\]
which completes the proof of the Lemma.

Note that \( V_P \) is a sum of independent random variables. Let \( \nabla_P = \frac{V_P}{(nH)^2} \). Then the limiting distribution of \( \nabla_P \) follows from the Central Limit Theorem.

**Corollary 2.3.** The limiting null distribution of \( \sqrt{2nH}V_P \) converges to a normal distribution with mean zero and variance \( \sigma^2 \), where
\[
\sigma^2 = \frac{2}{H} \left\{ \text{Var} \left[ \sum_{i=1}^{u} (1 - F(X_{[\alpha_i]}1) - \bar{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]}1) - \bar{\tau}_{\beta_k}) \right] + \text{Var} \left[ \sum_{i=1}^{H-u} (1 - F(X_{[\beta_i]}1) - \bar{\tau}_{\beta_i}) + \sum_{k=1}^{u} (F(Y_{[\alpha_i]}1) - \bar{\tau}_{\alpha_i}) \right] \right\}.
\]

**Proof.** Since \( \sum_{j=1}^{n} \left\{ \sum_{i=1}^{u} [1 - F(X_{[\alpha_i]}j) - \bar{\tau}_{\alpha_i}] + \sum_{k=1}^{H-u} [F(Y_{[\beta_k]}j) - \bar{\tau}_{\beta_k}] \right\} \) and \( \sum_{j=1}^{n} \left\{ \sum_{i=1}^{H-u} [1 - F(X_{[\beta_i]}j) - \bar{\tau}_{\beta_i}] + \sum_{k=1}^{u} [F(Y_{[\alpha_i]}j) - \bar{\tau}_{\alpha_i}] \right\} \) are independent, we easily draw the conclusion from the Central Limit Theorem (CLT).

In order to establish the asymptotic null distribution of \( T^* \), we need to show that the difference between \( T^* \) and \( V_P \) is negligible so that they have the same asymptotic distribution. This is equivalent to show that \( \lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(T^*) = \lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(V_P) \).

**Lemma 2.4.** Let \( T^* \) and \( V_P \) be defined as in equation (2.5) and Lemma 2.2. Then the following equality holds,
\[
\lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(T^*) = \lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(V_P).
\]
Proof. We note that $T^*$ can be rewritten as

$$T^* = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{[i]j} \leq Y_{[k]t}) - \tau_{ik}] = \sum_{k=1}^{H} \sum_{t=1}^{n} T_{[k]t},$$

where

$$T_{[k]t} = \sum_{i=1}^{H} \sum_{j=1}^{n} [I(X_{[i]j} \leq Y_{[k]t}) - \tau_{ik}] = \sum_{i=1}^{H} \sum_{j=1}^{n} C_{[i]j}([k]t),$$

and

$$C_{[i]j}([k]t) = I(X_{[i]j} \leq Y_{[k]t}) - \tau_{ik}.$$

With these notations, the variance of $T^*$ can be written as

$$Var(T^*) = \sum_{k=1}^{H} \sum_{t=1}^{n} Var(T_{[k]t}) + \sum_{k=1}^{H} \sum_{t=1}^{n} \sum_{u=1}^{n} \sum_{s=1}^{n} Cov(T_{[k]t}, T_{[u]s}).$$

We first compute $Var(T_{[k]t}).$

$$Var(T_{[k]t}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{u=1}^{n} \sum_{s=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t))$$

$$= \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{j \neq s} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t))$$

$$= \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) + o(n)$$

$$= \sum_{j=1}^{n} \sum_{s=1}^{n} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{j \neq s} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{s \neq t} \sum_{i=1}^{n} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{i \neq s} \sum_{s \neq t} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{j \neq s} \sum_{i \neq s} \sum_{u=1}^{n} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{j \neq s} \sum_{i \neq s} \sum_{u \neq s} E(C_{[i]j}([k]t)C_{[u]s}([k]t)) +$$

$$\sum_{j \neq s} \sum_{i \neq s} \sum_{u \neq s} E(C_{[i]j}([k]t)C_{[u]s}([k]t)).$$
\[
\sum_{j=1}^{n} \sum_{i=1}^{H} \sum_{u=1}^{H} E(C_{[i]j}([k]t)C_{[u]t}([k]t)) + o(n)
\]

\[
= (n - 1)(n - 2) \sum_{i=1}^{H} \sum_{u=1}^{H} E[(I(X_{[i]} \leq Y_{[k]3} - \tau_{ik}) (I(X_{[u]} \leq Y_{[k]3} - \tau_{uk})) + o(n)
\]

\[
= (n - 1)(n - 2) \sum_{i=1}^{H} \sum_{u=1}^{H} E[(F_{[i]}(Y_{[k]3}) - \tau_{ik})(F_{[u]}(Y_{[k]3}) - \tau_{uk})] + o(n)
\]

\[
= (n - 1)(n - 2)H^2E[(F(Y_{[k]3}) - \bar{\tau}_k)(F(Y_{[k]3}) - \bar{\tau}_k)] + o(n),
\]

where \(\bar{\tau}_k = \frac{\sum_{i=1}^{H} \tau_{ik}}{H}\). We now compute \(\text{Cov}(T_{[k]t}, T_{[u]s})\). Note that there are two distinct cases that we need to consider.

Case (1): Subscripts \(s = t\). Then,

\[
\text{Cov}(T_{[k]t}, T_{[u]s}) = \text{Cov}(T_{[k]t}, T_{[u]t}) = E[T_{[k]t}T_{[u]t}]
\]

\[
= \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{H} \sum_{s=1}^{n} E(C_{[i]j}([k]t)C_{[v]t}([u]t))
\]

\[
= \sum_{j=1}^{n} \sum_{s=1}^{H} \sum_{i=1}^{H} \sum_{v=1}^{n} E(C_{[i]j}([k]t)C_{[v]t}([u]t))
\]

\[
= \sum_{s=1}^{n} \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{j=1}^{n} E(C_{[i]s}([k]t)C_{[v]s}([u]t))
\]

\[
= \sum_{j=1}^{n} \sum_{s=1}^{H} \sum_{i=1}^{H} \sum_{v=1}^{n} E(C_{[i]j}([k]t)C_{[v]s}([u]t)) + o(n)
\]

\[
= (n - 1)(n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(F_{[i]}(Y_{[k]3}) - \tau_{ik})(F_{[v]}(Y_{[u]t}) - \tau_{uv})] + o(n)
\]

Case (2): Subscripts \(s \neq t\). Then,

\[
\text{Cov}(T_{[k]t}, T_{[u]s}) = E[T_{[k]t}T_{[u]s}]
\]

\[
= \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{[i]j}([k]t)C_{[v]m}([u]s))
\]

29
\[ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{j \neq t, s} E(C_{i,j}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{i,v}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{v=1}^{n} \sum_{m=1}^{H} E(C_{i,v}([k] t) C_{v,m}([u] s)). \]

That is,

\[ Cov(T_{[k] t}, T_{[u] s}) = \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{i,v}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{v=1}^{n} \sum_{m=1}^{H} E(C_{i,v}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{j \neq t, s} E(C_{i,j}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{m \neq t, s} E(C_{i,j}([k] t) C_{v,m}([u] s))). \]

\[ = A + B + C + D + \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{j \neq m, t, s} E(C_{i,j}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{m \neq t, s} E(C_{i,v}([k] t) C_{v,m}([u] s)) + \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{n} \sum_{m=1}^{H} \sum_{j \neq t, s} E(C_{i,j}([k] t) C_{v,m}([u] s)) \]

\[ = A + B + C + D + 0 + \sum_{i=1}^{H} \sum_{v=1}^{n} \sum_{j=1}^{n} \sum_{j \neq t, s} E(C_{i,j}([k] t) C_{v,j}([u] s)) \]

\[ = A + B + C + D + E. \]
Now we evaluate each term separately,

\[
A = \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{[i]t}([k]t)C_{[v]m}([u]s))
\]

\[
= \sum_{i=1}^{H} \sum_{v=1}^{H} E(C_{[i]t}([k]t)C_{[v]t}([u]s)) + \sum_{i=1}^{H} \sum_{m=1}^{n} \sum_{m \neq t} E(C_{[i]t}([k]t)C_{[m]t}([u]s))
\]

\[
= \sum_{i=1}^{H} \sum_{v=1}^{H} E(C_{[i]t}([k]t)C_{[v]t}([u]s)) + 0
\]

\[
= o(n).
\]

A similar computation yields

\[
B = \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{[i]s}([k]t)C_{[v]m}([u]s))
\]

\[
= \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{m=1}^{n} E(C_{[i]s}([k]t)C_{[v]m}([u]s)) + o(n)
\]

\[
= (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E([I(X_{[i]}s \leq Y_{[k]}t) - \tau_{ik})(I(X_{[v]}m \leq Y_{[u]}s) - \tau_{vu})] + o(n)
\]

\[
= (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E([1 - F_{[k]}(X_{[i]}s) - \tau_{ik})(1 - F_{[v]}(Y_{[u]}m) - \tau_{vu})] + o(n).
\]

The term \( C \) can be evaluated as

\[
C = \sum_{i=1}^{H} \sum_{j=1}^{H} \sum_{v=1}^{n} \sum_{j \neq t, s} E(C_{[i]j}([k]t)C_{[v]t}([u]s))
\]

\[
= \sum_{i=1}^{H} \sum_{j=1}^{H} \sum_{v=1}^{n} \sum_{j \neq t, s} E(C_{[i]j}([k]t)C_{[v]t}([u]s))
\]

\[
= (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E([I(X_{[i]}j \leq Y_{[k]}t) - \tau_{ik})(I(X_{[v]}t \leq Y_{[u]}m) - \tau_{vu})]
\]

\[
= (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E([F_{[i]}(Y_{[k]}t) - \tau_{ik})(1 - F_{[v]}(X_{[i]}t) - \tau_{vu})]
\]

\[
= (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E([F_{[i]}(Y_{[k]}t) - \tau_{ik})(1 - F_{[v]}(X_{[i]}t) - \tau_{vu})].
\]
Since all the terms in $D$ are independent, it follows that

\[ D = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{H} E(C_{[i]j}([k]t)C_{[v]s}([u]s)) \]

\[ = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{v=1}^{H} E(C_{[i]j}([k]t)C_{[v]s}([u]s)) \]

\[ = 0. \]

By expanding $E$ in a similar fashion, we obtain

\[ E = \sum_{i=1}^{H} \sum_{v=1}^{H} \sum_{j=1}^{n} E(C_{[i]j}([k]t)C_{[v]j}([u]s)) \]

\[ = (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E(C_{[i]j}([k]t)C_{[v]j}([u]s)) \]

\[ = (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(I(X_{[i]j} \leq Y_{[k]t}) - \tau_{ik})(I(X_{[v]j} \leq Y_{[u]s}) - \tau_{vu})] \]

\[ = (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(1 - F_{[k]}(X_{[i]j}) - \tau_{ik})(1 - F_{[u]}(X_{[v]j}) - \tau_{vu})]. \]

All these terms together yield

\[ \text{Cov}(T_{[k]t}, T_{[u]s}) = A + B + C + D + E \]

\[ = (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(1 - F_{[k]}(X_{[i]s}) - \tau_{ik})(F_{[v]}(Y_{[u]}s) - \tau_{vu})] + \]

\[ (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(F_{[i]}(Y_{[k]}t) - \tau_{ik})(1 - F_{[u]}(X_{[v]}t) - \tau_{vu})] + \]

\[ (n - 2) \sum_{i=1}^{H} \sum_{v=1}^{H} E[(1 - F_{[k]}(X_{[i]j}) - \tau_{ik})(1 - F_{[u]}(X_{[v]j}) - \tau_{vu})] + o(n). \]
Thus, after combining the above results, we get

\[
\text{Var}(T^*) = \sum_{k=1}^{H} \sum_{i=1}^{n} \text{Var}(T_{[k]i}) + \sum_{k=1}^{H} \sum_{u=1}^{n} \sum_{t=1}^{n} \sum_{s=1}^{n} \text{Cov}(T_{[k]t}, T_{[u]s})
\]

\[
= \sum_{u=1}^{H} \sum_{t=1}^{n} \text{Cov}(T_{[u]t}, T_{[u]t}) + \sum_{k=1}^{H} \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{u=1}^{n} \text{Cov}(T_{[k]t}, T_{[u]s})
\]

\[
= n \sum_{k=1}^{H} \sum_{u=1}^{H} \text{Cov}(T_{[k]1}, T_{[u]1}) + \sum_{k=1}^{H} \sum_{u=1}^{H} \sum_{t=1}^{n} \sum_{s=1}^{n} \text{Cov}(T_{[k]t}, T_{[u]s})
\]

\[
= n(n-1)(n-2)\left\{ \sum_{k=1}^{H} \sum_{u=1}^{H} \sum_{i=1}^{n} \sum_{v=1}^{n} E[(F_{[i]}(Y_{[k]1}) - \tau_{ik})(F_{[v]}(Y_{[u]1}) - \tau_{vu})] + \sum_{k=1}^{H} \sum_{u=1}^{H} \sum_{i=1}^{n} \sum_{v=1}^{n} E[(1 - F_{[i]}(X_{[i]1}) - \tau_{ik})(F_{[v]}(Y_{[u]1}) - \tau_{vu})] + \sum_{k=1}^{H} \sum_{u=1}^{H} \sum_{i=1}^{n} \sum_{v=1}^{n} E[(F_{[i]}(Y_{[k]1}) - \tau_{ik})(1 - F_{[u]}(X_{[i]1}) - \tau_{vu})] + \sum_{k=1}^{H} \sum_{u=1}^{H} \sum_{i=1}^{n} \sum_{v=1}^{n} E[(1 - F_{[i]}(X_{[i]1}) - \tau_{ik})(1 - F_{[v]}(X_{[v]1}) - \tau_{vu})] \right\} + o(n^{3}).
\]

By using (2.11), we simplify the variance of \( T^* \) to

\[
\text{Var}(T^*) = n(n-1)(n-2)H^2\left\{ \sum_{k=1}^{H} \sum_{u=1}^{H} E[(F(Y_{[k]1}) - \bar{\tau}_k)(F(Y_{[u]1}) - \bar{\tau}_u)] + \sum_{i=1}^{H} \sum_{u=1}^{H} E[(1 - F(X_{[i]1}) - \bar{\tau}_i)(F(Y_{[u]1}) - \bar{\tau}_u)] + \sum_{k=1}^{H} \sum_{v=1}^{H} E[(F(Y_{[k]1}) - \bar{\tau}_k)(1 - F(X_{[v]1}) - \bar{\tau}_v)] + \sum_{i=1}^{H} \sum_{v=1}^{H} E[(1 - F(X_{[i]1}) - \bar{\tau}_i)(1 - F(X_{[v]1}) - \bar{\tau}_v)] \right\} + o(n^{3}).
\]
The variance of $T^*$ in the above equation does not specify the design parameters $\alpha$ and $\beta$ within each set. If we partition the sums into the terms that contain $\alpha$ and $\beta$ parameters, $\text{Var}(T^*)$ can be rewritten as follows:

\[
\text{Var}(T^*) = n(n - 1)(n - 2)H^2 \{ \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{[\alpha_i]1}) - \bar{\alpha}_i)(F(Y_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{[\beta_i]1}) - \bar{\beta}_i)(F(Y_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(F(Y_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(F(Y_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(F(Y_{[\alpha_i]1}) - \bar{\alpha}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(F(Y_{[\beta_i]1}) - \bar{\beta}_i)(1 - F(X_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(1 - F(X_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(F(Y_{[\beta_j]1}) - \bar{\beta}_j)] + \\
2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(F(Y_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\beta_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + o(n^3).}
\]

By combining the third and the sixth, and the fourth and the fifth terms, we obtain

\[
\text{Var}(T^*) = n(n - 1)(n - 2)H^2 \{ \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{[\alpha_i]1}) - \bar{\alpha}_i)(F(Y_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{[\beta_i]1}) - \bar{\beta}_i)(F(Y_{[\beta_j]1}) - \bar{\beta}_j)] + \\
2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(F(Y_{[\beta_j]1}) - \bar{\beta}_j)] + \\
2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(F(Y_{[\alpha_j]1}) - \bar{\alpha}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \bar{\alpha}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + \\
\sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\beta}_i)(1 - F(X_{[\beta_j]1}) - \bar{\beta}_j)] + o(n^3).
\]
In limit, the variance of $T^*$ is reduced to

$$\lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(T^*) = \frac{1}{H} \{ \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{\alpha_i1}) - \overline{\tau}_{\alpha_i})(F(Y_{\alpha_j1}) - \overline{\tau}_{\alpha_j})] +$$

$$\sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{\beta_i1}) - \overline{\tau}_{\beta_i})(F(Y_{\beta_j1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(F(Y_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_j]1}) - \overline{\tau}_{\beta_j})(F(Y_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})] +$$

$$\sum_{\alpha_i} \sum_{\alpha_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(1 - F(X_{[\alpha_j]1}) - \overline{\tau}_{\alpha_j})] +$$

$$\sum_{\beta_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \overline{\tau}_{\beta_i})(1 - F(X_{[\beta_j]1}) - \overline{\tau}_{\beta_j})].$$

Note that the variance of the projection of $T^*$ can be written as

$$\text{Var}(V_P) = (nH)^2 \text{Var} \{ \sum_{i=1}^{n} \sum_{j=1}^{u} (1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \overline{\tau}_{\beta_k}) \} +$$

$$\sum_{j=1}^{n} \sum_{i=1}^{u} (1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \overline{\tau}_{\beta_k}) \} \}$$

$$= (nH)^2 n \{ \text{Var} \{ \sum_{i=1}^{u} (1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \overline{\tau}_{\beta_k}) \} +$$

$$\text{Var} \{ \sum_{i=1}^{u} (1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \overline{\tau}_{\beta_k}) \} \}.$$

After evaluating the variance and taking the limit as $n$ approaches to infinity, we obtain

$$\lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(V_P) = \frac{1}{H} \{ \text{Var} \{ \sum_{\alpha_i} (1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) + \sum_{\beta_i} (F(Y_{[\beta_i]1}) - \overline{\tau}_{\beta_i}) \} +$$

$$\text{Var} \{ \sum_{\beta_i} (1 - F(X_{[\beta_i]1}) - \overline{\tau}_{\beta_i}) + \sum_{\alpha_i} (F(Y_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i}) \} \}$$

$$= \frac{1}{H} \{ \sum_{\alpha_i} \sum_{\alpha_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(1 - F(X_{[\alpha_j]1}) - \overline{\tau}_{\alpha_j})] +$$

$$\sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{[\beta_i]1}) - \overline{\tau}_{\beta_i})(F(Y_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(F(Y_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(1 - F(X_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$\sum_{\alpha_i} \sum_{\alpha_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(1 - F(X_{[\alpha_j]1}) - \overline{\tau}_{\alpha_j})] +$$

$$\sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{[\beta_i]1}) - \overline{\tau}_{\beta_i})(F(Y_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(F(Y_{[\beta_j]1}) - \overline{\tau}_{\beta_j})] +$$

$$2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\alpha_i]1}) - \overline{\tau}_{\alpha_i})(1 - F(X_{[\beta_j]1}) - \overline{\tau}_{\beta_j})].$$
\[
\begin{align*}
\sum_{\beta_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_i]1}) - \bar{\tau}_{\beta_i})(1 - F(X_{[\beta_j]1}) - \bar{\tau}_{\beta_j})] + \\
\sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{[\alpha_i]1}) - \bar{\tau}_{\alpha_i})(F(Y_{[\alpha_j]1}) - \bar{\tau}_{\alpha_j})] + \\
2 \sum_{\alpha_i} \sum_{\beta_j} E[(1 - F(X_{[\beta_j]1}) - \bar{\tau}_{\beta_j})(F(Y_{[\alpha_i]1}) - \bar{\tau}_{\alpha_i})].
\end{align*}
\]

We finally observe that

\[
\lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(T^*) = \lim_{n \to \infty} \frac{1}{(nH)^3} \text{Var}(V_P)
\]

and this completes the proof. \qed

**Theorem 2.5.** The null distribution of \(\sqrt{2nH(T - \frac{1}{2})}\), as replication size \(n\) goes to infinity, converges to a normal distribution with mean 0 and variance \(\sigma^2\), where

\[
\sigma^2 = \frac{2}{H} \{ \text{Var} \sum_{i=1}^{u} (1 - F(X_{[\alpha_i]1}) - \bar{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \bar{\tau}_{\beta_k}) \} + \\
\text{Var} \sum_{i=1}^{H-u} (1 - F(X_{[\beta_i]1}) - \bar{\tau}_{\beta_i}) + \sum_{k=1}^{u} (F(Y_{[\alpha_k]1}) - \bar{\tau}_{\alpha_k}) \}.
\]

**Proof.** From Theorem 2.1 and Lemma 2.4, we establish that \(\sqrt{2nHT^*}\) converges to a normal distribution with mean zero and variance \(\sigma^2\). On the other hand, since \(T^* = T - E[T]\) and \(\overline{T}^* = \overline{T} - E[T]\), we need to compute \(E[T]\) under the null hypothesis to finish the proof. Notice that

\[
E[T] = E \left[ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{[i]j} \leq Y_{[k]t}) \right]
\]

\[
= E \left\{ \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{u} \sum_{t=1}^{n} I(X_{[\alpha_i]j} \leq Y_{[\alpha_k]t}) \right\} + \\
\sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(X_{[\alpha_i]j} \leq Y_{[\beta_k]t}) \} + \\
\sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{u} \sum_{t=1}^{n} I(X_{[\beta_i]j} \leq Y_{[\alpha_k]t}) \} + \\
\sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(X_{[\beta_i]j} \leq Y_{[\beta_k]t}) \} + \\
\sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{[\beta_i]j} \leq Y_{[\beta_k]t}) \}.
\]
\[ \sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} [I(X_{[\beta_i]j} \leq Y_{[\beta_k]t})]. \]

The evaluation of the expectations in the above equation yields

\[ E[T] = n^2 E\{ \sum_{i=1}^{u} \sum_{k=1}^{u} [I(X_{[\alpha_i]1} \leq Y_{[\alpha_k]2})] \} + \]
\[ n(n - 1) E\{ \sum_{i=1}^{u} \sum_{k=1}^{H-u} [I(X_{[\alpha_i]1} \leq Y_{[\beta_k]2})] \} + \]
\[ n(n - 1) E\{ \sum_{i=1}^{H-u} \sum_{k=1}^{u} [I(X_{[\beta_i]1} \leq Y_{[\alpha_k]2})] \} + \]
\[ n^2 E\{ \sum_{i=1}^{H-u} \sum_{k=1}^{H-u} [I(X_{[\beta_i]1} \leq Y_{[\beta_k]2})] \} + \]
\[ nE\{ \sum_{i=1}^{H} \sum_{k=1}^{H} [I(X_{[i]1} \leq Y_{[k]1})] \}. \]

In an ORRD, the observations from different sets are independent, we then express

\[ E[T] \]

as

\[ E[T] = n^2 E\{ \sum_{i=1}^{u} \sum_{k=1}^{u} [F_{[\alpha_i]1}(Y_{[\alpha_k]2})] \} + n(n - 1) E\{ \sum_{i=1}^{u} \sum_{k=1}^{H-u} [F_{[\alpha_i]1}(Y_{[\beta_k]2})] \} + \]
\[ n(n - 1) E\{ \sum_{i=1}^{H-u} \sum_{k=1}^{u} [F_{[\beta_i]1}(Y_{[\alpha_k]2})] \} + n^2 E\{ \sum_{i=1}^{H-u} \sum_{k=1}^{H-u} [F_{[\beta_i]1}(Y_{[\beta_k]2})] \} + \]
\[ o(n^2). \]

Again, according to (2.11), the expected value of \( T \) reduces to

\[ E[T] = n^2 H E\{ \sum_{k=1}^{u} [F(Y_{[\alpha_k]})] \} + n^2 H E\{ \sum_{k=1}^{H-u} [F(Y_{[\beta_k]})] \} + o(n^2) \]
\[ = n^2 H E\{ \sum_{k=1}^{H} [F(Y_{[k]})] \} + o(n^2) \]
\[ = \frac{n^2 H^2}{2} + o(n^2). \]

Consequently, \( E[T] = \frac{1}{2} + o(1) \). The proof is completed by using the Slutsky Theorem.
2.2 Asymptotic properties of $T$ under perfect ranking

The asymptotic theory in Section 2.1 holds under a general ranking scheme as long as we have a consistent ranking procedure. Under this ranking scheme, $X_{[i]}$’s and $Y_{[j]}$’s are judgment ranked order statistics from the $X$- and $Y$- samples. If the ranking information is complete (perfect), the judgment ranked order statistics become standard order statistics and denoted by $X_{(i)}$’s and $Y_{(j)}$’s. In this case, the two samples can be written as \{\{X_{(i)}\}, i = 1, ..., H; j = 1, ..., n\} and \{\{Y_{(j)}\}, k = 1, ..., H; t = 1, ..., n\}. They can also be rearranged in terms of design parameters $\alpha$ and $\beta$ as follows.

$$(X_{(\alpha_1)}1, ..., X_{(\alpha_u)}1, X_{(\beta_1)}1, ..., X_{(\beta_{H-u})}1, ..., X_{(\alpha_1)}n, ..., X_{(\alpha_u)}n, X_{(\beta_1)}n, ..., X_{(\beta_{H-u})}n),$$

and

$$(Y_{(\alpha_1)}1, ..., Y_{(\alpha_u)}1, Y_{(\beta_1)}1, ..., Y_{(\beta_{H-u})}1, ..., Y_{(\alpha_1)}n, ..., Y_{(\alpha_u)}n, Y_{(\beta_1)}n, ..., Y_{(\beta_{H-u})}n),$$

where sets $\alpha = (\alpha_1, ..., \alpha_u)$ and $\beta = (\beta_1, ..., \beta_{H-u})$ are two pre-determined, disjoint, nonempty subsets of $(1, ..., H)$. Let $F_{(i)}$ and $G_{(i)}$ be the c.d.f.s for $X_{(i)}$ and $Y_{(i)}$, respectively.

The proposed rank-sum test statistic $T$ under perfect ranking becomes

$$T = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{(i)j} \leq Y_{(k)t}).$$

In Section 2.1, under the null hypothesis we have shown that properly centered and scaled version of the test statistic $T$ has an asymptotically normal distribution with mean 0 and variance $\sigma^2$ regardless of the quality of ranking information. In this case, asymptotic null variance of $T$ depends on the underlying distribution and judgment.
ranking model. If the quality of ranking information is perfect, the asymptotic
distribution of $T$ is still normal, but $\sigma^2$ is distribution free and reduces to an explicit
expression that is a function of the design parameters $\alpha$ and $\beta$.

**Theorem 2.6.** Under perfect ranking, the asymptotic null distribution of $\sqrt{2nH(T - \frac{1}{2})}$ converges to a normal distribution with mean zero and variance $\sigma^2_p$, where

$$
\sigma^2_p = \frac{4}{H} \left\{ \sum_{i=1}^{u} \frac{\alpha_i (H+1-\alpha_i)}{(H+1)^2(H+2)} + 2 \sum_{i=1}^{u} \sum_{\alpha_i < \alpha_j} \frac{\alpha_i (H+1-\alpha_j)}{(H+1)^2(H+2)} + \sum_{k=1}^{H-u} \frac{\beta_k (H+1-\beta_k)}{(H+1)^2(H+2)} + 2 \sum_{\beta_k < \beta_l} \sum_{t=1}^{H-u} \frac{\beta_k (H+1-\beta_l)}{(H+1)^2(H+2)} \right\}.
$$

Proof. From Theorem 2.5, for an arbitrary but consistent ranking scheme, we know that $\sqrt{2nH(T - \frac{1}{2})}$ converges to a normal distribution with mean 0 and variance $\sigma^2$.

We then only need to compute the variance under perfect ranking.

Under the perfect ranking, this variance can be written as

$$
\sigma^2_p = \frac{2}{H} \left\{ \text{Var} \left[ \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)}) - \bar{\alpha}_{i^*}) + \sum_{k=1}^{H-u} (F(Y_{(\beta_k)}) - \bar{\beta}_{k^*}) \right] + \text{Var} \left[ \sum_{i=1}^{H-u} (1 - F(X_{(\beta_i)}) - \bar{\beta}_{i^*}) + \sum_{k=1}^{u} (F(Y_{(\alpha_k)}) - \bar{\alpha}_{k^*}) \right] \right\}.
$$

In the above equation we first consider $\text{Var} \left[ \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)}) - \bar{\alpha}_{i^*}) + \sum_{k=1}^{H-u} (F(Y_{(\beta_k)}) - \bar{\beta}_{k^*}) \right]
= \text{Var} \left( \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)}) \right) + \text{Var} \left( \sum_{k=1}^{H-u} F(Y_{(\beta_k)}) \right) - 2 \text{Cov} \left( \sum_{i=1}^{u} F(X_{(\alpha_i)}), \sum_{k=1}^{H-u} F(Y_{(\beta_k)}) \right).

39
The first term in the above equation can be written as

\[ \text{Var} \left( \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)})) \right) = \sum_{i=1}^{u} \text{Var}(F(X_{(\alpha_i)})) + \sum_{i=1}^{u} \sum_{j=1}^{u} \text{Cov}(F(X_{(\alpha_i)}), F(X_{(\alpha_j)})) \].

We now need to calculate \( \text{Var}(F(X_{(\alpha_i)})) \) and \( \text{Cov}(F(X_{(\alpha_i)}), F(X_{(\alpha_j)})) \). It is easy to show that

\[ \text{Var}(F(X_{(\alpha_i)})) = \int F^2(t) f_{\alpha_i; H}(t) dt - \left[ \int F(t) f_{\alpha_i; H}(t) dt \right]^2 \]

or

\[ \frac{\alpha_i (\alpha_i + 1)}{(H + 1)(H + 2)} - \left[ \frac{\alpha_i}{H + 1} \right]^2 = \frac{\alpha_i (H + 1 - \alpha_i)}{(H + 1)^2(H + 2)}. \]

A similar consideration yields that

\[ \text{Cov}(F(X_{(\alpha_i)}), F(X_{(\alpha_j)})) = \frac{\alpha_i (1 - \frac{\alpha_j + 1}{H + 1})}{H + 2} = \frac{\alpha_i (H + 1 - \alpha_j)}{(H + 1)^2(H + 2)} \text{ if } \alpha_i < \alpha_j. \]

By combining the above results, we obtain

\[ \text{Var} \left( \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)})) \right) = \sum_{i=1}^{u} \frac{\alpha_i (H + 1 - \alpha_i)}{(H + 1)^2(H + 2)} + 2 \sum_{i=1}^{u} \sum_{j=1}^{u} \frac{\alpha_i (H + 1 - \alpha_j)}{(H + 1)^2(H + 2)}. \]

Under the null hypothesis, \( X \) and \( Y \) distributions are the same, then it follows that

\[ \text{Var} \left( \sum_{k=1}^{H-u} F(Y_{(\beta_k)}) \right) = \sum_{k=1}^{H-u} \frac{\beta_k (H + 1 - \beta_k)}{(H + 1)^2(H + 2)} + 2 \sum_{k=1}^{H-u} \sum_{l=1}^{H-u} \frac{\beta_k (H + 1 - \beta_l)}{(H + 1)^2(H + 2)} \]

and

\[ \text{Cov} \left( \sum_{i=1}^{u} F(X_{(\alpha_i)}), \sum_{k=1}^{H-u} F(Y_{(\beta_k)}) \right) = \sum_{i=1}^{u} \sum_{k=1}^{H-u} \text{Cov}(F(X_{(\alpha_i)}), F(Y_{(\beta_k)})) \]

or

\[ \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\alpha_i (H + 1 - \beta_k)}{(H + 1)^2(H + 2)} + \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\beta_k (H + 1 - \alpha_i)}{(H + 1)^2(H + 2)}. \]
The combination of the variance and covariance expressions finally yields

\[
Var \left[ \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)1}) - \tilde{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{(\beta_k)1}) - \tilde{\tau}_{\beta_k}) \right]
\]

\[
= \sum_{i=1}^{u} \frac{\alpha_i (H + 1 - \alpha_i)}{(H + 1)^2 (H + 2)} + 2 \sum_{i=1}^{u} \sum_{\alpha_i < \alpha_j}^{u} \frac{\alpha_i (H + 1 - \alpha_j)}{(H + 1)^2 (H + 2)} + \sum_{k=1}^{H-u} \frac{\beta_k (H + 1 - \beta_k)}{(H + 1)^2 (H + 2)} + 2 \sum_{k=1}^{H-u} \sum_{\beta_k < \beta_t}^{H-u} \frac{\beta_k (H + 1 - \beta_t)}{(H + 1)^2 (H + 2)} - 2 \sum_{i=1}^{u} \sum_{\alpha_i < \beta_k}^{H-u} \frac{\alpha_i (H + 1 - \beta_k)}{(H + 1)^2 (H + 2)} - 2 \sum_{i=1}^{u} \sum_{\beta_k < \alpha_i}^{H-u} \frac{\beta_k (H + 1 - \alpha_i)}{(H + 1)^2 (H + 2)}.
\]

Again, under the null hypothesis we observe that

\[
Var \left[ \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)1}) - \tilde{\tau}_{\alpha_i}) + \sum_{k=1}^{u} (F(Y_{(\alpha_k)1}) - \tilde{\tau}_{\alpha_k}) \right]
\]

\[
= Var \left[ \sum_{i=1}^{u} (1 - F(X_{(\alpha_i)1}) - \tilde{\tau}_{\alpha_i}) + \sum_{k=1}^{H-u} (F(Y_{(\beta_k)1}) - \tilde{\tau}_{\beta_k}) \right].
\]

This completes the proof. \(\square\)

We wish to compare the efficiency of the proposed test with its competitors from the literature. The comparisons of the tests depend on how the type I and type II error probabilities are controlled and how the alternative hypotheses behave as the sample size gets large. One of the common criteria to compare nonparametric (as well as parametric) testing procedures is to compare the Pitman efficacies of the two competing tests. We first introduce regularity conditions for Pitman efficacy.

**Definition 2.1.** An estimating function \(S(\theta)\) is **Pitman regular** if the following four conditions C1-C4 hold.

**C1:** The estimating function \(S(\theta)\) is nonincreasing in \(\theta\);
C2: Let $\overline{S}(\theta) = S(\theta)/n^\gamma$ for some $\gamma > 0$. Then, there exist a function $\mu(\theta)$, such that

\[
\mu(0) = 0, \mu'(0) \text{ is continuous at } 0, \mu'(0) > 0 \text{ and }
\]

either $\overline{S}(0) \xrightarrow{p} \mu(\theta)$ or $E_\theta(\overline{S}(0) = \mu(\theta)$;

C3:

\[
\sup_{|b| \leq B} \left| \frac{\sqrt{n}S(b)}{\sqrt{n}} - \sqrt{n}S(0) + \mu'(0) b \right| \xrightarrow{p} 0 \text{ for any } B > 0;
\]

C4: There is a constant $\sigma(0)$ such that

\[
\sqrt{n} \left\{ \frac{\overline{S}(0)}{\sigma(0)} \right\} \xrightarrow{d_0} N(0, 1).
\]

Further, the quantity

\[
c = \frac{\mu'(0)}{\sigma(0)}
\]

is called the **efficacy** of $S(\theta)$.

The quantity $c$ in this definition has big impact on the properties of the test, estimator and confidence interval that are derived from the estimating equation $S(\theta)$. It measures the rate of change (in a standardized unit) in the asymptotic mean of $\overline{S}$ at the null hypothesis. A test with large $c$ responds quickly to small changes in the alternative hypothesis. We therefore expect that such a test has a higher local power.

**Theorem 2.7.** Let $X$ and $Y$ be two random variables with c.d.f. $F(x)$ and $G(y) = F(y - \Delta)$ respectively, where $\Delta$ is the location shift parameter between the distributions of $X$ and $Y$. Let $f(x)$ be the p.d.f. corresponding to $F(x)$. Assume that $\int f^2(x)dx$ is
bounded and $X$- and $Y$-samples are collected under ORRD with perfect ranking. Let

$$T^*(\Delta) = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X(i)_{j} \leq Y(k)_{t} - \Delta) - \tau_{ik}],$$

where $\tau_{ik}$ is defined as in (2.7). Then

(i). $T^*(\Delta)$ is Pitman regular.

(ii). The efficacy of $T^*(\Delta)$ is

$$c_{ORRD}^2 = \frac{(\int f^2(x)dx)^2}{\sigma_p^2},$$

where

$$\sigma_p^2 = \frac{4}{H} \left( \sum_{i=1}^{u} \alpha_i (H + 1 - \alpha_i) \right) + 2 \left( \sum_{i=1}^{u} \sum_{j=1}^{u} \frac{\alpha_i (H + 1 - \alpha_j)}{(H + 1)^2 (H + 2)} \right) +$$

$$\sum_{k=1}^{H-u} \frac{\beta_k (H + 1 - \beta_k)}{(H + 1)^2 (H + 2)} + 2 \sum_{k=1}^{H-u} \sum_{i=1}^{u} \frac{\beta_k (H + 1 - \beta_i)}{(H + 1)^2 (H + 2)} +$$

$$-2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\alpha_i (H + 1 - \beta_k)}{(H + 1)^2 (H + 2)} - 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\beta_k (H + 1 - \alpha_i)}{(H + 1)^2 (H + 2)}.$$
The last equation simplifies to

\begin{align*}
E_{\Delta}(T^*(0)) &= \frac{1}{(nH)^2} \Delta \left\{ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} \left[ I(X(i)_j \leq Y(k)_t) - \tau_{ik} \right] \right\} \\
&= \frac{1}{(nH)^2} \Delta \left\{ \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} \left[ I(X(\alpha)_j \leq Y(\beta)_t) - \tau_{\alpha\beta} \right] + \right. \\
&\left. \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} \left[ I(X(\beta)_j \leq Y(\alpha)_t) - \tau_{\beta\alpha} \right] + \right. \\
&\left. \sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} \left[ I(X(\beta)_j \leq Y(\beta)_t) - \tau_{\beta\beta} \right] \right\}. \\
\end{align*}

Combine the common terms and ignore the lower order terms, we then have

\begin{align*}
E_{\Delta}(T^*(0)) &= \frac{1}{(nH)^2} \left\{ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} E \left[ 1 - F(\alpha_k)(X(\alpha)_j - \Delta) - \tau_{\alpha\alpha} \right] + \right. \\
&\left. (n-1) \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} E \left[ 1 - F(\beta_k)(X(\alpha)_j - \Delta) - \tau_{\alpha\beta} \right] + \right. \\
&\left. (n-1) \sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} E \left[ 1 - F(\beta_k)(X(\beta)_j - \Delta) - \tau_{\beta\alpha} \right] + \right. \\
&\left. n \sum_{i=1}^{H-u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} E \left[ 1 - F(\beta_k)(X(\beta)_j - \Delta) - \tau_{\beta\beta} \right] \right\} + o(1) \\
&= \frac{1}{(nH)^2} \left\{ nH \sum_{i=1}^{u} \sum_{j=1}^{n} E \left[ 1 - F(X(\alpha)_j - \Delta) - \tau_{\alpha\alpha} \right] + \right. \\
&\left. nH \sum_{i=1}^{H-u} \sum_{j=1}^{n} E \left[ 1 - F(X(\beta)_j - \Delta) - \tau_{\beta\beta} \right] \right\} + o(1) \\
&= \frac{1}{(nH)^2} \left\{ n^2H \sum_{i=1}^{u} \left[ 1 - \int F(X(\alpha)_i - \Delta) dF(\alpha)(X(\alpha)_i) - \tau_{\alpha\alpha} \right] + \right. \\
&\left. n^2H \sum_{i=1}^{H-u} \left[ 1 - \int F(X(\beta)_i - \Delta) dF(\beta)(X(\beta)_i) - \tau_{\beta\beta} \right] \right\} + o(1).
\end{align*}

The last equation simplifies to

\begin{align*}
E_{\Delta}(T^*(0)) &= \left[ 1 - \int F(x - \Delta) f(x) dx \right] - \left( \sum_{i=1}^{u} \tau_{\alpha\alpha} + \sum_{i=1}^{H-u} \tau_{\beta\beta} \right) + o(1).
\end{align*}
\[ = \left[1 - \int F(x - \Delta)f(x)dx\right] - \left(\sum_{i=1}^{H} \sum_{i=1}^{H} \tau_{ij}/H^2\right) + o(1).\]

Hence, the limit of this expectation as \( n \) goes to infinity yields that

\[
\lim_{n \to \infty} E_\Delta(T^*(0)) = \left[1 - \int F(x - \Delta)f(x)dx\right] - \left(\sum_{i=1}^{H} \sum_{i=1}^{H} \tau_{ij}/H^2\right).
\]

Let

\[
\mu(\Delta) = \lim_{n \to \infty} E_\Delta(T^*(0)).
\]

We then observe that

\[
\mu(0) = \lim_{n \to \infty} E_0(T^*(0))
\]

\[
= \lim_{n \to \infty} E_0\left\{\sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{(i)j} \leq Y_{(k)t}) - \tau_{ik}]\right\}
\]

\[
= \lim_{n \to \infty} \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [E_0(I(X_{(i)j} \leq Y_{(k)t}) - \tau_{ik}]
\]

\[
= 0,
\]

and

\[
\mu'(0) = \frac{d}{d\Delta} \mu(\Delta) |_{\Delta=0} = -\int \frac{d}{d\Delta} F(x - \Delta)|_{\Delta=0} f(x)dx = \int f^2(x)dx > 0,
\]

which completes the proof of C2.

C3. Let \( N = 2nH \). We need to show that

\[
\sup_{|b| \leq B} \left|\sqrt{N}T^*(\frac{b}{\sqrt{N}}) - \sqrt{N}T^*(0) + \mu'(0) b\right| \overset{p_0}{\to} 0 \quad \forall B > 0.
\]

Let \( \overline{U} = \frac{T^*(\Delta) - T^*(0)}{\Delta} \), where \( \Delta = \frac{b}{\sqrt{N}} \). Notice that this \( \Delta \) depends on \( N \) and \( \Delta \to 0 \) as \( N \to \infty \). We take the following three steps to prove the result. We first show that \( \overline{U} \) converges to zero in probability for a fixed \( b \) through steps 1 and 2. Then the uniform convergence result is Step 3 from the monotonicity of \( T^*(\Delta) \) in \( \Delta \).
Step 1: We show that \( \lim_{N \to \infty} E_0(U) = - \int f^2(y)dy = -\mu'(0) \). There are two distinct cases \( b > 0 \) and \( b < 0 \). Assume that \( b > 0 \). It is not difficult to observe that the following equations hold:

\[
E_0(U) = \frac{1}{\Delta} E_0(\mathcal{T}^x(\Delta) - \mathcal{T}^x(0))
\]

\[
= \frac{1}{\Delta} E_0\left\{ \frac{1}{(nH)^2} \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{(i)j} \leq Y_{(k)t} - \Delta) - \tau_{ik}] - \frac{1}{(nH)^2} \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{(i)j} \leq Y_{(k)t}) - \tau_{ik}] \right\}
\]

\[
= \frac{1}{(nH)^2 \Delta} E_0\left\{ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{(i)j} \leq Y_{(k)t} - \Delta) - I(X_{(i)j} \leq Y_{(k)t})] \right\}
\]

\[
= \frac{1}{(nH)^2 \Delta} E_0\left\{ \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(Y_{(\alpha_k)t} - \Delta < X_{(\alpha_i)j} \leq Y_{(\alpha_k)t}) + \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(Y_{(\beta_k)t} - \Delta < X_{(\alpha_i)j} \leq Y_{(\beta_k)t}) + \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(Y_{(\alpha_k)t} - \Delta < X_{(\beta_i)j} \leq Y_{(\alpha_k)t}) + \sum_{i=1}^{u} \sum_{j=1}^{n} \sum_{k=1}^{H-u} \sum_{t=1}^{n} I(Y_{(\beta_k)t} - \Delta < X_{(\beta_i)j} \leq Y_{(\beta_k)t}) \right\}
\]

\[
= \frac{1}{(nH)^2 \Delta} \left\{ n \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{t=1}^{n} E[F_{(\alpha_i)}(y_{(\alpha_k)t}) - F_{(\alpha_i)}(y_{(\alpha_k)t} - \Delta)] + (n - 1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{t=1}^{n} E[F_{(\alpha_i)}(y_{(\beta_k)t}) - F_{(\alpha_i)}(y_{(\beta_k)t} - \Delta)] + (n - 1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{t=1}^{n} E[F_{(\beta_i)}(y_{(\alpha_k)t}) - F_{(\beta_i)}(y_{(\alpha_k)t} - \Delta)] + n \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{t=1}^{n} E[F_{(\beta_i)}(y_{(\beta_k)t}) - F_{(\beta_i)}(y_{(\beta_k)t} - \Delta)] \right\}
\]
\[
= \frac{-1}{(nH)^2} \left\{ \sum_{k=1}^{u} \sum_{t=1}^{n} nH E[F(y_{(\alpha_k)t}) - F(y_{(\alpha_k)t} - \Delta)] + \right.
\sum_{k=1}^{H-u} \sum_{t=1}^{n} nH E[F(y_{(\beta_k)t}) - F(y_{(\beta_k)t} - \Delta)] \left\} + o(1) \right.
\]

We take the limit in the last equation as \( N \) goes to infinity,

\[
\lim_{N \to \infty} E_0(U) = \lim_{N \to \infty} \frac{-1}{nH} \left\{ \sum_{k=1}^{u} \sum_{t=1}^{n} E \left[ \frac{F(y_{(\alpha_k)t}) - F(y_{(\alpha_k)t} - \Delta)}{\Delta} \right] + \right.
\sum_{k=1}^{H-u} \sum_{t=1}^{n} E \left[ \frac{F(y_{(\beta_k)t}) - F(y_{(\beta_k)t} - \Delta)}{\Delta} \right] \left\} \right. + o(1)
\]

\[
= \lim_{N \to \infty} \frac{-1}{nH} \left\{ \sum_{k=1}^{u} E \left[ \lim_{N \to \infty} \frac{F(y_{(\alpha_k)t}) - F(y_{(\alpha_k)t} - \Delta)}{\Delta} \right] + \right.
\sum_{k=1}^{H-u} E \left[ \lim_{N \to \infty} \frac{F(y_{(\beta_k)t}) - F(y_{(\beta_k)t} - \Delta)}{\Delta} \right] \left\} \right. + o(1)
\]

Since \( \Delta \) is a function of \( N \) and \( \Delta \to 0 \) as \( N \to \infty \),

\[
\lim_{N \to \infty} E_0(U) = \frac{-1}{H} \left\{ \sum_{k=1}^{u} E[f(y_{(\alpha_k)})] + \sum_{k=1}^{H-u} E[f(y_{(\beta_k)})] \right\}
\]

\[
= \frac{-1}{H} \left\{ \sum_{k=1}^{u} \int f(y_{(\alpha_k)}) f(y_{(\alpha_k)}) dy_{(\alpha_k)} + \sum_{k=1}^{H-u} \int f(y_{(\beta_k)}) f(y_{(\beta_k)}) dy_{(\beta_k)} \right\}
\]

\[
= \frac{-1}{H} \left\{ \sum_{k=1}^{u} \int f(y) f(y) dy + \sum_{k=1}^{H-u} \int f(y) f(y) dy \right\}
\]

\[
= \frac{-1}{H} \int f^2(y) dy = - \int f^2(y) dy.
\]

Similarly, if \( b < 0 \),

\[
\lim_{N \to \infty} E_0(U) = - \int f^2(y) dy.
\]
Thus, we conclude that
\[
\lim_{N \to \infty} E_0(\overline{U}) = -\mu'(0).
\]

Step 2: We show that the variance of \( \overline{U} \) converges to 0 as \( N \) goes to infinity. That is, \( \lim_{N \to \infty} \text{Var}_0(\overline{U}) = 0 \). Again, there are two possible cases, \( b > 0 \) and \( b < 0 \).

Assume that \( b > 0 \). The variance of \( \overline{U} \) can be written as
\[
\text{Var}_0(\overline{U}) = \text{Var}_0\left( \frac{T^*(\Delta) - T^*(0)}{\Delta} \right)
= \text{Var}_0\left\{ \frac{1}{(nH)^2} \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} [I(X_{(i)j} \leq Y_{(k)t} - \Delta) - I(X_{(i)j} \leq Y_{(k)t})] \right\}
= \frac{1}{(nH)^4\Delta^2} \text{Var}_0\left[ \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(Y_{(k)t} - \Delta < X_{(i)j} \leq Y_{(k)t}) \right].
\]

Let
\[
U_{[k]t} = \sum_{i=1}^{H} \sum_{j=1}^{n} I(Y_{(k)t} - \Delta < X_{(i)j} \leq Y_{(k)t})
= \sum_{i=1}^{H} \sum_{j=1}^{n} I(X_{(i)j} \leq Y_{(k)t} < X_{(i)j} + \Delta)
\]
and \( U = \sum_{k=1}^{H} \sum_{t=1}^{n} U_{[k]t} \). Then the variance of \( \overline{U} \) becomes
\[
\text{Var}_0(\overline{U}) = \frac{1}{(nH)^4\Delta^2} \text{Var}_0(U)
= \frac{1}{(nH)^4\Delta^2} \left[ \sum_{k=1}^{H} \sum_{t=1}^{n} \sum_{u=1}^{H} \sum_{s=1}^{n} \text{Cov}_0(U_{[k]t}, U_{[u]s}) + \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}) \right].
\]

The calculation of \( \text{Var}_0(U) \) is similar to the one given in \( \text{Var}(T^*) \). Only difference is that we need to replace \( F(y_{(k)}) - \overline{F}_k \) with \( F(y_{(k)}) - F(y_{(k)} - \Delta) \) and \( 1 - F(x_{(i)}) - \overline{F}_k \) with \( F(x_{(i)} + \Delta) - F(x_{(i)}) \).

With these new notations, we obtain that
\[
\lim_{N \to \infty} \text{Var}_0(\overline{U}) = \lim_{N \to \infty} \frac{1}{(nH)^4\Delta^2} \text{Var}_0(U)
= \lim_{N \to \infty} \frac{1}{(nH)^4\Delta^2} n(n-1)(n-2)H^2[A + B + C + D + E + F],
\]
where
\[
\begin{align*}
A &= \frac{H}{n} \sum_{k=1}^{H} \sum_{t=1}^{n} \sum_{u=1}^{H} \sum_{s=1}^{n} \text{Cov}_0(U_{[k]t}, U_{[u]s}), \\
B &= \frac{1}{(nH)^4\Delta^2} \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}), \\
C &= \frac{1}{(nH)^4\Delta^2} \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}), \\
D &= \frac{1}{(nH)^4\Delta^2} \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}), \\
E &= \frac{1}{(nH)^4\Delta^2} \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}), \\
F &= \frac{1}{(nH)^4\Delta^2} \sum_{k=1}^{H} \sum_{t=1}^{n} \text{Var}_0(U_{[k]t}).
\end{align*}
\]
where

\[
A = \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{\alpha_i}) - F(Y_{\alpha_j} - \Delta))(F(Y_{\alpha_j}) - F(Y_{\alpha_j} - \Delta))]
\]

\[
B = \sum_{\beta_i} \sum_{\beta_j} E[(F(Y_{\beta_i}) - F(Y_{\beta_j} - \Delta))(F(Y_{\beta_j}) - F(Y_{\beta_j} - \Delta))]
\]

\[
C = 2 \sum_{\alpha_i} \sum_{\alpha_j} E[(F(X_{\alpha_i} + \Delta) - F(X_{\alpha_i}))(F(Y_{\alpha_j}) - F(Y_{\alpha_j} - \Delta))]
\]

\[
D = 2 \sum_{\alpha_i} \sum_{\alpha_j} E[(F(X_{\beta_j} + \Delta) - F(X_{\beta_j}))(F(Y_{\alpha_i}) - F(Y_{\alpha_i} - \Delta))]
\]

\[
E = \sum_{\alpha_i} \sum_{\alpha_j} E[(F(X_{\alpha_i} + \Delta) - F(X_{\alpha_i}))(F(X_{\alpha_i} + \Delta) - F(X_{\alpha_i}))]
\]

\[
F = \sum_{\beta_i} \sum_{\beta_j} E[(F(X_{\beta_j} + \Delta) - F(X_{\beta_j}))(F(X_{\beta_j} + \Delta) - F(X_{\beta_j}))]
\]

We now show that \(\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} A = 0\),

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} A
\]

\[
= \lim_{N \to \infty} \frac{n(n-1)(n-2)}{n^4 H^2 \Delta^2} \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{\alpha_i}) - F(Y_{\alpha_i} - \Delta))]
\]

\[
(\Delta)
\]

\[
= \lim_{N \to \infty} \frac{n(n-1)(n-2)N}{n^4 H^2 \Delta^2} \sum_{\alpha_i} \sum_{\alpha_j} E[(F(Y_{\alpha_i}) - F(Y_{\alpha_i} - \Delta))]
\]

\[
(\Delta)
\]

\[
= 0.
\]

Similarly, we can show that

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} B = 0
\]

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} C = 0
\]

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} D = 0
\]

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} E = 0
\]

\[
\lim_{N \to \infty} \frac{n(n-1)(n-2)H^2}{(nH)^4 \Delta^2} F = 0.
\]
We finally conclude that \( \lim_{N \to \infty} \text{Var}_0(U) = 0 \).

A similar computations shows that if \( b < 0 \), \( \lim_{N \to \infty} \text{Var}_0(U) = 0 \).

Step 3: Let \( \varepsilon > 0 \) be an arbitrary positive constant. Then

\[
P_0 \left( \left| \frac{T^*(b)}{\sqrt{N}} - T^*(0) - (\mu'(0)) \right| > \varepsilon \right) \]
\[
= P_0 \left( \left| \frac{T^*(b)}{\sqrt{N}} - T^*(0) - E_0(U) + E_0(U) - (\mu'(0)) \right| > \varepsilon \right) \]
\[
\leq P_0 \left( \left| \frac{T^*(b)}{\sqrt{N}} - T^*(0) - E_0(U) \right| > \frac{\varepsilon}{2} \right) + P_0 \left( \left| E_0(U) - (\mu'(0)) \right| > \frac{\varepsilon}{2} \right) \]
\[
\leq \frac{\text{Var}_0(U)}{(\frac{\varepsilon}{2})^2} + P_0 \left( \left| E_0(U) - (\mu'(0)) \right| > \frac{\varepsilon}{2} \right).
\]

The last step in the above calculation used Chebyshev’s inequality. As \( n \) approaches

to infinity, we conclude that

\[
\sqrt{N} \left( \frac{T^*(b)}{\sqrt{N}} - T^*(0) + \mu'(0) \frac{b}{\sqrt{N}} \right) \overset{p}{\to} 0.
\]

To show the uniform convergence, we let

\[
W_N(b) = \sqrt{N} \left( \frac{T^*(b)}{\sqrt{N}} - T^*(0) + \mu'(0) \frac{b}{\sqrt{N}} \right).
\]

Let \( \varepsilon > 0 \) and \( \gamma > 0 \) be two arbitrary positive constants. We partition \([-B, B]\) into

\(-B = b_0 < b_1 < \ldots < b_m = B \) so that \( b_i - b_{i-1} \leq \varepsilon/(2|\mu'(0)|) \) for all \( i \).

Then, there exists an \( N' \) such that, when \( N \geq N' \), \( P(\max_i |W_N(b_i)| > \frac{\varepsilon}{2}) < \gamma \).

Suppose that \( W_N(b) \geq 0 \). Then

\[
|W_N(b)| = \sqrt{N} \left( \frac{T^*(b)}{\sqrt{N}} - T^*(0) \right) + \mu'(0) \frac{b}{\sqrt{N}} \]
\[
= \sqrt{N} \left( \frac{T^*(b)}{\sqrt{N}} - T^*(0) \right) + b_{i-1} \mu'(0) + (b - b_{i-1}) \mu'(0) \]
\[
\leq |W_N(b_{i-1})| + (b - b_{i-1})|\mu'(0)| \quad \text{(since } T \text{ is non-increasing in } b) \]
\[
\leq \max_i |W_N(b_i)| + \frac{\varepsilon}{2}.
\]
and we calculate that

\[ P_0(\sup_{|b| \leq B} |W_N(b)| > \varepsilon) \leq P_0(\max_i |W_N(b_i)| + \frac{\varepsilon}{2} > \varepsilon) < \gamma. \]

Similar result also holds when \( W_N(b) < 0 \). Then this yields the desired result

\[ \sup_{|b| \leq B} \left| \sqrt{N} T^*(b) - \sqrt{N} T^*(0) + \mu'(0) b \right| \rightarrow 0, \]

and this complete the proof of C3.

C4. We need to show that \( \sqrt{N} \frac{T^*(0)}{\sigma_p} \) converges in distribution to the standard normal distribution. Under the null hypothesis \( H_0 \), this follows from Theorem 2.6.

We finally conclude that \( T^*(\Delta) \) is Pitman regular and the asymptotic Pitman efficacy is given by

\[ c^2_{ORRD} = \left( \frac{\mu'(0)}{\sigma_p} \right)^2 = \left( \int f^2(x) dx \right)^2. \]

\[ \square \]

2.3 Optimal designs

Chapter 1 indicates that the ORRD is unique when \( H = 2 \), but the number of possible designs is \( 2^H - 1 \) for an arbitrary but fixed \( H > 1 \). The efficiency of the test depends on the design that generates the data. This efficiency can be improved by choosing a design that maximizes the Pitman asymptotic efficacy of the test. The efficacy factor is a function of the design parameters \( \alpha \) and \( \beta \) through the null variance of the test statistic. The maximization of the efficacy factor \( c^2 \) is equivalent to the minimization of the null variance of \( \sqrt{2nH(T - \frac{1}{T})} \) with respect to \( \alpha \) and \( \beta \).
The next theorem shows that there exists an optimal design that minimizes the null variance of the test statistic of all possible designs.

**Theorem 2.8.** Let $H$ be any fixed integer. The null variance of the test statistic $\sqrt{2nH(T - \frac{1}{2})}$, $\sigma_p^2$, is minimized when Set $\alpha$ takes odd orders only and Set $\beta$ takes even orders only, or vice versa.

**Proof.** The proof is divided into two steps. In Step I, we will show that any design in the form $(\ldots \beta \alpha \ldots \beta \ldots \beta \alpha \ldots)$ does not provide the smallest $\sigma_p^2$. The form $(\ldots \beta \alpha \ldots \beta \ldots)$ means that at least two consecutive orders somewhere in the middle of the integer set $(1, 2, \ldots, H)$ are assigned to Set $\alpha$. In Step II, we will show that the best design is of the form $(\alpha \beta \alpha \ldots \alpha \beta \ldots)$ or $(\beta \alpha \beta \ldots \beta \alpha \ldots)$. The form $(\alpha \beta \alpha \ldots \alpha \beta \ldots)$ means that the odd orders in the integer set $(1, 2, \ldots, H)$ are assigned to Set $\alpha$ and the even orders to Set $\beta$, while the form $(\beta \alpha \beta \ldots \beta \alpha \ldots)$ gives an opposite assignment. We note that, under perfect ranking,

$$
\sigma_p^2 = \frac{4}{H} \left\{ \sum_{i=1}^{u} \alpha_i \frac{(H + 1 - \alpha_i)}{(H + 1)^2(H + 2)} + 2 \sum_{i=1}^{u} \sum_{j=1}^{u} \frac{\alpha_i(H + 1 - \alpha_j)}{(H + 1)^2(H + 2)} + \right.
$$

$$
\sum_{k=1}^{H-u} \frac{\beta_k(H + 1 - \beta_k)}{(H + 1)^2(H + 2)} + 2 \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \frac{\beta_k(H + 1 - \beta_t)}{(H + 1)^2(H + 2)}
$$

$$
-2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\alpha_i(H + 1 - \beta_k)}{(H + 1)^2(H + 2)} - 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \frac{\beta_k(H + 1 - \alpha_i)}{(H + 1)^2(H + 2)} \bigg\},
$$

Step I. Consider a design $(\ldots \beta \alpha \ldots \beta \ldots)$ with $\alpha$’s consecutively located between two $\beta$’s and the number of $\alpha$’s is no less than 2. In this case, we consider two types of swaps.
Swap I. We interchange the position of the consecutive $\alpha$ and $\beta$ without changing the positions of the others to have $(...\beta \underbrace{\alpha_i \alpha_j} \beta_k ...) \to (...\beta \underbrace{\alpha_i' \alpha_j} \beta_k ...)$. Without loss of generality, let $\beta_k = p$, then $\alpha_i = p - 1, \beta_k' = p - 1$ and $\alpha_i' = p$.

Let the asymptotic variance of $2nH(T - \frac{1}{2})$ for the old $(...\beta \underbrace{\alpha_i \alpha_j} \beta_k ...)$ and new $(...\beta \underbrace{\alpha_i' \alpha_j} \beta_k ...)$. designs be $\sigma^2$ and $\sigma^2_{\text{new1}}$, respectively. Only difference between $\sigma^2$ and $\sigma^2_{\text{new1}}$ is due to the terms $(\alpha_i, \beta_k)$ and $(\alpha_i', \beta_k')$. All other terms in $\sigma^2$ and $\sigma^2_{\text{new1}}$ are exactly the same. By using this feature, we write the difference of between $\sigma^2$ and $\sigma^2_{\text{new1}}$ as

$$
\sigma^2_{\text{new1}} - \sigma^2 = \frac{4}{H(H+1)^2(H+2)} \left\{ \right.
\alpha_i'(H + 1 - \alpha_i') - \alpha_i(H + 1 - \alpha_i) + 2 \sum_{j=1}^{u} \alpha_j(H + 1 - \alpha_j) + 2 \sum_{j=1}^{u} \alpha_j'(H + 1 - \alpha_j) \\
-2 \sum_{\begin{subarray}{c} j=1 \\ \alpha_j < \alpha_i \end{subarray}}^{u} \alpha_j(H + 1 - \alpha_i) - 2 \sum_{\begin{subarray}{c} j=1 \\ \alpha_j > \alpha_i \end{subarray}}^{u} \alpha_j(H + 1 - \alpha_j) + \beta_k'(H + 1 - \beta_k') - \beta_k(H + 1 - \beta_k) \\
+ 2 H - u \sum_{\begin{subarray}{c} t=1 \\ \beta_t < \beta_k \end{subarray}}^{H-u} \beta_t(H + 1 - \beta_k') + 2 H - u \sum_{\begin{subarray}{c} t=1 \\ \beta_t > \beta_k \end{subarray}}^{H-u} \beta_t'(H + 1 - \beta_t) \\
- 2 H - u \sum_{\begin{subarray}{c} t=1 \\ \beta_t < \beta_k \end{subarray}}^{H-u} \beta_t(H + 1 - \beta_k) - 2 H - u \sum_{\begin{subarray}{c} t=1 \\ \beta_t > \beta_k \end{subarray}}^{H-u} \beta_t(H + 1 - \beta_t) - \\
2 \sum_{\begin{subarray}{c} j=1 \\ \alpha_j < \alpha_i \end{subarray}}^{u} \alpha_j(H + 1 - \beta_k') - \sum_{\begin{subarray}{c} j=1 \\ \alpha_j < \alpha_i \end{subarray}}^{u} \alpha_j(H + 1 - \beta_k) + \sum_{\begin{subarray}{c} j=1 \\ \beta_t > \beta_k \end{subarray}}^{H-u} \alpha_j'(H + 1 - \beta_k) - \sum_{\begin{subarray}{c} j=1 \\ \beta_t > \beta_k \end{subarray}}^{H-u} \alpha_j(H + 1 - \beta_t) \\
- \alpha_i'(H + 1 - \beta_k) - 2 H - u \sum_{\begin{subarray}{c} t=1 \\ \beta_t < \beta_k \end{subarray}}^{H-u} \beta_t(H + 1 - \alpha_i') + \beta_k'(H + 1 - \alpha_i') - \sum_{\begin{subarray}{c} t=1 \\ \beta_t < \beta_k \end{subarray}}^{H-u} \beta_t(H + 1 - \alpha_i) \\
+ \sum_{\begin{subarray}{c} j=1 \\ \alpha_j < \alpha_i \end{subarray}}^{u} \beta_k'(H + 1 - \alpha_j) - \sum_{\begin{subarray}{c} j=1 \\ \alpha_j < \alpha_i \end{subarray}}^{u} \beta_k(H + 1 - \alpha_j) \right\}.
$$
Since $\beta_k = \alpha'_i = p$ and $\alpha_i = \beta'_k = p - 1$, the above expression simplifies to

\[
\sigma^2_{new1} - \sigma^2 = \frac{8}{H(H+1)^2(H+2)} \left\{ \right.
- 2\left( \sum_{j=1}^{u} \alpha_j + \sum_{j=1}^{u} \alpha_j \right) + 2\left( \sum_{t=1}^{H-u} \beta_t + \sum_{t=1}^{H-u} \beta_t \right) +
2(H+1)\left( \sum_{j=1}^{u} I(\alpha_j > \alpha_i) - \sum_{t=1}^{H-u} I(\beta_t > \beta_k) \right) \left\} \right.
\]

\[
= \frac{8}{H(H+1)^2(H+2)} \left\{ -2\left( \sum_{j=1}^{u} \alpha_j - \alpha_i \right) + 2\left( \sum_{t=1}^{H-u} \beta_t - \beta_k \right) +
2(H+1)\left( \sum_{j=1}^{u} I(\alpha_j > \alpha_i) - \sum_{t=1}^{H-u} I(\beta_t > \beta_k) \right) \right\}.
\]

We rewrite this expression to have

\[
\sigma^2_{new1} - \sigma^2 = \frac{16}{H(H+1)^2(H+2)} \left[ -\sum_{j=1}^{u} \alpha_j + \sum_{t=1}^{H-u} \beta_t - 1 + (H+1)(A_p - B_p) \right]
\]

\[
\sigma^2_{new1} - \sigma^2 = \frac{16}{H(H+1)^2(H+2)} \left[ -\sum_{j=1}^{u} \alpha_j + \sum_{t=1}^{H-u} \beta_t - 1 + (H+1)(A_p - B_p) \right]
\]

where

\[ A_p = \sum_{j=1}^{u} I(\alpha_j > \alpha_i) = \text{number of } \alpha > \alpha_i = p - 1 \]

\[ B_p = \sum_{t=1}^{H-u} I(\beta_t > \beta_k) = \text{number of } \beta > \beta_k = p. \]

Since $A_p - B_p = -1$, we have

\[
\sigma^2_{new1} - \sigma^2 = \frac{16}{H(H+1)^2(H+2)} \left[ -\sum_{j=1}^{u} \alpha_j + \sum_{t=1}^{H-u} \beta_t - 1 - (H+1) \right].
\]

Swap II. We interchange the positions of the consecutive $\alpha_j$ and $\beta_t$ in the design ($\ldots, \beta_t, \alpha_j, \ldots, \alpha_i, \beta_k, \ldots$) without changing the positions of the other terms to obtain the design ($\ldots, \alpha'_j, \beta'_t, \alpha_i, \beta_k, \ldots$). Let $\alpha_j = q$, then $\beta_t = q - 1$, $\beta'_t = q$ and $\alpha'_j = q - 1$. 

54
Let $\sigma^2$ and $\sigma^2_{\text{new}}$ be the asymptotic variances of $\sqrt{2nH(T - \frac{1}{2})}$ for the old design ($\ldots \beta_t \ldots \alpha_j \ldots \alpha_i \beta_k \ldots$) and the new design ($\ldots \alpha'_j \beta'_t \ldots \alpha_i \beta_k \ldots$). Computations similar to the ones given in Swap I yield the following expression for the difference between the variances $\sigma^2$ and $\sigma^2_{\text{new}}$:

$$\sigma^2_{\text{new}} - \sigma^2 = \frac{4}{H(H + 1)^2(H + 2)} \left\{ \alpha'_j(H + 1 - \alpha'_j) - \alpha_j(H + 1 - \alpha_j) + 2 \sum_{s=1}^{u} \alpha_s(H + 1 - \alpha_j) + 2 \sum_{s=1}^{u} \alpha'_s(H + 1 - \alpha_s) - 2 \sum_{s=1}^{u} \alpha_s(H + 1 - \alpha_j) - 2 \sum_{s=1}^{u} \alpha_j(H + 1 - \alpha_s) + \beta'_t(H + 1 - \beta'_t) - \beta_t(H + 1 - \beta_t) + 2 \sum_{v=1}^{H-u} \beta_v(H + 1 - \beta'_t) + 2 \sum_{v=1}^{H-u} \beta'_v(H + 1 - \beta_v) - 2 \sum_{v=1}^{H-u} \beta_v(H + 1 - \beta_t) - 2 \sum_{v=1}^{H-u} \beta_t(H + 1 - \beta_v) - 2 \sum_{v=1}^{H-u} \beta_v(H + 1 - \beta'_t) - 2 \sum_{v=1}^{H-u} \beta'_v(H + 1 - \beta_v) - 2 \sum_{v=1}^{H-u} \beta_v(H + 1 - \beta_t) - 2 \sum_{v=1}^{H-u} \beta_t(H + 1 - \beta_v) + \alpha'_j(H + 1 - \alpha'_j) - \sum_{s=1}^{u} \alpha_s(H + 1 - \beta_t) + \sum_{v=1}^{H-u} \alpha'_v(H + 1 - \beta_v) - \sum_{v=1}^{H-u} \alpha_v(H + 1 - \beta_v) + \sum_{v=1}^{H-u} \beta'_v(H + 1 - \beta'_t) - \sum_{v=1}^{H-u} \beta'_v(H + 1 - \beta_v) - \sum_{v=1}^{H-u} \beta_v(H + 1 - \alpha_j) + \sum_{s=1}^{u} \beta'_s(H + 1 - \alpha_s) - \sum_{s=1}^{u} \beta_s(H + 1 - \alpha_s) \right\}.$$  

Again, since $\beta_t = \alpha'_j = q - 1$ and $\alpha_j = \beta'_t = q$, the above expression simplifies to

$$\sigma^2_{\text{new}} - \sigma^2 = \frac{8}{H(H + 1)^2(H + 2)} \left\{ - 2(\sum_{s=1}^{u} \alpha_s + \sum_{s=1}^{u} \alpha'_s) + 2(\sum_{v=1}^{H-u} \beta_v + \sum_{v=1}^{H-u} \beta'_v) + \right\}.$$
\[2(H + 1)[\sum_{s=1}^{u} I(\alpha_s > \alpha_j) - \sum_{v=1}^{H-u} I(\beta_v > \beta_t)]\]

\[= \frac{8}{H(H+1)^2(H+2)}\{ -2(\sum_{s=1}^{u} \alpha_s - \alpha_j) + 2(\sum_{v=1}^{H-u} \beta_v - \beta_t) + 2(H + 1)[\sum_{s=1}^{u} I(\alpha_s > \alpha_j) - \sum_{v=1}^{H-u} I(\beta_v > \beta_t)] \}.

This difference can be written as

\[\sigma_{new2}^2 - \sigma^2 = \frac{16}{H(H+1)^2(H+2)}\{ \sum_{s=1}^{u} \alpha_s - \sum_{v=1}^{H-u} \beta_v - 1 - (H + 1)(A_q - B_q) \},\]

where

\[A_q = \sum_{s=1}^{u} I(\alpha_s > \alpha_j) = \text{number of } \alpha > \alpha_j = q,\]

\[B_q = \sum_{v=1}^{H-u} I(\beta_v > \beta_t) = \text{number of } \beta > \beta_t = q - 1.\]

Between these two types of swaps, if \(\sigma_{new1}^2 - \sigma^2 < 0\), then the design \((\ldots \beta \alpha \ldots \beta' \alpha' \ldots)\) improves the design \((\ldots \beta \alpha \ldots \alpha \beta \ldots)\). On the other hand, if \(\sigma_{new1}^2 - \sigma^2 \geq 0\), then the following argument shows that \(\sigma_{new2}^2 - \sigma^2 < 0\), which leads to the conclusion that the design \((\ldots \alpha' \beta' \ldots \alpha \beta \ldots)\) improves the design \((\ldots \beta \alpha \ldots \alpha \beta \ldots)\). We now assume

\[\sigma_{new1}^2 - \sigma^2 \geq 0 \quad \Rightarrow \quad -\sum_{j=1}^{u} \alpha_j + \sum_{t=1}^{H-u} \beta_t \geq 1 - (H + 1)(A_p - B_p)\]

\[\Rightarrow \sum_{s=1}^{u} \alpha_s - \sum_{v=1}^{H-u} \beta_v \leq -1 + (H + 1)(A_p - B_p).\]
Then,

$$\sigma_{new}^2 - \sigma^2 = \frac{16}{H(H + 1)^2(H + 2)} \left[ \sum_{s=1}^{u} \alpha_s - \sum_{v=1}^{H-u} \beta_v - 1 - (H+1)(A_q - B_q) \right]$$

$$\leq \frac{16}{H(H + 1)^2(H + 2)} \left[ -1 + (H+1)(A_p - B_p) - 1 - (H+1)(A_q - B_q) \right]$$

$$\leq \frac{16}{H(H + 1)^2(H + 2)} \left[ -2 + (H+1)(A_p - A_q + B_q - B_p) \right]$$

$$\leq \frac{16}{H(H + 1)^2(H + 2)} \left[ -2 + (H+1)(-1 + 1) \right]$$

$$\leq \frac{-32}{H(H + 1)^2(H + 2)} < 0.$$  

The third inequality is due to the fact that \( A_p - A_q \leq -1, \ B_q - B_p = 1. \)

These two swaps show that we can always find a design that improves the designs of the form \((\cdots \beta_i \alpha_j \cdots \beta_k \cdots)\). This completes the first part of the proof. As a result, we can only look for the optimal design in those with no consecutive \( \alpha \)'s or \( \beta \)'s in the middle ranks \((\cdots \beta \alpha \beta \alpha \cdots)\). (There can still be consecutive \( \alpha \)'s or \( \beta \)'s at the both ends).

Step II: We will show that two or more \( \alpha \) or \( \beta \) can not be at the end of the sequence.

Proof: Let \( u \) and \( H - u \) be the number of elements in Set \( \alpha \) and \( \beta \), respectively. Without loss of generality, assume that \( u \leq H - u \).

We again need to consider two cases.

Case I: Consider a design of the form \((\beta_1 \beta_2 \cdots \beta_1 \alpha_1 \alpha_2 \cdots \beta \alpha \cdots)\) with at least two \( \beta \)'s consecutively located at the beginning. In this case, the lower end of the set is a \( \beta \), ie, \( \beta_1 = 1 \). The higher end of the set could be an \( \alpha \) in the form of \((\cdots \beta \alpha \beta \alpha \cdots \beta \alpha \cdots)\) or \((\cdots \beta \alpha \beta \alpha \cdots \beta \alpha \cdots \alpha \alpha \cdots)\). The higher end of the set could also be a \( \beta \) in the form of \((\cdots \beta \alpha \beta \alpha \cdots \beta \alpha \beta \cdots \beta \beta \cdots)\) or \((\cdots \beta \alpha \beta \alpha \cdots \beta \alpha \beta \cdots \beta \alpha \beta \cdots)\). 

57
We generate a new design \((\alpha' \beta \alpha \beta \alpha \ldots)\) through replacing \(\beta_1\) by \(\alpha'\) at the first position. That is, the first position is assigned to Set \(\alpha\) instead of Set \(\beta\). So in the new design, we include an extra \(\alpha' = 1\), but exclude \(\beta_1 = 1\) without touching the others. In short, the old design \((\beta_1 \beta_2 \ldots \beta \alpha \beta \alpha \ldots)\) is changed to the new design \((\alpha' \beta_2 \ldots \beta \alpha_1 \beta \alpha_2 \ldots \beta \alpha \ldots)\).

Let the asymptotic variances of \(\sigma_{2\text{old}}^2\) and \(\sigma_{2\text{new}}^2\) for the old and new designs be \(\sigma_{\text{old}}^2\) and \(\sigma_{\text{new}}^2\), respectively. Computations similar to those given in step I produce a difference statement between \(\sigma_{\text{old}}^2\) and \(\sigma_{\text{new}}^2\)

\[
\sigma_{\text{new}}^2 - \sigma_{\text{old}}^2 = \frac{4}{H(H+1)^2(H+2)} \left\{ \sum_{i=1}^{u} \alpha'(H+1-\alpha_i) - \sum_{i=1}^{u} \beta_1(H+1-\alpha_i) \right\} - 2 \sum_{t=2}^{H-u} \beta_1(H+1-\beta_t) - 2 \sum_{k=2}^{H-u} \alpha'(H+1-\beta_k) - 2\left[ - \sum_{i=1}^{u} \beta_1(H+1-\alpha_i) \right]
\]

\[
= \frac{16}{H(H+1)^2(H+2)} \left[ \sum_{i=1}^{u} (H+1-\alpha_i) - \sum_{t=2}^{H-u} (H+1-\beta_t) \right].
\]

Note that the last equality follows from the fact that \(\alpha' = \beta_1 = 1\). After further simplification,

\[
\sigma_{\text{new}}^2 - \sigma_{\text{old}}^2 = \frac{16}{H(H+1)^2(H+2)} [(H+1)(2u - H + 1) + \sum_{t=2}^{H-u} \beta_t - \sum_{i=1}^{u} \alpha_i].
\]

Since old design can have three different types of end on the right, it can have one of the form \((\beta_1 \beta_2 \ldots \beta \alpha_1 \beta \alpha \ldots \beta_{H-u} \alpha_u)\), \((\beta_1 \beta_2 \ldots \beta \alpha_1 \beta \alpha \ldots \beta \alpha_u \beta \ldots \beta_{H-u})\) or \((\beta_1 \beta_2 \ldots \beta \alpha_1 \beta \alpha \ldots \beta_{H-u} \alpha \ldots \alpha_u)\). We notice that in all three forms, we can separate Set \(\beta\) into three disjoint subsets. The first one is \(\{\beta_1 = 1\}\). The second one is \(\{\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_u\}\), where all \(u\) elements are distinct and \((\tilde{\beta}_i - \alpha_i) \leq -1\) for \(i = 1, \ldots, u\) (i.e., for each \(\alpha_i\) we can find a \(\tilde{\beta}_i\) with a smaller rank). The last one is \((\tilde{\beta}_{u+1}, \tilde{\beta}_{u+2}, \ldots, \tilde{\beta}_{H-u-1})\), where all \(H - 2u - 1\) elements are distinct and \(\tilde{\beta}_i < H + 1\) for \(i = u + 1, \ldots, H - u\). Then it
follows that
\[
\sum_{t=2}^{H-u} \beta_t - \sum_{i=1}^{u} \alpha_i < (H + 1)(H - 2u - 1) - u
\]
and we conclude that
\[
\sigma_{new3}^2 - \sigma_{old1}^2 < \frac{16}{H(H + 1)^2(H + 2)}[(H + 1)(2u - H + 1) + (H + 1)(H - 2u - 1) - u]
\]
\[= - \frac{16u}{H(H + 1)^2(H + 2)} < 0.
\]
This indicates that the new design of the form \((\alpha' \beta_2 \alpha_1 \beta \alpha \ldots)\) improves the old design of the form \((\beta_1 \beta_2 \alpha_1 \beta \alpha \ldots)\).

Case II: We now consider the design of the form \((\ldots \beta \alpha \alpha_1 \beta \alpha_{u-1} \beta_{H-u})\) with at least two \(\beta\)'s consecutively located at the higher end. In this case, \(\beta_{H-u} = H\) and the lower end of the set could be an \(\alpha\) such as the form \((\alpha_1 \alpha_2 \ldots \alpha \beta \alpha \alpha_{u-1} \beta)\) or a \(\beta\) such as the form \((\beta_1 \beta_2 \alpha_1 \ldots \alpha \beta \alpha \beta \ldots)\) or \((\beta_1 \alpha_1 \beta \alpha \ldots)\).

We create a new design \((\ldots \beta \alpha \alpha_{u-1} \beta_{H-u-1} \alpha')\) through replacing \(\beta_{H-u}\) at the higher end position with \(\alpha'\) so that the new design contains an extra \(\alpha' = H\), but excludes \(\beta_{H-u} = H\) without changing the others.

Let the asymptotic variances of \(\sqrt{2nH(T - \frac{1}{2})}\) for the old and new design be \(\sigma_{old2}^2\) and \(\sigma_{new4}^2\), respectively. The difference between \(\sigma_{old2}^2\) and \(\sigma_{new4}^2\) is
\[
\sigma_{new4}^2 - \sigma_{old2}^2 = \frac{4}{H(H + 1)^2(H + 2)}\left\{\alpha'(H + 1 - \alpha') + 2 \sum_{i=1}^{u} \alpha_i(H + 1 - \alpha') - \\
\beta_{H-u}(H + 1 - \beta_{H-u}) - 2 \sum_{t=1}^{H-u-1} \beta_t(H + 1 - \beta_{H-u})
\right\}
\]
\[-2[- \sum_{i=1}^{u} \alpha_i(H + 1 - \beta_{H-u})] - 2\left[ \sum_{t=1}^{H-u-1} \beta_t(H + 1 - \alpha')\right]\]
\[= \frac{16H}{H(H + 1)^2(H + 2)}\left(\sum_{i=1}^{u} \alpha_i - \sum_{t=1}^{H-u-1} \beta_t\right).
\]
The last equality follows from the fact that \(\alpha' = \beta_{H-u} = H\).
Note that the old design can have three different forms depending on lower end point convention. It would be of the form $(\beta_1 \beta_2 \beta \alpha_1 \beta \alpha_2 \beta \alpha_3 \beta \alpha_4 \beta \alpha_5 \beta \alpha_u \beta \alpha_{H-u})$, $(\beta_1 \alpha_1 \beta \alpha_2 \beta \alpha_3 \beta \alpha_4 \beta \alpha_5 \beta \alpha_u \beta \alpha_{H-u})$ or $(\alpha_1 \alpha_2 \beta \alpha_3 \beta \alpha_4 \beta \alpha_5 \beta \alpha_u \beta \alpha_{H-u})$. We again separate Set $\beta$ into three different disjoint subsets. The first set contains $\{\beta_{H-u} = H\}$. The second set contains $\{\tilde{\beta}_1, \tilde{\beta}_2, ..., \tilde{\beta}_u\}$, where all $u$ elements are distinct and $(\alpha_i - \tilde{\beta}_i) \leq -1$ for $i = 1, ..., u$ (i.e., for each $\alpha_i$ we can find a $\tilde{\beta}_i$ with a larger rank). The last one is $(\tilde{\beta}_{u+1}, \tilde{\beta}_{u+2}, ..., \tilde{\beta}_{H-u-2})$, where all $H-2u-1$ elements are distinct and positive. Then by using the properties of these sets we obtain

$$\sum_{i=1}^{u} \alpha_i - \sum_{t=1}^{H-u-1} \beta_t < -u - \sum_{i=u+1}^{H-2} \tilde{\beta}_i < 0.$$ 

This indicates that the new design $(\ldots \beta \alpha \beta \alpha_1 \beta \alpha_2 \beta \alpha_3 \beta \alpha_4 \beta \alpha_5 \beta \alpha_u \beta \alpha_{H-u-1} \alpha')$ improves the old design of the form $(\ldots \beta \alpha \beta \alpha_1 \beta \alpha_2 \beta \alpha_3 \beta \alpha_4 \beta \alpha_5 \beta \alpha_u \beta \alpha_{H-u-1} \beta \alpha_{H-u})$.

The results of Case I and II together imply that the best design can not be of the form with at least two $\beta$ (or two $\alpha$) consecutively located at the beginning or end of the set if $u \leq H - u$ or $u \geq H - u$.

Step I and II together imply that the best design must be of the form that contains alternate $\alpha$ and $\beta$ terms. There are two choices to construct such a design. One choice is that Set $\alpha$ takes odd integers only and Set $\beta$ takes even integers only. The other is that Set $\alpha$ takes even integers only and Set $\beta$ takes odd integers only. Furthermore, both of these choices produce the same design. This completes the proof of the Theorem.

\[\square\]

For the optimal design, the asymptotic null variance of $\sqrt{2nH(T - \frac{1}{2})}$ can further be reduced to a value that is not a function of $\alpha$ or $\beta$. It only depends on the set size $H$. 

60
Theorem 2.9. For any fixed, but arbitrary set size \( H \geq 2 \), the asymptotic null variance of \( \sqrt{2nH(T - \frac{1}{2})} \), \( \sigma_p^2 \), under perfect ranking and optimal design, reduces to \( \frac{1}{(H+1)^2} \) when \( H \) is even and \( \frac{1}{H(H+2)} \) when \( H \) is odd.

Proof. We rewrite the expression of \( \sigma_p^2 \) as follows

\[
\sigma_p^2 = \frac{4}{H(H+1)^2(H+2)} \left\{ \sum_{i=1}^{u} \alpha_i(H + 1 - \alpha_i) + 2 \sum_{i=1}^{u} \sum_{j=1}^{u} \alpha_i(H + 1 - \alpha_j)I(\alpha_i < \alpha_j) + \sum_{k=1}^{H-u} \beta_k(H + 1 - \beta_k) + 2 \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \beta_k(H + 1 - \beta_k)I(\beta_k < \beta_t) - 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i(H + 1 - \beta_k)I(\alpha_i < \beta_k) - 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \beta_k(H + 1 - \alpha_i)I(\beta_k < \alpha_i) \right\}.
\]

This expression can be expanded as

\[
\sigma_p^2 = \frac{4}{H(H+1)^2(H+2)} \left\{ (H+1) \sum_{i=1}^{u} \alpha_i - \sum_{i=1}^{u} \alpha_i^2 + 2(H+1) \sum_{i=1}^{u} \sum_{j=1}^{u} \alpha_i I(\alpha_i < \alpha_j) - 2 \sum_{i=1}^{u} \sum_{j=1}^{u} \alpha_i \alpha_j I(\alpha_i < \alpha_j) + (H+1) \sum_{k=1}^{H-u} \beta_k - \sum_{k=1}^{H-u} \beta_k^2 + 2(H+1) \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \beta_k I(\beta_k < \beta_t) - 2 \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \beta_k \beta_t I(\beta_k < \beta_t) - 2(H+1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i I(\alpha_i < \beta_k) + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i \beta_k I(\alpha_i < \beta_k) - 2(H+1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \beta_k I(\beta_k < \alpha_i) + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \beta_k \alpha_i I(\beta_k < \alpha_i) \right\}.
\]

Let

\[
C_i = \sum_{j=1}^{u} I(\alpha_i < \alpha_j), \quad D_i = \sum_{k=1}^{H-u} I(\alpha_i < \beta_k), \quad E_k = \sum_{t=1}^{H-u} I(\beta_k < \beta_t), \quad F_k = \sum_{i=1}^{u} I(\beta_k < \alpha_i).
\]

We then reduce \( \sigma_p^2 \) as

\[
\sigma_p^2 = \frac{4}{H(H+1)^2(H+2)} (A + B + C),
\]

61
where

\[ A = (H + 1) \sum_{i=1}^{u} \alpha_i + (H + 1) \sum_{k=1}^{H-u} \beta_k \]
\[ = (H + 1)(\sum_{i=1}^{u} \alpha_i + \sum_{k=1}^{H-u} \beta_k) \]
\[ = \frac{(H + 1)^2 H}{2}, \]

\[ B = -\sum_{i=1}^{u} \alpha_i^2 - 2 \sum_{i=1}^{u} \alpha_i \alpha_j I(\alpha_i < \alpha_j) - \sum_{k=1}^{H-u} \beta_k^2 \]
\[ -2 \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \beta_k \beta_t I(\beta_k < \beta_t) + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i \beta_k I(\alpha_i < \beta_k) \]
\[ + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \beta_k \alpha_i I(\beta_k < \alpha_i) \]
\[ = -\left[ \left( \sum_{i=1}^{u} \alpha_i \right)^2 + \left( \sum_{k=1}^{H-u} \beta_k \right)^2 - 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i \beta_k \right] \]
\[ = -\left[ \left( \sum_{i=1}^{u} \alpha_i \right)^2 + \left( \sum_{k=1}^{H-u} \beta_k \right)^2 - 2 \left( \sum_{i=1}^{u} \alpha_i \right) \left( \sum_{k=1}^{H-u} \beta_k \right) \right] \]
\[ = -\left( \sum_{i=1}^{u} \alpha_i - \sum_{k=1}^{H-u} \beta_k \right)^2 = -\left( \frac{H(H + 1)}{2} - 2 \sum_{k=1}^{H-u} \beta_k \right)^2, \]

and

\[ C = 2(H + 1) \sum_{i=1}^{u} \sum_{j=1}^{u} \alpha_i I(\alpha_i < \alpha_j) + 2(H + 1) \sum_{k=1}^{H-u} \sum_{t=1}^{H-u} \beta_k I(\beta_k < \beta_t) \]
\[ -2(H + 1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \alpha_i I(\alpha_i < \beta_k) - 2(H + 1) \sum_{i=1}^{u} \sum_{k=1}^{H-u} \beta_k I(\beta_k < \alpha_i) \]
\[ = 2(H + 1) \sum_{i=1}^{u} C_i \alpha_i + 2(H + 1) \sum_{k=1}^{H-u} E_k \beta_k \]
\[ -2(H + 1) \sum_{i=1}^{u} D_i \alpha_i - 2(H + 1) \sum_{k=1}^{H-u} F_k \beta_k \]
\[ = 2(H + 1) \left[ \sum_{i=1}^{u} (C_i - D_i) \alpha_i + \sum_{k=1}^{H-u} (E_k - F_k) \beta_k \right]. \]
Without loss of generality, assume that in the optimal design all $\alpha$’s are odd while $\beta$’s are even. We calculate $\sigma_p^2$ when $H$ is even and odd separately.

Case I. Assume that $H$ is even. Then the number of elements in the Set $\alpha$ is $u = H - u = \frac{H}{2}$ and the expression of $B$ simplifies to

$$B = -(\frac{H(H+1)}{2} - 2 \sum_{k=1}^{H-u} \beta_k)^2 = -(\frac{H(H+1)}{2} - 2(\frac{H}{2})(\frac{H}{2} + 1))^2 = -\frac{H^2}{4}.$$ 

Since $C_i - D_i = -1$ and $E_k - F_k = 0$, we have

$$C = -2(H+1) \sum_{i=1}^{u} \alpha_i = -2(H+1)(\frac{H}{2})^2 = -\frac{H^2(H+1)}{2}.$$ 

By putting these together, we obtain

$$\sigma_p^2 = \frac{4}{H(H+1)^2(H+2)}[\frac{(H+1)^2H}{2} - \frac{H^2}{4} - \frac{H^2(H+1)}{2}] = \frac{1}{(H+1)^2}.$$ 

Case II. Assume that $H$ is odd. Then in the optimal design, the number of elements in Set $\alpha$ and $\beta$ are $u = \frac{H-1}{2} + 1$ and $H - u = \frac{H-1}{2}$, respectively. This leads to

$$B = -(\frac{H(H+1)}{2} - 2 \sum_{k=1}^{H-u} \beta_k)^2$$

$$= -(\frac{H(H+1)}{2} - 2(\frac{H-1}{2})(\frac{H-1}{2} + 1))^2 = -\frac{(H+1)^2}{4}.$$ 

Again since $C_i - D_i = 0$ and $E_k - F_k = -1$, we have

$$C = -2(H+1) \sum_{k=1}^{H-u} \beta_k = -2(H+1)(\frac{H-1}{2})(\frac{H-1}{2} + 1) = -\frac{(H+1)^2(H-1)}{2}.$$ 

These together yields

$$\sigma_p^2 = \frac{4}{H(H+1)^2(H+2)}[\frac{(H+1)^2H}{2} - \frac{(H+1)^2}{4} - \frac{(H+1)^2(H-1)}{2}]$$

$$= \frac{1}{H(H+2)}.$$ 

$\square$
We compare the proposed testing procedure with its competitors in literature with respect to the asymptotic Pitman efficiency. Let $T_1$ and $T_2$ be two tests. Then the asymptotic Pitman relative efficiency (ARE) is defined as $\text{ARE}(T_1, T_2) = \frac{c_1^2}{c_2^2}$, where $c_1^2$ and $c_2^2$ are the asymptotic Pitman efficacies of the tests $T_1$ and $T_2$, respectively. If the $\text{ARE}(T_1, T_2)$ is larger than 1, the test $T_1$ is superior to the test $T_2$ in Pitman efficacy.

**Definition 2.2.** Let the asymptotic Pitman efficacies of $T$, $U_{RSS}$ and Mann-Whitney-Wilcoxon (MWW) statistic be $c_{\text{ORRD}}^2$, $c_{RSS}^2$ and $c_{SRS}^2$, respectively. We will say the **asymptotic relative efficiency (ARE)** of $T$ relative to $U_{RSS}$ and MWW statistic is $c_{\text{ORRD}}^2/c_{RSS}^2$ and $c_{\text{ORRD}}^2/c_{SRS}^2$, respectively. The asymptotic relative efficiency (ARE) of $U_{RSS}$ relative to MWW statistic is $c_{RSS}^2/c_{SRS}^2$.

Based on previous results, under the optimal ORRD with perfect ranking the asymptotic Pitman efficacy of $T$ is given by

$$c_{\text{ORRD}}^2 = \begin{cases} (H + 1)^2(\int f^2(x)dx)^2 & \text{if } H \text{ is even;} \\ H(H + 2)(\int f^2(x)dx)^2 & \text{if } H \text{ is odd.} \end{cases}$$

The asymptotic Pitman efficacy of the MWW test is available in standard text books such as Hettmansperger and McKean [18] and Randles and Wolfe [40]. When the sample sizes of $X$- and $Y$-samples are equal, it is given by

$$c_{SRS}^2 = 3(\int f^2(x)dx)^2.$$  

The Pitman efficacy of the ranked set sample rank sum test, $U_{RSS}$, given in Bohn and Wolfe [4], can be simplified to

$$c_{RSS}^2 = \frac{3(H + 1)}{2}(\int f^2(x)dx)^2.$$
In order to match the fully measured observations in all three designs, the equal set and cycle sizes are selected in each treatment groups. We summarize the ARE results for the three designs (SRS, RSS and ORRD) in Table 2.1. It is clear that the ARE depends on whether the set size $H$ is even or odd. Another observation is that ARE does not depend on the underlying distributions. This is not a surprising result since all three tests use the same test statistics and a balanced design. The differences in the ARE are due to the way that data are collected. Table 2.1 shows that the proposed test outperforms its competitors for all set sizes $H \geq 2$. For the set sizes $H = 2, 3, 4$, the numerical values of these ARE results are given in Table 2.2. It is clear from Table 2.2 that the ORRD provides substantial improvement in the Pitman relative efficiency over its competitors. The results in Table 2.2 are also consistent with the results given in Bohn and Wolfe [3].

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
$H$ & ARE(RSS, SRS) & ARE(ORRD, SRS) & ARE(ORRD, RSS) \\
\hline
& $\frac{H+1}{2}$ & $\frac{(H+1)^2}{3}$ & $\frac{2(H+1)}{3}$ \\
\hline
\text{even} & & & \\
\hline
& $\frac{H^2+1}{2}$ & $\frac{H(H+2)}{3}$ & $\frac{2H(H+2)}{3(H+1)}$ \\
\hline
\text{odd} & & & \\
\hline
\end{tabular}
\caption{The asymptotic relative efficiency (ARE)}
\end{table}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
$H$ & ARE(RSS, SRS) & ARE(ORRD, SRS) & ARE(ORRD, RSS) \\
\hline
2 & 1.5 & 3 & 2 \\
3 & 2 & 5 & 2.5 \\
4 & 2.5 & $\frac{25}{3}$ & $\frac{10}{3}$ \\
\hline
\end{tabular}
\caption{Examples for ARE}
\end{table}
2.4 Simulation study for the ORRD

In Section 2.1 and 2.2, we showed that the test statistic has an asymptotically normal distribution and desirable efficiency properties. In this section, a simulation study is performed to investigate the convergence rate of the test statistic for finite sample sizes under different settings. The simulation settings consist of different set ($H$) and replication ($n$) sizes, various degree of ranking information, and some common underlying distributions ($F$). The main purpose of the simulation is to show how the type I error rates behave under these settings when the sample size is relatively small.

The ORRD uses the natural variation among the experimental units through a judgment ranking process to create artificial covariates. If this ranking information is reliable, then the artificial covariates act as variance reduction techniques. Otherwise, any error in ranking process yields a loss of efficiency in an ORRD design. Thus, modeling the quality of ranking information is important to evaluate the performance of the test. Any judgment ranking model should be practical and flexible so that it can be applied to a wide range of quality of ranking information. In the literature, there are mainly three classes of judgment ranking models. The first class of models generates judgment ranked observations from a mixture distribution of actual order statistics (Bohn and Wolfe [4]; Frey [13]). In this model, quality of ranking information is controlled by mixing parameters in the mixture distribution. The $j$-th judgment class distribution is modeled as

$$F_{[j]}(y) = \sum_{i=1}^{H} p_{ij} F_{(i)}(y),$$
where \( p_{ij} \) is the probability that the \( i \)-th order statistic is judged to be in the \( j \)-th judgment class distribution. The matrix that contains these probabilities must be a doubly-stochastic matrix. The second class uses the monotone likelihood ratio principle to rank the units in a set (Fligner and McEachern [12]).

The third class selects the units based on their perceived values that are tied to unmeasured true value of the units (Dell and Clutter [10]; David and Levire [9]). In this dissertation, we use Dell and Clutter model to rank the experimental units. This model states that the error term \( \varepsilon_i \), that is the property of the experimental units, is modeled with

\[
\begin{align*}
  u_i = \epsilon_i + \omega_i, & \quad i = 1, \ldots, H, \\
\end{align*}
\]

(2.12)

where \( \omega_i \)'s are i.i.d. draws form a suitably chosen distribution with mean zero and variance \( \tau^2 \). The model above creates a set of vectors \((u_i, \epsilon_i)\) for \( i = 1, \ldots, H \). The units in these sets are ranked based on the first components of \((u_i, \epsilon_i)\) for \( i = 1, \ldots, H \) and the second components are selected as the judgment ranked units. The quality of ranking in this model is controlled by the noise variable through its variance \( \tau^2 \). This dependence can be expressed in terms of the correlation coefficient between \( u \) and \( \epsilon \),

\[
\rho = \frac{\phi}{\sqrt{\phi^2 + \tau^2}},
\]

where \( \phi^2 \) is the variance of \( \epsilon_i \). It is clear that if \( \omega \) has a degenerate distribution, its variance is 0 and \( \rho = 1 \). This leads to perfect ranking \( \epsilon_i \equiv u_i \). Otherwise, ranking process will contain some ranking error. The magnitude of this error depends on the size of \( \tau^2 \). For example, when \( \phi^2 = 1 \), the selection of \( \tau^2 = 7/9 \) or 3, leads to \( \rho = 0.75 \) or 0.5.
The simulation parameters include the set size \( H = 2, 3, 4, 5 \), replication size \( n = 3, 5, 7, 8, 10 \), correlation coefficient \( \rho = 1, 0.9, 0.75, 0.5 \) and three different underlying distributions. These distributions are the standard normal distribution \((N(0,1))\), Student’s t-distribution with 3-degrees of freedom \((t(3))\), and the log-normal distribution \((LN(0,1))\). The simulation size is taken as 5000 replicates. Throughout the simulation, judgment ranked residuals are generated from the following algorithm for each simulation parameter combination.

**Step I:** Generate \( \epsilon_i \) and \( \omega_i \) independently from the underlying distribution \( F \) with mean 0 and variance \( \phi^2 \) and the normal distribution with mean 0 and variance \( \tau^2 \), respectively, for \( i = 1, 2, ..., 2nH \). Compute \( u_i = \epsilon_i + \omega_i \) for \( i = 1, 2, ..., 2nH \).

**Step II:** Randomly separate these \( u_i \)'s into \( 2n \) sets each of size \( H \).

**Step III:** Rank \( u_i \)'s in each set from 1 to \( H \).

**Step IV:** Randomly separate the \( 2n \) sets into two groups, \( n \) sets of each. In one of the groups, perform a randomization to decide whether the control or treatment regime is assigned to units that have ranks in Set \( \alpha \) or in Set \( \beta \). In the other group, perform an opposite allocation without a randomization. The residuals in the control group and the treatment group are \( \epsilon_{[h]1i} \) and \( \epsilon_{[h]2i} \) for \( h = 1, ..., H \) and \( j = 1, ..., n \).

**Step V:** Let the data in the control group and the treatment group be \( X \) and \( Y \), respectively. Then \( X_{[h]i} = \epsilon_{[h]1i} \) and \( Y_{[h]i} = \Delta + \epsilon_{[h]2i} \) for \( h = 1, ..., H \) and \( j = 1, ..., n \).

**Step VI:** Under the null hypothesis, \( \Delta = 0 \), compute \( z = \sqrt{2nH(T - \frac{1}{2})/\sigma_p} \).

The Type I error rates are estimated as the proportion of \( |z| \) values that exceed the critical value 1.96 for a 5% test. Table 2.3 illustrates these estimated Type I error rates under perfect ranking. The entries in Table 2.3 indicate that the estimated Type I error rates are reasonably close to nominal size (0.05) for replication size as
small as $n = 3$. It appears that heavy tailed and skewed distributions require slightly larger sample size (replication size $n \geq 5$).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>$H$</th>
<th>$n=3$</th>
<th>$n=5$</th>
<th>$n=7$</th>
<th>$n=8$</th>
<th>$n=10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(0,1)$</td>
<td></td>
<td>0.0493</td>
<td>0.0466</td>
<td>0.0431</td>
<td>0.0472</td>
<td>0.0479</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0031)</td>
<td>(0.0030)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0291</td>
<td>0.0437</td>
<td>0.0490</td>
<td>0.0439</td>
<td>0.0527</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0024)</td>
<td>(0.0029)</td>
<td>(0.0031)</td>
<td>(0.0029)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0354</td>
<td>0.0435</td>
<td>0.0451</td>
<td>0.0454</td>
<td>0.0475</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0026)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0318</td>
<td>0.0420</td>
<td>0.0414</td>
<td>0.0447</td>
<td>0.0488</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0025)</td>
<td>(0.0028)</td>
<td>(0.0028)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>$t(3)$</td>
<td></td>
<td>0.0463</td>
<td>0.0468</td>
<td>0.0434</td>
<td>0.0432</td>
<td>0.0480</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0030)</td>
<td>(0.0030)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0310</td>
<td>0.0396</td>
<td>0.0496</td>
<td>0.0451</td>
<td>0.0501</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0025)</td>
<td>(0.0028)</td>
<td>(0.0031)</td>
<td>(0.0029)</td>
<td>(0.0031)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0359</td>
<td>0.0443</td>
<td>0.0417</td>
<td>0.0452</td>
<td>0.0445</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0026)</td>
<td>(0.0029)</td>
<td>(0.0028)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0319</td>
<td>0.0441</td>
<td>0.0413</td>
<td>0.0448</td>
<td>0.0475</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0025)</td>
<td>(0.0029)</td>
<td>(0.0028)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>$LN(0,1)$</td>
<td></td>
<td>0.0474</td>
<td>0.0475</td>
<td>0.0430</td>
<td>0.0477</td>
<td>0.0460</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0030)</td>
<td>(0.0030)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0298</td>
<td>0.0410</td>
<td>0.0449</td>
<td>0.0437</td>
<td>0.0460</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0024)</td>
<td>(0.0028)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0357</td>
<td>0.0477</td>
<td>0.0461</td>
<td>0.0426</td>
<td>0.0432</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0026)</td>
<td>(0.0030)</td>
<td>(0.0030)</td>
<td>(0.0029)</td>
<td>(0.0029)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0318</td>
<td>0.0437</td>
<td>0.0386</td>
<td>0.0459</td>
<td>0.0478</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0025)</td>
<td>(0.0029)</td>
<td>(0.0027)</td>
<td>(0.0030)</td>
<td>(0.0030)</td>
</tr>
</tbody>
</table>

Table 2.3: The estimated type I error rates under perfect ranking. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.

We also investigate the Type I error rates under imperfect ranking. Table 2.4 provides the estimated type I error rates when $\rho = 0.9, 0.75, \text{ and } 0.5$. The data sets for the simulation results in Table 2.4 are generated from normal distribution $N(0,1)$ when $n = 5$. The entries of Table 2.4 show that although the performance of the test is excellent under the perfect ranking, the true Type I error rates are inflated
seriously under imperfect ranking. It is obvious that even a small ranking error, such as \( \rho = 0.9 \), has a big impact on the type I error rates. This suggests that there is a need to calibrate the effect of imperfect ranking on the type I error rate of the proposed test. The next chapter develops such a procedure.

<table>
<thead>
<tr>
<th>( H )</th>
<th>( \rho = 1 ) (0.003)</th>
<th>( \rho = 0.9 ) (0.005)</th>
<th>( \rho = 0.75 ) (0.005)</th>
<th>( \rho = 0.5 ) (0.006)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.042</td>
<td>0.112</td>
<td>0.178</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>3</td>
<td>0.038</td>
<td>0.161</td>
<td>0.259</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>4</td>
<td>0.043</td>
<td>0.242</td>
<td>0.371</td>
<td>0.469</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>5</td>
<td>0.044</td>
<td>0.283</td>
<td>0.447</td>
<td>0.523</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
</tbody>
</table>

Table 2.4: Type I error rates under imperfect ranking when \( n = 5 \) and underlying distribution is standard normal \( N(0,1) \). Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.

### 2.5 The point and interval estimates for location shift parameter \( \Delta \)

In order to complete the inferential procedures associated with the ORRD design, in this section we develop a point estimator and a distribution-free confidence interval for the location shift parameter \( \Delta \). Using the notation introduced in Chapter 1, we define the Hodges-Lehmann estimator of \( \Delta \) as the median of the pairwise difference of \( X \)- and \( Y \)-sample observations,

\[
\hat{\Delta} = \text{median}\{Y_{[k]t} - X_{[i]j}; i = 1, ..., H; j = 1, ..., n; k = 1, ..., H; t = 1, ..., n\}.
\]
An alternative representation for this point estimator can be obtained from the $L_1$-based norm $D(\Delta)$,

$$D(\Delta) = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} |X_{ij} - (Y_{kt} - \Delta)|.$$  

Since $D(\Delta)$ is a metric, it has a minimizer. We then define our point estimator, $\widehat{\Delta}$, as the minimizer of $D(\Delta)$. This minimizer also solves the estimating equation $S(\Delta) = \frac{\partial}{\partial \Delta} D(\Delta)$. The point estimate of $\Delta$ is then defined as $\widehat{\Delta}$ value such that $S(\widehat{\Delta}) = 0$. We note that this point estimate is equivalent to the Hodges-Lehmann estimator.

Theorem 2.7 indicates that $S(\Delta)$ is Pitman regular with efficacy factor $c_{DRRD}^2 = (\int f^2(x)dx)^2/\sigma^2_\nu$. Then it follows that $\sqrt{2nH\widehat{\Delta}}$ converges in distribution to a normal distribution with mean zero and variance $c^{-2}$.

Note that the proposed test statistic is $T = \sum_{i=1}^{n} \sum_{j=1}^{H} \sum_{k=1}^{H} \sum_{t=1}^{n} I(X_{ij} \leq Y_{kt})$. Let $T' = \sum_{i=1}^{H} \sum_{j=1}^{n} \sum_{k=1}^{H} \sum_{t=1}^{n} I(Y_{kt} \leq X_{ij})$. We can establish that $T' = (nH)^2 - T$. Under the null hypothesis, $T$ and $T'$ are stochastically equivalent and the following argument shows that $T$ is symmetric around $(nH)^2/2$.

$$P(T = t) = P(T' = t) = P(T = (nH)^2 - t),$$

which is equivalent to say $P(T - (nH)^2/2) = P(nH)^2/2 - T)$.

The distribution-free confidence interval of $\Delta$ follows directly from the inversion of the null distribution of $T$ according to the symmetric property of $T$. Let $D_{(1)} \leq \ldots \leq D_{((nH)^2)}$ be the ordered differences of $Y_{kt} - X_{ij}$ for $i = 1, \ldots, H$; $j = 1, \ldots, n$; $k = 1, \ldots, H$; $t = 1, \ldots, n$. If $P_0(T \leq k^*) = \alpha/2$, then since the distribution of $T$ is symmetric under the null hypothesis that is $\Delta = 0$, we have that

$$[D_{(k^*+1)}, D_{((nH)^2-k^*)}]$$

(2.13)
is a $100(1 - \alpha)\%$ confidence interval for $\Delta$. For large $n$, $k^*$ can be approximated from the asymptotic null distribution of $T$,

$$k^* = \frac{(nH)^2}{2} - 0.5 - z_{\alpha/2}\sigma_T,$$  \hspace{1cm}(2.14)

where the variance of the asymptotic null distribution of $T$ is $\sigma_T^2 = (nH)^3\sigma_p^2/2$. If we use the optimal ORRD, then under perfect ranking,

$$\sigma_T^2 = \begin{cases} 
\frac{(nH)^3}{2(H+1)^2} & \text{if } H \text{ is even;} \\
\frac{n^3H^2}{2(H+2)} & \text{if } H \text{ is odd.}
\end{cases}$$

Similarly, we conclude that the one-sided $100(1 - \alpha)\%$ lower (upper) confidence bound for $\Delta$ is given by $[D_{(k^*+1)}, \infty) \ ((-\infty, D_{((nH)^2 - k^*)})$ if $P(T \leq k^*) = \alpha$. Another confidence interval for $\Delta$ can be constructed based on the asymptotic distribution of $\hat{\Delta}$, $\hat{\Delta} \pm z_{\alpha/2}\sqrt{2nHc_{ORRD}}$ where $z_{\alpha/2}$ is the $100(1 - \alpha/2)\%$ percentile. This confidence interval is not very useful since $c_{ORRD}$ depends on the underlying density function, $f$. 

72
CHAPTER 3

ASYMPTOTIC PROPERTIES OF THE RANK-SUM STATISTIC UNDER IMPERFECT RANKING

As we have discussed in Chapter 2, the quality of ranking information has a big impact on the null distribution of the test statistic. Even a slight departure from a perfect judgment ranking inflates the size of the test considerably and the test loses its distribution free property. This is not a unique phenomenon for ORRD. It may happen in other sampling designs that rely on subjective ranking of experimental units. Most of the inferential procedures in ranked set sampling literature are developed under perfect ranking assumption. If this assumption is correct or nearly correct, procedures based on judgment ranking offer the most efficient inference. On the other hand, if the ranking is not perfect, the procedures may not be only inefficient, but may also be invalid. To address this problem, the researchers took two different approaches. The first approach develops inference without putting any restriction on the ranking process. These procedures may lose some efficiency under perfect ranking, but produce valid inference under imperfect ranking. Fligner and MacEachern [12], MacEachern et. al. [25] and McIntyre [27] are some examples in this category.
The second approach reduces the effect of imperfect ranking by estimating the quality of the judgment information from the data under a suitably chosen ranking model. In testing statistical hypotheses in Chapter 2, we develop the null distribution of the test statistic under perfect ranking. If the ranking information is perfect, then the null variance of the test statistic is equal to the actual variance of the test statistic. On the other hand any imperfection in ranking process inflates the actual variance of the test statistic, and yields larger p-values since the null variance is not calibrated for ranking error.

In order to control the size of the type I error rates, we calibrate the test by estimating the true null variance of the test statistic from the data. Throughout this chapter, we assume that we have an arbitrary but consistent imperfect ranking process. The consistency indicates that the same imperfect ranking model is used throughout the experiment in all cycles.

In Section 3.1, we construct a consistent estimator for the the null variance of the test statistic and show that suitably standardized version of it asymptotically converges to a normal distribution. In Section 3.2, we illustrate that the Student’s $t$-distribution provides a better approximation to the limiting null distribution of the test statistic for small sample sizes.
3.1 Asymptotic null distribution of $T$ under imperfect ranking

Recall that the variance of the asymptotic null distribution of $\sqrt{2nH(T - \frac{1}{2})}$ is given by

$$\sigma^2 = \frac{2}{H} \{ Var[\sum_{i=1}^{u} (1 - F(X_{[\alpha_i]1}) - \tau_{[\alpha_i]})] + \sum_{k=1}^{H-u} (F(Y_{[\beta_k]1}) - \tau_{[\beta_k]}) \} +$$

$$Var[\sum_{i=1}^{H-u} (1 - F(X_{[\beta_i]1}) - \tau_{[\beta_i]})] + \sum_{k=1}^{u} (F(Y_{[\alpha_k]1}) - \tau_{[\alpha_k]}) \}.$$

This expression holds irrespective of judgment ranking error as long as we have a consistent ranking process. Since the two terms in $\sigma^2$ are equal, we have

$$\sigma^2 = \frac{4}{H} \{ Var[\sum_{i=1}^{u} F(X_{[\alpha_i]1})] + Var[\sum_{k=1}^{H-u} F(Y_{[\beta_k]1})] -$$

$$2Cov[\sum_{i=1}^{u} F(X_{[\alpha_i]1}), \sum_{k=1}^{H-u} F(Y_{[\beta_k]1})] \}.$$

This expression can further be expanded as

$$\sigma^2 = \frac{4}{H} \{ \sum_{i=1}^{u} Var[F(X_{[\alpha_i]1})] + 2 \sum_{i=1}^{u} \sum_{j=1}^{\alpha_i < \alpha_j} Cov[F(X_{[\alpha_i]1}), F(X_{[\alpha_j]1})]$$

$$+ \sum_{k=1}^{H-u} Var[F(Y_{[\beta_k]1})] + 2 \sum_{k=1}^{H-u} \sum_{\beta_k < \beta_s} Cov[F(Y_{[\beta_k]1}), F(Y_{[\beta_s]1})] -$$

$$2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} Cov[F(X_{[\alpha_i]1}), F(Y_{[\beta_k]1})] - 2 \sum_{i=1}^{H-u} \sum_{\alpha_i > \beta_k} Cov[F(X_{[\alpha_i]1}), F(Y_{[\beta_k]1})] \}.$$
Let \( \mathbf{Z}_1 = (Z_{[1]}, \ldots, Z_{[H]}) = (X_{[\alpha_1]}, \ldots, X_{[\alpha_u]}, Y_{[\beta_1]}, \ldots Y_{[\beta_{H-u}]}) \). Then,

\[
\sigma^2 = \frac{4}{H} \left\{ \sum_{i=1}^{H} \operatorname{Var}[F(Z_{[i]}))] + 2 \sum_{i=1}^{u} \sum_{j=1}^{u} \operatorname{Cov}[F(X_{[\alpha_i]}), F(X_{[\alpha_j]}))] + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))] \right. \\
2 \sum_{k=1}^{H-u} \sum_{s=1}^{u} \operatorname{Cov}[F(Y_{[\beta_s]}), F(Y_{[\beta_k]}))] + 2 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))] - 2 \sum_{k=1}^{H-u} \sum_{s=1}^{u} \operatorname{Cov}[F(Y_{[\beta_s]}), F(Y_{[\beta_k]}))] \\
4 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{\alpha_i < \beta_k} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))] - 4 \sum_{i=1}^{u} \sum_{k=1}^{H-u} \sum_{\alpha_i > \beta_k} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))].
\]

Regrouping certain terms simplifies \( \sigma^2 \) to

\[
\sigma^2 = \frac{4}{H} \left\{ \sum_{i=1}^{H} \operatorname{Var}[F(Z_{[i]}))] + 2 \sum_{i=1}^{H} \sum_{i<j} \operatorname{Cov}[F(Z_{[i]}), F(Z_{[j]}))] - 2 \sum_{i=1}^{H} \sum_{j=1}^{H-i} \operatorname{Cov}[F(Z_{[i]}), F(Z_{[j]}))] \right. \\
4 \sum_{i=1}^{H} \sum_{k=1}^{H-u} \sum_{\alpha_i < \beta_k} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))] - 4 \sum_{i=1}^{H} \sum_{k=1}^{H-u} \sum_{\alpha_i > \beta_k} \operatorname{Cov}[F(X_{[\alpha_i]}), F(Y_{[\beta_k]}))].
\]

We note that

\[
\operatorname{Var}[F(Z_{[i]}))] = E[F^2(Z_{[i]}))] - (E[F(Z_{[i]}))]^2
\]

\[
= \int F^2(t) dF_{[i]}(t) - (E[F(Z_{[i]}))]^2
\]

and under a consistent ranking scheme

\[
\sum_{i=1}^{H} \operatorname{Var}[F(Z_{[i]}))] = \sum_{i=1}^{H} \int F^2(t) dF_{[i]}(t) - \sum_{i=1}^{H} \{E[F(Z_{[i]}))]^2
\]

\[
= H \int F^2(t) dF(t) - \sum_{i=1}^{H} \{E[F(Z_{[i]}))]^2
\]

\[
= \frac{H}{3} - \sum_{i=1}^{H} \{E[F(Z_{[i]}))]^2
\]

Let

\[
\mu_{[i]} = E[F(Z_{[i]}))] \text{ and } \nu_{[i,j]} = E[F(Z_{[i]}))F(Z_{[j]})).
\]

76
Then
\[ \sum_{i=1}^{H} \text{Var}[F(Z[i])] = \frac{H}{3} - \sum_{i=1}^{H} \mu_{[i]}^{2} \]
and
\[ \sum_{i=1}^{H} \sum_{j=1}^{H} \text{Cov}[F(Z[i]), F(Z[j])] = \sum_{i=1}^{H} \sum_{j=1}^{H} \{E[F(Z[i])F(Z[j])] - E[F(Z[i])]E[F(Z[j])]\} \]
\[ = \sum_{i=1}^{H} \sum_{j=1}^{H} (\nu_{[i,j]} - \mu_{[i]}\mu_{[j]}). \]
Hence, \( \sigma^2 \) can be expressed in \( \mu_{[i]} \) and \( \nu_{[i,j]} \) as
\[ \sigma^2 = \frac{4}{H} \left( \frac{H}{3} - \sum_{i=1}^{H} \mu_{[i]}^{2} + 2 \sum_{i=1}^{H} \sum_{j=1}^{H} (\nu_{[i,j]} - \mu_{[i]}\mu_{[j]}) - \right. \]
\[ \left. 4 \sum_{i=1}^{u} \sum_{k=1}^{H-u} (\nu_{[\alpha_i,\beta_k]} - \mu_{[i]}\mu_{[j]}) - 4 \sum_{i=1}^{u} \sum_{k=1}^{H-u} (\nu_{[\alpha_i,\beta_k]} - \mu_{[i]}\mu_{[j]}). \right] \tag{3.1} \]
Therefore to construct a consistent estimator for \( \sigma^2 \) we need to find consistent estimators for \( \mu_{[i]}, i = 1, ..., H, \) and \( \nu_{[i,j]}, i = 1, ..., H, j = 1, ..., H \) under an arbitrary but consistent ranking procedure.

**Lemma 3.1.** Under an arbitrary but consistent ranking procedure the consistent and unbiased estimator of \( \mu_{[i]} \) is given by
\[ \hat{\mu}_{[i]} = \frac{1}{2n(2n - 1)H} \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} I(Z[s]k \leq Z[i,j]). \]
Proof. It is easy to show that \( \hat{\mu}_{[i]} \) is an unbiased estimator of \( \mu_{[i]} \)

\[
E[\hat{\mu}_{[i]}] = \frac{1}{2n(2n-1)H} \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} E[I(Z_{[s]k} \leq Z_{[i]j})]
\]

\[
= \frac{1}{2n(2n-1)H} \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} E[F_{[s]}(Z_{[i]j})]
\]

\[
= \frac{1}{2n(2n-1)H} \sum_{j=1}^{2n} \sum_{k=1}^{2n} HE[F(Z_{[i]j})]
\]

\[
= \mu_{[i]}.
\]

To prove the consistency of \( \hat{\mu}_{[i]} \), we first show that \( \text{Var}(\hat{\mu}_{[i]}) \to 0 \) as \( n \to \infty \). First note that

\[
\text{Var}\left[ \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} \text{Cov}\left( \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right),
\]

\[
\sum_{w=1}^{H} I(Z_{[w]q} \leq Z_{[i]p})\].
\]

These covariances can be partitioned into four different sums:

Case I: \( j \neq k \neq p \neq q \); note that all \( Z_{[s]k} \)'s and \( Z_{[i]j} \)'s in this case are independent and it follows that

\[
\text{Var}\left[ \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = 0.
\]

Case II: \( j \neq k, p \neq q, k = q, j \neq p \);

\[
\text{Var}\left[ \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} \text{Cov}\left( \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right),
\]

\[
\sum_{w=1}^{H} I(Z_{[w]k} \leq Z_{[i]p})\].
\]
Case III: \( j \neq k, p \neq q, k \neq q, j = p; \)

\[
Var\left[ \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = \sum_{k=1}^{2n} \sum_{q=1}^{2n} \sum_{j=1 \atop k \neq j}^{2n} \sum_{q \neq k}^{2n} Cov\left[ \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}), \sum_{w=1}^{H} I(Z_{[w]q} \leq Z_{[i]j}) \right].
\]

Case IV: \( j \neq k, p \neq q, k = q, j = p; \)

\[
Var\left[ \sum_{j=1}^{2n} \sum_{k=1}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = \sum_{j=1}^{2n} \sum_{k=1 \atop k \neq j}^{2n} \sum_{s=1}^{H} Cov\left[ \sum_{k=1}^{2n} I(Z_{[s]k} \leq Z_{[i]j}), \sum_{w=1}^{H} I(Z_{[w]k} \leq Z_{[i]j}) \right].
\]

Putting these four sums together, we obtain that

\[
Var\left[ \sum_{j=1}^{2n} \sum_{k=1 \atop k \neq j}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right] = \\
2n(2n-1)(2n-2)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]2} \leq Z_{[i]1}) \right] + \\
2n(2n-1)(2n-2)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]3} \leq Z_{[i]1}) \right] + \\
2n(2n-1)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]2} \leq Z_{[i]1}) \right].
\]

Hence, the variance of \( \hat{\mu}_{[i]} \) can be written as

\[
Var(\hat{\mu}_{[i]}) = \frac{1}{4n^2(2n-1)^2H^2} Var\left[ \sum_{j=1}^{2n} \sum_{k=1 \atop k \neq j}^{2n} \sum_{s=1}^{H} I(Z_{[s]k} \leq Z_{[i]j}) \right]
\]

\[
= \frac{1}{4n^2(2n-1)^2H^2} \left\{ \\
2n(2n-1)(2n-2)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]2} \leq Z_{[i]1}) \right] + \\
+2n(2n-1)(2n-2)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]3} \leq Z_{[i]1}) \right] + \\
+2n(2n-1)Cov\left[ \sum_{s=1}^{H} I(Z_{[s]2} \leq Z_{[i]1}), \sum_{w=1}^{H} I(Z_{[w]2} \leq Z_{[i]1}) \right] \right\}
\]

\[
= o(1).
\]
We complete the proof by observing that

\[ P(\vert \hat{\mu}_{ij} - \mu_{ij} \vert > \varepsilon) \leq \frac{\text{Var}(\hat{\mu}_{ij})}{\varepsilon^2} \to 0 \text{ as } n \to \infty. \]

\[ \square \]

**Lemma 3.2.** Under the conditions of Lemma 3.1, the consistent and unbiased estimator of \( \nu_{ij} \) is given by

\[ \hat{\nu}_{ij} = \frac{1}{4n(2n-1)(n-1)H^2} \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{s=1}^{H} I(Z_{s|k} \leq Z_{ij|l}) \sum_{s=1}^{H} I(Z_{s|t} \leq Z_{ij|l}). \]

**Proof.**

\[ E[\hat{\nu}_{ij}] = \frac{1}{4n(2n-1)(n-1)H^2} \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{s=1}^{H} E[\sum_{s=1}^{H} I(Z_{s|k} \leq Z_{ij|l}) \sum_{s=1}^{H} I(Z_{s|t} \leq Z_{ij|l})] \]

\[ = \frac{1}{4n(2n-1)(n-1)H^2} \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{s=1}^{H} E[\sum_{s=1}^{H} I(Z_{s|k} \leq Z_{ij|l}) I(Z_{s|t} \leq Z_{ij|l})] \]

\[ = \frac{2n(2n-1)(2n-2)}{4n(2n-1)(n-1)H^2} \sum_{s=1}^{H} \sum_{v=1}^{H} E[I(Z_{s|2} \leq Z_{ij|1}) I(Z_{v|3} \leq Z_{ij|1})] \]

\[ = \frac{1}{H^2} \sum_{s=1}^{H} \sum_{v=1}^{H} E[E(I(Z_{s|2} \leq Z_{ij|1}) I(Z_{v|3} \leq Z_{ij|1})|Z_{ij|1})] \]

\[ = \frac{1}{H^2} \sum_{s=1}^{H} \sum_{v=1}^{H} E[E(I(Z_{s|2} \leq Z_{ij|1})|Z_{ij|1})] E[I(Z_{v|3} \leq Z_{ij|1})|Z_{ij|1})] \]

\[ = \frac{1}{H^2} \sum_{s=1}^{H} \sum_{v=1}^{H} E[F_s(Z_{ij|1}) F_v(Z_{ij|1})] \]

\[ = \frac{1}{H^2} \sum_{s=1}^{H} F_s(Z_{ij|1}) \sum_{v=1}^{H} F_v(Z_{ij|1}) = E[F(Z_{ij|1})] \]

Thus, \( \hat{\nu}_{ij} \) is an unbiased estimator of \( \nu_{ij} \).
For the consistency, it is sufficient to show that $\text{Var}(\hat{\nu}_{i,j})$ goes to zero as $n$ gets large. Let

$$U_{ilk} = \sum_{s=1}^{H} I(Z_{s}k \leq Z_{s}l), \quad U_{jlt} = \sum_{s=1}^{H} I(Z_{s}l \leq Z_{s}t),$$

and $a = \frac{1}{4n(2n-1)(n-1)H^2}$. Then, the variance of $\hat{\nu}_{i,j}$ can be written as

$$\text{Var}(\hat{\nu}_{i,j}) = \text{Var}(a \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} U_{ilk} U_{jlt})$$

$$= a^2 \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{p=1}^{2n} \sum_{q=1}^{2n} \sum_{w=1}^{2n} \sum_{q'\neq p} \sum_{w'\neq q,p} \text{Cov}(U_{ilk} U_{jlt}, U_{ipq} U_{jpw}) .$$

If $l \neq p$, then $\text{Cov}(U_{ilk} U_{jlt}, U_{ipq} U_{jpw}) = 0$. Hence we only need to consider the case that $l = p$

$$\text{Var}(\hat{\nu}_{i,j}) = a^2 \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{q=1}^{2n} \sum_{w=1}^{2n} \sum_{q'\neq p} \sum_{w'\neq q,p} \text{Cov}(U_{ilk} U_{jlt}, U_{ilq} U_{jlt}) .$$

Furthermore $\text{Var}(\hat{\nu}_{i,j})$ can be partitioned into four cases:

Case I: $k = q$ and $t = w$;

Case II: $k = q$ and $t \neq w$;

Case III: $k \neq q$ and $t = w$;

Case IV: $k \neq q$ and $t \neq w$.

The variance of $\hat{\nu}_{i,j}$ can be written as

$$\text{Var}(\hat{\nu}_{i,j}) = a^2 \left\{ \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{k \neq l} \sum_{t \neq k} \text{Cov}(U_{ilk} U_{jlt}, U_{ilk} U_{jlt}) \right. \right.$$

$$+ \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{w=1}^{2n} \sum_{w' \neq l} \sum_{w' \neq l} \text{Cov}(U_{ilk} U_{jlt}, U_{ilk} U_{jtw}) +$$

$$+ \sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{q=1}^{2n} \sum_{q' \neq l} \sum_{q' \neq l} \text{Cov}(U_{ilk} U_{jlt}, U_{ilq} U_{jlt}) \right.$$

$$81$$
\[
\sum_{l=1}^{2n} \sum_{k=1}^{2n} \sum_{t=1}^{2n} \sum_{q=1}^{2n} \sum_{w=1}^{2n} \text{Cov}(U_{ilk}U_{jlt}, U_{dlq}U_{jlw})
\]

and

\[
\text{Var}(\widehat{\nu}_{[i,j]}) = a^2 \{ 2n(2n-1)(2n-2)\text{Var}(U_{112}U_{113}) \\
+ 2n(2n-1)(2n-2)(2n-3)\text{Cov}(U_{113}U_{114}, U_{i13}U_{j15}) \\
+ 2n(2n-1)(2n-2)(2n-3)\text{Cov}(U_{i13}U_{114}, U_{i15}U_{j14}) \\
+ 2n(2n-1)(2n-2)(2n-3)\text{Cov}(U_{i12}U_{j13}, U_{i14}U_{j15}) \} \\
= o\left(\frac{1}{n^2}\right) + o\left(\frac{1}{n}\right) + o\left(\frac{1}{n}\right) + o\left(\frac{1}{n}\right).
\]

Thus, \(\text{Var}(\widehat{\nu}_{[i,j]}) \to 0\) as \(n \to \infty\) and we conclude that \(\widehat{\nu}_{[i,j]}\) is a consistent estimator of \(\nu_{[i,j]}\).

\[\square\]

Note that \(\sigma^2\) is a finite sum and a continuous function of \(\mu_{[i]}\) and \(\nu_{[i,j]}\). Then an unbiased and consistent estimator of \(\sigma^2\) is obtained by inserting \(\widehat{\mu}_{[i]}\) and \(\widehat{\nu}_{[i,j]}\) in (3.1), which yields

\[
\widehat{\sigma}^2 = \frac{4}{H} \left( \frac{H}{3} - \sum_{i=1}^{H} \widehat{\mu}^2_{[i]} + 2 \sum_{i=1}^{H} \sum_{j=1}^{H} (\widehat{\nu}_{[i,j]} - \widehat{\mu}_{[i]} \widehat{\mu}_{[j]}) - \\
4 \sum_{i=1}^{H-u} \sum_{k=1}^{u} \sum_{\alpha_i < \beta_k} (\widehat{\nu}_{[\alpha_i, \beta_k]} - \widehat{\mu}_{[\alpha_i]} \widehat{\mu}_{[\beta_k]}) - 4 \sum_{i=1}^{H-u} \sum_{k=1}^{H-u} \sum_{\alpha_i > \beta_k} (\widehat{\nu}_{[\alpha_i, \beta_k]} - \widehat{\mu}_{[\alpha_i]} \widehat{\mu}_{[\beta_k]}) \right). \tag{3.2}
\]

**Theorem 3.3.** For a fixed set size \(H\), under an arbitrary and consistent ranking model, the asymptotic null distribution of \(\sqrt{2nH(T - \frac{1}{2})}/\widehat{\sigma}\) converges to a standard normal distribution as \(n\) goes to infinity.

**Proof.** The proof simply follows from the Slutsky Theorem. \(\square\)
The asymptotic null distribution shown in Theorem 3.3 is standard normal and distribution-free. The type I error rates therefore hold under imperfect rankings when the sample size is large enough. For the implementation of the proposed test, the null variance of \( T \) must be estimated under the null hypothesis. In order to achieve this, we first center \( X \)- and \( Y \)-sample observations by subtracting their medians. Let \( M_x \) be the median of all \( X_{i[j]}, \ i = 1, ..., H, \ j = 1, ..., n \) and \( M_y \) be the median of \( Y_{i[j]}, \ i = 1, ..., H, \ j = 1, ..., n \). We then center \( X \) and \( Y \) observations with \( \tilde{X}_{i[j]} = X_{i[j]} - M_x \) and \( \tilde{Y}_{i[j]} = Y_{i[j]} - M_y \). The estimator \( \tilde{\sigma} \) is constructed through the estimators \( \tilde{\mu}_{[i]} \) and \( \tilde{\nu}_{[i,j]} \) obtained from the centered data \( \tilde{Z}_i = [\tilde{X}_{[i][1]}, ..., \tilde{X}_{[i][H]}, \tilde{Y}_{[i][1]}, ..., \tilde{Y}_{[i][H]}] \) for \( i = 1, ..., H \).

We use the Hodges-Lehmann estimator, \( \tilde{\Delta} \), to estimate the location shift parameter \( \Delta \). Its variance under imperfect ranking is obtained by replacing \( \sigma_p^2 \) with \( \tilde{\sigma}^2 \) in the efficacy factor, where \( \tilde{\sigma}^2 \) is from (3.2). The associated confidence intervals can be constructed similarly except that \( k^* \) should be approximated through \( \tilde{\sigma} \) instead of \( \sigma \), that is

\[
k^* = (nH)^2/2 - 0.5 - z_\alpha/2\tilde{\sigma}_T,
\]

where \( \tilde{\sigma}_T \) is the estimated variance of the asymptotic null distribution of \( T \) and

\[
\tilde{\sigma}_T^2 = (nH)^3\tilde{\sigma}^2/2.
\]

### 3.2 Simulation study for the ORRDD with imperfect rankings

In this section, we reinvestigate the Type I error rate of the test. The test statistic is calibrated by estimating the null variance through \( \tilde{\sigma}^2 \). To estimate the
error rates, the data sets are generated from models (1.2) and (1.3) under the Dell-Clutter judgment ranking model. The residuals are generated through steps 1-4 in Chapter 2 with the exception of Step VI. In Step VI, we compute $\hat{\sigma}$ and construct the test statistic based on it.

From the normal approximation in the previous section, the Type I error rates are calculated for a 0.05 test. The simulation parameters are chosen as $H = 2, 3, 4, 5, n = 4, 5, 7, 10, \rho = 1.0, 0.9, 0.75, 0.5$. The residuals are generated from the following underlying distributions including the standard normal distribution (N(0,1)), $t$-distribution ($t(3)$), or lognormal distribution (LN(0,1)), which represent symmetric, heavy-tailed and skewed distributions, respectively. Tables 3.1, 3.2, 3.3 and 3.4 present the estimated Type I error rates for $\rho = 1.0, 0.9, 0.75, 0.5$, respectively. Each table provides two estimates, one based on normal approximation (Normal), the other based on Student’s $t$-distribution ($t$). We noticed that estimated Type I error rates slowly converge to nominal value 0.05 under normal approximation. This is not surprising result since the estimation of $\sigma$ introduces additional variation and this inflates the Type I error rates for small sample sizes. Thus, for the small sample sizes, Student’s $t$-distribution does provide a better approximation to the null distribution of the test statistic. The degrees of freedom of the Student’s $t$-distribution is computed from the Satterthwaite approximation.

Based on these tables, it is clear that there is a big difference between the normal and $t$ approximation. The estimated Type I error rates based on $t$-approximation are much closer to the nominal value 0.05 for all sample sizes. Nevertheless, as we expect, this difference gets smaller as the sample size increases. Another observation from the simulation results in Table 3.1-3.4 is that the $t$-approximation works quite
well for both heavy-tailed (t(3)) and skewed (LN(0,1)) distributions when replication size is as small as \( n = 4 \).

In conclusion, the Type I error rate based on the proposed test can be appropriately controlled even when there is an error in judgment ranking process. The test is a valid test irrespective of ranking errors.
Table 3.1: The estimated Type I error rates when $\rho = 1.0$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.
<table>
<thead>
<tr>
<th>$H$</th>
<th>$n$</th>
<th>$t$</th>
<th>Normal $t$</th>
<th>$t$</th>
<th>Normal $t$</th>
<th>$t$</th>
<th>Normal $t$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$N(0,1)$</td>
<td>$t(3)$</td>
<td>$LN(0,1)$</td>
<td>$t(3)$</td>
<td>$LN(0,1)$</td>
<td>$t(3)$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.0606</td>
<td>0.1362</td>
<td>0.0634</td>
<td>0.1445</td>
<td>0.0640</td>
<td>0.1437</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0034)</td>
<td>(0.0049)</td>
<td>(0.0034)</td>
<td>(0.0050)</td>
<td>(0.0035)</td>
<td>(0.0050)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0510</td>
<td>0.1106</td>
<td>0.0565</td>
<td>0.1158</td>
<td>0.0556</td>
<td>0.1144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0031)</td>
<td>(0.0044)</td>
<td>(0.0033)</td>
<td>(0.0045)</td>
<td>(0.0032)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0582</td>
<td>0.0932</td>
<td>0.0568</td>
<td>0.0900</td>
<td>0.0520</td>
<td>0.0903</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0041)</td>
<td>(0.0033)</td>
<td>(0.0040)</td>
<td>(0.0031)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0548</td>
<td>0.0754</td>
<td>0.0534</td>
<td>0.0737</td>
<td>0.0497</td>
<td>0.0730</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td>(0.0031)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.0698</td>
<td>0.1420</td>
<td>0.0633</td>
<td>0.1453</td>
<td>0.0658</td>
<td>0.1439</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0036)</td>
<td>(0.0049)</td>
<td>(0.0034)</td>
<td>(0.0050)</td>
<td>(0.0035)</td>
<td>(0.0050)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0618</td>
<td>0.1180</td>
<td>0.0609</td>
<td>0.1184</td>
<td>0.0640</td>
<td>0.1229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0034)</td>
<td>(0.0046)</td>
<td>(0.0034)</td>
<td>(0.0046)</td>
<td>(0.0035)</td>
<td>(0.0046)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0580</td>
<td>0.0908</td>
<td>0.0615</td>
<td>0.0947</td>
<td>0.0601</td>
<td>0.0956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0041)</td>
<td>(0.0034)</td>
<td>(0.0041)</td>
<td>(0.0034)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0484</td>
<td>0.0704</td>
<td>0.0577</td>
<td>0.0785</td>
<td>0.0524</td>
<td>0.0724</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0030)</td>
<td>(0.0036)</td>
<td>(0.0033)</td>
<td>(0.0038)</td>
<td>(0.0032)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.0578</td>
<td>0.1340</td>
<td>0.0599</td>
<td>0.1435</td>
<td>0.0531</td>
<td>0.1317</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0048)</td>
<td>(0.0034)</td>
<td>(0.0050)</td>
<td>(0.0032)</td>
<td>(0.0048)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0578</td>
<td>0.1134</td>
<td>0.0610</td>
<td>0.1207</td>
<td>0.0489</td>
<td>0.1039</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0045)</td>
<td>(0.0034)</td>
<td>(0.0046)</td>
<td>(0.0030)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0560</td>
<td>0.0882</td>
<td>0.0592</td>
<td>0.0951</td>
<td>0.0541</td>
<td>0.0886</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0040)</td>
<td>(0.0033)</td>
<td>(0.0041)</td>
<td>(0.0032)</td>
<td>(0.0040)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0546</td>
<td>0.0754</td>
<td>0.0587</td>
<td>0.0787</td>
<td>0.0512</td>
<td>0.0727</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td>(0.0033)</td>
<td>(0.0038)</td>
<td>(0.0031)</td>
<td>(0.0037)</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.0536</td>
<td>0.1332</td>
<td>0.0565</td>
<td>0.1432</td>
<td>0.0492</td>
<td>0.1221</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0032)</td>
<td>(0.0048)</td>
<td>(0.0033)</td>
<td>(0.0050)</td>
<td>(0.0031)</td>
<td>(0.0046)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0596</td>
<td>0.1148</td>
<td>0.0530</td>
<td>0.1092</td>
<td>0.0509</td>
<td>0.1012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0045)</td>
<td>(0.0032)</td>
<td>(0.0044)</td>
<td>(0.0031)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0596</td>
<td>0.0954</td>
<td>0.0560</td>
<td>0.0893</td>
<td>0.0489</td>
<td>0.0808</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0042)</td>
<td>(0.0033)</td>
<td>(0.0040)</td>
<td>(0.0030)</td>
<td>(0.0039)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0568</td>
<td>0.0766</td>
<td>0.0621</td>
<td>0.0845</td>
<td>0.0501</td>
<td>0.0703</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0033)</td>
<td>(0.0038)</td>
<td>(0.0034)</td>
<td>(0.0039)</td>
<td>(0.0031)</td>
<td>(0.0036)</td>
</tr>
</tbody>
</table>

Table 3.2: The estimated Type I error rates when $\rho = 0.9$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.
<table>
<thead>
<tr>
<th>$\rho = .75$</th>
<th>$N(0,1)$</th>
<th>$t(3)$</th>
<th>LN$(0,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>$n</td>
<td>$t$</td>
<td>Normal</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.0642 (0.0035)</td>
<td>0.1434 (0.0050)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0535 (0.0032)</td>
<td>0.1140 (0.0045)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0571 (0.0033)</td>
<td>0.0931 (0.0041)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0537 (0.0032)</td>
<td>0.0736 (0.0037)</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.0595 (0.0033)</td>
<td>0.1476 (0.0050)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0583 (0.0033)</td>
<td>0.1175 (0.0046)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0607 (0.0034)</td>
<td>0.0948 (0.0041)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0536 (0.0032)</td>
<td>0.0722 (0.0037)</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.0582 (0.0033)</td>
<td>0.1360 (0.0048)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0542 (0.0032)</td>
<td>0.1108 (0.0044)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0598 (0.0034)</td>
<td>0.0952 (0.0042)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0598 (0.0034)</td>
<td>0.0798 (0.0038)</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.0518 (0.0031)</td>
<td>0.1358 (0.0048)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0559 (0.0032)</td>
<td>0.1150 (0.0045)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0562 (0.0033)</td>
<td>0.0927 (0.0041)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.0556 (0.0032)</td>
<td>0.0768 (0.0038)</td>
</tr>
</tbody>
</table>

Table 3.3: The estimated Type I error rates when $\rho = 0.75$. Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.
<table>
<thead>
<tr>
<th>( \rho = .5 )</th>
<th>N(0,1)</th>
<th>t(3)</th>
<th>LN(0,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>n</td>
<td>Normal</td>
<td>t</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>0.0613</td>
<td>0.1370</td>
</tr>
<tr>
<td>5</td>
<td>0.0529</td>
<td>0.1160</td>
<td>0.0586</td>
</tr>
<tr>
<td>7</td>
<td>0.0557</td>
<td>0.0898</td>
<td>0.0535</td>
</tr>
<tr>
<td>10</td>
<td>0.0513</td>
<td>0.0730</td>
<td>0.0542</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.0608</td>
<td>0.1485</td>
</tr>
<tr>
<td>5</td>
<td>0.0599</td>
<td>0.1175</td>
<td>0.0586</td>
</tr>
<tr>
<td>7</td>
<td>0.0585</td>
<td>0.0942</td>
<td>0.0515</td>
</tr>
<tr>
<td>10</td>
<td>0.0549</td>
<td>0.0736</td>
<td>0.0550</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.0494</td>
<td>0.1309</td>
</tr>
<tr>
<td>5</td>
<td>0.0525</td>
<td>0.1117</td>
<td>0.0540</td>
</tr>
<tr>
<td>7</td>
<td>0.0573</td>
<td>0.0889</td>
<td>0.0568</td>
</tr>
<tr>
<td>10</td>
<td>0.0564</td>
<td>0.0744</td>
<td>0.0541</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.0507</td>
<td>0.1362</td>
</tr>
<tr>
<td>5</td>
<td>0.0554</td>
<td>0.1092</td>
<td>0.0552</td>
</tr>
<tr>
<td>7</td>
<td>0.0573</td>
<td>0.0913</td>
<td>0.0587</td>
</tr>
<tr>
<td>10</td>
<td>0.0611</td>
<td>0.0828</td>
<td>0.0506</td>
</tr>
</tbody>
</table>

Table 3.4: The estimated Type I error rates when \( \rho = 0.5 \). Standard errors are given in parentheses. The simulation study is based on 5,000 replicates.
CHAPTER 4

SIMULATION STUDY FOR THE EMPIRICAL POWER OF THE TEST

In this chapter, we investigate the empirical power of the proposed test through a simulation study. Simulation contains four parameters, set size $H$, replicate size $n$, the correlation coefficient between ranking and response variables $\rho$, location shift parameter $\Delta$ and underlying distribution $F$. The parameters take the following values: $H = 2, 3, 4, 5$, $n = 5, 10$, $\rho = 1.0, 0.9, 0.75, 0.5$, $\Delta = 0(0.1)1$, and the underlying distribution follows the standard normal distribution (N(0,1)), $t$-distribution (t(3)), or lognormal distribution (LN(0,1)). For each one of the factorial combination of $H$, $n$, $\rho$, $\Delta$ and $F$, five thousand data sets are generated through steps I-V in Chapter 2, the location shift parameter is tested at a 5% level based on each data set and the percent of tests rejecting the null hypothesis is taken as the empirical power of the test. Simulation is divided into two parts, perfect and imperfect ranking. In perfect ranking, the critical region is approximated by normal distribution, and in imperfect ranking, it is approximated by Student’s $t$-distribution, where the degree of freedom is computed from the Satterthwaite’s approximation.

Figures 4.1-4.3 demonstrate the empirical power curves for a 5% test when $\rho = 1.0$ and the underlying distributions are N(0,1), t(3) and LN(0,1), respectively. In each
figure, the left column shows the empirical power curves of the test based on $T$, the middle column shows the empirical power curves of the test based on $U_{RSS}$ and the right column shows the empirical power curves of the test based on the MWW. Additionally, the power curves for the three test statistics are matched for sample size $2nH$ in each row. The numbers on each curve indicate the set size $H$ and the values of $n$ are provided in the sub-title at each panel. Although the MWW test does not involve the design parameter $H$, we use it in our graphs to match the sample sizes in all three tests.

We notice from the figures 4.1-4.3 that the power curves for all three tests based on $T$, $U_{RSS}$ and the MWW statistics are getting steeper as $H$ increases for a fixed $n$. Similarly, when $H$ is fixed, the power curves are getting steeper as $n$ increases. This result is consistent with a well known fact that the power of the test gets larger as the sample size increases. The power curves also demonstrate that the tests based on $T$ has higher power than the tests based on $U_{RSS}$ and the MWW statistics with matched sample sizes. For example, the power curve of $T$ with $H = 2$ and $n = 5$ is steeper than and lies above the power curves of $U_{RSS}$ and the MWW statistics. This power curve is almost equivalent to the power curve of $U_{RSS}$ with $H = 3$ and the power curve of the MWW statistic with $H = 5$. In other words, to achieve the same power, the test based on $T$ requires a much smaller sample size than the sample size required for the $U_{RSS}$ and the MWW statistics. Therefore, we conclude that, under perfect ranking, the proposed test statistic $T$ outperforms the other two competitors with the matched sample sizes.

Figures 4.4-4.15 demonstrate the power curves of proposed test under imperfect ranking along with its competitor test proposed by Fligner and MacEachern [12] with
matched sample sizes. Fligner and MacEachern [12] developed a test statistic based on ranked set samples, denoted by

\[ FM = \sum_{i=1}^{H} T_{ii}, \text{ where } T_{ii} = \sum_{t=1}^{m} \sum_{j=1}^{n} I(X_{[i]t} < Y_{[i]j}). \]

Under the null hypothesis, each \( T_{ii} \) is a MWW statistic from the \( i \)-th judgment class of \( X \)- and \( Y \)-samples with sample size \( m \) and \( n \), respectively.

The construction of the FM statistic is appealing for two reasons. The first reason is that the exact null distribution of \( T_{ii} \), for \( i = 1, \ldots, H \), is distribution-free and does not depend on whether the judgment ranking is perfect, imperfect or random. This follows from the fact that \( F_{[i]}(x) = G_{[i]}(x) \) under the null hypothesis. Therefore, the null distribution remains unchanged even when ranking information is incomplete as long as both treatment populations have the same judgment ranking mechanism. Since \( T_{ii} \)'s, for \( i = 1, \ldots, H \), are all mutually independent and the FM statistic is a convolution of these \( H \)-independent test statistics, it also preserves the distribution-free property of its components and does not require perfect ranking information.

Even though the exact null distribution of the FM statistic can be constructed, it would be computationally expensive for large \( n \). In this case, one can approximate the null distribution of the FM statistic with a normal distribution. For large \( n \), the null distribution of the FM statistic is normal with mean \( 
\frac{Hmn}{2} \) and standard deviation \( \sqrt{\frac{Hmn(m+n-1)}{12}} \).

The second appealing reason is that the test based on the FM statistic does not suffer from loss of power due to stratifications. Close inspection of the FM statistic shows that it is constructed based on a total of \( Hnm \) comparisons among the treatment and control response, while the statistic \( U_{RSS} \) is constructed based on \( (nH)^2 \) comparisons. Under perfect ranking, one may intuitively think that the
FM statistic may lose power due to reduction in the number of comparisons between treatment and control responses. On the other hand, Fligner and MacEachern [12] showed that loss of power is minimal.

Due to these appealing features of the FM statistic, we take it as a competitor to our statistic $T$ under imperfect ranking. To match the sample sizes, the FM statistic is constructed based on balanced ranked set samples with X-sample size $n_H$ and Y-sample size $n_H$. Therefore, the asymptotic null distribution of FM statistic is normal with mean $\frac{Hn^2}{2}$ and standard deviation $\sqrt{\frac{Hn^2(2n-1)}{12}}$.

Generally speaking, the power curves of the test based on $T$ are similar to the power curves of the test based on the MWW statistics if $\rho = 0.5$. If $\rho = 1, 0.9, 0.75$, however, the power curves of $T$ are steeper than the power curves of the MWW statistics. Therefore, if the quality of ranking is reasonably good ($\rho \geq 0.75$) the performance of $T$ is superior to the MWW statistic.

When we compare the power curves between $T$ and FM statistic, we have the following findings.

- The power curves of $T$ are much steeper than those of FM statistics for $\rho = 1, 0.9$. For example, the power curves of $T$ with $H = 4$ are above the power curves of the FM statistic with $H = 5$.

- If $\rho = 0.75$, the power curves of the two statistics are very close, but the curves of $T$ are slightly steeper.

- If $\rho = 0.5$, the power curves of the two statistics are essentially the same.
• Part of the differences between the two test statistics may be the result of slightly larger estimated type I error rates of the proposed test compared to the FM test.

• The above patterns are consistent under the three underlying distributions, $N(0,1)$, $t(3)$ and $LN(0,1)$.

Hence we conclude that the tests based on both $T$ and the FM statistics are robust to the quality of judgment ranking and the underlying distributions. The performance of $T$ is superior to the FM statistic when the ranking quality is reasonably good ($\rho \geq 0.75$). When $\rho = 0.5$, the performance of the tests based on $T$ or the FM statistic is equivalent to the MWW statistic (random ranking case).
Figure 4.1: Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size ($H$) and replicate size ($n$). Data sets are generated from $N(0,1)$. Perfect ranking is used and set sizes ($H$) are marked on each curve.
Figure 4.2: Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size ($H$) and replicate size ($n$). Data sets are generated from $t(3)$. Perfect ranking is used and set sizes ($H$) are marked on each curve.
Figure 4.3: Power curves of $T$, $U_{RSS}$ and MWW statistics for selected set size ($H$) and replicate size ($n$). Data sets are generated from LN(0,1). Perfect ranking is used and set sizes ($H$) are marked on each curve.
Figure 4.4: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 5 and selected set sizes ($H$) marked on each curve. Data sets are generated from standard normal distribution ($N(0,1)$).
Figure 4.5: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($N(0,1)$).
Figure 4.6: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 10 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($N(0,1)$).
Figure 4.7: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 10 and selected set sizes ($H$) marked on each curve. Data sets are generated from standard normal distribution ($N(0,1)$).
Figure 4.8: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$).
Figure 4.9: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 5 and selected set sizes ($H$) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$).
Figure 4.10: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 10 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution ($t(3)$).
Figure 4.11: Power curves of \( T \) (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient \( \rho = 0.75 \) or 0.5, replicate size \( (n) = 10 \) and selected set sizes \( (H) \) marked on each curve. Data sets are generated from standard normal distribution \( (t(3)) \).
Figure 4.12: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (LN(0,1)).
Figure 4.13: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size ($n$) = 5 and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (LN(0,1)).
Figure 4.14: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 1$ or 0.9, replicate size ($n$) = 10 and selected set sizes ($H$) marked on each curve. Data sets are generated from standard normal distribution (LN(0,1)).
Figure 4.15: Power curves of $T$ (indicated by solid lines) and FM (indicated by dotted lines) statistics for correlation coefficient $\rho = 0.75$ or 0.5, replicate size $(n) = 10$ and selected set sizes (H) marked on each curve. Data sets are generated from standard normal distribution (LN(0,1)).
In this chapter, we illustrate the use of the proposed test in a data set. Ideally, we should design an experiment based on an ORRD and collect data to apply the procedure. This kind of experiments, however, is not available currently. We therefore apply the proposed test to a well known AIDS Clinical Trial Group Protocol 320 (ACTG 320) study performed by Hammer et. al. [16].

The researchers in ACTG 320 clinical trial compare two types of treatments for human immunodeficiency virus type I (HIV-1). One treatment is a two-drug regime involving two nucleoside analogues (lamivudine and zidovudine/stavudine), the other is a three-drug regime including the protease inhibitor (indinavir) along with the two nucleoside analogues. The study is designed as a randomized, double-blind, placebo-controlled trial with 1156 HIV-infected patients who have no more than 200 CD4 cells per cubic millimeter at the screening stage and are not previously treated by lamivudine and indinavir, but by zidovudine for at least three months. As a little introduction, CD4 cells mainly refer to T cells expressing CD4 (cluster of differentiation 4), a glycoprotein expressed on the surface of T helper cells, regulatory T cells, etc. They play an important role in human immune system. Since HIV attacks
the CD4 cells, the CD4 cell counts in blood can tell us the strength of the immune system and the stage of HIV disease. The Centers for Disease Control and Prevention considers HIV-infected persons who have CD4 counts below 200 to have AIDS, regardless of whether they are sick or not. In ACTG 320 study, according to the CD4 cell counts at the screening stage, patients were separated into two strata, no more than 50 CD4 cells per cubic millimeter and between 51 and 200 CD4 cells per cubic millimeter. The patients within each stratum were randomly assigned to two treatments. For each eligible patient, researchers recorded the demographic information, CD4 cell counts and Karnof scores at the screening stage; and monitored the new AIDS-defining events, death, adverse events and CD4 cell counts at week 4, 8, 24 and 40 during the therapeutic period. Plasma HIV-1 RNA concentrations of 190 patients randomly selected from the 1156 eligible ones were measured retrospectively. Statistical analyses were performed on the time to the development of AIDS or death, the changes in CD4 cell counts, and the changes in HIV-1 RNA concentrations over time to assess the effect of two- or three-drug treatment regimes. Their conclusions show that the three-drug regime is better than the two-drug regime in term of slowing the development of HIV-1 disease.

In our study, we select an ORRD sample of size 30 from all available patients in this study to compare the CD4 cell counts at week 4 between two treatment regimes. We denote two-drug regime as treatment A and three-drug regime as treatment B. We predetermine the set size $H = 3$ and the replication size $n = 5$, and use the design $\alpha = \{1, 3\}$ and $\beta = \{2\}$. It is reasonable to assume that patients with larger CD4 cell counts at the screening stage have relatively larger CD4 cell counts at week 4. Figure 5.1 shows the scatter plots of the CD4 cell counts at the screening stage versus
at week 4 for each treatment regime and the pooled data. The two variables are positively correlated and there is a potential outlier in treatment A. The correlation coefficients between the two variables are 0.843, 0.758 and 0.798 for treatment A, B and the pooled data, respectively. Figure 5.2 demonstrates the histograms of the CD4 cell counts at week 4 for each treatment regime and the pooled data. All three histograms are skewed to the right and have similar shape. The difference of medians of the CD4 cell counts from the two treatment regime is $M_B - M_A = 30$ per cubic millimeter. The potential outlier in treatment A and the skewness of distributions suggest that use of the proposed test would be appropriate for this data.

The above discussion indicates that it is reasonable to choose the CD4 cell counts at the screening stage to judgment rank the subjects prior to experiment. For the purpose of illustration, we treat all patients from treatment A and B as potential population by excluding the missing values in the CD4 cell counts at the screening stage and week 4. After excluding the missing values, the total number of patients available for sampling is 1077. Thirty patients are taken from the population through the following steps that resembles the ORRD as closely as possible.

Step I: We only allow the rankers to see the CD4 cell counts at the screening stage and the treatment variable. On the other hand, the rankers do not know what treatments A and B stand for.

Step II: We first assign a random number to each patient in the population and then sort the data by these random numbers so that the patients in the data set are listed in a random order.

Step III: We select the first three patients from the list and sort them based on their CD4 cell counts at the screening stage. If the first and third patients in this
sorted sample are from treatment A and the second patient is from treatment B, we then select these three patients as type I set in replicate 1. However, if the first and third patients are from treatment B and the second patient is from treatment A, we then select these three patients as type II set in replicate 1. If the first three patients are in neither of the two forms, we discard them and take the next three patients from the remaining population. We continue to this process until we have one set of type I and one set of type II for the first replicate.

Step IV: We repeat step III in the remaining population until we have five replicates.

We note that in Step III, three patients are selected at random from the list of all available patients. Since the sample size three can be considered relatively small with respect to the number of all available patients, 1077, between-set responses are all mutually independent. Although the experiment is not originally designed as an ORRD, sampled data resemble a valid ORRD structure with $H = 3$ and design parameters $\alpha = 1, 3$ and $\beta = 2$. The sampled data set is listed in Table 5.1.

We apply the proposed test to decide if the medians of CD4 cell counts at week 4 for the two treatment regimes are different. The test statistic for the data in Table 5.1 yields that $T = 83$ and p-value = 0.018. The value of the Hodges Lehmann estimator of the parameter is 31 CD4 cell counts per cubic millimeter with an associated 95% confidence interval from 24 to 37 CD4 cell counts per cubic millimeter. The confidence interval contains the true value 30. Based on the statistical evidence, we conclude that the three-drug treatment regime provides a substantial improvement over the two-drug treatment regime.
Figure 5.1: Scatter plots of the CD4 cell counts at the screening stage (Baseline CD4) versus at week 4.
Figure 5.2: Histograms of the CD4 cell counts at week 4 for each treatment regime and the pooled data.
<table>
<thead>
<tr>
<th>REGIME</th>
<th>PATIENT ID</th>
<th>BASELINECD4</th>
<th>WK4CD4</th>
<th>REPLICATE</th>
<th>SET</th>
<th>RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>61440</td>
<td>16.0</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>242216</td>
<td>29.0</td>
<td>98</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>10333</td>
<td>95.0</td>
<td>100</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>250896</td>
<td>25.0</td>
<td>120</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>240309</td>
<td>157.5</td>
<td>161</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>580773</td>
<td>206.0</td>
<td>430</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>212100</td>
<td>85.5</td>
<td>81</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>320941</td>
<td>115.0</td>
<td>176</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>320890</td>
<td>172.0</td>
<td>189</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>320723</td>
<td>26.0</td>
<td>85</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>580398</td>
<td>55.0</td>
<td>61</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>253060</td>
<td>159.0</td>
<td>207</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>580325</td>
<td>132.5</td>
<td>180</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>21074</td>
<td>156.5</td>
<td>199</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>250901</td>
<td>160.0</td>
<td>170</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>231668</td>
<td>14.0</td>
<td>35</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>150053</td>
<td>59.0</td>
<td>102</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>111347</td>
<td>204.0</td>
<td>227</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>181575</td>
<td>107.0</td>
<td>109</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>260237</td>
<td>139.5</td>
<td>181</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>760083</td>
<td>219.0</td>
<td>238</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>90927</td>
<td>21.5</td>
<td>50</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>610386</td>
<td>50.0</td>
<td>91</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>253079</td>
<td>58.5</td>
<td>135</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>570097</td>
<td>4.5</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>300109</td>
<td>39.0</td>
<td>107</td>
<td>5</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>161036</td>
<td>225.5</td>
<td>261</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>760084</td>
<td>5.0</td>
<td>126</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>540183</td>
<td>10.5</td>
<td>24</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>321044</td>
<td>209.5</td>
<td>220</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.1: The sampled ORRD data for week 4 CD4 cell counts
CHAPTER 6

DISCUSSION

6.1 Conclusions

The two-sample location shift is one of the common research problems in experimental studies. Main interest in this problem is either to estimate the contrast parameter, construct confidence interval or to test a hypothesis. The standard inferential procedures could be either in parametric or nonparametric framework. In nonparametric framework, Hodges Lehmann estimator and MWW test are well known rank-based inferential procedures in the literature. The MWW statistic is constructed based on two independent random samples from two populations. These samples are collected from a completely randomized design in which some useful information inherent with experimental units is ignored. This information may be subjective, imprecise, or even biased, but it is very helpful to explain the heterogeneity (or homogeneity) among the experimental units.

Ranked set sampling procedure exploits this kind of information through judgment ranking process and consequently creates highly structured data that generates an informative sample. Due to this highly structured data, the rank-sum test statistic \( U_{RSS} \) for a two-sample location problem based on two independent ranked set
samples is superior to MWW statistic if the judgment ranking is perfect. There are some technical and practical concerns in the use of rank-sum test in ranked set sampling. At the technical level, under imperfect ranking the test loses its distribution free property. Its size is inflated and it becomes an invalid test. At the practical level, it is not clear how to conduct an experiment with a proper randomization scheme which is essential for any randomized experiment.

The order restricted randomized design addresses these concerns by using not only the judgment ranking information, but also a proper randomization scheme in the design of experiments. The data collected under the ORRD are presumably positively correlated through judgment ranking process. These positive correlations are transformed into negative ones for the estimation of the contrast parameters with a proper randomization. This leads to a reduction in the variance of contrast parameter and improve the inference. We proposed a rank-sum test based on the ORRD data to test a two-sample location shift parameter. Under the null hypothesis, the test statistic converges to a normal distribution as the sample size increases regardless of the quality of judgment ranking. If the judgment ranking is perfect, the variance of the asymptotic null distribution has a simple form as a function of the set size $H$ and the design parameters $\alpha$ and $\beta$. If the judgment ranking is imperfect, the test is calibrated by estimating the null variance of the test statistic. When the sample size is moderately large, the asymptotic null distribution of the test statistic is approximated by Student’s $t$-distribution. The test is asymptotically distribution-free regardless of the judgment ranking quality.

Chapter 2 shows that the test statistic is Pitman regular. The Pitman efficacy is a function of the underlying distribution and the null variance of $T$. For a fixed
set size $H$, it is shown that there exist an optimal design that maximizes the Pitman efficacy among all possible order restricted randomized designs.

The simulation results illustrate that type I error rates are close to the nominal size of the test regardless of the judgment ranking quality. The proposed test outperforms its competitor tests based on $U_{RSS}$ and the MWW statistics. This result is proved analytically in terms of asymptotic relative efficiencies when the ranking information is perfect and verified by the simulation study when the sample size is small and ranking information is imperfect. In addition to the testing procedure, the location shift parameter is estimated and the associated confidence interval is constructed.

6.2 Future directions

Under imperfect ranking, the true cumulative distribution functions (c.d.f.s) for the judgment rank order statistics are unknown. In our study, we use empirical c.d.f.s to estimate the true c.d.f.s so that we can develop a consistent estimator for the variance of the asymptotic null distribution of the proposed test statistic $T$. The consistent estimator may not be unique for these true c.d.f.s. One may use the structure of the ORRD to develop better estimators for these c.d.f.s. This may ultimately improve the performance of the proposed test.

The order restricted randomized design for two treatment regimes can be extended to multiple treatment regimes. There would be two approaches in this direction. The first approach could use the pair-wise comparisons among treatment regimes by conducting several rank-sum tests. This approach lends itself to a multiple comparison procedure. In this case, the overall type I error rate needs to be controlled. The second approach is to use rank regression models to develop inference on the regression
parameter. This approach is appealing since it provides a comprehensive treatment of all parameters (including the parameters that are the property of a ranking model) in the model.

In this study we assume that the underlying distribution of the response variable is continuous. The categorical responses are however quite common in scientific studies. The basic idea of this dissertation can be extended to categorical responses to develop inferential procedures.
BIBLIOGRAPHY


